

DEEP LEARNING FOR CLINICAL IMAGE CLASSIFICATION OF GENITAL LESIONS CAUSED BY SEXUALLY TRANSMITTED DISEASES

APRENDIZAJE PROFUNDO PARA LA CLASIFICACIÓN DE IMÁGENES CLÍNICAS DE LESIONES GENITALES CAUSADAS POR ENFERMEDADES DE TRANSMISIÓN SEXUAL

Raquel González-Alday¹; Francois Peinado²; Daniel Carrillo¹; Víctor Maojo¹

1. Biomedical Informatics Group, Dpt of Artificial Intelligence, Universidad Politécnica de Madrid

2. Jefe de Urología. Hospital Ruber Juan Bravo 39, Madrid

Keywords:

Artificial Intelligence; Sexually Transmitted Diseases; Clinical Image Classification.

Palabras clave:

Inteligencia Artificial; Enfermedades de Transmisión Sexual; Clasificación de Imágenes Clínicas.

Abstract

Sexually transmitted diseases (STDs) are one of the world's major health emergencies. Given its incidence and prevalence, particularly in developing countries, it is necessary to find new methods for early diagnosis and treatment. However, this can be complicated in geographical areas where medical care is limited. In this article, we present the basis of a deep learning-based system for image classification of genital lesions caused by these diseases, built using a convolutional neural network model and methods such as transfer learning and data augmentation. In addition, an explainability method (GradCam) is employed to enhance the interpretability of the obtained results. Finally, we developed a web framework to facilitate additional data collection and annotation. This work aims to be a starting point, a "proof of concept" to test various different approaches, for the development of more robust and trustworthy Artificial Intelligence approaches for medical care in STDs, which could substantially improve medical assistance in the near future, particularly in developing regions.

Resumen

Las enfermedades de transmisión sexual (ETS) son una de las mayores emergencias de salud a nivel mundial. Debido a su incidencia y prevalencia, particularmente en países en desarrollo, es necesario encontrar nuevos métodos para diagnóstico y tratamiento precoz. Sin embargo, esto puede ser complicado en áreas geográficas en las que la asistencia médica es limitada. En este artículo, presentamos una prueba de concepto de un sistema basado en aprendizaje profundo para la clasificación de imágenes de lesiones genitales causadas por estas enfermedades, construido utilizando un modelo de red neuronal convolucional y métodos como *transfer learning* y *data augmentation*. Además, incorpora un método de explicabilidad (GradCam) para mejorar la interpretabilidad de los resultados obtenidos, y se ha desarrollado un servicio web para facilitar la recogida de datos adicionales y su anotación. Este trabajo pretende ser un punto de partida y una prueba de concepto para valorar diferentes enfoques en el desarrollo de modelos de Inteligencia Artificial más robustos y fiables para la asistencia médica en ETS, que podría mejorar sustancialmente la asistencia médica en un futuro próximo, particularmente en regiones en desarrollo.

1. INTRODUCTION

1.1. Sexually transmitted diseases: prevalence and lesions

Sexually transmitted diseases (STDs), also referred as sexually transmitted infections (STIs), are one of the biggest health issues worldwide, with nearly 1 million estimated people becoming infected every

day, according to the WHO (1). This is a major problem in many countries. At the Universidad Politécnica de Madrid, we addressed this problem in a European Commission-funded project, called Africa Build (2010-2013), which we coordinated with partners such as WHO, various European institutions and colleagues from Ghana, Mali, Egypt and Cameroon. We carried out various initiatives such as teaching how to use medical informatics methods and tools, as well as distance learning in various topics, which included STDs.

Autor para la correspondencia

Víctor Maojo

Universidad Politécnica de Madrid

Escuela Técnica Superior de Ingenieros Informáticos. Campus de Montegancedo UPM

Tlf.: +34 910 672 898 | E-Mail: vmaojog@gmail.com

The burden of these diseases is most notable in low and middle-income countries. However, the health, social and economic problems caused by the high prevalence of STDs are a significant concern all around the world (2,3), and even more in the last few years, with higher rates of transmission, emerging outbreaks and antimicrobial resistances increasing (4).

