

Apoyo al Diagnóstico de Cáncer de Piel con Inteligencia Artificial: Estudio Piloto

<https://doi.org/10.32635/2176-9745.RBC.2026v72n1.5443ES>

Suporte ao Diagnóstico de Câncer de Pele com Inteligência Artificial: Estudo-Piloto

Support for Skin Cancer Diagnosis Using Artificial Intelligence: Pilot Study

Eduarda Menezes da Silveira¹; Helena Cargnelutti Grimaldi²; Sandro da Silva Camargo³

RESUMEN

Introducción: El cáncer de piel constituye una de las neoplasias de mayor incidencia en el Brasil, siendo el diagnóstico temprano un factor determinante para el éxito terapéutico y para la reducción de la morbilidad asociada. Ante este escenario epidemiológico, se observa un contexto particularmente propicio para la incorporación de herramientas computacionales complementarias a la evaluación clínica tradicional, con énfasis en enfoques basados en inteligencia artificial. **Objetivo:** Desarrollar y validar un modelo basado en redes neuronales convolucionales para la clasificación automática de lesiones cutáneas malignas y benignas. **Método:** Se utilizaron 2639 imágenes de la base pública *International Skin Imaging Collaboration* (ISIC), con anotaciones validadas por biopsia. El sistema computacional incluyó etapas de preprocesamiento y entrenamiento supervisado con la arquitectura YOLOv11. El rendimiento fue evaluado mediante validación interna y validación externa. **Resultados:** El modelo alcanzó una exactitud media del 80,53% y una sensibilidad media del 80,44% en la identificación de ocho clases de lesiones: melanoma, nevo, carcinoma de células basales, queratosis actínica, queratosis benigna, dermatofibroma, lesión vascular y carcinoma espinocelular. La implementación también resultó en una base de imágenes anotadas y en un flujo de análisis reproducible. **Conclusión:** La aplicación de inteligencia artificial en el apoyo al diagnóstico del cáncer de piel demostró un desempeño prometedor, con potencial aplicación en tamizajes clínicos. Estudios futuros deberán considerar la ampliación de la base de datos y el desarrollo de interfaces para su uso por profesionales de la salud.

Palabras clave: Neoplasias Cutáneas/clasificación; Avulsiones cutáneas/clasificación; Aprendizaje Profundo; Redes Neuronales Convolucionales; Procesamiento de Imagen Asistido por Computador.

RESUMO

Introdução: O câncer de pele constitui uma das neoplasias de maior incidência no Brasil, e o diagnóstico precoce representa um fator determinante para o êxito terapêutico e para a redução da morbimortalidade associada. Diante desse cenário epidemiológico, observa-se um contexto particularmente propício à incorporação de ferramentas computacionais complementares à avaliação clínica tradicional, com destaque para abordagens baseadas em inteligência artificial. **Objetivo:** Desenvolver e validar um modelo baseado em redes neurais convolucionais para a classificação automática de lesões cutâneas malignas e benignas. **Método:** Foram utilizadas 2.639 imagens da base pública *International Skin Imaging Collaboration* (ISIC), com anotações validadas por biópsia. O sistema computacional incluiu etapas de pré-processamento e treinamento supervisionado com arquitetura YOLOv11. O desempenho foi avaliado por validação interna e validação externa. **Resultados:** O modelo alcançou acurácia média de 80,53% e sensibilidade média de 80,44% na identificação de oito classes de lesões: melanoma, nevo, carcinoma de células basais, queratose actínica, queratose benigna, dermatofibroma, lesão vascular e carcinoma espinocelular. A implementação também resultou em uma base de imagens anotadas e em um fluxo de análise reproduzível. **Conclusão:** A aplicação de inteligência artificial no suporte ao diagnóstico de câncer de pele demonstrou desempenho promissor, com potencial aplicação em triagens clínicas. Estudos futuros devem considerar a expansão da base de dados e o desenvolvimento de interfaces para uso por profissionais da saúde.

Palavras-chave: Neoplasias Cutâneas/classificação; Avulsões Cutâneas/classificação; Aprendizagem Profunda; Redes Neurais Convolucionais; Processamento de Imagem Assistido por Computador.

ABSTRACT

Introduction: Skin cancer is one of the most prevalent neoplasms in Brazil, and early diagnosis is a key determinant of therapeutic success and reduction of associated morbidity and mortality. In this epidemiological context, there is a particularly favorable scenario for the incorporation of computational tools to complement traditional clinical evaluation, with emphasis on artificial intelligence-based approaches. **Objective:** To develop and validate a convolutional neural network-based model for the automatic classification of malignant and benign skin lesions. **Method:** A total of 2,639 images from the public *International Skin Imaging Collaboration* (ISIC) database, with biopsy-validated annotations, were used. The computational system included preprocessing steps and supervised training using the YOLOv11 architecture. Performance was assessed through internal and external validation. **Results:** The model achieved a mean accuracy of 80.53% and a mean sensitivity of 80.44% in the identification of eight classes of lesions: melanoma, nevus, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma, vascular lesion, and squamous cell carcinoma. The implementation also resulted in an annotated image dataset and a reproducible analysis pipeline. **Conclusion:** The application of artificial intelligence to support skin cancer diagnosis demonstrated promising performance, with potential clinical screening applications. Future studies should consider expanding the dataset and developing user interfaces for healthcare professionals.

Key words: Skin Neoplasms/classification; Degloving Injuries/classification; Deep Learning; Convolutional Neural Networks; Image Processing, Computer-Assisted.

