

Cristopher Jose Rodolfo Barrios Solis - 18207

Nikolas Dimitrio Badani Gasdaglis - 20092

Diego Andre Castellanos Castro - 20082

Osmin Josué Sagastume Orellana - 18173

# ALGORITMOS DE APRENDIZAJE DE MÁQUINAS

---

## ¿Qué es el Machine Learning y para qué sirve?

Según los expertos, el machine learning es conocido como la inteligencia artificial que sustenta sobre la computación y las matemáticas. Esto le permite tener la facultad de resolver problemas cotidianos mediante el análisis crítico y el pensamiento continuo. La forma en la que esta inteligencia artificial puede operar es gracias al uso de algoritmos. Los cuales a simple vista pueden parecer un simple conjunto de operaciones. Pero el trabajo de estos algoritmos tiene grandes ventajas en lo que la búsqueda de patrones, la clasificación, segmentación y regresión de un conjunto de datos.

Es importante señalar que estos algoritmos utilizan los datos como unidad básica de conocimiento. De ahí es donde proviene el término aprendizaje automático. El cual no solo es conocido como una rama principal de la inteligencia artificial. Sino que también posee como objetivo principal el desarrollo de técnicas que permitan que una computadora pueda aprender de manera autorizada.

## CNN

Según Data Scientist, el modelo CNN que pertenece a la subcategoría de redes neuronales es considerado como uno de los modelos de clasificación de imágenes más eficientes que existen. Una de las aplicaciones más importantes de este modelo es para la captura de imágenes médicas. Lo cual puede ser de mucha utilidad para la detección de coágulos de sangre. Debido a que el modelo CNN permitirá que las imágenes sean analizadas y así poder determinar la presencia o ausencia de coágulos de sangre en los pacientes.

## YOLO

De acuerdo con Microsoft, YOLO es una red neuronal que predice cuadros delimitadores y probabilidades de clase a partir de una imagen. Su procesamiento de fotogramas permite fácilmente identificar objetos en imágenes. Al realizar su análisis en tiempo real, este modelo sería muy efectivo para poder detectar coágulos de sangre en imágenes médicas. Debido a que el algoritmo podría identificar y ubicar la cantidad de coágulos que se encuentran presentes en la imagen del paciente.

## Fast R-CNN

Conocido por ser un algoritmo de detección de objetos, el Fast R-CNN resulta ser de mucha utilidad para el análisis de imágenes médicas. Con su función de detección de objetos, el algoritmo es capaz de detectar si hay coágulos de sangre en las imágenes médicas de la base de datos analizada.

## Algoritmo de clustering

Se encarga de agrupar todas las observaciones en grupos parecidos. Solamente que se basa en características particulares para la formación de los grupos y la consiguiente separación de las observaciones. Al utilizar el algoritmo de clustering como el k-means, las imágenes similares se agrupan, y su vez el algoritmo permite detectar patrones en los datos. Esto sería de mucha utilidad para la predicción de coágulos de sangre debido a que el algoritmo se encargaría de agrupar todas las imágenes de la base de datos que demuestren que un paciente tiene coágulos de sangre. Y por lo tanto, se puede encontrar un patrón para poder determinar la cantidad de pacientes que sufrieron un accidente cerebrovascular.

## Selección de algoritmos a probar:

### Elección del algoritmo HOG

El algoritmo Histogram of Oriented Gradients (HOG) fue seleccionado por varias razones clave:

**Naturaleza de la Imagen:** Las imágenes en el conjunto de datos representan coágulos sanguíneos, y es esencial capturar las estructuras y formas presentes en las imágenes. HOG es eficaz para capturar los contornos y la estructura de las imágenes.

**Invarianza a las transformaciones:** Las características HOG son relativamente invariantes a las transformaciones locales y a cambios en la iluminación, lo que es esencial dada la variedad de imágenes en nuestro conjunto de datos.

**Simplicidad y eficiencia:** En comparación con redes neuronales profundas, como las CNN, HOG es más simple y computacionalmente menos intensivo, lo que facilita su implementación y adaptación.

## Elección del algoritmo CNN

Seleccionar un algoritmo de Redes Neuronales Convolucionales (CNN) para la detección de coágulos de sangre en imágenes de accidentes cerebrovasculares se justifica por varias razones fundamentales.

En primer lugar, las CNN están diseñadas específicamente para procesar imágenes y son altamente efectivas en la extracción de características visuales complejas de los datos médicos. Su capacidad para aprender automáticamente jerarquías de características, desde bordes simples hasta características más abstractas, las convierte en una elección natural para identificar las sutilezas de los coágulos de sangre en imágenes de alta resolución.

Además, las CNN son robustas frente a la variabilidad de las imágenes médicas, incluyendo diferencias en iluminación, orientación y escala, lo que es esencial en un entorno clínico donde las imágenes pueden variar ampliamente. Además, las CNN pueden ser entrenadas en grandes conjuntos de datos médicos, lo que permite aprovechar la abundancia de información disponible.

