

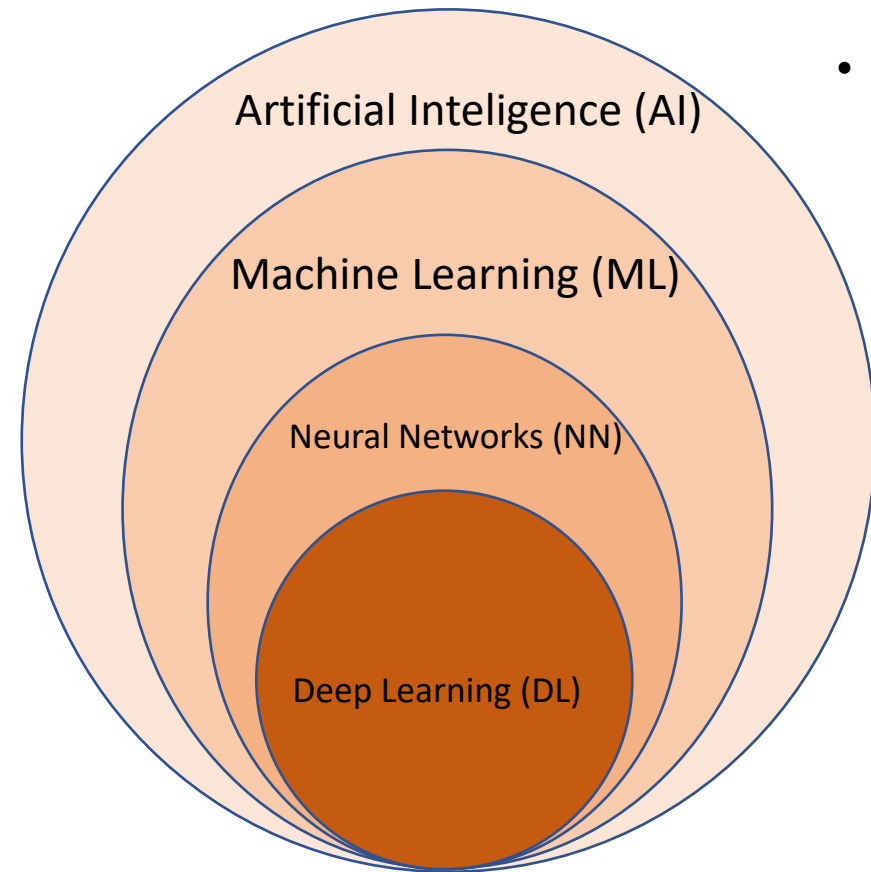
Anatomía e implementación de un método numérico de 'Deep Learning' por medio de optimización estocástica y sus aplicaciones industriales

Dr. Alberto Santamaria Pang

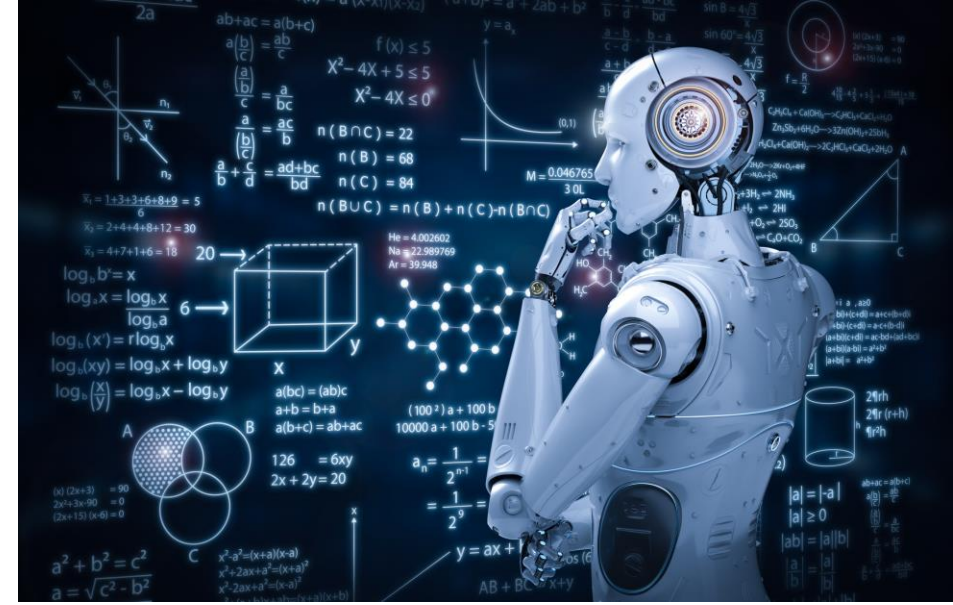
Contenido

- Introducción
- Estructura de algoritmos de aprendizaje
- Aplicaciones
 - Manufactura
 - Turbinas de avión
 - Análisis de imágenes medicas y biomédicas
 - Radiología
 - Patología Molecular