¹Universidade Federal do Pampa (Unipampa). Bagé (RS), Brasil. Serviço Nacional de Aprendizagem Industrial do Rio Grande do Sul (Senai-RS), *Center for Embedded Devices and Research in Digital Agriculture* (Cedra). São Leopoldo (RS) Brasil. E-mail: eduardasilveira.aluno@unipampa.edu.br. Orcid iD: <https://orcid.org/0009-0004-7829-9273>

²Pesquisadora-autônoma. Bagé (RS), Brasil. E-mail: hgrimaldi87@gmail.com. Orcid iD: <https://orcid.org/0009-0002-2762-4094>

³Unipampa, Programa de Pós-Graduação em Computação Aplicada (PPGAP). Bagé (RS), Brasil. Senai-RS, Cedra. São Leopoldo (RS) Brasil. E-mail: sandrocamargo@unipampa.edu.br. Orcid iD: <https://orcid.org/0000-0001-8871-3950>

Dirección para correspondencia: Sandro da Silva Camargo. Unipampa. Av. Maria Anunciação Gomes Godoy, 1650, Gabinete 3139 – Malafaia. Bagé (RS), Brasil. CEP 96413-172. E-mail: sandrocamargo@unipampa.edu.br



INTRODUCCIÓN

El cáncer de piel es la neoplasia maligna más incidente en el Brasil y en el mundo, representando un 30% de todos los tumores diagnosticados en el país¹. Está clasificado en dos grandes grupos: el cáncer de piel no melanoma, que comprende principalmente el carcinoma basocelular (CBC) y el carcinoma espinocelular (CEC), y el melanoma, menos frecuente, pero responsable por la mayoría de las muertes por causa de su agresividad y potencial metastásico^{2,3}. En el Brasil, la incidencia estimada es de aproximadamente 60 casos por cada 100 000 habitantes, siendo el CBC el más prevalente (70-75%), seguido por el CEC (15-20%)². Aunque presenten baja mortalidad, estas neoplasias tienen carácter invasivo local y elevada tasa de recurrencia. Ya el melanoma, a pesar de corresponder a solo el 3% de los casos, concentra la mayor parte de las muertes relacionadas con el cáncer de piel³. Además de las neoplasias malignas, otras lesiones cutáneas, como nevos, queratosis actínicas, queratosis benignas, dermatofibromas y lesiones vasculares, también representan desafíos diagnósticos, pues pueden mimetizar tumores malignos⁴⁻⁶.

La detección temprana es fundamental para reducir morbilidad, especialmente en los casos de melanoma. En este contexto, la dermatoscopia ha ampliado la precisión diagnóstica a cerca del 90%, en comparación al 75-80% obtenido por la inspección clínica aislada^{7,8}. No obstante, su efectividad depende de la experiencia del examinador, lo que resulta en variabilidad en la interpretación. Pero la limitación del acceso a especialistas, sobre todo en áreas remotas, refuerza las desigualdades en el diagnóstico y tratamiento oportunos⁹. En este escenario, soluciones basadas en inteligencia artificial (IA) surgen como herramientas complementarias a la práctica clínica. La IA ha demostrado potencial en acelerar diagnósticos, apoyar decisiones médicas y ampliar el acceso al cuidado en salud¹⁰. En particular, las redes neuronales convolucionales (RNC) han alcanzado un desempeño comparable al de dermatólogos experimentados en la clasificación automática de lesiones cutáneas¹¹⁻¹⁸. Entre los algoritmos disponibles, el *You Only Look Once* (YOLO) se destaca por la detección en tiempo real con elevada precisión, configurándose como alternativa promisor para tamizaje automatizado y apoyo al diagnóstico¹⁹.

La aplicación de modelos de IA como el propuesto en este estudio tiene especial relevancia para el contexto del Sistema Único de Salud (SUS) y de las políticas nacionales de control del cáncer. El Brasil presenta una distribución geográfica desigual de dermatólogos, concentrados en las regiones Sur y Sudeste, lo que limita el acceso a diagnósticos especializados en gran parte del

territorio⁹. Esta disparidad se refleja en diagnósticos tardíos y sobrecarga de servicios de alta complejidad. Además, la disponibilidad limitada de dermatoscopios en unidades básicas de salud y la diversidad étnico-racial de la población brasileña representan desafíos adicionales, puesto que la mayoría de las bases públicas de imágenes dermatológicas contiene muestras predominantemente de fototipos claros, lo que puede comprometer la precisión de los modelos en tonos de piel más oscuros.

En este sentido, iniciativas recientes han explorado el uso de la IA como herramienta de apoyo para la Atención Primaria y para el tamizaje en oncología¹⁰. Algunas investigaciones brasileñas ya demuestran el creciente interés de la comunidad científica nacional en integrar sistemas inteligentes a los flujos de cuidado del SUS, ampliando la detección temprana y reduciendo desigualdades regionales^{10,17}. Tales esfuerzos dialogan directamente con las directrices del Plan de Acciones Estratégicas para el Enfrentamiento de las Enfermedades Crónicas no Transmisibles y con las metas de control del cáncer establecidas por el Instituto Nacional del Cáncer (INCA)/Ministerio de Salud (MS).

Así, este estudio tiene como objetivo desarrollar, hacer pruebas y evaluar un modelo computacional basado en la arquitectura YOLO para la detección automática de lesiones cutáneas malignas y benignas en imágenes dermatoscópicas, buscando proporcionar soporte para el diagnóstico médico y, consecuentemente, contribuir para la identificación temprana del cáncer de piel.

MÉTODO

Estudio piloto que se configura como una investigación aplicada, de enfoque cuantitativo y delineamiento experimental, dirigida hacia el desarrollo y la evaluación de un sistema computacional de soporte al diagnóstico de cáncer de piel con uso de IA. La investigación se fundamentó en el análisis de imágenes dermatoscópicas para la detección automática de lesiones cutáneas, mediante RNC. El desempeño de los modelos fue evaluado por métricas cuantitativas estandarizadas, incluyendo precisión y sensibilidad, con el fin de medir su eficacia en la identificación de patrones morfológicos asociados a diferentes tipos de lesiones cutáneas.

Fueron obtenidas imágenes dermatoscópicas extraídas de la base pública *International Skin Imaging Collaboration* (ISIC) *Archive* 2019¹⁹, compuesta por 25 331 imágenes con diagnóstico histopatológico confirmado, según la distribución presentada en la Tabla 1. Los criterios de inclusión involucraron calidad visual adecuada, ausencia de interferencias clínicas visibles (como cicatrices de biopsia o suturas) y presencia de un único diagnóstico

por imagen. Todas las imágenes están anonimizadas, en conformidad con las directrices éticas internacionales para el uso secundario de datos clínicos.

La anotación de las imágenes es el proceso que consiste en delimitar la lesión e indicar su diagnóstico. Estas anotaciones fueron realizadas manualmente utilizando la herramienta LabelImg²⁰, con marcación de las lesiones mediante cajas delimitadoras (*bounding boxes*) y asociación de etiquetas de clase compatibles con el formato requerido por la arquitectura YOLO²¹. Las clases de las lesiones fueron definidas con base en los informes histopatológicos puestos a disposición junto con la base de imágenes¹⁹.