Su capacidad de generalización y adaptación a nuevas imágenes, combinada con la posibilidad de transferencia de aprendizaje desde modelos preentrenados, acelera y mejora significativamente el proceso de desarrollo de modelos de detección de coágulos de sangre en el contexto de accidentes cerebrovasculares, lo que convierte a las CNN en una elección poderosa y confiable en esta aplicación crítica de la atención médica.

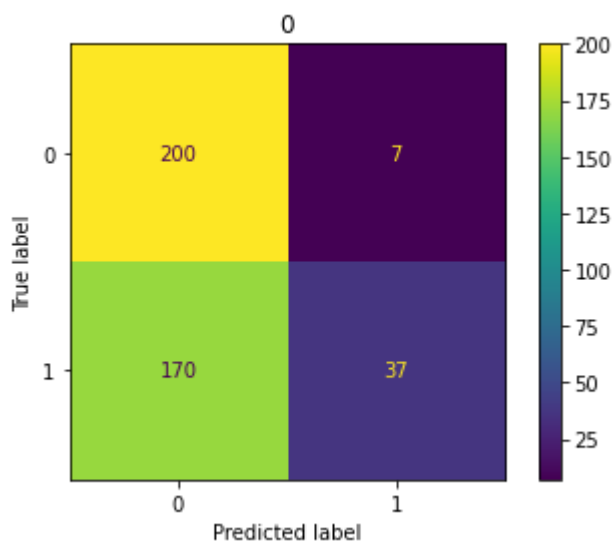
# Resultados

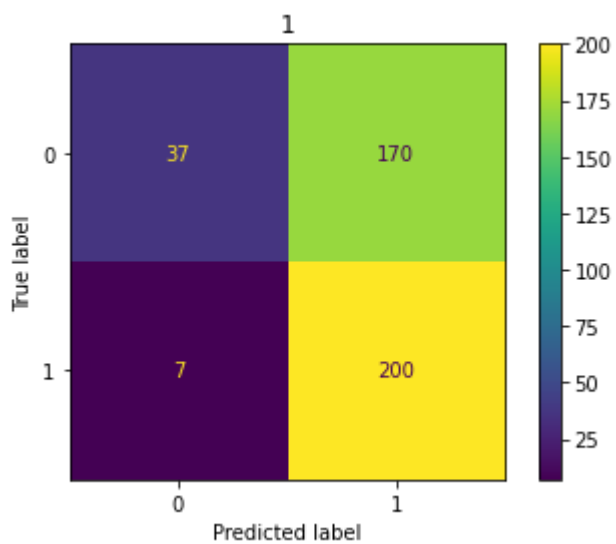
## Modelo HOG:



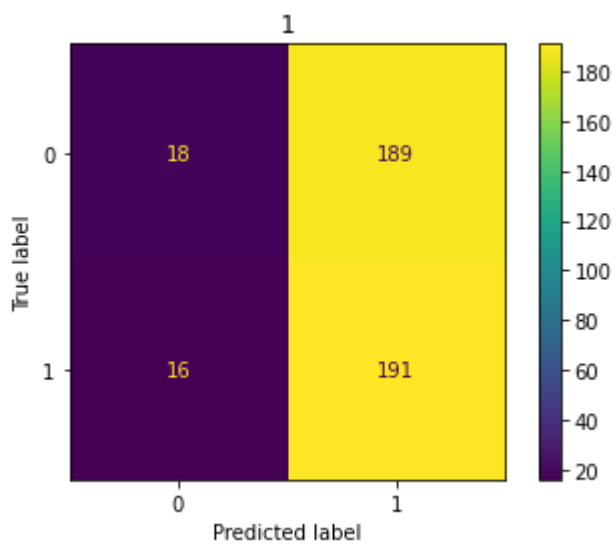
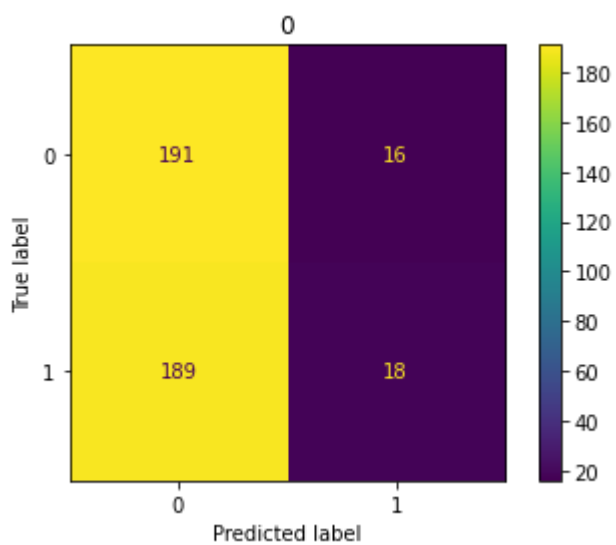
## Modelo CNN :

