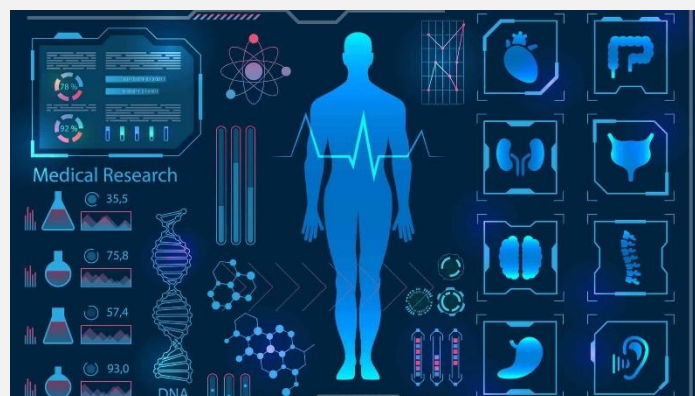


PRÁCTICA 2

IA en salud

Aplicaciones Avanzadas de la IA

**Máster Universitario en Ingeniería Informática
(2023/24)**



Versión 3.0 -23 de Febrero de 2024

Introducción

Los algoritmos de aprendizaje automático están en uso en medicina desde hace décadas: los primeros algoritmos para analizar o ayudar a interpretar imágenes de radiografía datan de los años 1960s. Los algoritmos de soporte a la decisión para detección y diagnóstico avanzaron en los años 1980s, primero con algoritmos para el diagnóstico de cáncer en radiografías de tórax y mamografías, ampliando después su alcance a otras modalidades como la tomografía computerizada (CT) y los ultrasonidos¹. Los Sistemas de Apoyo a la decisión en esa época usaban un enfoque basado en los datos (data-driven) igual que hacen actualmente los sistemas basados en aprendizaje profundo (deep learning). La principal diferencia es que los primeros sistemas dependían principalmente de la selección de características (feature engineering): el flujo de trabajo típico para diseñar un algoritmo consistía en entender qué tipo de imágenes y evidencia clínica estaban usando los médicos para una determinada tarea, trasladar ese conocimiento a un programa para extraer las características más relevantes, y usar después algoritmos de aprendizaje automático para combinar las características y generar un resultado.

En paralelo, a lo largo de los años 1990s y 2000s, algunos investigadores empezaron a trabajar en el uso de redes neuronales convolucionales y otros tipos de redes complejas para tareas de detección y caracterización en imagen médica. Todos estos métodos compartían algunas características de las redes convolucionales profundas actuales: los datos se propagaban a través de la red vía convoluciones, las redes aprendían los filtros, y los métodos no se basaban en la extracción de características, sino que la entrada de la red era directamente la imagen (los píxeles). Sin embargo, debido a la capacidad computacional disponible en ese momento, la mayoría de las redes tenían solo una o dos capas y se entrenaban con conjuntos mucho más pequeños de los que se usan en la actualidad, por lo que esas primeras CNNs no supusieron una transformación masiva en el aprendizaje automático para imagen médica.

A medida que se generalizaron los conjuntos de datos más grandes, y los procesadores gráficos (GPUs) se hicieron más asequibles, el entrenamiento de arquitecturas profundas de forma razonablemente rápida se volvió una posibilidad real. En ello influyó también diferentes innovaciones y buenas prácticas como métodos mejorados para la inicialización y actualización de los pesos², nuevas funciones de activación³, nuevos métodos de regularización como cortar conexiones o poner a cero los pesos

¹ Kanazawa K, Kawata Y, Niki N, et al. Computer-aided diagnosis for pulmonary nodules based on helical CT images. *Comput Med Imaging Graph.* 1998;22:157–167.

² Sutskever I, Martens J, Dahl G, Hinton G. On the importance of initialization and momentum in deep learning. In: Sanjoy D, David M, eds. *Proceedings of the 30th International Conference on Machine Learning (PMLR, Proceedings of Machine Learning Research)*, Vol. 28; 2013:1139–1147.

³ Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In: Geoffrey G, David D, Miroslav D, eds. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (PMLR, Proceedings of Machine Learning Research)*, Vol. 15; 2011:315–323.

durante el entrenamiento⁴ y estrategias de “data augmentation” que hacen que las redes sean más robustas frente a transformaciones simples de los datos de entrada. El momento de inflexión que abrió el camino al uso generalizado de redes convolucionales se produjo en la competición ImageNet de 2012 en el que se preparó una base de datos con un millón de imágenes de mil clases de objetos diferentes, y que ganó la red denominada AlexNet⁵. A partir de ese momento se propusieron otras arquitecturas que se han popularizado mucho en el análisis de imágenes naturales como VGGNet⁶, GoogleNet o ResNet⁷ y que se han popularizado también en el área de la imagen médica donde las aplicaciones han crecido de forma exponencial, incluyendo aplicaciones clásicas de aprendizaje automático como segmentación, clasificación y detección de lesiones, pero también otras nuevas como la reconstrucción de imágenes a partir de la proyecciones o la reducción de artefactos.

La transferencia de aprendizaje⁸ es un enfoque que propone que las representaciones de los datos aprendidos para un conjunto de datos o un problema determinado siguen siendo útiles cuando se aplican a un conjunto diferente. Aunque no es un enfoque específico de Deep learning, la gran necesidad de datos de entrenamiento propia de Deep learning lo hace especialmente útil en este contexto donde no se dispone de suficientes datos relevantes para una tarea concreta.

En el campo de la imagen médica es posible utilizar arquitecturas entrenadas con imágenes naturales y entrenarlas de nuevo para la tarea específica a desarrollar. También es posible partir de una red entrenada con imágenes de una modalidad y transferir ese aprendizaje a otra modalidad o problema⁹. Un enfoque común es entrenar la red solo para que aprenda los pesos de la capa de clasificación (fully connected) o también de algunas capas convolucionales¹⁰, manteniendo fijos los pesos de las capas convolucionales de extracción de características (o las de bajo nivel).

⁴ Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15:1929–195

⁵ Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, eds. *Advances in Neural Information Processing Systems*, Vol. 25. New York, NY: Curran Associates, Inc.; 2012:1097–1105.

⁶ Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition; 2014. arXiv:1409.1556.

⁷ He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition; 2016:770–778.

⁸ Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng.* 2010;22:1345–1359.

⁹ Azizi S, Mousavi P, Yan P, et al. Transfer learning from RF to B-mode temporal enhanced ultrasound features for prostate cancer detection. *Int J Comput Assist Radiol Surg.* 2017;12:1111–1121.

¹⁰ Tajbakhsh N, Shin JY, Gurudu SR, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging.* 2016;35:1299–1312

Objetivo

El objetivo básico de esta práctica es entrenar un modelo que permita clasificar imágenes médicas uno de los siguientes dominios

- Imágenes de radiografía de tórax, para identificar pacientes con COVID-19.
- Imágenes de histología de tejido colorrectal, para clasificar el tipo de tejido, incluyendo células tumorales.
- Imágenes de fondo de ojo para identificar lesiones asociadas con retinopatía diabética.

Para cada dominio se debe entrenar un **clasificador binario (sano vs no sano)** y uno multiclase (que depende del problema).