Considerando que el proceso de anotación de las lesiones en las imágenes es manual y realizado individualmente en cada muestra, este estudio piloto fue realizado sobre una fracción del conjunto de imágenes disponible, también presentada en la Tabla 1. El trabajo se encuentra en progreso, con la anotación de más imágenes y la ejecución de nuevos ciclos de entrenamiento y validación, con el objetivo de perfeccionar progresivamente los indicadores de desempeño del modelo.

Tabla 1. Distribución de las imágenes disponibles por clase de lesión cutánea (ISIC 2019)¹⁹ y distribución de la cantidad de imágenes por clase utilizadas en los experimentos

Clase de lesión cutánea	Número de imágenes disponibles	Número de imágenes utilizadas
Melanoma	4522	600
Nevo	12 875	600
Carcinoma basal	3323	200
Queratosis actínica	867	200
Queratosis benigna	2624	200
Dermatofibroma	239	239
Lesión vascular	253	200
Carcinoma espinocelular	628	400
Total	25 331	2639

El *pipeline* de modelado, representado en la Figura 1, incluye los siguientes pasos:

1) Preprocesamiento de las imágenes: todas las imágenes fueron estandarizadas en cuanto al tamaño y a las anotaciones de las lesiones, garantizando consistencia en el análisis y facilitando la comparación entre diferentes tipos de lesiones. Esta etapa es equivalente a la preparación cuidadosa de exámenes clínicos antes de la interpretación diagnóstica.

2) Entrenamiento del modelo, con ajuste automático de hiperparámetros, buscando optimizar la capacidad de clasificación de las lesiones. Este proceso permite que

el modelo aprenda de forma optimizada a distinguir características clínicas relevantes, de manera similar a la experiencia adquirida por un dermatólogo al revisar múltiples casos.

3) Diagnóstico, o validación interna del desempeño: durante el entrenamiento, el modelo fue evaluado continuamente en subconjuntos de datos reservados, midiendo precisión, sensibilidad y especificidad. Esta etapa garantiza que el desempeño del sistema sea confiable y consistente, similar a la revisión por pares en un contexto clínico antes de la implementación de un nuevo protocolo diagnóstico.

Se entrenaron arquitecturas de la familia YOLOv11 (versión 11)²⁰, con el objetivo de evaluar la precisión y la eficiencia de la detección automática de lesiones cutáneas. Esta arquitectura fue elegida debido a sus características de precisión, velocidad y capacidad de generalización, especialmente en aplicaciones médicas^{22,23}. Tal elección está alineada al objetivo de explorar redes neuronales profundas como soporte para el diagnóstico temprano del cáncer de piel. Las arquitecturas de la familia YOLOv11 empleadas en este estudio fueron:

- YOLOv11n (*nano*): versión compacta, optimizada para ejecución en dispositivos con capacidad computacional reducida, tal como *smartphones*, *tablets* y sistemas embebidos en dispositivos médicos portátiles, donde hay restricciones de memoria y procesamiento.
- YOLOv11s (*small*), YOLOv11m (*medium*), YOLOv11l (*large*): arquitecturas progresivamente más complejas, con mayor número de parámetros y, por lo tanto, mayor capacidad de representación, a costa del aumento del tiempo de procesamiento. Estas versiones son más indicadas para la ejecución en estaciones de trabajo con unidades de procesamiento gráfico (GPU) dedicadas, servidores clínicos y plataformas de investigación, contextos en donde la precisión diagnóstica se prioriza con relación al tiempo de procesamiento.

Todos los modelos fueron entrenados mediante prueba de diferentes hiperparámetros, como número de épocas, con *batch size* adaptativo, parada anticipada y validación interna continua durante el entrenamiento, lo que permite optimizar el desempeño del modelo. Este rigor metodológico, alineado con las mejores prácticas para entrenamiento y uso de modelos de clasificación²⁴, asegura que el modelo aprenda patrones relevantes y minimice sesgos en la identificación de las lesiones.

La validación de los modelos de IA se realizó en dos etapas complementarias²⁵:

1) Validación interna, utilizando una división estratificada aleatoria del 90% de las imágenes para el

