



Grado en Ingeniería Informática

Curso 2019-2020

Aprendizaje Automático

Trabajo Final

Autores:

Alba Reinders Sánchez	100383444
Alejandro Valverde Mahou	100383383

Índice

1. Definición del Problema	3
2. Datos y Obtención de Datos	4
3. Preproceso de Datos	4
4. Modelado	4
5. Evaluación	4
6. Uso	4

1. Definición del Problema

El problema a resolver es la clasificación de imágenes de muestras víricas para detectar el virus *SARS-CoV-2*, también conocido como *COVID-19*.

Esta idea nace de una propuesta realizada para el *Hackathon* de la Comunidad de Madrid '*Vence al virus*', el pasado 1 y 2 de Abril. La idea es identificar los pacientes infectados a través de las muestras víricas observadas con un microscopio TEM (*Transmission Electron Microscopes*), para así conseguir una prueba rápida, efectiva y de bajo coste.

Esta prueba deberá ser capaz de reconocer si el paciente está infectado por algún virus de la familia *Coronaviridae* y, dado que existen muy pocos virus de esta familia que afectan a humanos, se podría asumir con cierta probabilidad que el paciente estaría infectado con el *COVID-19*.

Se ha pensado resolver este problema a través de técnicas de aprendizaje automático porque se trata de un problema de clasificación. Dado que las imágenes usadas para el entrenamiento tendrán una clase asociada, se podrán usar técnicas de **aprendizaje supervisado**. De esta forma, se reduciría el tiempo necesario para clasificar cada una de las muestras, porque eliminaría la necesidad de los especialistas de tener que revisar cada una de las muestras para comprobar si el paciente está o no infectado.

Respecto a la viabilidad de la solución propuesta, se cree que supondría una reducción de coste y tiempo respecto a los medios actuales usados para detectar este virus. Sin embargo, la obtención de las imágenes necesarias para el entrenamiento puede ser complicada, ya que el microscopio que se requiere para obtenerlas no se encuentra en todos los laboratorios de los hospitales.

La fiabilidad que se espera obtener de esta prueba es bastante elevada puesto que se conoce que los virus de la familia *Coronaviridae* poseen una característica distintiva: suelen tener forma esférica y unas púas alrededor de todo su cuerpo, de forma que facilitará la diferenciación de estos virus respecto a otros.

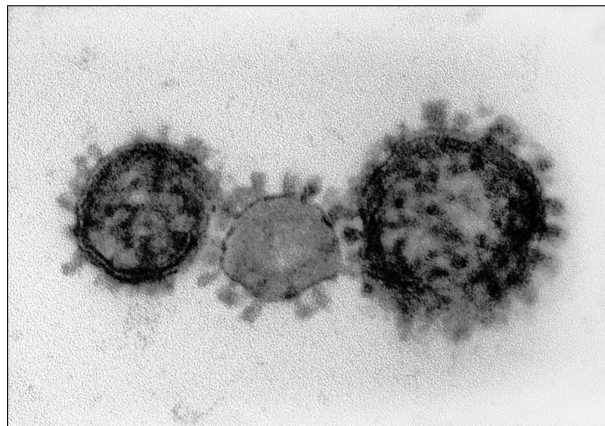


Figura 1: Muestra del virus *COVID-19*

El proceso que habría que llevar a cabo sería:

1. Recopilar un número de imágenes de muestras considerable entre las que se encuentren todo tipo de virus, entre ellos, el *COVID-19*. Estas imágenes deberán estar previamente etiquetadas.
2. Realizar un preprocesado a las imágenes.
3. Entrenar el modelo de aprendizaje automático.
4. Evaluar el modelo generado.
5. Utilizar este modelo para clasificar nuevas imágenes.

Dado que el conjunto de datos son imágenes, se va a usar una **red de neuronas** para realizar la tarea de aprendizaje automático. Esta red tendrá que ser capaz de diferenciar muestras con *COVID-19*, muestras con otros virus, y muestras vacías.

2. Datos y Obtención de Datos

El conjunto de datos que se usan para entrenar y evaluar el modelo son imágenes de muestras víricas clasificadas. Estas imágenes son tomadas por el microscopio TEM, son en blanco y negro y de un tamaño variado. Estarán clasificadas en tres categorías: **COVID-19**, **otros virus** y **vacía**.

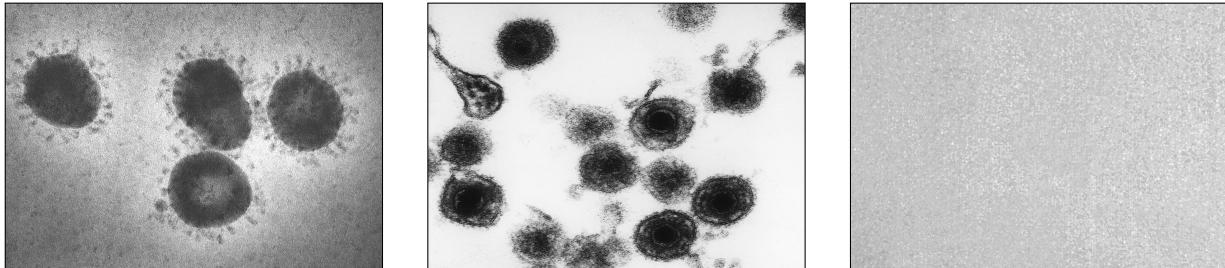


Figura 2: Muestras de *COVID-19*, otros virus y vacía respectivamente

Los datos se obtienen del *NIAID (National Institute of Allergy and Infectious Diseases)* y de la *Biblioteca Sanitaria Pública de EEUU*. También, y en base al *Hackathon*, se estableció contacto con el *CSIC* para conseguir mayor número de imágenes, aunque todavía no se han conseguido.

El problema principal de estas imágenes es su poca cantidad (alrededor de 50 imágenes de cada clase), por eso ha sido necesario contactar con el *CSIC*. Ya que es de vital importancia conseguir un conjunto de imágenes lo suficientemente grande como para poder entrenar a la red de la mejor forma posible.

3. Preproceso de Datos

4. Modelado

5. Evaluación

6. Uso