Al final de la práctica se proponen otros trabajos más avanzados para los grupos que ya tengan experiencia previa en el diseño y entrenamiento de redes neuronales convolucionales

Se plantean **diferentes niveles para la práctica** en función de los conocimientos previos de los alumnos. La nota máxima posible dependerá de la complejidad del trabajo, de acuerdo con lo siguiente:

	Nota máxima
RX de tórax, Anatomía patológica	9
Retinopatía	9,5
Incluyendo interpretabilidad (CAM u otros)	10
Incluyendo segmentación	10

En el siguiente apartado se describe con más detalle cada uno de los problemas y dónde conseguir los datos. También tenéis compartido un conjunto de los datasets en Google drive

<https://drive.google.com/drive/u/0/folders/1uv1oO6UnuuLvK0zS61NSKu4uwDUKwJKV>

El dominio más sencillo es del de RX de tórax, y el más complejo el de retinopatía, ya que se parte de un conjunto de datos sin pre-procesar.

Resumen de los datos de partida

Problema	Tamaño del dataset original	Origen	Subconjunto en Google drive
Covid	1 Gb	Kaggle / paper ieee	480 Mb
Anatomía patológica	11.7Gb + 800Mb	Open Aire	600 Mb
Retinopatía	88 Gb	kaggle	5 Gb

Covid-19

Las imágenes de Radiografía de Tórax han sido recopiladas de dos repositorios diferentes:

- Covid 19: Repositorio del [Dr. Joseph Cohen](#), postdoctoral fellow de la Universidad de Montreal. Accessible en [ieee8023/covid-chestxray-dataset](#)
- Pacientes sanos y neumonía bacteriana: [Kaggle's Chest X-Ray Images \(Pneumonia\) dataset](#)

Un subconjunto de este dataset está accesible en la carpeta compartida de Google Drive

El objetivo es clasificar una imagen en covid /no covid (problema binario) o en tres clases (covid / neumonía bacteriana / paciente sano)

Clasificación de imágenes de patología digital

La anatomía patológica es la rama de la medicina que estudia los efectos que produce la enfermedad en los órganos del cuerpo. Una parte fundamental es el estudio del corte histológico: una rodaja fina de un tejido adherida sobre un portaobjetos y generalmente coloreada con alguna tinción específica para resaltar una parte de la estructura.

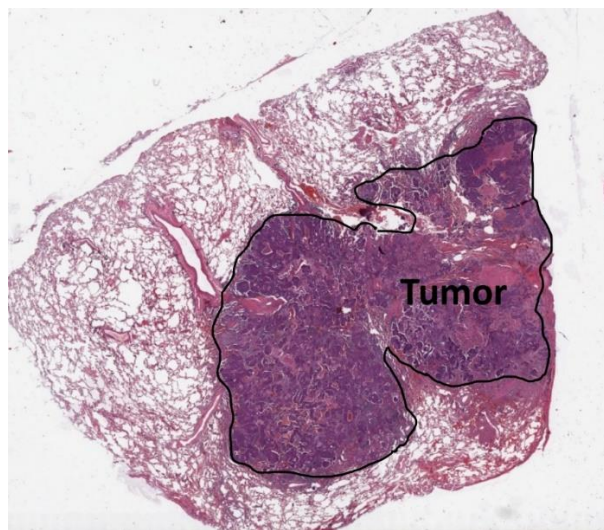


Figura 1. Whole slide image. (Images courtesy of National Cancer Institute)

Vamos a usar un conjunto de datos publicado en abierto que contiene 100.000 imágenes histológicas de tejido colorrectal. Los datos están accesibles en Zenodo, un repositorio de acceso abierto de propósito general desarrollado bajo el programa europeo OpenAIRE (iniciativa avalada por la comisión europea para promover la ciencia abierta). Zenodo está operado por el CERN y permite a los investigadores depositar artículos de investigación, conjuntos de datos, software de investigación, informes y cualquier otro artefacto digital relacionado con la investigación. Para cada

envío, se genera un identificador de objeto digital persistente (DOI), lo que hace que los elementos almacenados se puedan citar fácilmente.

Las imágenes histológicas se toman con microscopio, típicamente con una resolución de 0,5 micras por pixel (amplificación 20x), por lo que una imagen original que contiene todo el corte tiene un tamaño de aproximado de 3,5 GB, diez veces más que una imagen típica de radiología. A este tipo de imagen se la llama Whole slide image (WSI).