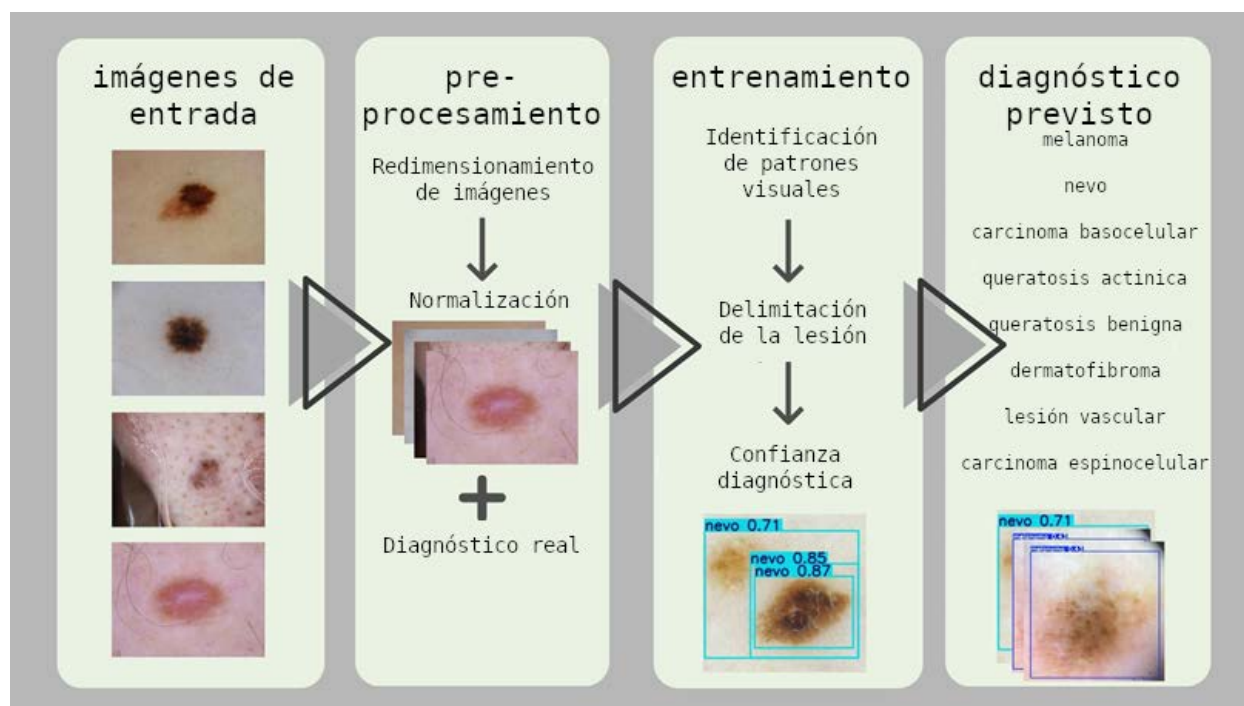


Figura 1. Pipeline de detección de lesiones cutáneas con YOLOv11

entrenamiento del modelo y 10% para probarlo. Esta técnica consiste en evaluar el desempeño del modelo de forma controlada, utilizando los propios datos del conjunto original, permitiendo estimar la capacidad predictiva del modelo y garantizando que todas las clases de lesiones cutáneas fuesen representadas proporcionalmente en los conjuntos de entrenamiento y prueba.

2) Validación externa, realizada con imágenes inéditas, es decir, no utilizadas anteriormente en el entrenamiento, capturadas en un ambiente clínico real. Estas imágenes fueron procesadas con el modelo entrenado, permitiendo verificar si el modelo mantiene su precisión en situaciones prácticas, fuera del ambiente de prueba inicial.

En la validación interna, fueron evaluadas las siguientes métricas de desempeño: precisión (*mean Average Precision* – mAP), sensibilidad media, precisión media y tabla de desempeño diagnóstico.

La muestra de la validación externa estuvo compuesta por 58 imágenes dermatoscópicas obtenidas en un servicio ambulatorio de dermatología de un servicio privado de salud. Las imágenes fueron capturadas con dermatoscopio DermLite DL4, con iluminación polarizada y no polarizada de 24 leds blancos y *Pigment Boost Plus* (luz naranja), siguiendo la rutina de tamizaje clínico de lesiones sospechosas. Imágenes con artefactos importantes (v.g. curativos, sangre, pelos gruesos) fueron excluidas.

No hubo balanceo de clases en la validación externa, de modo que la distribución de las lesiones refleja la prevalencia observada en la práctica clínica (mayor presencia de lesiones

benignas, como nevos y queratosis). Esta ausencia de balanceo puede favorecer el desempeño del modelo en las clases más frecuentes y perjudicar la evaluación en las menos representadas, como melanoma y CEC, resultando en mayor probabilidad de falsos negativos en estas categorías. Esta limitación será tratada en estudios futuros mediante ampliación de la muestra clínica y estrategias de balanceo.

El entrenamiento fue realizado en el ambiente de alto desempeño Saturno²⁶, equipado con ocho GPU NVIDIA A100 y dos TB de memoria. El sistema fue implementado en *Python 3.10*, con uso de las bibliotecas *PyTorch*, *Ultralytics*, *Pandas*, *NumPy*, *OpenCV* y *Matplotlib*, entre otras.

Considerando los aspectos éticos de la investigación, se destaca que no hubo necesidad de remitir este estudio al Comité de Ética en Pesquisa (CEP). Esto porque la investigación utilizó exclusivamente datos secundarios de acceso público, disponibles en base de datos abierta en internet, sin posibilidad alguna de identificación directa o indirecta de individuos. Tal encuadramiento encuentra respaldo en la Resolución n.º 466/2012²⁷ del Consejo Nacional de Salud (CNS), que establece directrices y normas para investigaciones involucrando a seres humanos y, de modo más específico, en la Resolución n.º 510/2016²⁸, la cual dispone sobre las normas aplicables a las investigaciones en ciencias humanas y sociales y manifiesta que no necesitan de apreciación ética los estudios que utilizan informaciones de dominio público. De esta forma, la realización de esta investigación está en conformidad con los principios éticos vigentes en el Brasil.

RESULTADOS

Se realizaron pruebas preliminares con diferentes arquitecturas de la familia YOLOv11 (*nano*, *small*, *medium* y *large*) para evaluar la detección de lesiones cutáneas en imágenes dermatoscópicas. Entre estas variantes, la arquitectura YOLOv11s (*small*) presentó los mejores resultados, los cuales están aquí informados. La Tabla de desempeño diagnóstico (Figura 2) ilustra los aciertos y errores por clase, considerando en el eje *x* las clases reales y en el eje *y*, las clases predichas. El análisis de esta matriz permitió identificar patrones de acierto y error del modelo en diferentes tipos de lesiones cutáneas.

El desempeño varió según la clase evaluada. Para lesiones vasculares, la tasa de acierto fue del 93%, con 7% de fallas de reconocimiento, manifestando elevada especificidad, pero sensibilidad limitada en casos marginales. En la queratosis actínica, la tasa de acierto fue del 90%; no obstante, el 10% fue incorrectamente clasificado como melanoma, lo que denota un riesgo clínico relevante en función del carácter agresivo de esta neoplasia.

En los nevos melanocíticos, el 83% fue correctamente clasificado, mientras que el 16% fue reconocido equivocadamente como melanoma, sugiriendo dificultad del modelo en diferenciar lesiones benignas y malignas. Para los dermatofibromas, la tasa de acierto fue del 78%, con el 6% de errores como melanoma y el 6% como nevo.

En cuanto al melanoma, el 75% fue identificado correctamente, pero el 12% fue confundido con nevo y el 7% con otras categorías, revelando la necesidad de mayor precisión en esta clase. En los carcinomas, se observó desempeño heterogéneo: el 67% de precisión para el espinocelular y solo el 45% para el basocelular. La queratosis benigna presentó la menor tasa de acierto (33%), posiblemente en función de la escasez de ejemplos en el conjunto de entrenamiento y de la semejanza morfológica con otras lesiones.

Los casos en que el modelo atribuyó la clase *background* correspondieron a falsos negativos, o sea, situaciones en que la lesión real no fue reconocida. Este tipo de falla tiene relevancia clínica crítica, pues puede contribuir para atrasar el diagnóstico y comprometer el tratamiento terapéutico.