### Primer Modelo :





Segundo Modelo :



# Discusión sobre el desempeño de los modelos

## HOG

El modelo, después de ser entrenado con las características HOG y usando un clasificador SVM, mostró una precisión del 71.5%. Esto indica que el modelo acertó en aproximadamente 3 de cada 4 predicciones. Si bien el modelo mostró una precisión general decente, es esencial desglosar el desempeño por clase:

Clase CE: El modelo tuvo un desempeño notablemente mejor en la predicción de la clase CE, con una precisión del 76% y un recall del 90%. Esto indica que el modelo es muy bueno detectando coágulos de tipo CE cuando están presentes, pero tiende a clasificar erróneamente algunos coágulos de tipo LAA como CE.

Clase LAA: En contraste, la precisión para la clase LAA fue del 45% con un recall del 22%. Esto sugiere que el modelo tiene dificultades para identificar correctamente los coágulos de tipo LAA.

Las posibles razones de este desequilibrio en el desempeño podrían ser la distribución desequilibrada de clases dado que si hay muchas más imágenes de una clase que de otra, el modelo podría estar sesgado hacia la clase predominante, como también es posible que las características de la clase LAA sean más sutiles o variadas que las de la clase CE, lo que hace que sea más difícil para el modelo distinguir entre las dos.

## CNN

Modelo 1 :

Accuracy (precisión) : 77.54%

Loss (pérdida) : 0.4155

El Modelo 1 muestra un rendimiento bastante sólido con un 77.54% de precisión en los datos de prueba. Esto significa que aproximadamente el 77.54% de las predicciones son correctas. La pérdida (loss) de 0.4155 es relativamente baja, lo que sugiere que el modelo está aprendiendo bien de los datos de entrenamiento y generaliza razonablemente bien a los datos de prueba. Dado que se ha llegado a 15 épocas y se decidió detener el entrenamiento debido a las limitaciones de la memoria, es posible que el modelo hubiera continuado mejorando si se le hubiera permitido entrenar durante más épocas. Esto es especialmente relevante ya que se observó que el accuracy seguía aumentando antes de detener el entrenamiento.

## Modelo 2 :

Accuracy (precisión) : 49.76%

Loss (pérdida) : 0.6952

El Modelo 2 muestra un rendimiento bastante bajo en comparación con el Modelo 1. Con un accuracy del 49.76%, está prácticamente en el rango de una predicción aleatoria (50% en un problema binario). La pérdida de 0.6952 es alta, lo que indica que el modelo no está aprendiendo bien de los datos de entrenamiento y tiene dificultades para generalizar.

La diferencia significativa en el rendimiento entre los dos modelos puede atribuirse a varios factores, como la arquitectura de la red, los hiperparámetros, la cantidad y calidad de los datos de entrenamiento y la inicialización de los pesos. Podría ser valioso revisar la arquitectura y los hiperparámetros del Modelo 2 y considerar posibles ajustes para mejorar su rendimiento.

## **Bibliografía :**

- Nodriz Tech *Tipos de Algoritmos Machine Learning* Recuperado de : <https://nodrizatech.com/knowhow/algoritmos-machine-learning/#:~:text=Existen%20dos%20grandes%20tipologías%20de,los%20de%20aprendizaje%20no%20supervisado.>
- Pep Canals (2 de marzo de 2022) *Algoritmos de clustering : el futuro del marketing sin cookies* Recuperado de : <https://www.cyberclick.es/numerical-blog/algoritmos-de-clustering-el-futuro-del-marketing-sin-cookies#:~:text=Un%20algoritmo%20de%20clustering%20es,cuenten%20objetos%20similares%20entre%20sí.>
- The Black Box Lab *Machine Learning : Diferencias entre algoritmos de clasificación y regresión* Recuperado de : <https://theblackboxlab.com/2022/05/06/machine-learning-diferencias-entre-algoritmos-clasificacion-regresion/>
- MathWorks *¿Qué son las redes neuronales convencionales ?* Recuperado de : <https://es.mathworks.com/discovery/convolutional-neural-network-matlab.html>
- Data Scientist (16 de diciembre de 2021) *Convolutional Neural Network : Definición y funcionamiento* Recuperado de : <https://datascientest.com/es/convolutional-neural-network-es>
- Microsoft (11 de julio de 2023) *Entrenamiento de un modelo con detección de objetos con TensorFlow* Recuperado de : <https://learn.microsoft.com/es-es/windows/ai/windows-ml/tutorials/tensorflow-train-model>
- Microsoft (17 de septiembre de 2023) *Detección de objetos mediante Fast R-CNN* Recuperado de : <https://learn.microsoft.com/es-es/cognitive-toolkit/object-detection-using-fast-r-cnn>