Las imágenes que vamos a usar están accesibles en <http://dx.doi.org/10.5281/zenodo.1214456>.

Hay dos conjuntos de datos diferentes

- CRC-VAL-HE-7K.zip – 800 Megas, un conjunto de prueba
- NCT-CRC-HE-100K.zip - 11.7 GB

En este caso cada imagen completa está recortada en imágenes más pequeñas de 122 micrómetros de lado, que han sido etiquetadas por un patólogo en función del tipo de tejido:

- ADI = tejido adiposo
- BACK = fondo (no hay tejido)
- DEB = debris
- LYM = linfocitos,
- MUC = mucus,
- MUS = músculo
- NORM = mucosa normal
- STR = stroma
- TUM = epitelio tumoral

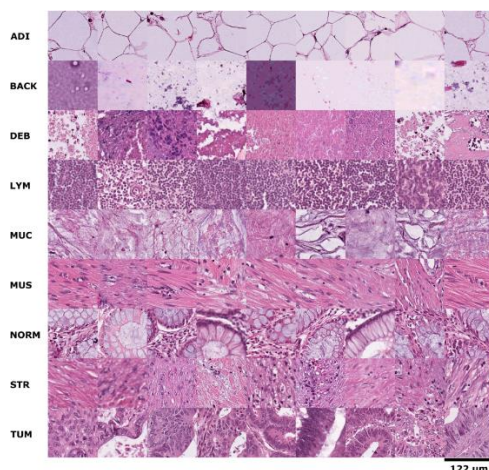


Figura 2. Ejemplos de imágenes de diferentes tipos de tejido

El objetivo de este trabajo es clasificar cada imagen en una de las nueve clases posibles. Además, se debe crear también un clasificador binario que pueda clasificar la imagen en tumor o cualquier otro tejido

Para una descripción más detallada del problema y cada tipo de tejido, podéis consultar el artículo original:

Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Kather, J.N., Pearson, A.T., Halama, N. et al. Nature Medicine 25,1054–1056 (2019). <https://doi.org/10.1038/s41591-019-0462-y>

Para un tutorial paso a paso de implementación con Matlab:

<https://blogs.mathworks.com/deep-learning/2019/07/24/deep-learning-for-medical-imaging/>

Retinopatía diabética

La **retinopatía diabética** es la principal causa de ceguera en la población en edad laboral en el mundo desarrollado. Se estima que afecta a más de 93 millones de personas en el mundo. Es una enfermedad relacionada con la diabetes, el 40% de los diabéticos acaban desarrollando algún grado de retinopatía a lo largo de la vida. Es una enfermedad tratable si se detecta a tiempo.

Actualmente todos los enfermos diabéticos son revisados anualmente en busca de signos de la enfermedad, en un proceso largo que requiere que un médico experto examine y evalúe la imagen de **retinografía** (fotografía de la retina). La retinopatía diabética se identifica por la presencia de lesiones asociadas a anomalías vasculares causadas por la enfermedad.

El gran volumen de personas que necesitan este seguimiento y la escasez de profesionales preparados, especialmente en algunos ámbitos, ha hecho que se planteen nuevos modelos de seguimiento basados en un triaje (screening) previo por personal menos especializado, pero disponible localmente (por ejemplo, en un centro de salud u hospital comarcal), evitando desplazamientos al paciente. El reto actual es realizar este triaje de manera semi-automatizado.