Deep Learning



- Habilidad de una maquina de imitar inteligencia humana
- Automáticamente aprende de 'errores'
- Inspirados en modelos biológicos
- Redes neuronales de aprendizaje profundo
 - 'Deep Neural Networks'



[2159x1389 \(parentology.com\)](https://www.parentology.com/2159x1389)

Si pudiéramos modelar el funcionamiento neurona.....

AUTOMATIC THREE-DIMENSIONAL MORPHOLOGICAL RECONSTRUCTION OF NEURONS

A Dissertation

Presented to

the Faculty of the Department of Computer Science

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Alberto Santamaría-Pang

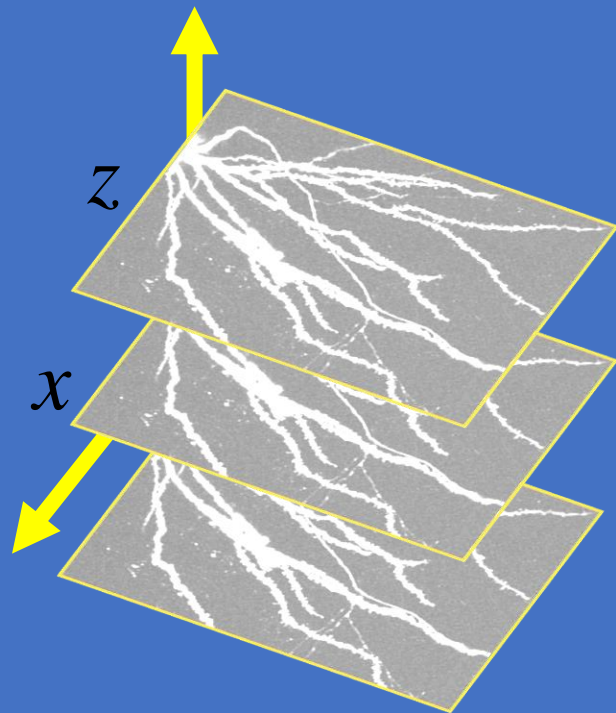
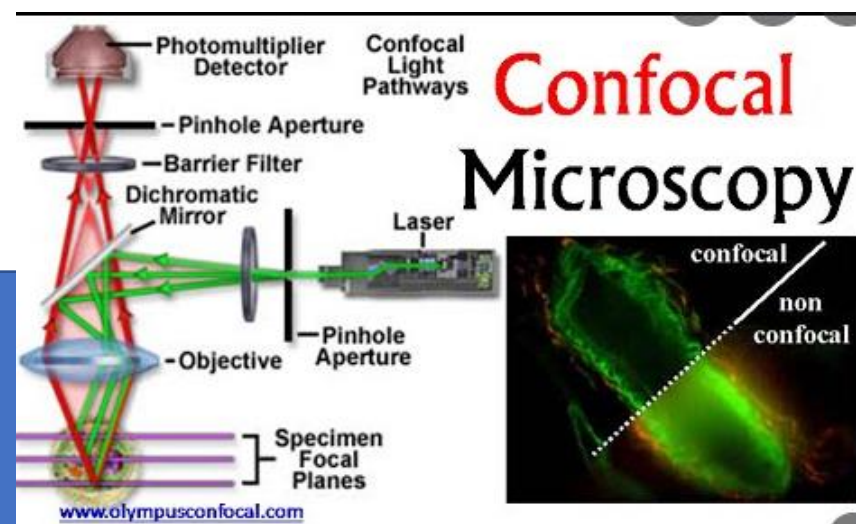
December 2007