Además de la evaluación cuantitativa, se realizó un análisis cualitativo de las predicciones visuales en imágenes dermatoscópicas. La Figura 3 muestra ejemplos representativos, con destaque para la correcta localización y clasificación de diferentes tipos de lesiones, evidenciando la aplicabilidad práctica del modelo.

Para facilitar la visualización del desempeño discriminativo del modelo, las principales métricas por clase fueron consolidadas en la Tabla 2, permitiendo

observar diferencias de sensibilidad, precisión y perfiles de error entre los tipos de lesiones.

DISCUSIÓN

Los resultados evidencian el potencial del modelo propuesto como herramienta de apoyo al diagnóstico médico en oncología cutánea, sobre todo en contextos de tamizaje y fundamento para la decisión clínica. A pesar de los avances observados, el desempeño heterogéneo entre clases señala limitaciones relevantes. Clases con patrones morfológicos semejantes, como nevos y melanomas, todavía presentaron confusión significativa, indicando la necesidad de mejoramiento en la capacidad discriminativa del sistema.

El bajo desempeño en categorías específicas, como CBC y queratosis benigna, sugiere impacto tanto de la desproporción muestral como de la complejidad intrínseca de estas lesiones. Estrategias de expansión y diversificación de la base de datos surgen como caminos prioritarios para aumentar la robustez del modelo.

Desde el punto de vista clínico, la ocurrencia de falsos negativos representa el mayor desafío, puesto que el no reconocimiento de una lesión maligna puede tener repercusiones severas para el paciente. Así, futuros entrenamientos deben priorizar la reducción de este tipo de error, buscando elevar la precisión global para niveles superiores al 95%.

Cuando se compara con estudios internacionales, el desempeño obtenido en este trabajo es compatible con investigaciones recientes que usaron redes neuronales profundas para detección de melanoma. Estudios realizados con RNC o modelos YOLO adaptados reportan sensibilidades entre el 72% y el 89% y precisión media entre el 78% y el 91% en la identificación de melanoma y CBC^{11,18}. En el presente estudio, la sensibilidad para melanoma fue del 75% y la precisión media general fue del 80,53%, valores dentro del rango observado en la literatura, lo que refuerza la viabilidad técnica del modelo incluso utilizando un conjunto reducido de imágenes anotadas manualmente.

Además, trabajos como SkinSage XAI de Munjal et al.¹² y el estudio prospectivo multicéntrico de Heinlein et al.¹⁸ demuestran que sistemas basados en IA pueden lograr un desempeño semejante o superior al de dermatólogos en escenarios de tamizaje, reforzando que la aplicación de modelos de visión computacional puede reducir costos y acelerar el diagnóstico en contextos clínicos. Así, el desempeño alcanzado por este estudio se aproxima a la tendencia global, reforzando su relevancia científica.

La perspectiva de implementación del modelo en dispositivos móviles o plataformas web constituye un avance estratégico, pues amplía significativamente la