El objetivo de este ejercicio, como los anteriores es entrenar un modelo multiclase y uno binario. El modelo binario se deberá diferenciar los pacientes sanos de los pacientes con retinopatía. El modelo multiclase, buscará identificar los diferentes estadios de la enfermedad. En la práctica este tipo de modelos se utilizan para triaje (distinguir los pacientes que no necesitan más pruebas de aquellos que deben ser estudiados con más detalle), y es especialmente interesante ser capaz de identificar los estados más iniciales de la enfermedad

Las imágenes que vamos a usar están disponibles en la plataforma Kaggle, una empresa subsidiaria de Google en la que se publican conjuntos de datos, modelos y concursos. Pertenecen a un concurso se celebró en 2015 patrocinado por la California Healthcare Foundation y una empresa del sector (eyePacs)

<https://www.kaggle.com/c/diabetic-retinopathy-detection/data>

Si queréis trabajar sólo con un subconjunto, lo tenéis disponible en Google Drive

<https://drive.google.com/drive/folders/1rGpePSJOBhC4gc2mPs6ogzJ7S00h6B59?usp=sharing>

De los tres problemas presentados, éste es el más cercano a un problema real, ya que es necesario seleccionar y preprocesar las imágenes. El preprocesamiento puede incluir reducir la resolución de las imágenes y recortarlas al 90% del tamaño original para reducir el efecto de borde. Para homogeneizarlas, se puede normalizar el color (por ejemplo, restar la media local del color y mapearlo a 50% de gris). Para más información sobre el preprocesado consultar las referencias. Puede ser recomendable no usar el conjunto completo de datos.



Figure 5: Healthy eye and stages of DR with increasing severity.

Figura 3. Ojo sano y diferentes fases del progreso de retinopatía diabética a) leve b) moderada 3) severa 4) proliferativa

Referencias

B. Graham. "Kaggle Diabetic Retinopathy Detection competition report". Enlazado desde <https://www.kaggle.com/c/diabetic-retinopathy-detection/discussion/15801>
Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316:2402-2410. doi:10.1001/jama.2016.17216

Prácticas avanzadas

Si algún grupo ya tiene experiencia con redes convoluciones y quiere probar nuevos retos, os proponemos dos ideas:

- 1) Extender la práctica ofreciendo una interpretación de los resultados con técnicas como class activation mapping o Layer-wise relevance propagation (LRP).

Referencias:

<https://es.mathworks.com/help/deeplearning/ug/investigate-network-predictions-using-class-activation-mapping.html;jsessionid=f10ce464aca0b01e8c75a03f1f99>

Bach, Sebastian, et al. "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation." PloS one 10.7 (2015)

- 2) Segmentación de imágenes médicas

Podéis elegir las imágenes que queráis. Un ejemplo es la segmentación de células o núcleos de células. Un dataset accesible directamente es [éste](#), donde el objetivo es segmentar las células blancas (White blood cells)

La red que recomendamos usar es UNet

Herramientas

Para alcanzar el objetivo de esta práctica, el alumno podrá utilizar tanto la técnica de IA como la herramienta que considere más adecuada.

Si tenéis ordenadores con poca memoria, puede ser interesante trabajar con Python-keras-tensorflow y ejecutar el entrenamiento en GoogleColab.

En ese caso os recomiendo ir guardando los resultados después de cada epoch, para no perderlo en caso de desconexión.

Referencias para implementación en python

1. Deep Learning with Python. Francois Chollet
2. https://keras.io/examples/vision/image_classification_from_scratch/
3. <https://www.tensorflow.org/tutorials/images/classification>

Referencias para implementación en Matlab

Deep Network designer

<https://es.mathworks.com/help/deeplearning/gs/get-started-with-deep-network-designer.html>

Train Deep Learning Network to Classify New Images

<https://es.mathworks.com/help/deeplearning/ug/train-deep-learning-network-to-classify-new-images.html>

Algunos consejos

Debéis valorar las siguientes estrategias para construir el modelo, y seleccionar las más adecuadas para vuestro caso, justificando las decisiones