Diagnosis of these diseases is usually done through a physical examination —either by a general practitioner or a specialist—, with further tests if needed and available. While in many cases STDs might be asymptomatic, common symptoms include abnormal discharge, lower abdominal pain and genital skin lesions such as ulcers, warts or lumps.

A correct and early diagnosis is vital for the proper treatment and information of the patient, and therefore to avoid transmission (3). However, it can be not so obvious for a non-specialist health professional such as a general practitioner or a nurse, or even more complicated in cases when access to healthcare is difficult or when the patient is reluctant to seek medical attention due to social stigma —e.g., social, economic, or religious.

In this article we present a prototype of an Artificial Intelligence (AI)-based approach: a deep learning model to automatically classify images of different genital manifestations of STDs. This kind of tool and digital environment could be used to build systems to support medical diagnosis in these developing countries, aiming to create new ways to help non-specialist health providers as well as, in the near or mid future, the patients themselves by providing them with an accurate, easy-to-use medical decision support system. In this study, we decided to choose two kinds of typical STD lesions that will be classified in: typical genital herpes ulcerative lesions as well as genital warts and condylomas, which are usually caused by human papillomaviruses (HPV). These are very common problems within the area of STDs.

1.2. Genital herpes

Genital herpes is a very common STD caused by either the herpes simplex virus type 1 (HSV-1) or type 2 (HSV-2). HSV infections are spread through contact with herpes lesions, mucosal surfaces, genital or oral secretions from an infected individual. Most infections are asymptomatic or have very mild symptoms that go unnoticed or that can be easily confused as another skin condition (5).

For that reason, even though the highest risk of transmission of HSV happens in periods when visible lesions are present, it is common to become infected by contact with an asymptomatic partner without visible manifestations who might not be aware of their infection.

In the case that symptoms do manifest, herpes lesions usually consist of one or more vesicles or small blisters that appear on the genital, rectum or

mouth area. These vesicles break and leave painful ulcers that can take two to four weeks to heal after the initial infection, and can be accompanied by other symptoms like fever, body aches, swollen lymph nodes, or headache.

Recurrent episodes of genital herpes are common, especially for HSV-2 infections, and are usually preceded by prodromal symptoms such as localized pain or tingling. However, recurrent herpes outbreaks tend to be milder than the first one, which is often associated with a longer duration of herpetic lesions and systemic symptoms, as well as increased viral shedding that increases the risk of transmission (6).

1.3. Human papillomavirus

Human papilloma virus (HPV) is one of the most frequently sexually transmitted viruses. In fact HPV is so widespread that, if unvaccinated, almost every sexually active person will get HPV at some point in their lifetime (7).

This virus is usually transmitted by direct skin-to-skin or mucosal contact, commonly via vaginal, oral or anal sex. The virus is usually cleared out in a period of about two years without causing health issues, however, in other cases HPV can cause significant problems. These include genital warts, which are small bumps in the anogenital area, or condylomas, which are more raised, cauliflower-like growths, as well as other complications such as cervical, penile, anal and oropharyngeal cancer.

There are HPV tests that can screen for cervical cancer, used for women above 30 years of age, however, there is currently no approved test for HPV in men, nor is it recommended for younger women, so in order to diagnose it, a physical examination to identify HPV lesions should be carried out, with an additional biopsy in if the diagnosis is uncertain or if cancer is suspected.

1.4. Deep learning in medicine

While AI has been applied to medicine since the 1970s, the recent emphasis on data-driven techniques in the area of machine learning have shown the substantial AI possibilities in a wide variety of fields, with countless applications. Its capacity to facilitate —and even automate— many types of tasks, with a performance sometimes surpassing human performance, has transformed many areas of medicine. If AI was focused in the period from 1970 to 1990 in knowledge-based systems — with the clear example of expert systems—, since the 1990s the predominance of AI in medicine has been transferred to data-based systems. The recent explosion of medical data available from many sources has fueled such a predominance. In this context, a technique called deep learning (8) — whose roots can be traced back to the 1940s, with

the design of the so called “artificial neurons”—has come promising to revolutionize issues such as medical diagnosis, therapy, patient management, prognosis or public health and surveillance systems, among others.