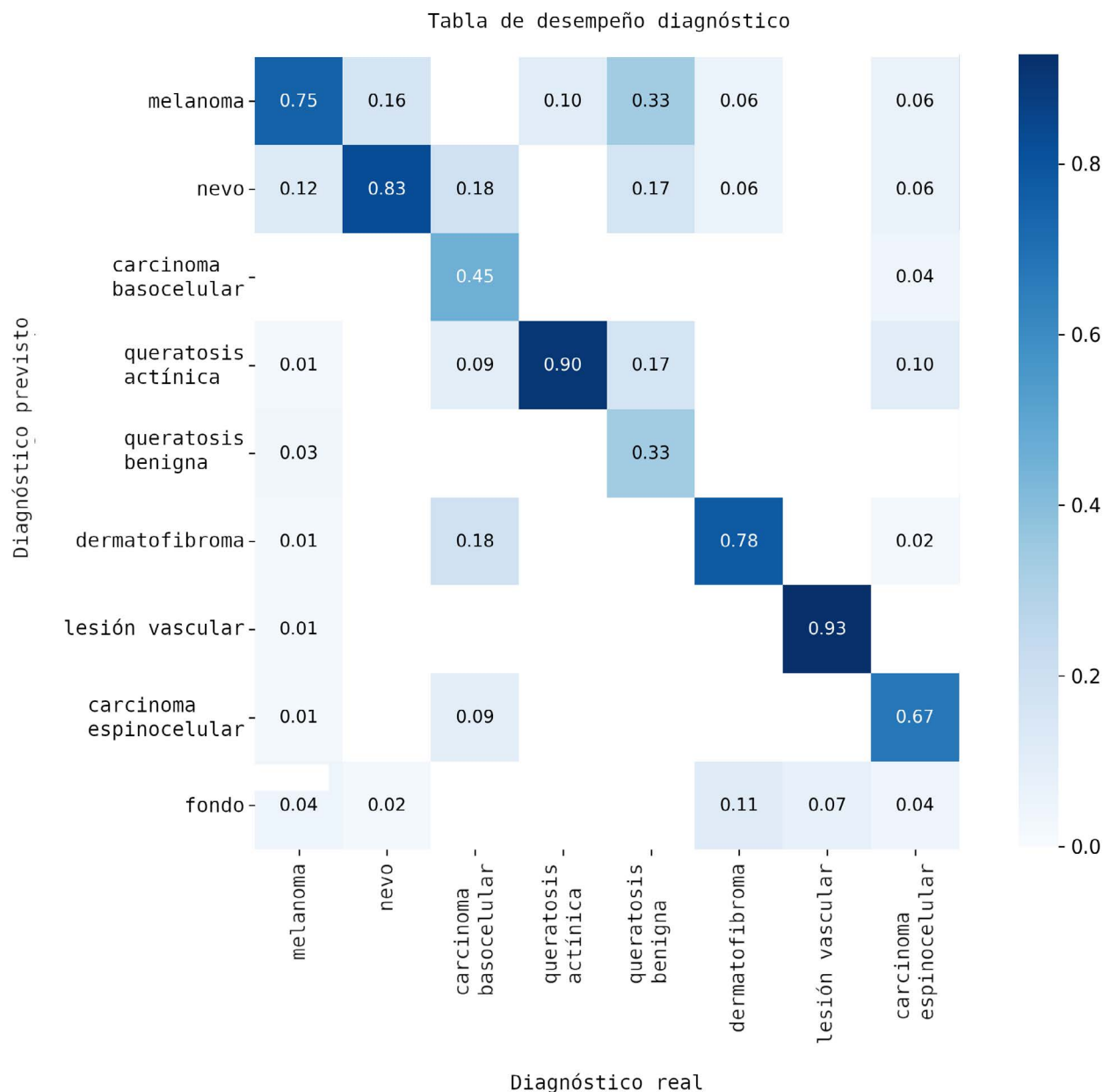


Figura 2. Tabla de desempeño diagnóstico presentando los diagnósticos reales y los diagnósticos predichos por el modelo
Nota: En el eje horizontal, se encuentran las clases reales de las lesiones y, en el eje vertical, las predicciones del modelo. La diagonal principal representa los aciertos. Cuando el modelo clasifica una lesión maligna como benigna, ocurre un falso negativo, situación de mayor riesgo clínico por atrasar el diagnóstico. Cuando una lesión benigna es clasificada como maligna, ocurre un falso positivo, pudiendo llevar a biopsias o encaminamientos innecesarios.

posibilidad de validación externa y aplicación práctica. Esta portabilidad favorece la realización de pruebas colaborativas en diferentes regiones del país, permitiendo que especialistas de distintas instituciones y contextos clínicos evalúen la herramienta de forma sistemática. Tal enfoque potencia la generación de evidencias sobre la robustez y la generalización del modelo, además de contribuir para la consolidación de su utilidad en escenarios de telemedicina y oncología de precisión.

Así, este estudio se caracteriza como una “prueba de concepto” (*proof of concept*), demostrando la viabilidad técnica y científica del uso del modelo propuesto como

herramienta de apoyo al diagnóstico dermatológico. La creación de un *pipeline* completo (anotación → entrenamiento → inferencia) y la validación con imágenes reales del ambiente clínico refuerzan el potencial de transferencia tecnológica del modelo para el SUS, especialmente en localidades con baja disponibilidad de dermatólogos y acceso limitado a la dermatoscopia.

A pesar de las limitaciones, sensibilidad aún insuficiente en algunas clases y baja representación de tonos de piel más oscuros, los resultados muestran potencial clínico, especialmente como herramienta de tamizaje en atención primaria a la salud.



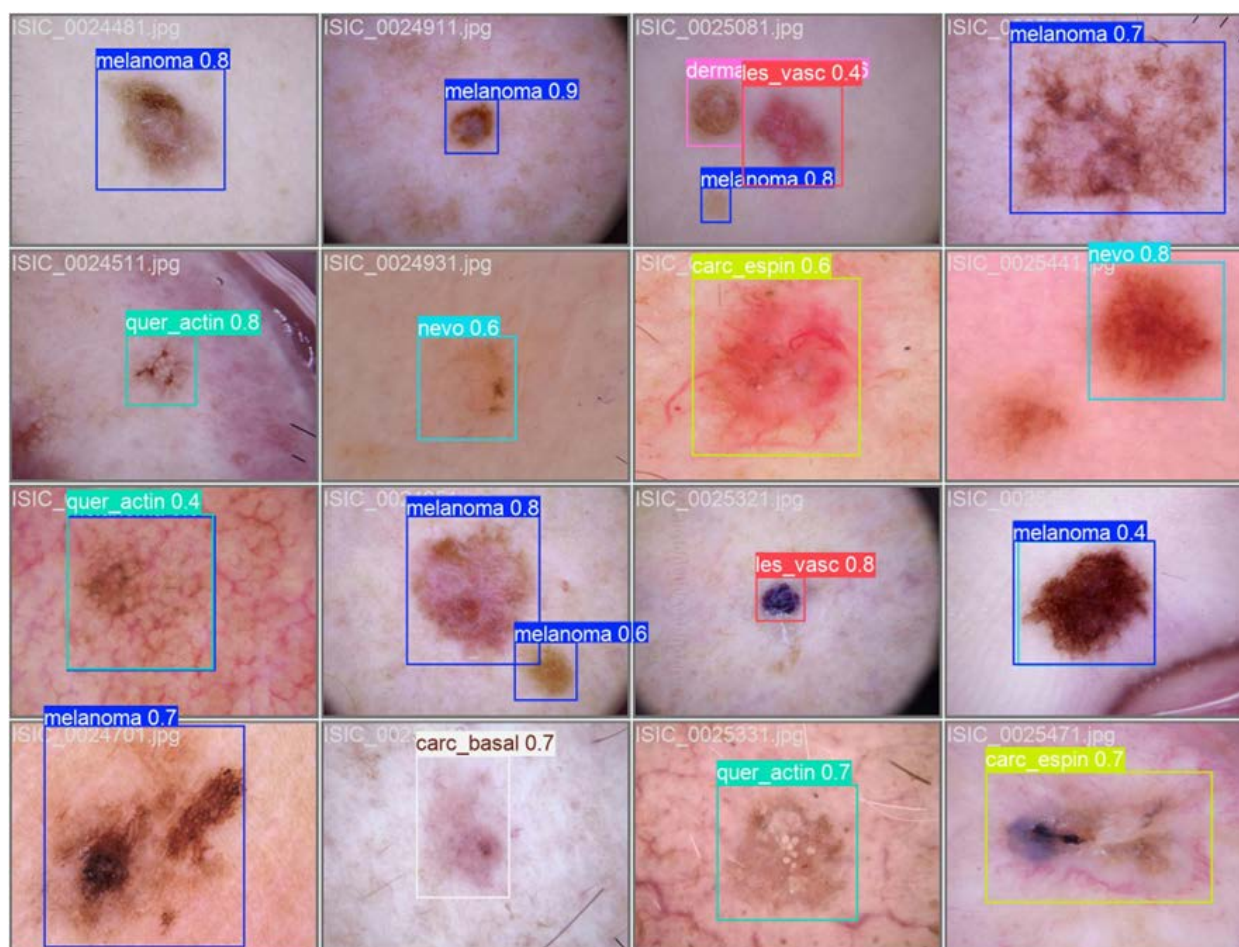


Figura 3. Ejemplos de detección automatizada de lesiones cutáneas por el modelo desarrollado, acompañados de los respectivos niveles de confianza atribuidos por el sistema al diagnóstico previsto

Nota: Cada imagen presenta la clase prevista y el nivel de confianza (probabilidad) de la predicción. (a) Verdadero positivo: el modelo identifica correctamente la lesión y delimita su región. (b) Falso positivo: el modelo clasifica una lesión benigna como maligna, lo que puede llevar a investigación innecesaria. (c) Falso negativo: el modelo falla en identificar una lesión maligna, lo que representa riesgo clínico por posible atraso en el diagnóstico. Estos ejemplos ilustran cómo responde el modelo a diferentes patrones dermatoscópicos, como bordes irregulares, asimetría de pigmentación y presencia de vasos sanguíneos.