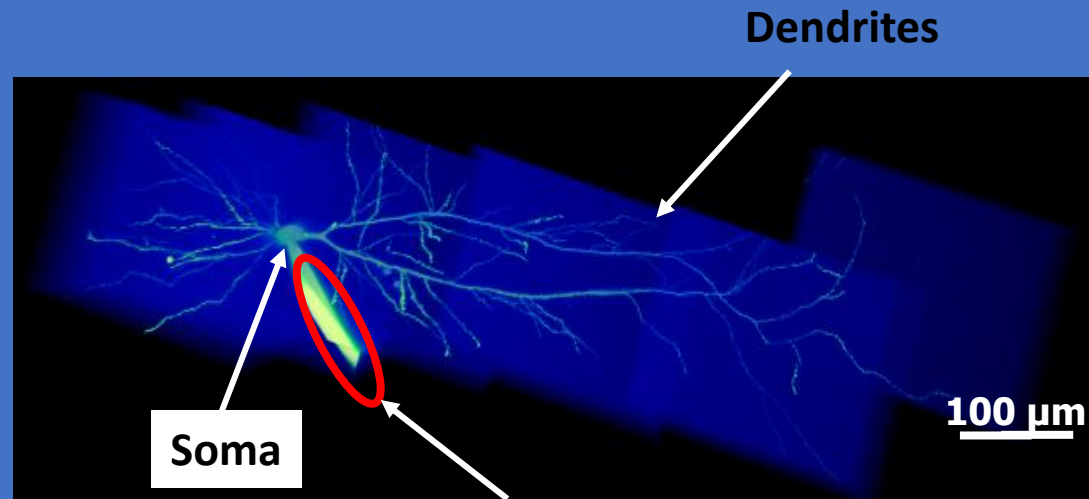
Computational Biomedicine Lab

Changing the Way People Look at Computers
Computers *People*