Deep learning is a branch of Machine Learning that encompasses algorithms and models that allow computers to automatically learn complex patterns and relationships from data, thanks to their architectures indirectly inspired by neuroscience, which are composed of multiple stacked processing layers (9), each one containing artificial “neurons”, which are basically simulated processors that can carry out mathematical processes. The most popular example of this area is deep neural networks.

Neural networks can take as an input different kinds of data, like values for several variables, images, text or audio, for example. Data is passed through the various layers of artificial neurons, and inside each of these processing units, the input data (which originally could be the value of a pixel in an image, signals, or clinical data if we are trying to classify or predict a disease state as an outcome) is pondered using a weight. The final aggregation of the serial operations made by the network is used to calculate the probability of pertinence to a class, using a mathematical function.

The training process of these networks consists of modifying the weights for each of the artificial neurons that lead to the best classification/prediction accuracy. To do so, roughly, in one of the most classical examples in the area, a set of data is used to iteratively test the network and actualize the weights by back propagating the derivative of the error obtained comparing the actual class of each instance and the prediction of the model.

The development of these methods has allowed computers to perform with great accuracy tasks such as, to mention two common cases, object detection or speech recognition, which can seem easy for humans but have been difficult to formalize by computers (10,11).

In this paper, we focus our work on a particular deep learning technique for computer vision, convolutional neural networks (CNNs). We base our approach in this technique to build an image analysis system that can aid clinical diagnosis in STDs.

2. METHODS

2.1. Image dataset handling

The main requirement of deep learning models is a dataset of good quality and sufficient size to be trained on. For this work, we had available a series of images of both male and female external genitalia and perianal area with visible lesions of three types: herpes, warts and condylomas.

Before being able to train and test the classification model of choice, some preprocessing was done: all the images were labeled with the corresponding condition, and as they were taken from varied angles, lighting and positions, they were manually cropped so that the lesions were more or less centered, and they were also normalized and rescaled to a standard size. The result of this process were images of 224x224 pixels and 3 color channels with values scaled between 0 and 1.

The dataset contained 261 images in total, of which 42 belonged to the herpes class, 34 to genital warts and 185 to condylomas.

2.2. Deep learning model

As mentioned before, the deep learning model used for this work is a CNN. However, as is the case with any other deep learning techniques, this model acts like a “black-box”, it classifies images but does not provide any direct explanation for its performance. For that reason, we decided to explore the use of a technique from the field of Explainable Artificial Intelligence (XAI) that helps the assessment of the model’s functioning by producing visual explanations.

2.2.1 Convolutional Neural Network

CNNs are a particular kind of neural network that are particularly adequate for computer vision tasks. They are inspired by the vision processes in living beings -its goal is not to imitate nature but to optimize image processing by computers-, as they consist of processing units that compute the input of a set of pixels (by performing a mathematical operation called convolution), the same way as animal cells in the visual cortex detect light in receptive fields (12). They receive their name from the basic mathematical operation performed by the processing units: convolutions.

These processing units are grouped in stacked layers (see Figure 1), with different architectures depending on each case, that end up learning different features, invariant to position, that lead to the final classification. This means, for example, that a layer inside of the whole network might learn to detect edges, specific shapes or more abstract features that will then be detected in other images (13). Specifically, the network architecture used for this prototype system is the state-of-the-art EfficientNet (14), with some minor adjustments to properly fit the image dataset available.

Given the limited quantity of data available for the network training, transfer learning was used. This method consists of using a network previously trained on another large dataset, unrelated to the problem at hand (in this case, the ImageNet object recognition dataset (15)), and then retraining the network with the desired data. Thus, the generic deep features learned by the neural network are exploited, and the model is then refined for the problem at hand, achieving better performance than if it was training only on the small dataset (16).