Tabla 2. Desempeño del modelo por clase de lesión

Clase de lesión cutánea	Sensibilidad (%)	Precisión (%)	Principales errores observados
Melanoma	75	78	Falso negativo: confundido con nevo
Nevo	83	81	Falso positivo: confundido con melanoma
Carcinoma basal	45	72	No detectado en lesiones pequeñas o poco pigmentadas
Queratosis actínica	90	84	Confundida con melanoma en bordes irregulares
Queratosis benigna	33	68	Confundida con otras queratosis
Dermatofibroma	78	70	Confundido con nevo o melanoma
Lesión vascular	93	89	Baja tasa de error: patrón fácilmente identificado
Carcinoma espinocelular	67	74	Confundido con queratosis actínica

En síntesis, el modelo desarrollado presenta los siguientes potenciales principales:

- viabilidad técnica comprobada para detección automática de múltiples tipos de lesiones;
- desempeño compatible con estudios internacionales que utilizan *deep learning* en dermatoscopia;
- potencial de aplicación en tamizaje clínico, especialmente en el contexto del SUS;
- establecimiento de bases para estudios futuros con mayor diversidad de datos, balanceo y validación multicéntrica. Esta actividad presentaría potencial significativo mediante expansión y fortalecimiento por medio de alianzas institucionales estratégicas, especialmente con el INCA/MS. El apoyo de estas entidades podría viabilizar mayor alcance, integración de bases de datos nacionales y perfeccionamiento metodológico, contribuyendo para la consolidación de una red colaborativa orientada hacia el monitoreo y la investigación en oncología en el país.

CONCLUSIÓN

Este estudio demostró avances relevantes en la aplicación de la IA para el diagnóstico asistido del cáncer de piel, con énfasis en la utilización de la arquitectura YOLOv11. La construcción de un *pipeline* funcional, que abarca desde el preprocesamiento de las imágenes hasta el entrenamiento del modelo con datos públicos anotados manualmente, evidencia la viabilidad técnica de la propuesta.

Entre las contribuciones más significativas, se destacan: la implementación de un modelo de detección automática de lesiones cutáneas con múltiples clases, a partir de imágenes dermatoscópicas, ampliando su potencial de aplicación clínica; el desarrollo de un conjunto de datos con anotaciones manuales, priorizando patrones visuales relevantes como vasos y bordes, con impacto directo en la precisión del modelo; la integración de tecnologías de visión computacional e IA en un flujo de trabajo reproducible y escalable.

A pesar de los avances, el estudio presenta limitaciones importantes: la precisión del modelo, aunque prometedora con un valor del 80,53%, no alcanza todavía los niveles necesarios para uso clínico directo. Esta limitación será abordada mediante nuevos ciclos de entrenamiento, que estarán en ejecución en los próximos seis meses, con bases de datos expandidas y más diversificadas; el conjunto de imágenes utilizado tiene baja representatividad de tonos de piel más oscuros, lo que compromete la generalización del modelo para la población brasileña. Tal vacío evidencia la necesidad de iniciativas nacionales orientadas a la recolección sistemática de imágenes dermatológicas con

diversidad étnico-racial; la validación externa, aunque iniciada con imágenes inéditas capturadas en ambiente clínico, requiere un nivel de detalle adicional respecto al muestreo, a los criterios de inclusión y a los parámetros clínicos evaluados, para asegurar la extrapolación de los resultados a los servicios reales de salud.

Finalmente, estas consideraciones demuestran el potencial transformador del proyecto, al mismo tiempo que evidencian los desafíos superados y trazan caminos para un resultado más robusto y aplicable en la práctica clínica.

AGRADECIMIENTOS

A NVIDIA, por el soporte computacional puesto a disposición, fundamental para la ejecución de los experimentos y análisis de este estudio.

APORTES

Todos los autores contribuyeron substancialmente en la concepción y en la planificación del estudio; en la obtención, análisis e interpretación de los datos; en la redacción y revisión crítica; y aprobaron la versión final a publicarse.

DECLARACIÓN DE CONFLICTOS DE INTERÉS

Nada a declarar.

DECLARACIÓN DE DISPONIBILIDAD DE DATOS

Los conjuntos de datos generados y analizados durante el estudio actual están disponibles en el repositorio <https://github.com/Sandrocamargo/publications/tree/main/rbc25a>

FUENTES DE FINANCIAMIENTO

No hay.

REFERENCIAS

1. Delgado LP, Regis MN, Caldas D, et al. Epidemiologia e manejo clínico do câncer de pele melanoma: uma revisão narrativa. J Soc Issues Health Sci. 2024;1(7):1-11. doi: <https://doi.org/10.5281/zenodo.17473590>
2. Rezende Filho AV, Yamamoto HG, Macedo JLS et al. Perfil epidemiológico de pacientes portadores de câncer de pele atendidos no Hospital Regional da Asa Norte/DF - Brasil. Rev Bras Cir Plást. 2020;35(3):316-21. doi: <https://doi.org/10.5935/2177-1235.2020RBCP0056>