3D Optical Imaging

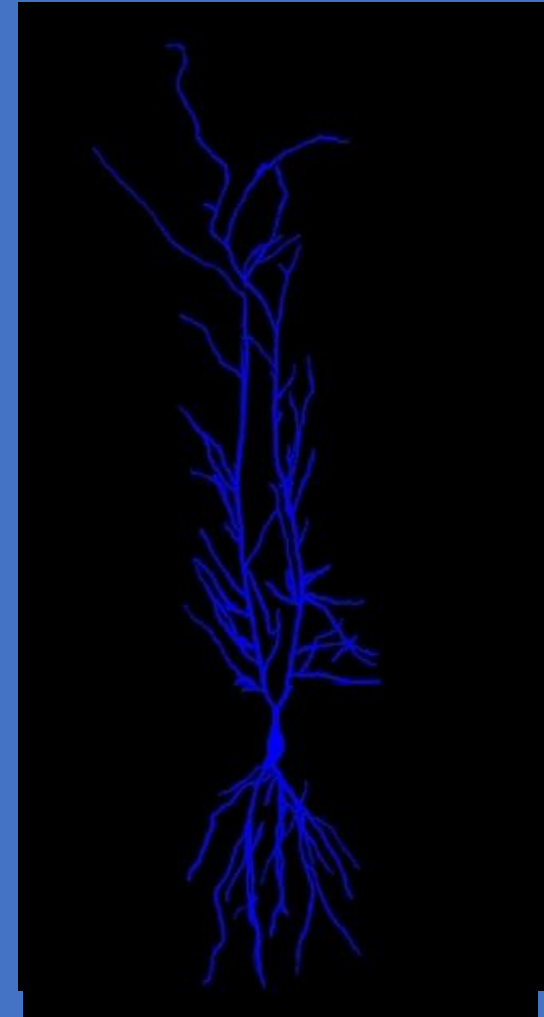
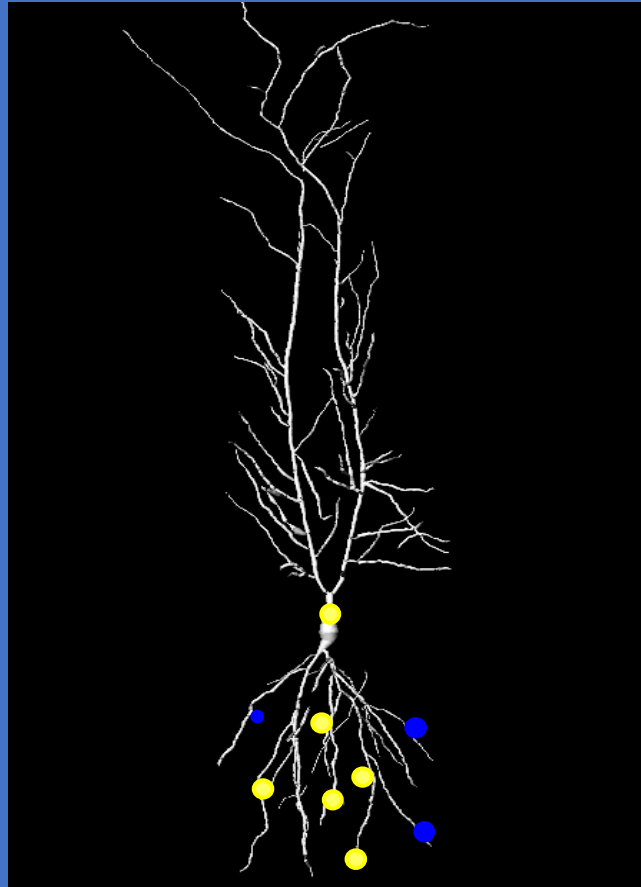
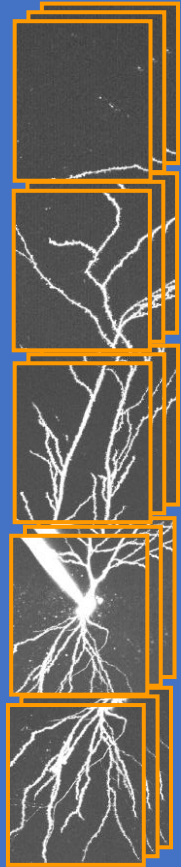