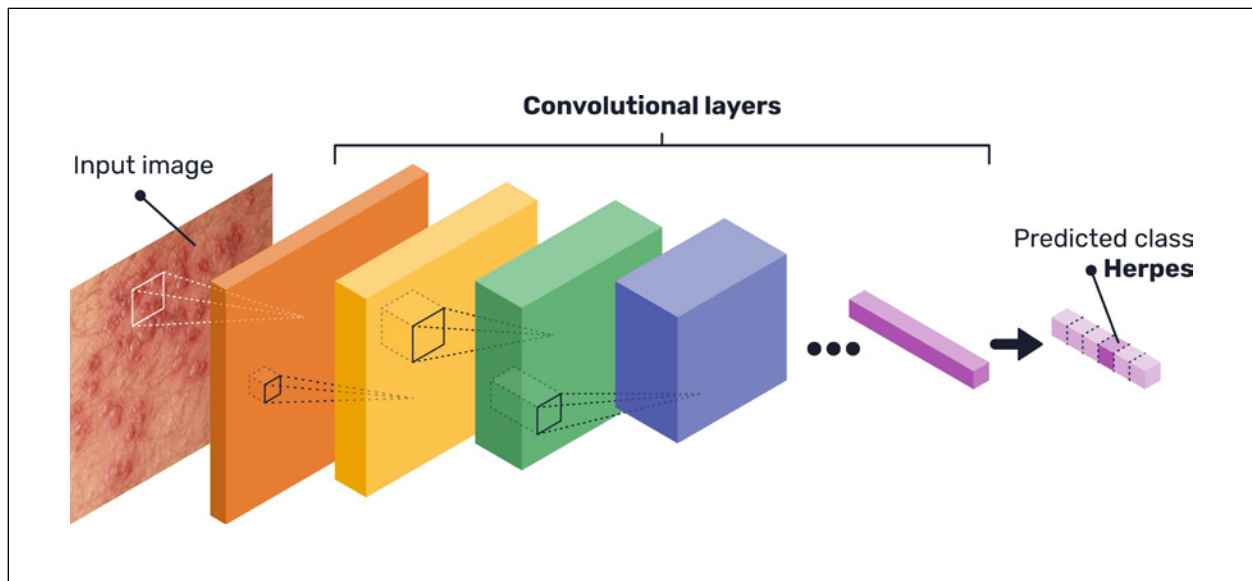


Figure 1. Schematic representation of a convolutional neural network.

In this case, given the scarcity of the data, we carried out a k-fold cross-validation process, with 5 folds. The dataset images were randomly distributed in 5 sections, and having the pre-trained model, 4 of these sections were used for re-training the network while the remaining one was saved for testing. This process was repeated 5 times, rotating the sections.

Moreover, during the training of the CNN, data augmentation was performed, as another way to palliate the problems of having limited data. This means that some random subtle transformations were made to the training images (tweaks in color or size, rotations, translations...), so that they differ every time and the model does not suffer from overfitting (17).

Presenting varying images instead of the same images over and over again helps the model to learn to ignore differences caused by illumination or perspective and focus on the important features that define the class shown on the image. For this work in particular, we tweaked the image orientation and included random shifts in height and width.

2.2.2 Explainability methods

To enhance the interpretability of the CNN, and therefore enable a better evaluation of its functioning, a post-hoc explainability method was used. Techniques of this type, which fall within the field of XAI, allow explanations to be obtained for the predictions of a model by either visualizing the inner workings of it, creating surrogate simplified models that are more understandable or fiddling with the inputs and predictions to identify important features (18).

For CNNs and other image analysis models, the most popular methods currently are those that produce

different kinds of heat maps showing what are the parts of the image that are more influential in the final outcome of the network, or where the network's attention is focused on.

In particular, the method of choice for this work is GradCam (19), a technique that visualizes the last convolutional layer of the CNN. Usually, while the first layers of these networks recognise simple features, the deepest layers identify the most complex concepts, and the final one is the one that directly leads to the final classification, by transferring its values to a traditional shallow neural network classifier. Moreover, the artificial neurons' position inside the convolutional layers is directly related to the original position of pixels in the image.

These two properties imply that, by visualizing the resulting values of the artificial neurons' operations of the last convolutional layer, a heatmap can be produced indicating the regions of the image that are important to predict the chosen class. To do so, the final classifier loss for a particular class is differentiated with respect to these raw values from the convolutional layer to obtain a series of importance weights. These weights are used to ponderate and pool these same values from the CNN last layer and obtain a gradient map. In this map, specific to the chosen class, higher gradients mean more influence on that prediction.

2.3. Data collection and annotation web tool

As an addition to the deep learning model, it was decided to develop a web tool to enable further data collection for this project. This tool consists in a web service that can be accessed through a web page where, after registering as an user, clinicians can upload images and label with the respective condition shown and additional data if necessary.

