

VALIDACIÓN DE MODELO BASADO EN INTELIGENCIA ARTIFICIAL PARA LA EVALUACIÓN DE ÚLCERAS GENITALES, ESTUDIO COMPARATIVO CON MÉDICOS ESPECIALISTAS

VALIDATION OF THE MODEL BASED ON ARTIFICIAL INTELLIGENCE FOR GENITAL ULCERS ASSESSMENT; SPECIALISTS COMPARATIVE STUDY

Gallegos Morán, H.^{1,2}; Aclé Fossa, R.³; Bravo I., J.⁴; Sepúlveda C., F.⁴; Zúñiga A., Á.⁵; San Francisco R., I.⁶; Mery, D.⁷.

RESUMEN

Introducción: Existen distintos virus que causan úlceras genitales. El diagnóstico inicial es difícil de diferenciar médicos no expertos. La inteligencia artificial (IA) es un elemento importante, permitiendo una aproximación certera, comparable con personas expertas. El objetivo del trabajo es mostrar la aplicación y capacidad diagnóstica de un modelo basado en redes neuronales que permite diferenciar entre las lesiones antes mencionadas a través de una fotografía y síntomas del paciente.

Materiales y métodos: Se utilizó una base de datos de imágenes de los 5 tipos de úlceras genitales más comunes. Se entrenó un modelo de Redes Neuronales Convolucionales, concatenado con modelos de machine learning para diferenciar entre las 5 diferentes tipos de lesiones. Se validó utilizando 15 casos clínicos, comparando médicos generales, dermatólogos y urólogos. Se análisis estadísticos mediante modelos de regresión Poisson, SPSS 22.0 y R Studio.

Resultados: Se incluyeron 104 médicos generales y especialistas. El porcentaje de acierto para el diagnóstico más probable fue de 54.6%, 55.1%, 70.4% y 86.7% para Médicos generales, Urólogos, Dermatólogos y Modelo de IA, respectivamente ($p=0.006$). Mientras que, para los 2 diagnósticos más probables, el porcentaje de acierto fue de 71.4%, 70.9%, 81.5% y 100% para Médicos generales, Urólogos, Dermatólogos y Modelo de IA, respectivamente ($p=0.031$).

Conclusión: Presentamos la validación de un modelo de Inteligencia Artificial para el diagnóstico de úlceras genitales, sin reportes disponibles en la literatura. Existen diferencias significativas en la probabilidad de obtener un diagnóstico adecuado, siendo superior a la de los médicos especialistas gracias al aprendizaje desarrollado en redes neuronales.

ABSTRACT

Introduction: There are different viruses that cause genital ulcers. The initial diagnosis is difficult to differentiate between non-expert doctors. Artificial intelligence (AI) is an important element, allowing an accurate approach, comparable with experts. The objective of this study is to show the application and diagnostic capacity of a model based on neural networks that allows differentiating between the injuries previously mentioned through a photograph and symptoms of the patient.

Materials and methods: A database of images of the 5 most common types of genital ulcers was used. A Convolutional Neural Networks model was trained, concatenated with machine learning models to differentiate between the 5 different types of injuries. It was validated using 15 clinical cases, comparing general physicians, dermatologists and urologists. Statistical analysis was performed using Poisson regression models, SPSS 22.0 and R Studio.

Results: 104 general physicians and specialists were included. The percentage of accuracy for the most probable diagnosis was 54.6%, 55.1%, 70.4% and 86.7% for General physicians, Urologists, Dermatologists and IA Model, respectively ($p = 0.006$). While for the 2 most probable diagnoses, the percentage of accuracy was 71.4%, 70.9%, 81.5% and 100% for General physicians, Urologists, Dermatologists and AI Model, respectively ($p = 0.031$).

¹Urología Redsalud CChC, Santiago, Chile; ²Servicio de Urología, Hospital Dr. Sótero del Río, Santiago, Chile; ³Servicio de Dermatología, CRS Hospital Provincia Cordillera, Santiago, Chile; ⁴Facultad de Medicina. Escuela de Postgrado, Pontificia Universidad Católica de Chile, Santiago, Chile; ⁵Servicio de Urología, Clínica Santa María, Santiago, Chile; ⁶Departamento de Urología, Pontificia Universidad Católica de Chile, Santiago, Chile; ⁷Departamento de Ciencias de la Computación, Pontificia Universidad Católica de Chile, Santiago, Chile.