Individual optical sections



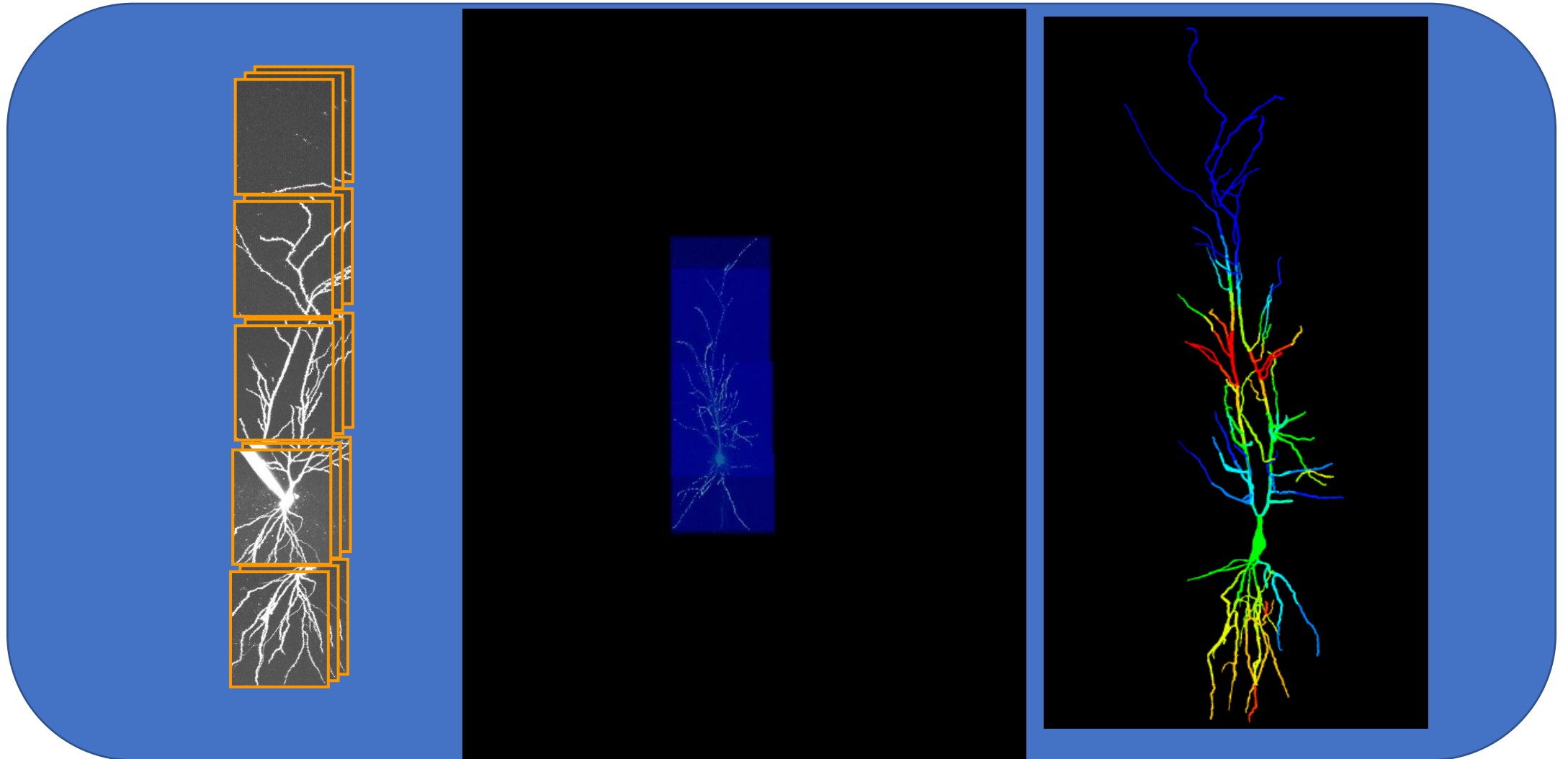
Stacks of optical sections

Motivación – Modelado Neuronal



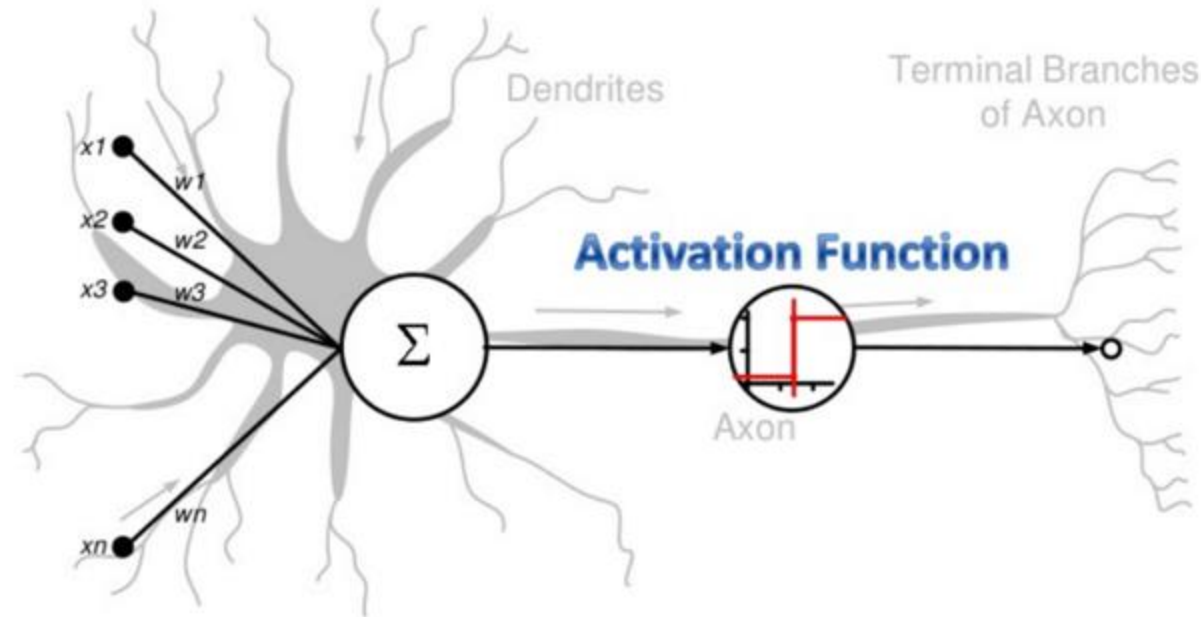
Credits: C.M. Colbert

Motivación – Modelado Neuronal



Credits: C.M. Colbert

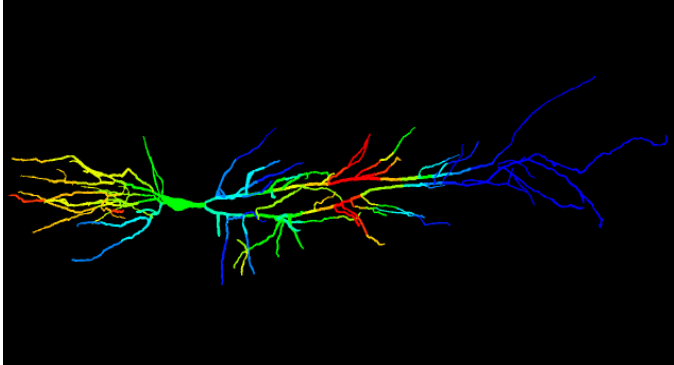