Using HTTPS, a secure transference protocol, we ensure a secure channel between client and server, ensuring the confidentiality of the communication, as well as the integrity of the data, detecting and avoiding confusion errors in transmission. Moreover, a web service offers the possibility of easily including numerous upgrades if needed in the future, such as automating image sending, further security functionalities, time logs and so on. This web service includes user authentication, so users need to be registered with an e-mail address, username and password, to access their particular profile where they can upload images, annotate and manage them.

Besides specifying the image class, users can access the web page interface to draw bounding boxes over the images to properly locate the lesions, making future preprocessing work easier, or to enable the use of CNN architectures for object detection and not only classification. Moreover, they can add additional patient symptoms or notes that could be of help for the model's development.

After uploading the images, they can be collected by the developing team to expand the database used to train the deep learning model.

3. RESULTS

The training process of the CNN model consisted of 100 iterations (called epochs) over the training dataset images (with the data augmentation transformations described before) and starting from the pre-trained network weights. This training process was manually supervised, ensuring that the accuracy consistently improved and no weird oscillations occurred.

As mentioned above, for carrying out the 5-folds cross-validation, this process was repeated 5 times rotating the training and testing images. Then, we evaluated the model's performance over the complete dataset as well as its generalization capabilities, as the testing was always made on images that were "unseen" during training.

The final testing result was an accuracy of **86.6%** after conducting the whole cross-validation of the model. Figure 2 shows the confusion matrix for the different classes, to better visualize in which cases the CNN model is getting its predictions right and in which not.

