

Procesamiento de imágenes biomédicas mediante IA

1. El uso de las imágenes biomédicas en salud

Las imágenes biomédicas son representaciones visuales de estructuras y funciones del cuerpo humano, obtenidas mediante técnicas de imagen como las radiografías, las resonancias magnéticas, las ecografías... entre muchas otras. Estas imágenes permiten a los profesionales de la salud y científicos visualizar y comprender mejor las enfermedades que afectan al organismo de una persona. Proporcionan información muy valiosa a la hora de tomar decisiones acerca del diagnóstico y tratamiento de un paciente por lo que la interpretación de los datos que nos proporcionan puede ser de vital importancia.

Gracias al apoyo en herramientas de inteligencia artificial, analizar, comprender y tomar decisiones en base a las imágenes biomédicas puede ser un proceso más eficiente y preciso. Además, puede facilitar los avances en investigación y desarrollo de tratamientos médicos. Para lograr esto, las herramientas de IA necesitan una gran cantidad de datos, con el objetivo de extraer información de ellos y usarla en el futuro.

2. Open Data

El concepto Open Data, o Datos Abiertos, hace referencia a una filosofía de acumulación de datos electrónicos que persigue su divulgación, con pocas o ninguna restricción y permitiendo su uso por parte de terceros, con el simple objetivo de ayudar a cualquier persona, entidad u organización que los necesite para desarrollar una actividad, un proyecto de investigación, un negocio o con cualquier otra finalidad.

El movimiento de Datos Abiertos se consolidó cuando en diciembre de 2007 una treintena de expertos del gobierno abierto de los Estados Unidos estableció “los 8 principios del Open Data”, que se resumen en lo siguiente:

1. **Libertad.** Los datos abiertos serán públicos y accesibles, y bajo ningún concepto se restringirá su acceso por privacidad, seguridad, privilegios o derechos de autor.
2. **Originalidad.** Los datos abiertos serán presentados como originalmente se encontraban cuando se recolectaron, lo más detallados posible y sin haber sido modificados ni transformados. En definitiva, los datos abiertos son datos en bruto.
3. **Actualidad.** Los datos abiertos serán presentados lo antes posible, de forma que sean útiles en el momento de su publicación y no “caduquen” por el paso del tiempo.
4. **Accesibilidad.** Los datos abiertos estarán disponibles al mayor número de personas para el mayor número de propósitos. Para su publicación se debe de utilizar Internet y permitir la descarga de estos de manera sencilla, para que todos podamos hacerlo.
5. **Claridad.** Los datos abiertos tendrán una estructura que permitirá a programas informáticos su procesamiento automático. Esto significa que los datos estarán en

formato texto limpio (“Bloc de notas” de Windows) y con suficientes explicaciones de qué significa cada característica de los datos recogidos.

6. **Libre descarga.** Los datos abiertos se podrán descargar sin registrarse en ningún formulario ni dejar constancia de quién ha accedido a ellos. Es decir, se garantizará el acceso anónimo a los datos.
7. **Dominio público.** Los datos abiertos no serán propiedad de nadie; serán propiedad de todos. De esta manera, ninguna persona ni organismo podrá restringir quién puede ver y utilizar los datos.
8. **Sin licencia.** Los datos abiertos no pueden ser licenciados por copyright, patentes o marcas comerciales. Pueden someterse a ciertas medidas de seguridad y privacidad, pero nunca pueden ser vendidos o intercambiados “en secreto”.

En resumen, los datos abiertos son un material que los proyectos de investigación, instituciones gubernamentales y negocios pueden utilizar libremente y sin ninguna restricción, salvo la de mantener y perpetuar el “trato” o “acuerdo” de los datos abiertos (es decir, que sigan estando disponibles para todos y los demás principios descritos anteriormente).

En este programa, vosotros, los estudiantes, vais a beneficiaros de esta filosofía de los datos abiertos para realizar un pequeño proyecto de investigación en el área sanitaria. Todo proyecto de investigación, especialmente si utiliza Inteligencia Artificial, necesita de muchos datos, como se ha discutido con anterioridad. Es por ello por lo que cada vez más gobiernos europeos y empresas están impulsando este movimiento de compartir datos, pues con ello estamos también compartiendo conocimiento en beneficio de todos.

3. Las imágenes en Inteligencia Artificial

Con la suficiente cantidad de imágenes, podemos entrenar a un ordenador o sistema para extraer información que nos interesa de dichas imágenes, encontrar rápidamente detalles relevantes e incluso tomar decisiones basadas en lo que “ve” en ellas. Para lograr todo esto, son necesarias no solo una gran cantidad de información, sino sistemas con una gran potencia informática para procesarla toda. Algunos ejemplos de tareas que podemos resolver gracias al uso de imágenes en IA [\[1\]](#) son:

- Clasificación de imágenes, que nos permite identificar a qué “categoría” o clase pertenece una imagen. Un ejemplo puede ser un algoritmo que distinga e identifique si un hueso está roto en base a una radiografía.
- Rastreo de objetos, siguiendo un objeto por ejemplo, en un video, para calcular su trayectoria, como sucede en los vehículos autónomos a la hora de evitar colisiones.
- Recuperación de imágenes de grandes bases de datos, utilizando el contenido de las imágenes en vez de sus etiquetas o características del propio archivo, podemos buscar y recuperar imágenes automáticamente. Un ejemplo que conocemos casi todos es la búsqueda por imágenes en Google.