Fecha de Recepción: 30-11-2021

Fecha de aceptación: 17-02-2022

Contacto: isanfrancisco@med.puc.cl

Conclusion: We present the validation of an Artificial Intelligence model for the diagnosis of genital ulcers, without previous reports available in the literature. There are significant differences in the probability of obtaining an adequate diagnosis, being higher than that of specialist doctors thanks to the learning developed in neural networks.

Introducción

La incidencia global de úlceras genitales está estimada en más de 20 millones de casos anualmente a nivel global (1,2), estas lesiones al ser comunes llevan a un impacto en la salud pública (3,4). El virus herpes simple y la sífilis son de las causas más habituales de úlceras genitales. Otras causas frecuentes infecciosas son el Chancroide, Linfogranuloma venéreo, Donovanosis, infecciones bacterianas secundarias y Hongos. En general el tratamiento inicial de estas úlceras está basado en la presentación clínica, posteriormente confirmado por los test serológicos, de secreción o cultivos (5). Es por esto que el diagnóstico adecuado es fundamental, siendo para los clínicos no especializados un desafío difícil de lograr por la similitud que presentan estas úlceras.

Para los urólogos o médicos no dermatólogos, que ven comparativamente menos casos de estas enfermedades muco-cutáneas, sería útil contar con una herramienta sobre la base de la morfología clínica de las lesiones, que, a través de la utilización de elementos simples y ampliamente disponibles como el celular, pudieran diferenciar entre las diferentes lesiones.

La inteligencia artificial se está convirtiendo en un elemento cada vez más importante. El aprendizaje profundo aprovecha las redes neuronales artificiales, que aprenden asignaciones complejas entre entradas (por ejemplo, imágenes) y salidas (por ejemplo, diagnósticos) sin ingeniería humana explícita. Inspiradas por el cerebro, las neuronas artificiales dispuestas en capas profundas adaptan la fuerza de sus conexiones entre sí a medida que el modelo aprende por sí mismo las características de la entrada, como los patrones visuales, que son más relevantes para predecir la salida (Figura 1 y 2). Es así como la aplicación de estos modelos al análisis de imágenes ha permitido desarrollar redes neuronales que permitan una buena aproximación diagnóstica para diferenciar entre diferentes lesiones (6,7). Existiendo metaanálisis que describen que la capacidad de diferenciar sería comparable o superior a la de un dermatólogo experto (8).

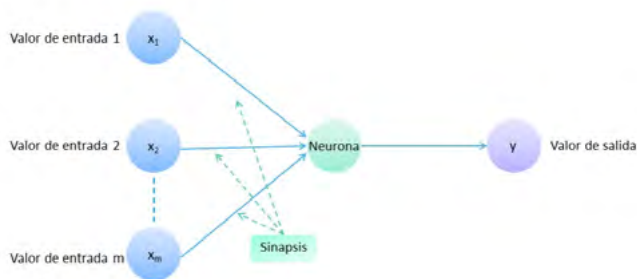


Figura 1. Esquema de unidad fundamental de red neuronal (perceptrón), donde los distintos valores de entrada (variables independientes), interactúan entre sí, como si fueran las dendritas, para finalmente, luego de un proceso algorítmico, dar un valor de salida.

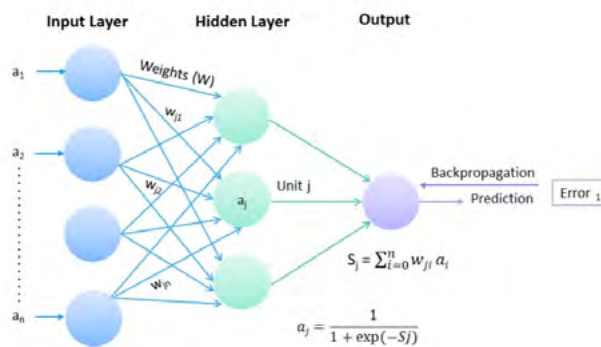


Figura 2. Esquema de red neuronal artificial. Se puede apreciar la complejidad matemática del proceso de aprendizaje, para lograr finalmente una mejor predicción.

El objetivo de nuestro trabajo es mostrar la aplicación y capacidad diagnóstica de un modelo basado en redes neuronales que permite diferenciar entre las lesiones antes mencionadas a través de una fotografía y síntomas del paciente, con el fin de poder mejorar el diagnóstico y tratamiento de estas patologías.