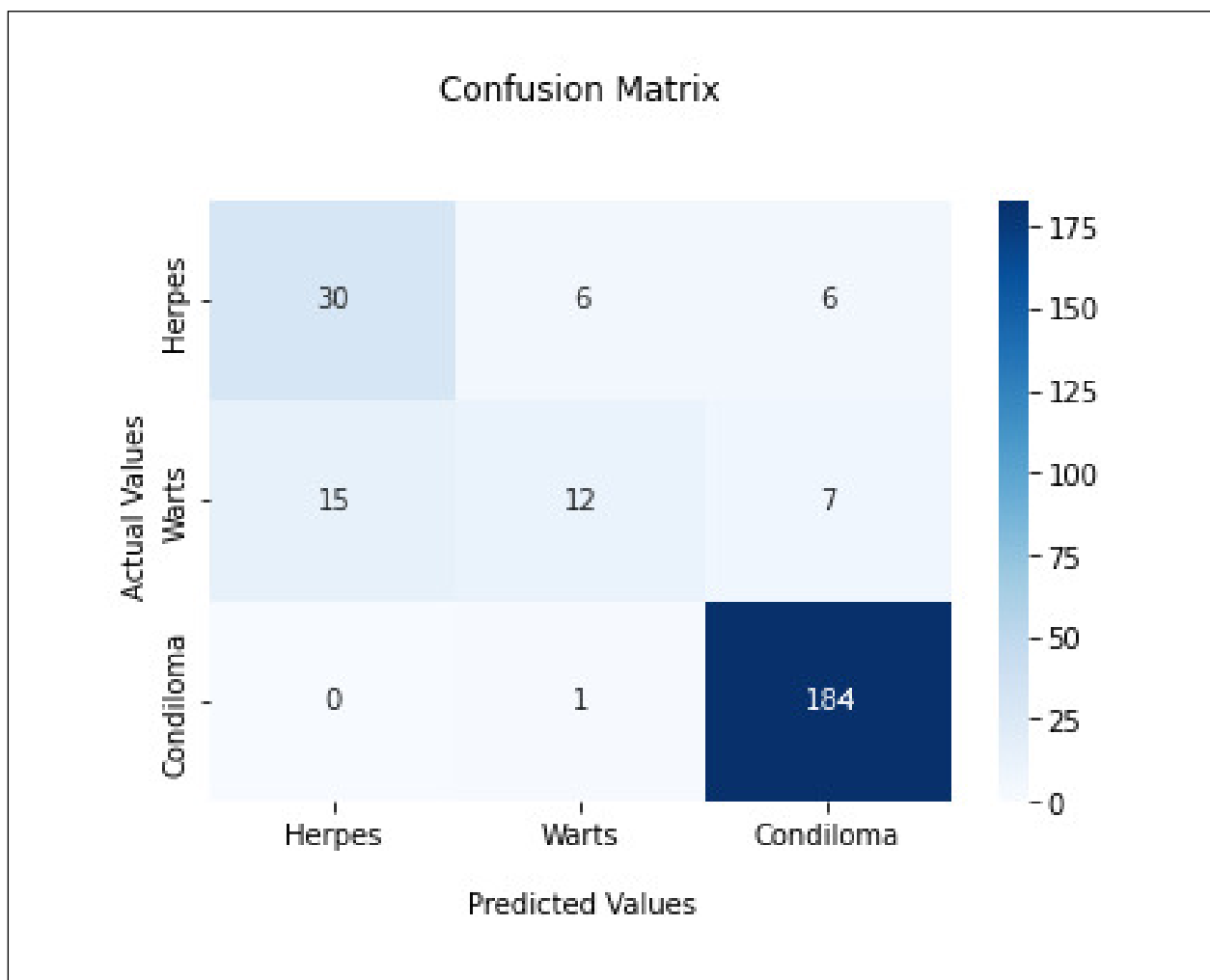


Figure 2. Confusion matrix of the model's results during testing.

Figure 3 shows some examples of the GradCam explanations produced for the classified instances, some of them from accurate predictions and others from incorrect predictions, in order to illustrate how

these heatmaps can help assess the model's functioning. Warmer colors in these maps indicate higher gradients and therefore more influence of those regions on the final model's prediction.

