Algoritmo basado en aprendizaje de máquina para detección de cáncer de piel

Machine learning based algorithm for skin cancer detection

Verónica Angélica Villalobos Romo[a], Luis Carlos Méndez González[a]*, José David Díaz Román[b]

[a]Departamento de Industrial y Manufactura, Maestría en Tecnología, Universidad Autónoma de Ciudad Juárez, México

[b]Departamento de Eléctrica y Computación, Maestría en Ingeniería Eléctrica, Universidad Autónoma de Ciudad Juárez, México.

*Autor de correspondencia. Correo: luis.mendez@uacj.mx

No. de resumen
2CP21-76

Formato
Ponencia

Evento
2.º Coloquio de Posgrados IIT

Presentador
Verónica Angélica Villalobos Romo

Tema
Procesos Tecnológicos

Estatus
Estudio en curso

Fecha de la presentación
Noviembre 12, 2021

RESUMEN

El cáncer de piel es el segundo más frecuente en México desde 2001, afirmó el director de la Clínica Oncodermatológica de la Facultad de Medicina de la UNAM en el año 2017. Por lo cual, desarrollar un algoritmo basado en herramientas de inteligencia artificial para el diagnóstico de cáncer de piel es una excelente opción para la detección temprana, ya que es crítica para el tratamiento del cáncer tipo melanoma, que es el más agresivo. Las redes neuronales artificiales y el aprendizaje de máquina son a menudo utilizados en el reconocimiento y clasificación de imágenes. Para desarrollar el algoritmo para detección de cáncer de piel se utilizó la base de datos HAM10000, que cuenta con 10 015 imágenes de casos que conforman una colección representativa de todas las categorías de diagnóstico importantes en el ámbito de las lesiones pigmentadas. Las características de este conjunto de imágenes son: tamaño de 600x450 pixeles, color RGB y formato JPG. La organización de las imágenes consistió en separarlas por categorías en carpetas por medio de un programa realizado en Python 3.8, el cual obtiene la información detallada del archivo HAM10000_metadata.csv y guarda cada imagen del conjunto en la carpeta correspondiente a su categoría. Con esto, crear los conjuntos de datos entrenamiento, de validación y de prueba con los cuales permiten la construcción del modelo para el diseño del algoritmo y el entrenamiento de la red neuronal que realiza la clasificación de las lesiones.

Palabras clave: cáncer de piel; aprendizaje de máquina; aprendizaje profundo; redes neuronales.

ABSTRACT

Skin cancer is the second most frequent in Mexico since 2001, stated the director of the Oncodermatology Clinic of the UNAM Faculty of Medicine in 2017. Therefore, developing an algorithm based on artificial intelligence tools for the diagnosis of skin cancer is an excellent option for early detection as it is essential to treat melanoma cancer, which is the most aggressive. Artificial neural networks and machine learning are often used in image recognition and classification. The HAM10000 database was used for algorithm for skin cancer detection, which has 10,015 images of cases that make up a representative collection of all the critical diagnostic categories in the field of pigmented lesions. The characteristics of this set of images are 600x450 pixel size, RGB color and JPG format. The organization of the images consisted of separating the images into categories and folders using a program made in Python 3.8, which obtains the detailed information from the HAM10000_metadata.CSV file and saves each image in the set in the folder corresponding to its category. With this, the training, validation and test data sets are created with which they allow the construction of the model, the design of the algorithm and the training of the neural network that performs the classification of skin cancer lesions.

Keywords: skin cancer, machine learning, deep learning, neural networks.

Entidad legal responsable del estudio

Universidad Autónoma de Ciudad Juárez.

Financiamiento

CONACYT, CVU: 272489.

Conflictos de interés

Los autores declaran que no existe conflicto de intereses.