Materiales y métodos

Se utilizó una base de datos de imágenes de los 5 tipos de úlceras genitales más comunes, imágenes extraídas de repositorios virtuales, validadas por Dermatólogos y Urólogos de la Pontificia Universidad Católica de Chile. Se entrenó un modelo de Redes Neuronales Convolucionales (Figura 3 y 4), concatenado con modelos de machine learning para poder diferenciar entre las 5 diferencias tipo de lesiones. El modelo fue desarrollado en un trabajo en conjunto entre urólogos, dermatólogos y el laboratorio XCV del Departamento de Ciencias de la Computación de la Pontificia Universidad Católica de Chile. Finalmente se realizó un estudio de validación utilizando 15 casos clínicos, desconocidos para el modelo, comparando sus resultados con un grupo de médicos generales, del área dermatológica y urológica.

Se realizó estadística descriptiva y analítica mediante modelos de regresión Poisson. Análisis estadístico realizado mediante los Software Python v7.0, SPSS v22.0 y R Studio. Se consideró significativo un valor- $p < 0,05$.

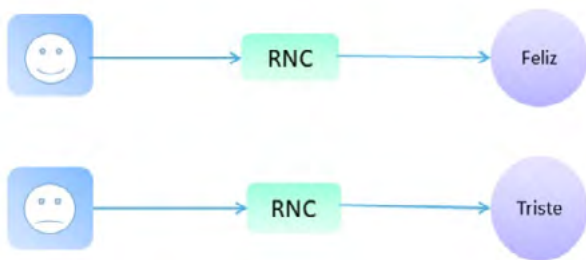


Figura 3. Ejemplo que funcionamiento de Red Neuronal Convolutiva (RNC). En este tipo de redes neuronales, la información de los valores de entrada, tienen relación a los pixeles de la imagen (blanco/negro o color), dando finalmente un valor de salida.

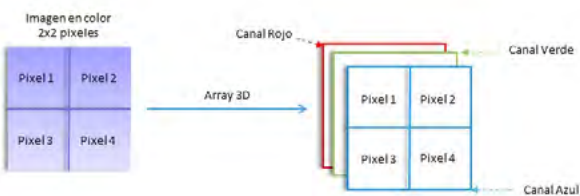


Figura 4. Esquema de valores de entrada de RNC en imagen a color. Se observa que valores de entrada tiene relación a intensidad de pixeles en cada uno de los 3 canales de color RGB.

Resultados

En relación al entrenamiento del modelo de Inteligencia Artificial, se recopilamos y verificamos un total de 346 imágenes de los 5 tipos de úlceras genitales.

Se entrenó una Red Neuronal Convolutiva (RNC) para evaluar el diagnóstico solamente viendo la imagen: 99.3 % de acierto en el conjunto de entrenamiento, 62.5 % de acierto en el conjunto de testing. Se concatenó un algoritmo de Machine Learning para agregar los síntomas a la evaluación de imagen realizada por la RNC: 97.8 % de acierto en el conjunto de entrenamiento, 85.2 % de acierto en el conjunto de testing.

En relación al estudio de validación. Se incluyó un total de 104 médicos generales y especialistas (Urología y Dermatología). Se analizaron 15 casos clínicos: Historia clínica e imagen de lesión. Se realizó test a los 3 grupos de médicos, finalmente se aplicó test a modelo de IA (casos clínicos e imágenes nunca antes utilizadas en entrenamiento de modelo de IA). Las respuestas fueron anónimas.

Al momento de hacer el análisis de regresión (Tabla 1), destaca el hecho que la tasa de acierto del grupo de urólogos no presentó diferencias estadísticamente significativas con el grupo de médicos generales, ajustados por tipo de lesión evaluada. Por el contrario, tanto el grupo de dermatólogos como el modelo de IA mejoraron la tasa de acierto de diagnóstico de úlceras genitales (ajustando por tipo de úlcera), con una mejoría en la tasa de acierto de 29 % y 59 %, respectivamente.