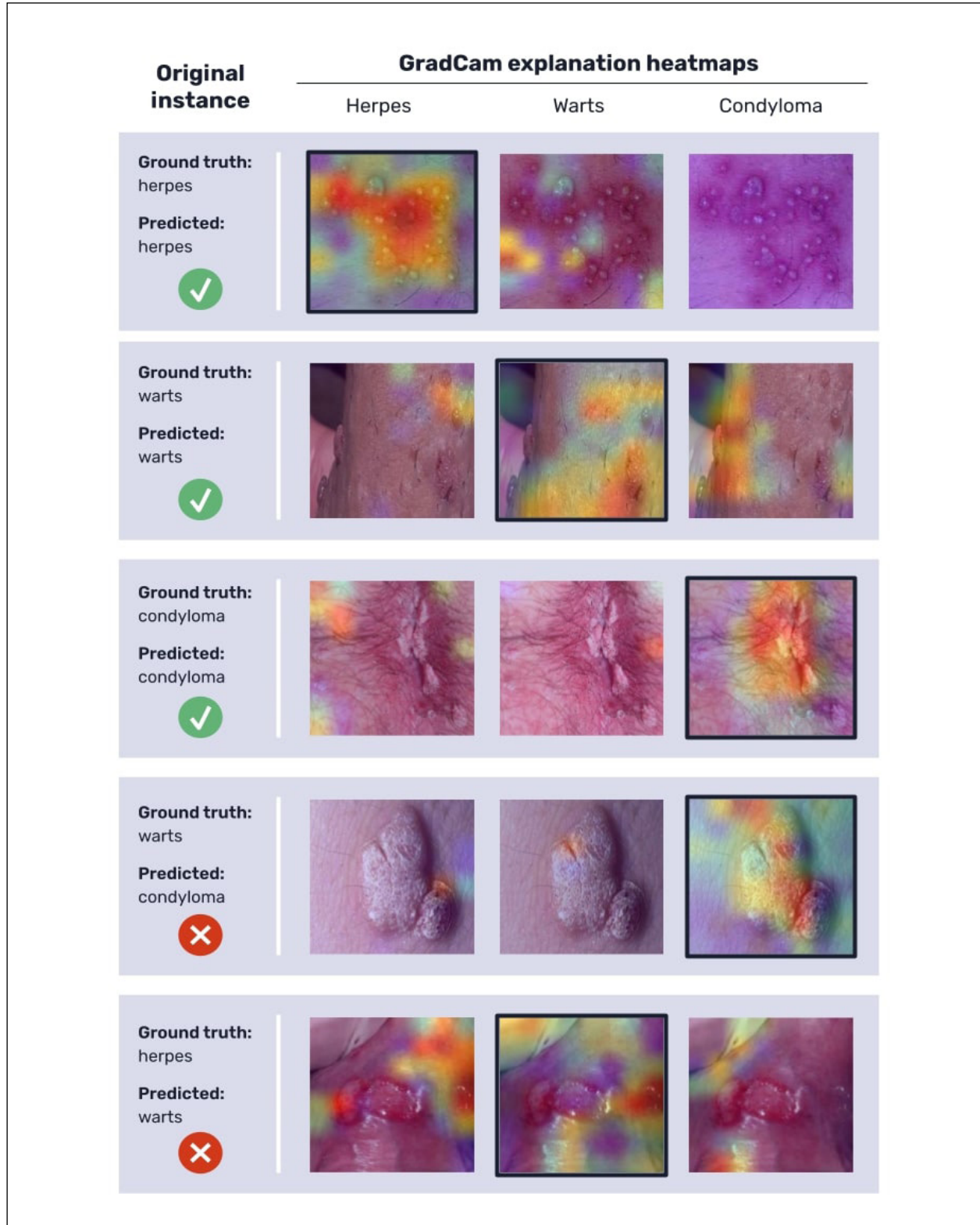


Figure 3. Produced GradCam explanations for some examples during testing. Outlined images remark the predicted class in each case.)

4. DISCUSSION

4.1. Model's results and performance

The overall accuracy is highly positive, although we must comment various conditions. In particular, the model classifies correctly all the instances of condylomas, the most represented class in the dataset, but has some errors in the classification of herpes and warts. This particular result is a direct consequence of the relatively low number of examples in the used image dataset.

On the one hand, the produced GradCam explanations (figure 3) for correctly classified instances confirm that the network's attention is focused correctly, as higher gradients are located over the lesions.

On the other hand, for wrongly classified instances, the heatmaps can also provide some interesting insights. In some cases, some lesions such as warts and condylomas can be confused due to their similar appearance (in fact, condylomas and warts essentially differ in size, and some herpes vesicles can be easily mistaken for wart-like lumps), and therefore the explanations seem reasonable. But in other cases —the most common among wrongly classified instances—, the error is the result of an erroneous focus process of the network, which can be detected in heatmaps, when the network's gradients are not focused on the lesions (see figure 3).

These results corroborate that, most likely, if more images were available, especially for the herpes and wart classes (that only had around 30 to 40 images, which is very few for a deep learning model), the model should improve the learning process for classification and then the model's performance would be significantly better.

We consider that the number of images of the available dataset was sufficient for the proof-of-concept presented in this article, suggesting that Deep Learning can be a successful approach for early diagnosis and prevention in these STDs, particularly for places where specialists are scarce. We can expect that a larger dataset could improve robustness, performance and generalization of the trained model, with a more balanced distribution between classes and adding other kinds of lesions such as ulcers. The usually large datasets needed for many Deep Learning applications is a major limiting factor, particularly in medicine. In addition, external datasets from different hospitals and countries should be required for a comprehensive evaluation before a tool of this kind could be definitively introduced into clinical practice.

In developing countries —e.g., in subsaharian Africa— the deployment of AI-based tools might be a dramatic challenge to improve medical care in places where specialists and advanced technology are scarce.

It should be noted also the promise of including explainability methods —such as GradCam in this

case—, in these AI-based tools. While explainability has been a major limitation of most AI systems in medicine already since the 1970s, recent advances might facilitate an improved, detailed analysis of the model, facilitating to identify when and how the system fails, instead of behaving as, for instance, the classical black-box characteristic of machine learning systems. Such a feature will be finally necessary to ensure the acceptance of medical professionals.

5. CONCLUSIONS

Artificial Intelligence and deep learning models in particular have a great potential to improve and transform healthcare, as well as to bring it closer to places where medical expertise is difficult to access. Under this scope, the work presented in this paper shows a promising starting point to implement a CNN for assisting the diagnosis of genital lesions caused by STDs.

The presented prototype aimed to show various alternatives already available in AI-based medical imaging applications. The presented system shows such possibilities with a limited image dataset with a few common and well-differentiated genital lesions. These performance results are encouraging and confirm the feasibility and promise of developing a robust application using the same methods while relying on a larger variety of images.

