

CREACIÓN DE UNA RED NEURONAL ARTIFICIAL PARA LA DETECCIÓN DE CUATRO ENFERMEDADES PULMONARES

Marta Díez Asensio

INS Josep Lladonosa

Introducción

Debido a la pandemia de la COVID-19, son muchos los casos de pacientes que padecen enfermedades pulmonares. Para la detección precoz de estas se necesita un gran número de médicos titulados con la capacidad de diagnosticarlas y dado que la demanda de médicos en unidades dedicadas únicamente a los infectados por el coronavirus, hay enfermedades que han quedado, en cierto modo, desatendidas o faltas de personal. Es por este motivo y por el importante auge de la tecnología y la inteligencia artificial que me propongo en esta investigación el objetivo de diseñar un sistema de inteligencia artificial capaz de diagnosticar enfermedades a partir de imágenes de rayos-X. En este caso en particular me centraré en imágenes de rayos-X torácicas pero el mismo programa se puede usar y modificar para la detección de enfermedades distintas. Para comprobar la veracidad de los diagnósticos dados haré una comparación de los resultados adquiridos por el programa y el diagnóstico real de médicos titulados.

Hipótesis

Imagino que esta práctica será efectiva y además, el porcentaje de acierto del programa será superior al 90%.

Objetivos

Los objetivos de esta investigación son desarrollar un programa capaz de detectar las enfermedades de: Cardiomegalia, tumor, edema y neumotórax. Todo ello mediante el lenguaje de programación python, lenguaje natural, la biblioteca keras, las herramientas de Deeplearning, Densenet121 y GradCAM entre otros.

Otro objetivo de la investigación es entender en profundidad los mecanismos de los rayos X así como su naturaleza e historia con la idea de “enseñar” de una manera más efectiva al programa qué y cómo mirar las imágenes de rayos-X por tal de ver en ellas eso que se busca: las anomalías, las enfermedades.

Metodología

Esta investigación se basa en una construcción práctica pero se divide en dos partes. En la primera se explora la naturaleza de los rayos X lo cual le da al trabajo un carácter teórico y, en la segunda, se efectúa y evalúa el trabajo práctico.

Para hacer la práctica, además de todas las herramientas tecnológicas mencionadas anteriormente, así como algunas más específicas (ej. Estudio de la curva ROC y AUROC, paquetes como matplotlib.pyplot y seaborn o Util) se requerirán imágenes de rayos-X torácicas reales. He obtenido las mismas de la base de datos ChestX-ray8. Esta es pública y abierta, proveniente de Estados Unidos y fue publicada el 1 de septiembre de 2017 por el

“National Institutes of Health - Clinical centre”. Es de escala hospitalaria y está constituida por 108.948 imágenes de rayos X de vista frontal de 32.717 pacientes distintos con las cuatro enfermedades etiquetadas en este estudio.

La práctica consistirá en construir y evaluar el modelo clasificador de Deep Learning. Los pasos que seguiré para hacerlo serán:

- Procesar y preparar un conjunto de datos de rayos X del mundo real.
- Utilizar el aprendizaje por transferencia (transfer learning) para volver a capacitar un modelo DenseNet para la clasificación de imágenes de rayos X.
- Usar técnicas para el manejo de desequilibrio de clases o “Class imbalance”.
- Medir el rendimiento de diagnóstico calculando el AUC (área bajo la curva) para la curva ROC (característica de funcionamiento del receptor).
- Visualizar la actividad del modelo usando GradCAMs.

Además también haré un estudio evaluativo de los resultados, comparándolos con los dados por médicos titulados.

Resultados

En primer lugar he comparado los resultados del algoritmo con el diagnóstico dado por 9 radiólogos. Esta información la he extraído de un archivo adjunto a la base de datos. Con esta información me he centrado en el rendimiento del modelo. El porcentaje de acierto y error lo he calculado comparándolo con el diagnóstico de los radiólogos.

Analizados los resultados, observo que tanto el diagnóstico proporcionado por los médicos experimentados como el algoritmo tienen una media de acierto del 95%.

Conclusión

Dados los resultados, afirmo que la diferencia de fiabilidad entre un diagnóstico dado por un médico o por el algoritmo solo se ve decantado hacia los resultados de los médicos en el caso de la Cardiomegalia, en el resto tienen la misma fiabilidad un médico que el sistema creado. Por lo que queda confirmada mi hipótesis inicial.

Este estudio ha sido riguroso y significativo en la época actual, por este motivo y por la perseverancia que requiere el montaje de un sistema de inteligencia artificial (aparición de múltiples errores que dificultan la construcción del mismo) pienso que mediante este estudio he alcanzado los objetivos del trabajo de investigación los cuales, a mi parecer son: desarrollar un pensamiento crítico, aprender (términos y conductas (perseverancia y organización)) y, mediante el esfuerzo, llegar a resultados satisfactorios.

Bibliografía

- Ripoll, M. M. (2010). *Los rayos X*. Consultado el 8/2/2020, en https://www.xtal.iqfr.csic.es/Cristalografia/parte_02.htm
- *Bases físicas de la radiación ionizante*. (2020, Marzo). Consultado el 7/5/2020, en <https://www.sciencedirect.com/science/article/pii/S0120563320300085>
- Elert, G. (2020). *X-ray*. The physics Hypertextbook. Consultado el 9/3/2020, en <https://physics.info/x-ray/>
- Tacto, V., & Jones, J. (n.d.). *Pulmonary mass*. Consultado el 8/5/2020, en <https://radiopaedia.org/articles/pulmonary-mass>
- Mayo Clinic. (2019, Octubre 20). *Pulmonary Edema*. Consultado el 8/2/2020, en <https://www.mayoclinic.org/diseases-conditions/pulmonary-edema/symptoms-causes/syc-20377009>
- Mayo Clinic. (2019, Febrero 28). *Pneumothorax*. Consultado el 8/2/2020, en <https://www.mayoclinic.org/diseases-conditions/pneumothorax/symptoms-causes/syc-20350367>
- Ujjwalkarn. (2016, Agosto 9). *A quick Introduction to Neural Networks*. The data science blog. Consultado el 8/4/2020, en <https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks/>
- Ujjwalkarn. (2016, Agosto 11). *An Intuitive Explanation of Convolutional Neural Networks*. The data science blog. Consultado el 8/5/2020, en <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>
- Instituto Nacional del Cáncer. (n.d.). *Radiación dispersa*. Consultado el 8/2/2020, en <https://www.cancer.gov/espanol/publicaciones/diccionario/def/radiacion-dispersa>
- Ramírez, A. F., & Martínez, E. (2018). *Impacto del desequilibrio de clases en el entrenamiento de redes neuronales convolucionales en problemas multi-clase*. Semantic Scholar. Consultado el 8/2/2020, en <https://www.semanticscholar.org/paper/IMPACTO-DEL-DESEQUILIBRIO-DE-CLASE-S-EN-EL-DE-REDES-Ram%C3%ADrez-Mart%C3%ADnez/7c65b79f57618132abe24228d24b99596ca829db>
- Allum, J., & Talbot, C. (2016). OPCIÓN C - Imágenes. In *Física - Ib diploma* (1r ed., p. 592). Vicens Vives. http://www.vicensvives.com/vvweb/view/webwidgets/ibextras/docs/15_IB_Fisica_015.pdf?ext=.pdf