El porcentaje de acierto para el diagnóstico más probable (top-1) fue de 54.6 % , 55.1 % , 70.4 % y 86.7 % para Médicos generales, Urólogos, Dermatólogos y Modelo de IA, respectivamente (p=0.006), **Figura 5.**

Variables	Multivariate	
	RR [95% CI]	P-value
Work area:		
General practitioner	0*	-
Urology	1.01 [0.84 – 1.22]	0.931
Dermatology	1.29 [1.12 – 1.48]	<0.001
AI model	1.59 [1.14 – 2.21]	0.006
Disease:		
Chancre (Syphilis)	0*	-
Chancroid	0.62 [0.51 – 0.75]	<0.001
Donovanosis	0.26 [0.19 – 0.33]	<0.001
Genital herpes	1.18 [1.00 – 1.39]	0.049
LG Venereum	0.74 [0.62 – 0.89]	0.002

Tabla 1. Análisis multivariado (Regresión de Poisson) para predicción de tasa de acierto en el diagnóstico de úlceras genitales

Mientras que, para los 2 diagnósticos más probables, el porcentaje de acierto fue de 71.4 % , 70.9 % , 81.5 % y 100 % para Médicos generales, Urólogos, Dermatólogos y Modelo de IA, respectivamente (p=0.031), **Figura 6.**

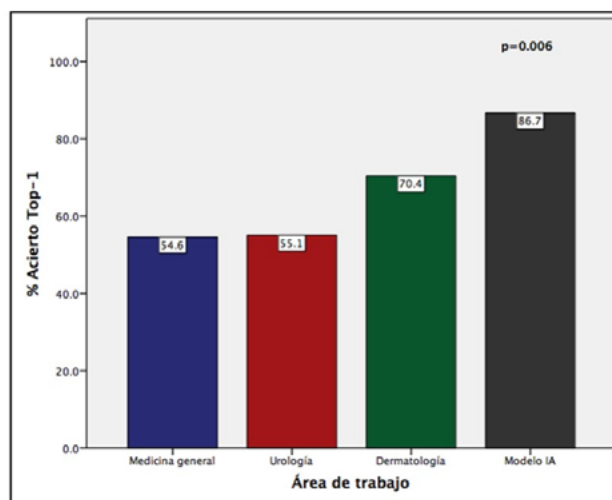


Figura 5. Tasa de acierto al diagnóstico más probable (top-1) según área médica de trabajo.

Discusión

El desarrollo de la Inteligencia Artificial (IA) en los últimos años ha permitido su difusión en distintos ámbitos de la medicina. La IA permite crear herramientas basadas en la construcción de redes neuronales, capaces de realizar procesos de aprendizaje automático, en inglés Machine Learning, mediante análisis multifactoriales pudiendo clasificar información, ya sea mediante imágenes o datos clínicos, siempre y cuando estos sean conocidos por la red neuronal. Existen diferentes técnicas para el entrenamiento de la red neuronal,

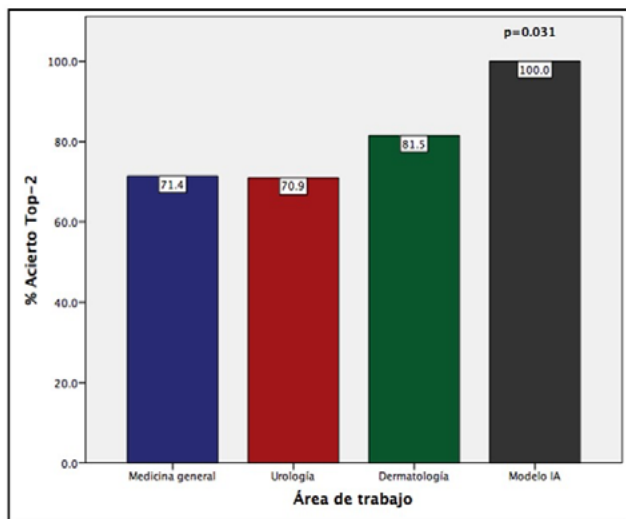


Figura 6. Tasa de acierto a los 2 diagnósticos más probable (top-2, que alguno de los 2 diagnósticos sea el correcto) según área médica de trabajo.

siendo una de las más utilizadas la retropropagación, en la cual existe una primera Fase de Entrenamiento en la que se compara el resultado obtenido por el modelo de IA con el resultado deseado, calculando una tasa de error que se reduce mediante modificaciones en la red. Este proceso se repite hasta obtener un porcentaje de error deseado, luego del cual se procede a un Proceso de Validación mediante el ingreso de nuevos datos de entrada (previamente desconocidos por la red) que permiten objetivar el aprendizaje de la red (10,11).