Motivación – Modelado Neuronal



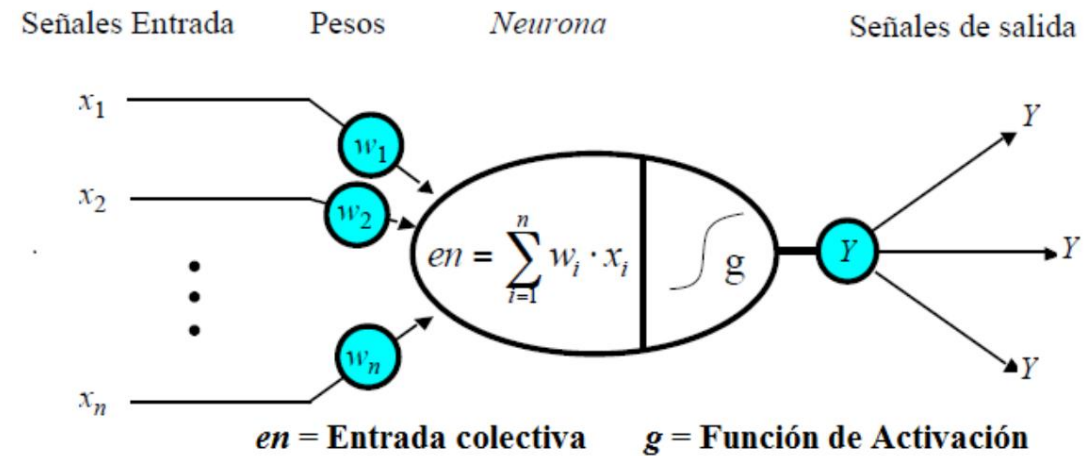
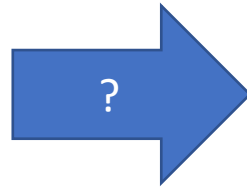
McCulloch, Warren; Walter Pitts (1943). "A Logical Calculus of Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics*. 5 (4): 115–133. [doi:10.1007/BF02478259](https://doi.org/10.1007/BF02478259).