3. Borges AL, Zalaudeck I, Longo C, et al. Melanocytic nevi with special features: clinical-dermoscopic and reflectance confocal microscopic-findings. *JEADV*. 2014;28(7):833-45. doi: <https://doi.org/10.1111/jdv.12291>
4. Piccolo V, Russo T, Moscarella E, et al. Dermatoscopy of vascular lesions. *Dermatol Clin*. 2018;36(4):389-95. doi: <https://doi.org/10.1016/j.det.2018.05.006>
5. Álvarez-Salafranca M, Gómez-Martíns I, Bañuls J, et al. Dermoscopy of inflamed seborrheic keratosis: a great mimic of malignancy. *Australas J Dermatol*. 2022;63(1):53-61. doi: <https://doi.org/10.1111/ajd.13781>
6. Thamm JR, Welzel J, Schuh S. Diagnosis and therapy of actinic keratosis. *J Dtsch Dermatol Ges*. 2024;22(5):675-90. doi: <https://doi.org/10.1111/ddg.15288>
7. Azulay RA, Azulay DR, Azulay-ABULAFIA L. *Dermatologia*. 7.ed. Rio de Janeiro: Guanabara Koogan; 2017.
8. Botton DV, Barbosa DGR, Cavalcante Junior CAC, et al. Relevância da dermatoscopia para o diagnóstico precoce de melanomas: uma revisão de literatura. *RISE*. 2020;1(2):159-74. doi: <https://doi.org/10.56344/2675-4827.v1n2a20209>
9. Schmitt JV, Miot HA. Distribution of Brazilian dermatologists according to geographic location, population and HDI of municipalities: an ecological study. *An Bras Dermatol*. 2014;89(6):1013-5. doi: <https://doi.org/10.1590/abd1806-4841.20143276>
10. Siqueira ASE, Santos Neto MF, Ferreira CBT, et al. Inteligência artificial nas ações de controle do câncer: solução ou problema? *Rev Bras Cancerol*. 2025;71(3):e-005291. doi: <https://doi.org/10.32635/2176-9745.RBC.2025v71n3.5291>
11. Phillips M, Marsden H, Jaffe W, et al. Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions. *JAMA Netw Open*. 2019;2(10):e1913436. doi: <https://doi.org/10.1001/jamanetworkopen.2019.13436>
12. Munjal G, Bhardwaj P, Bhargava V, et al. SkinSage XAI: an explainable deep learning solution for skin lesion diagnosis. *Health Care Sci*. 2024;3(6):438-55. doi: <https://doi.org/10.1002/hcs2.121>
13. Cui X, Wei R, Gong L, et al. Assessing the effectiveness of artificial intelligence methods for melanoma: a retrospective review. *J Am Acad Dermatol*. 2019;81(5):1176-80. doi: <https://doi.org/10.1016/j.jaad.2019.06.042>
14. Jairath N, Pahalyants V, Shah R, et al. Artificial intelligence in dermatology: a systematic review of its applications in melanoma and keratinocyte carcinoma diagnosis. *Dermatol Surg*. 2024;50(9):791-8. doi: <https://doi.org/10.1097/DSS.0000000000004223>
15. Wang HH, Wang YH, Liang CW, et al. Assessment of deep learning using nonimaging information and sequential medical records to develop a prediction model for nonmelanoma skin cancer. *JAMA Dermatol*. 2019;155(11):1277-83. doi: <https://doi.org/10.1001/jamadermatol.2019.2335>
16. Chu YS, An HG, Oh BH, et al. Artificial intelligence in cutaneous oncology. *Front Med (Lausanne)*. 2020;7:318. doi: <https://doi.org/10.3389/fmed.2020.00318>
17. Silva TAM, Souza EP, Bambil D, et al. Aprendizado de máquina aplicado ao diagnóstico de câncer de pele. In: 16 Anais do Congresso Brasileiro de Inteligência Computacional [Internet]; 2021 out 3-6; Joinville. Joinville: UDESC; 2021. doi: <http://dx.doi.org/10.21528/CBIC2021-55>
18. Heinlein L, Maron RC, Hekler A, et al. Clinical melanoma diagnosis with artificial intelligence: insights from a prospective multicenter study [Preprint]. *arXiv*. 2024. doi: <https://doi.org/10.48550/arXiv.2401.14193>
19. ISIC 2019: Skin lesion analysis towards melanoma detection [Internet]. Versão 2019. [sem local]: International Skin Imaging Collaboration; 2019 [acesso 2024 set 18]. Disponível em: <https://challenge2019.isic-archive.com/>
20. LabelImg [Internet]. Versão 1.8.6. [sem local]: Python Software Foundation; ©2025 [acesso 2024 set 18]. Disponível em: <https://pypi.org/project/labelImg/>
21. Ultralytics. Ultralytics YOLO [repositório GitHub] [acesso 2025 abr 10]. Disponível em: <https://github.com/ultralytics/ultralytics>
22. Terven J, Córdova-Esparza DM, Romero-González JA. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS. *Mach Learn Knowl Extr*. 2023;5(4):1680-716. doi: <https://doi.org/10.3390/make5040083>
23. Esteva A, Chou K, Yeung S, et al. Deep learning-enabled medical computer vision. *npj Digit Med*. 2021;4:5. doi: <https://doi.org/10.1038/s41746-020-00376-2>
24. Cechinel C, Camargo SS. Mineração de dados educacionais: avaliação e interpretação de modelos de classificação [Internet]. In: Jaques P, Pimentel M, Siqueira S, et al, editores. *Metodologia de pesquisa científica em informática na educação: abordagem quantitativa*. Porto Alegre: SBC; 2020. Disponível em: https://ceie.sbc.org.br/metodologia/wp-content/uploads/2019/11/livro2_cap12.pdf
25. Steyerberg EW, Harrell Jr FE. Prediction models need appropriate internal, internal-external, and external



- validation. J Clin Epidemiol. 2016;69:245-7. doi: <https://doi.org/10.1016/j.jclinepi.2015.04.005>
26. Saturn Cloud [Internet]. Nova Iorque: Saturn Cloud, ©2025 [acesso 2024 set 18]. Disponível em: <https://saturncloud.io/>
27. Conselho Nacional de Saúde (BR). Resolução n° 466, de 12 de dezembro de 2012. Aprova as diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos [Internet]. Diário Oficial da União, Brasília, DF. 2013 jun 13 [acesso 2025 ago 27]; Seção 1:59. Disponível em: https://bvsmms.saude.gov.br/bvs/saudelegis/cns/2013/res0466_12_12_2012.html
28. Conselho Nacional de Saúde (BR). Resolução n° 510, de 7 de abril de 2016. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais cujos procedimentos metodológicos envolvam a utilização de dados diretamente obtidos com os participantes ou de informações identificáveis ou que possam acarretar riscos maiores do que os existentes na vida cotidiana, na forma definida nesta Resolução [Internet]. Diário Oficial da União, Brasília, DF. 2016 maio 24 [acesso 2024 abr 7]; Seção 1:44. Disponível em: http://bvsmms.saude.gov.br/bvs/saudelegis/cns/2016/res0510_07_04_2016.html

Recebido em 1/9/2025
Aprovado em 11/11/2025