Otro ejemplo muy interesante y reciente es el procesado de radiografías torácicas para la detección de COVID-19 [\[2\]](#). Estos modelos emplean colecciones de imágenes disponibles en fuentes de datos abiertos [\[3\]](#) de radiografías en las que los pacientes habían dado

positivo en COVID-19 entre otras neumonías bacterianas y virales. Tras entrenarlos con estas fuentes de imágenes, los modelos son capaces de apoyar a un especialista médico a la hora de decidir si un paciente presenta COVID-19, acelerando y facilitando el diagnóstico.

4. Fuentes de datos abiertos

La búsqueda de datos abiertos no es una tarea compleja en un principio, pues podemos rápidamente llegar a muchas fuentes de datos con una búsqueda en internet. Sin embargo, debemos asegurarnos de que efectivamente los datos que estamos viendo pueden ser usados para la finalidad que nos interesa. Para identificar qué datos podemos usar, debemos fijarnos en su Licencia y en sus Condiciones. Podemos encontrar fuentes de datos abiertos si nos fijamos en que den su consentimiento expreso a su libre uso o en que estén licenciadas bajo aquellas licencias “Creative Commons” que no prohíban su distribución y uso.

En nuestro caso, estamos buscando fuentes de datos que contengan imágenes y una descripción e identificación de la misma, de forma que podamos entrenar modelos de inteligencia artificial para que relacionen el contenido de las imágenes con su descripción. Esto es lo que llamamos imágenes etiquetadas o anotadas. Algunas de las fuentes más conocidas [\[4\]](#) (también llamados “repositorios”) son:

- [MNIST](#), que contiene más de 70.000 imágenes en blanco y negro de números escritos a mano para entrenar algoritmos de reconocimiento de números.
- [ImageNET](#), que recoge palabras o conjuntos de palabras con un significado asociado a cientos o miles de imágenes.

Sin embargo existen cientos de fuentes de las que podemos extraer imágenes que nos serán útiles. Lo más importante es asegurarse de que su uso esté permitido y que efectivamente se trata de imágenes biomédicas que podemos usar.

5. Preguntas fundamentales

Como primer objetivo de proyecto, queremos contestarnos las siguientes preguntas:

- ¿Existen casos conocidos, que hayan salido en noticias o artículos, del éxito en el uso de la Inteligencia Artificial con imágenes biomédicas en un entorno sanitario real? ¿Dónde y qué se ha conseguido?
- ¿Qué tipos de imágenes biomédicas existen, además de las conocidas radiografías?
- ¿En qué formato y por qué canales se distribuyen las imágenes biomédicas de datos abiertos?
- ¿De qué forma se suelen obtener imágenes biomédicas de los bancos de datos?
- ¿Consideráis ético el almacenamiento de imágenes biomédicas de los pacientes de un hospital?
- **¿Estáis conformes con el uso de Inteligencia Artificial en los hospitales?**

6. Trabajo y entrega

A la hora de abordar este proyecto de investigación, queremos que tengáis la libertad de explorar y buscar información por vuestra cuenta de forma que gracias a lo que habeis aprendido, elaboréis un trabajo que explore aquellas secciones del procesamiento de imágenes biomédicas con IA que os parezcan más interesantes u os hayan llamado más la atención. El objetivo no es solo que vayáis más allá e **investiguéis más** a cerca del tema por vuestra cuenta, sino que además tengáis la posibilidad de **demostrar todo lo aprendido** de la forma que veáis que se adapte más al tema y a vosotros mismos. Podéis usar las preguntas del punto anterior como puntos de partida que os sirvan de ayuda a la hora de empezar pero valoraremos que seáis originales y demostréis vuestra creatividad.

A continuación os dejamos un ejemplo de una posible estructura del trabajo para que os sirve de posible guía:

- Definir **imagen biomédica** y descubrir de qué tipos existen, para qué se utilizan tradicionalmente en el mundo de la medicina y qué interés tienen en el mundo de la IA.
- Investigar y especular un poco sobre el impacto que puede llegar a tener la inclusión de análisis de imágenes biomédicas mediante IA en los hospitales. No hace falta dar detalles concretos; sólo divagar un poco en qué se podría lograr con ello.
- Estudiar las principales dificultades a la hora de utilizar imágenes biomédicas en IA para los hospitales, tanto si son tecnológicas, como administrativas (falta de personal especializado...), como éticas y legales.
- Investigar y nombrar modelos de IA se pueden usar para analizar imágenes biomédicas, y explicar sus conceptos más básicos.
- Responder a las preguntas fundamentales listadas en la sección 5.

Referencias:

[1] What is Computer Vision? | IBM. <https://www.ibm.com/es-es/topics/computer-vision>

[2] Saiz, F. A., & Barandiarán, Í. (2020). COVID-19 Detection in Chest X-ray Images using a Deep Learning Approach. International Journal of Interactive Multimedia and Artificial Intelligence, 6(2), 4. <https://doi.org/10.9781/ijimai.2020.04.003>

[3] Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. arXiv preprint arXiv:2003.11597. <https://doi.org/10.48550/arXiv.2003.11597>

[4] datos.gob.es. (2022, October 19). Repositorios abiertos de imágenes para entrenamiento de modelos de Inteligencia Artificial. datos.gob.es. <https://datos.gob.es/es/blog/repositorios-abiertos-de-imagenes-para-entrenamiento-de-modelos-de-inteligencia-artificial>