Motivación – Modelado Neuronal

- Modelo simplificado de una neurona



Biología



Activation Functions

Sigmoid
 $\sigma(x) = \frac{1}{1+e^{-x}}$



Leaky ReLU
 $\max(0.1x, x)$



tanh
 $\tanh(x)$



Maxout
 $\max(w_1^T x + b_1, w_2^T x + b_2)$



ReLU
 $\max(0, x)$



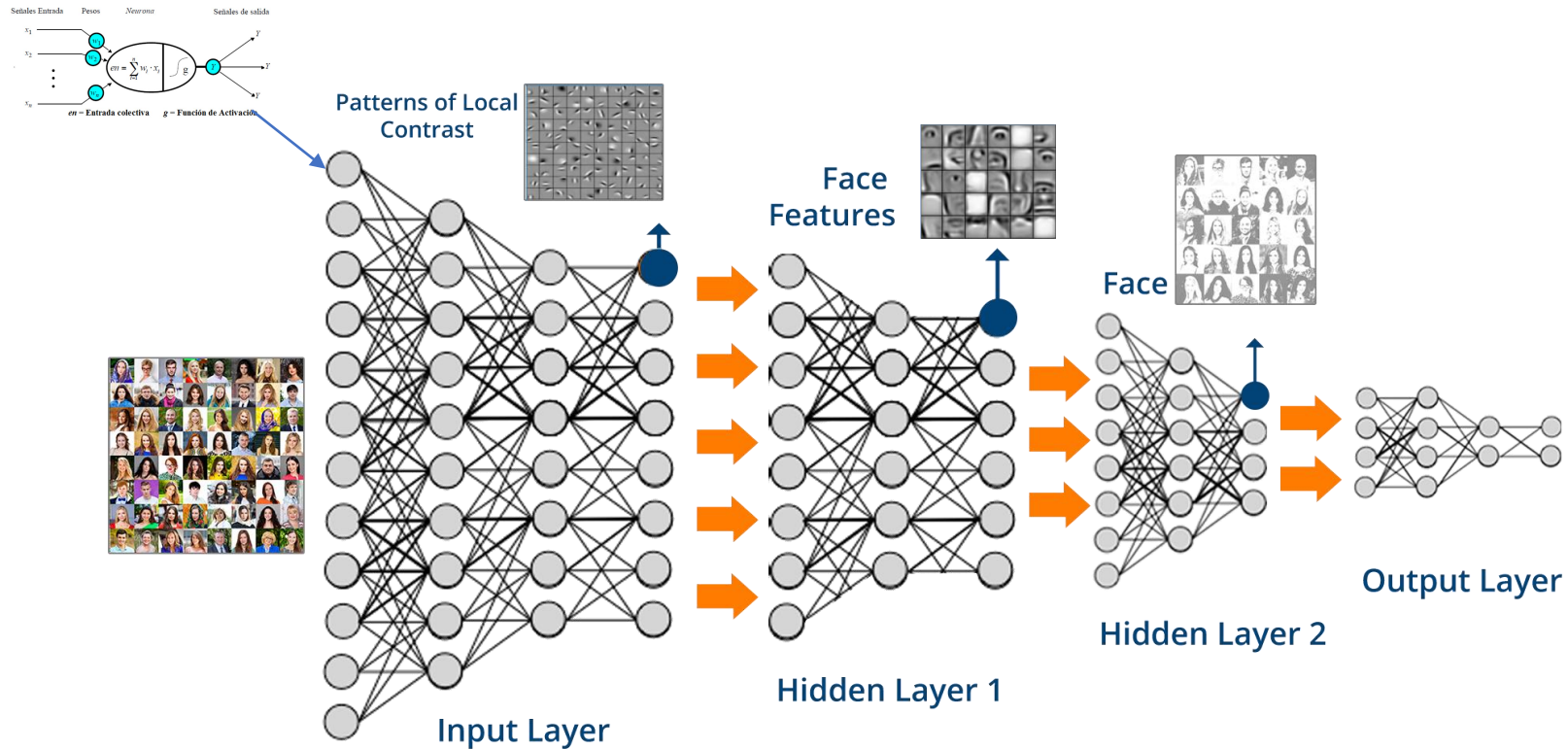
ELU
 $\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$



En general no lineal

[red neuronal, activacion de una neurona - Bing images](#)

Concepto de Red Neuronal



Modelo:
 $f_w(X) \approx Y$

ResNet-50 tiene mas 23 millones de parámetros
¿Cómo optimizar?

Descenso por Gradiente



Modelo:

$$f_w(X) \approx Y$$

Función de costo J :

$$J(w) = \frac{1}{N} \sum_{i=1}^N ||f_w(x_i) - y_i||^2$$

dos veces continuamente diferenciable

Minimizar:

$$\min_w J(w)$$

Descenso por gradiente:

$$w^{t+1} = w^t - \alpha \nabla J(w^t)$$

Dos problemas:

- i) convexidad
- ii) limitación de procesamiento y memoria





Descenso por gradiente descenso de gradiente estocástico (SGD)

Idea: Vamos a seleccionar un subconjunto de todos los datos y vamos a calcular el gradiente:

$$\mathbb{E}_{v \sim D} [J_v(w^t)]$$

v es una muestra aleatoria de una distribución de probabilidad.