Hasta la fecha el uso de esta tecnología se ha implementado en diversas áreas de la medicina; con respecto a urología, en Uro-oncología se ha utilizado mediante el desarrollo de modelos para predecir outcomes post prostatectomía radical en Cáncer de Próstata, estratificación de pacientes con carcinoma urotelial, análisis de imágenes de seguimiento, entre otros. Con aplicaciones no oncológicas en urolitiasis, trasplante renal, infecciones del tracto urinario y urología funcional (9, 11). Respecto al impacto de uso, una revisión sistemática encontró que un 71.8 % de los estudios incluidos concluye que la IA es superior en diagnóstico y predicción de resultados en comparación a la estadística tradicional. En cuanto al diagnóstico por imágenes se ha demostrado que mediante la asistencia de un modelo de inteligencia artificial para el análisis de biopsias prostáticas la sensibilidad del patólogo aumenta significativamente de un 74 a un 90 %, sin disminución significativa en la especificidad, con resultados similares en cáncer renal. Se han desarrollado otros modelos que mediante hallazgos radiológicos en RM prostáticas predicen un score de Gleason para lesiones sospechosas, los cuales al ser comparados con los resultados obtenidos por biopsias han mostrado un área bajo la curva de 0.72 para predecir Gleason ≥ 8 (9,10). Sin embargo, la mayoría de los modelos fueron desarrollados con una sola entrada de datos (imágenes), lo que podría disminuir su eficacia respecto a nuestro modelo.

Finalmente, existen reportes sobre el desarrollo de modelos de inteligencia artificial para el diagnóstico ITU que incorporan datos de la anamnesis, examen físico y análisis de orina con una tasa de precisión de un 98.3 % (10,11). Hasta el momento no se han encontrado estudios sobre la existencia de modelos para identificar úlceras genitales mediante el uso de imágenes y datos clínicos, por lo cual nuestro trabajo sería pionero en dicha área.

Conclusión

Presentamos la validación de un modelo de IA para el diagnóstico de úlceras genitales, sobre el cual no existen reportes disponibles en la literatura. El modelo de IA muestra diferencias significativas en la probabilidad de obtener un diagnóstico adecuado, siendo superior incluso a la de los médicos especialistas, gracias al proceso de aprendizaje desarrollado en redes neuronales.

Bibliografía

1. Low N, Broutet N, Adu-Sarkodie Y, Barton P, Hossain M, Hawkes S. Global control of sexually transmitted infections. *Lancet*. 2006;368(9551):2001-2016.
2. Centers for Disease Control and Prevention. 2010 STD Treatment Guidelines. 2010. <http://www.cdc.gov/std/treatment/2010/>
3. Gabrielson A, Le T, Fontenot C, et al. Male Genital Dermatology: A Primer for the Sexual Medicine Physician. *Sex Med Rev* 2018;1e13. <https://doi.org/10.1016/j.sxmr.2018.09.004>
4. Sobhinf. Informe: Situación epidemiológica de las infecciones de transmisión sexual en Chile, 2017. *Rev Chilena Infectol* 2019; 36 (2): 221-233
5. Dick V, Sinz C, Mittlböck M, Kittler H, Tschandl P. Accuracy of computer-aided diagnosis of melanoma: a meta-analysis. *JAMA Dermatol* 2019;155:1291e9.
6. Roett M, Mayor M, Uduhiri K. Diagnosis and Management of Genital Ulcers. *American Family Physician*. February 1, 2012, Volume 85, Number 3
7. Young A, Xiong M, Pfau J. Artificial Intelligence in Dermatology: A Primer. *Journal of Investigative Dermatology* (2020), Volume
8. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Heal* 2019a;1:e271e97.
9. Suarez-Ibarrola R, Hein S, Reis G, et al. Current and future applications of machine and deep learning in urology: a review of the literature on urolithiasis, renal cell carcinoma, and bladder and prostate cancer, *World Journal of Urology* 2020 38:2329-2347. <https://doi.org/10.1007/s00345-019-03000-5>
10. Chen J, Remulla D, H.Nguyen J, et al. Current status of artificial intelligence applications in urology and their potential to influence clinical practice, *BJU International* 2019; 124: 567-57, DOI:10.1111/bju.14852. PMID: 31219658.
11. Shah M, Naik N, K. Somani B, et al. Inteligencia artificial (IA) en urología: uso actual y direcciones futuras: un estudio iTRUE, *Turk Journal of Urology* 2020 Nov; 46 (Supl. 1): S27 - S39. PMC7731952