- Usar una red pre-entrenada (transfer learning)
- Balancear las clases
- Usar estrategias de data-augmentation
- Normalizar las entradas de la red
- Usar GPUs para entrenar si se dispone de ellas
- Usar una técnica de parada (early stopping)
- Preprocesar las imágenes para que tengan el tipo de fichero y resolución más adecuada. En el caso de anatomía patológica el formato original de las imágenes es en TIFF. Podéis cambiarlo a JPG mediante la librería de Python PIL Image, para que el conjunto de datos ocupe menos
- Hacer las pruebas iniciales con un número reducido de imágenes
- Puede ser útil asegurarse de que la proporción de las distintas clases es igual los conjuntos de entrenamiento, validación y test (balancear el conjunto de datos de entrenamiento y usar muestreo estratificado para el conjunto de test)

Como orientación el tiempo de entrenamiento para esta práctica oscila entre los 15 minutos y las 6 horas, en función de la arquitectura de la red, el conjunto de datos elegidos, el hardware utilizado y la configuración del entrenamiento

Entregables

La práctica constará de **dos entregables**

- Una **memoria** en pdf
- El **código** utilizado, por ejemplo en formato ipython notebook. El código debe incluir el preprocesado de la imagen para que se puedan reproducir los experimentos a partir del conjunto de datos original

La memoria contendrá una página de **resumen ejecutivo** en la que se presente

- 1) Un **abstract gráfico** que explique de manera visual cómo se ha realizado el pre procesado y el entrenamiento
- 2) Una tabla **resumen de los resultados**

El resto de la memoria incluirá los datos que se consideren importantes para la valoración del trabajo, por ejemplo :

- Enfoque del problema, antecedentes y documentación consultada para diseñar la solución (no es necesario realizar una revisión del estado del arte, pero si consultáis alguna referencia, debéis citarla)
- Decisiones de diseño e implementación, incluyendo lenguaje de programación, librerías y frameworks utilizados
- Tipo de clasificador (no es obligatorio que sean redes neuronales)
- Arquitectura de la red: número de capas, tamaño de los kernels, tipo de funciones de activación, tipo de regularización, función de pérdida
- Entrenamiento de la red: número de epochs, tamaño del mini-batch, tipo de inicialización de los pesos, learning rate
- Recursos usados: Hardware utilizado para entrenar y tiempo de entrenamiento
- Pruebas realizadas
- Resultados (según se describe a continuación)

Los **resultados** deben presentarse eligiendo las medidas de rendimiento más adecuadas para cada problema. Se debe considerar el tipo de problema clínico a la hora de ofrecer los resultados (por ejemplo, ¿tiene el mismo impacto confundir dos tipos de tejidos sanos o un tejido sano y un tumor?). Se recomienda el uso de un **conjunto de test independiente** (no usado en el entrenamiento), y especificar siempre si los resultados son en un conjunto de test o en validación cruzada.

Para el problema binario es obligatorio presenta el valor del **área bajo la curva ROC (AUC-ROC) y la sensibilidad para un 85% de especificidad**. Se recomienda incluir también la curva ROC.

Por último, se debe añadir una breve reflexión sobre la **utilidad clínica** de los resultados obtenidos. ¿se podrían trasladar directamente a la práctica clínica? ¿Qué

otros pasos sería necesario dar? Si necesitáis inspiración sobre qué temas tratar podéis mirar la Tabla 1 de este artículo.

[Table 1 A summary of some general questions readers of an A.I. medical paper might want to ask themselves to evaluate the quality of the results of a research. \(nature.com\)](#)

La extensión máxima de la memoria es de 10 páginas (15 si se trabaja en alguna de las extensiones)

Entrega de la práctica

- La práctica deberá entregarse únicamente por uno de los integrantes del grupo.
- La entrega de esta Práctica deberá realizarse por Aula Global en la fecha programada en el entregable

Evaluación de la práctica

La calificación se asignará tras la valoración de estos aspectos

- Planteamiento y desarrollo del problema: 30%
- Resultados del problema: 25%
- Análisis de resultados y conclusiones: 25%
- Presentación: 20%