The explainability method included in this prototype shows the potential capabilities of these tools. While explainability has been a major limitation of most AI systems in medicine already since the 1970s, recent advances facilitates the detailed analysis of the model, allowing to identify when and how the system fails, instead of behaving as the classical black-box characteristic of machine learning systems. This way, the trustworthiness of the model, regarding both its development and application, can be notably increased and might facilitate their acceptance by health professionals, which has been a traditional drawback of many AI-based systems.

ACKNOWLEDGEMENTS

STD Images were provided by Dr. Francois Peinado, urologist —a coauthor of this manuscript— and Dr. Álvaro Vives, urologist at Fundació Puigvert, Barcelona, as part of a joint collaboration for a future enhanced development. This work was partially supported by the *Proyecto colaborativo de integración de datos genómicos (CICLOGEN)* (No. PI17/01561), funded by the Carlos III Health Institute from the Spanish National Plan for Scientific and Technical Research and Innovation 2017-2020 and the European Regional Development Fund (FEDER).

CONFLICT OF INTEREST STATEMENT

The authors of this article declare that they have no conflict of interest with respect to what is expressed in this work.

BIBLIOGRAPHY

1. World Health Organization. Sexually transmitted infections (STIs). Fact Sheets [Internet]. 2019 [cited 2022 Jun 20]; Available from: [https://www.who.int/news-room/fact-sheets/detail/sexually-transmitted-infections-\(stis\)](https://www.who.int/news-room/fact-sheets/detail/sexually-transmitted-infections-(stis))
2. Chesson HW, Mayaud P, Aral SO. Sexually transmitted infections: Impact and cost-effectiveness of prevention. En: Holmes KK, Bertozzi S, Bloom BR, Jha P, editors. Major infectious diseases. 3rd ed. Washington (DC): The International Bank for Reconstruction and Development / The World Bank; 2017.
3. Unemo M, Bradshaw CS, Hocking JS et al. Sexually transmitted infections: challenges ahead. *Lancet Infect Dis.* 2017; 17(8): e235-279.
4. Williamson DA, Chen MY. Emerging and re-emerging sexually transmitted infections. *New Engl J Med.* 2020; 382(21): 2023-2032.
5. Centers for disease control and prevention. Std Fact Sheets - Genital herpes [Internet]. 2022 [cited 2022 Jul 11]. Available from: <https://www.cdc.gov/std/herpes/stdfact-herpes.htm>
6. Kimberlin DW, Rouse DJ. Genital herpes. *New Engl J Med.* 2004; 350(19): 1970-1977.
7. Centers for disease control and prevention. Std Fact Sheets - Human papillomavirus (HPV) [Internet]. 2022 [cited 2022 Jul 11]. Available from: <https://www.cdc.gov/std/hpv/stdfact-hpv.htm>
8. Sejnowski TJ. The deep learning revolution. MIT press; 2018.
9. LeCun Y, Bengio Y, Hinton G. Deep learning. *nature.* 2015; 521(7553): 436-444.
10. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press; 2016.
11. Esteva A, Robicquet A, Ramsundar B et al. A guide to deep learning in healthcare. *Nat Med.* 2019; 25(1): 24-29.
12. Gu J, Wang Z, Kuen J et al. Recent advances in convolutional neural networks. *Pattern Recogn.* 2018; 77: 354-377.
13. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. En: 2017 international conference on engineering and technology (ICET). IEEE; 2017. p. 1-6.
14. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. En: International conference on machine learning. PMLR. 2019. p. 6105-6114.
15. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. En: 2009 IEEE conference on computer vision and pattern recognition. 2009. p. 248-255.
16. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering.* 2009; 22(10): 1345-1359.
17. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *Journal of big data.* 2019; 6(1): 1-48.
18. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access.* 2018; 6: 52138-52160.
19. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. En: Proceedings of the IEEE international conference on computer vision. 2017. p. 618-626.

Si desea citar nuestro artículo:

González-Alday R, Peinado F, Carrillo D, Maojo V. Deep Learning for Clinical Image Classification of Genital Lesions Caused by Sexually Transmitted Diseases. *An RANM.* 2022;139(03): 266-273. DOI: 10.32440/ar.2022.139.03.rev07
