

Proyecto Final

Clasificación de variantes de covid-19

Integrantes

- Luis Manuel Gonzalez Hidalgo
- Kevin Ariel Merino Peña
- Marco Antonio Rivera Silva

Introducción

El COVID-19 ha tenido un gran impacto en la sociedad, la economía y la salud de todo el mundo. Ha provocado la imposición de restricciones y bloqueos, el cierre de negocios y escuelas, la cancelación de eventos y viajes, y ha llevado a una crisis sanitaria global. A medida que se desarrolla la investigación científica y se implementan estrategias de salud pública se han creado distintos modelos matemáticos, los cuales tienen como objetivo clasificar las múltiples variantes del COVID-19 utilizando el aprendizaje automatizado, siendo una de las herramientas más poderosas que se tienen actualmente para analizar datos y trabajar con ellos. y así poder facilitar el estudio de estas variantes.

Objetivo

El objetivo de este proyecto es analizar la efectividad de un modelo de clasificación en la identificación y clasificación de variantes del virus COVID-19. Dado que el virus SARS-CoV-2 ha experimentado diferentes mutaciones y variantes a lo largo del tiempo, es importante comprender cómo estas variantes pueden afectar la propagación de la enfermedad, la gravedad de los síntomas y la eficacia de las vacunas y tratamientos existentes.

El proyecto busca analizar algunas secuencias genéticas del virus obtenidas de muestras clínicas y epidemiológicas. Estas secuencias se analizarán utilizando técnicas de bioinformática y algoritmos de aprendizaje automático para determinar la precisión de un modelo de clasificación capaz de identificar y distinguir entre las distintas variantes de COVID-19.

Dataset

El conjunto de datos que utilizamos para entrenar el modelo fué obtenido de la plataforma *Huggin Face*, para ser más exactos, de [este repositorio](#). En el cual se proporciona también el modelo.

El conjunto de datos está conformado por varias secuencias genéticas del COVID-19 obtenidas en un laboratorio.

Método

El método utilizado para analizar la precisión del modelo fué entrenarlo en un notebook de google colab, utilizando el lenguaje de programación python.

Una vez que el modelo esté entrenado, se evaluará su efectividad utilizando datos adicionales y comparándolos con las clasificaciones realizadas por expertos en genética viral. El objetivo es determinar si el modelo puede proporcionar una clasificación precisa y confiable de las variantes del virus y si puede identificar correctamente las mutaciones relevantes asociadas con un mayor riesgo de transmisión o resistencia a los tratamientos existentes.

Resultados

En conclusión, los resultados obtenidos en este proyecto revelaron que el modelo de clasificación desarrollado para identificar las variantes de COVID-19 alcanzó una precisión del 52%. Aunque este resultado indica que el modelo fue capaz de clasificar correctamente la mitad de las variantes evaluadas, también demuestra la necesidad de mejoras para lograr una mayor precisión.

Es importante destacar que la identificación de variantes de COVID-19 es un desafío complejo debido a la diversidad genética del virus y las mutaciones en constante evolución. Esto puede dificultar la tarea de clasificación y requerir enfoques más sofisticados y actualizados en el análisis de los datos.

Para mejorar la precisión del modelo, se pueden considerar diferentes estrategias. En primer lugar, es necesario aumentar el tamaño y la diversidad de los conjuntos de datos utilizados para el entrenamiento del modelo. Esto permitirá una mayor representatividad de las diferentes variantes y, por lo tanto, una mejor capacidad de clasificación.

Además, se pueden explorar técnicas avanzadas de aprendizaje automático y algoritmos más sofisticados para mejorar la capacidad del modelo de capturar patrones sutiles y características distintivas de las variantes del virus. Esto podría incluir el uso de redes neuronales profundas, enfoques de transferencia de aprendizaje y técnicas de aumento de datos.