Se puede expresar como:

$$w^{t+1} = w^t - \alpha \nabla J_{v_t}(w^t)$$



Teorema de convergencia de SGD

1) J_v es L-'smooth' (convexidad)

$$\mathbb{E}_{v \sim D} [||\nabla J_v(w) - \nabla J_v(w^*)||] \leq 2L [J_v(w) - J_v(w^*)]$$

2) El 'ruido' estocástico del gradiente es finito:

$$\sigma^2 = \mathbb{E}_{v \sim D} [||J_v(w^{t+1})||]$$

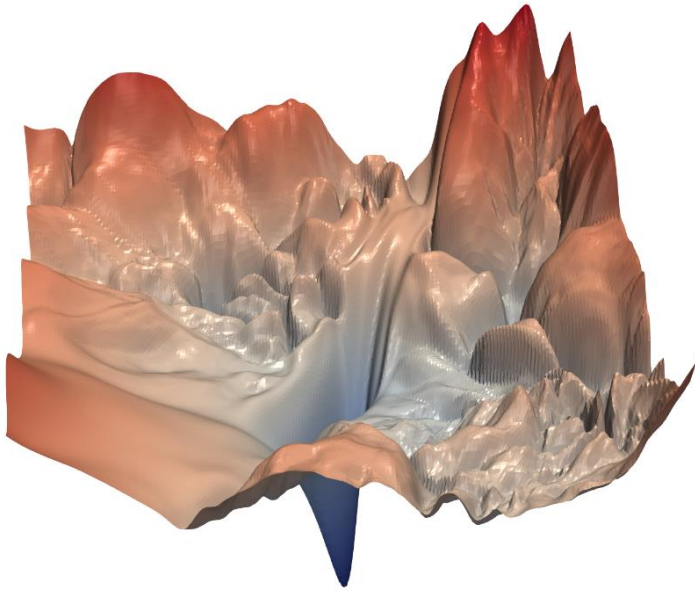
Entonces:

$$\mathbb{E}_{v \sim D} [J_v(w^{t+1}) | w^t] \leq \frac{1}{\alpha(t+t^0)}, \quad \alpha = \frac{2\mu}{L^2 + \sigma^2}, \mu \text{ constante}$$

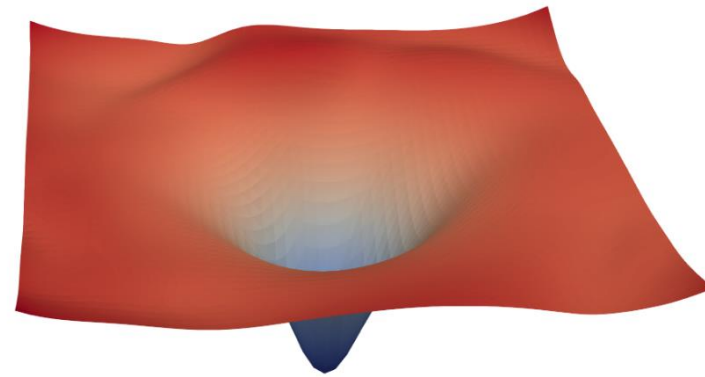
Interpretación: Cuanto más fuerte es la convexidad, más rápida es la convergencia.



Convexidad



(a) without skip connections



(b) with skip connections

Figure 1: The loss surfaces of ResNet-56 with/without skip connections. The proposed filter normalization scheme is used to enable comparisons of sharpness/flatness between the two figures.

Li, “Visualizing the Loss Landscape of Neural Nets”, NeurIPS 2018



Demo



Clasificación de imágenes de dígitos del 0 al 9

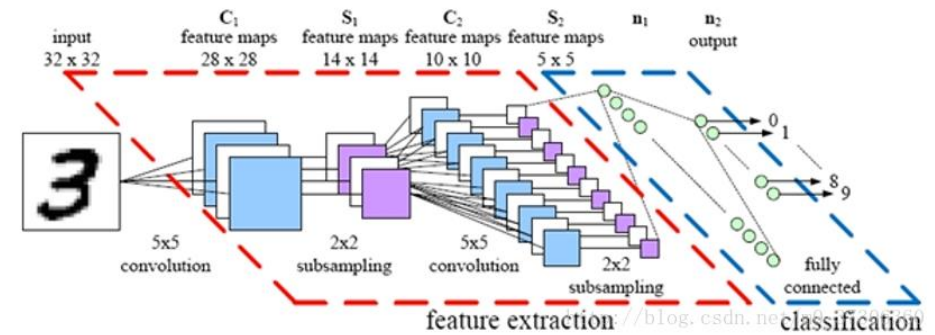
Datos:

- Creados en 1998
- Modified National Institute of Standards and Technology (MNIST)
- 60,000 imágenes de entrenamiento
- 10,000 imágenes de prueba



Imágenes monocromáticas de 8bits:
cada pixel tiene 256 valores

CNN

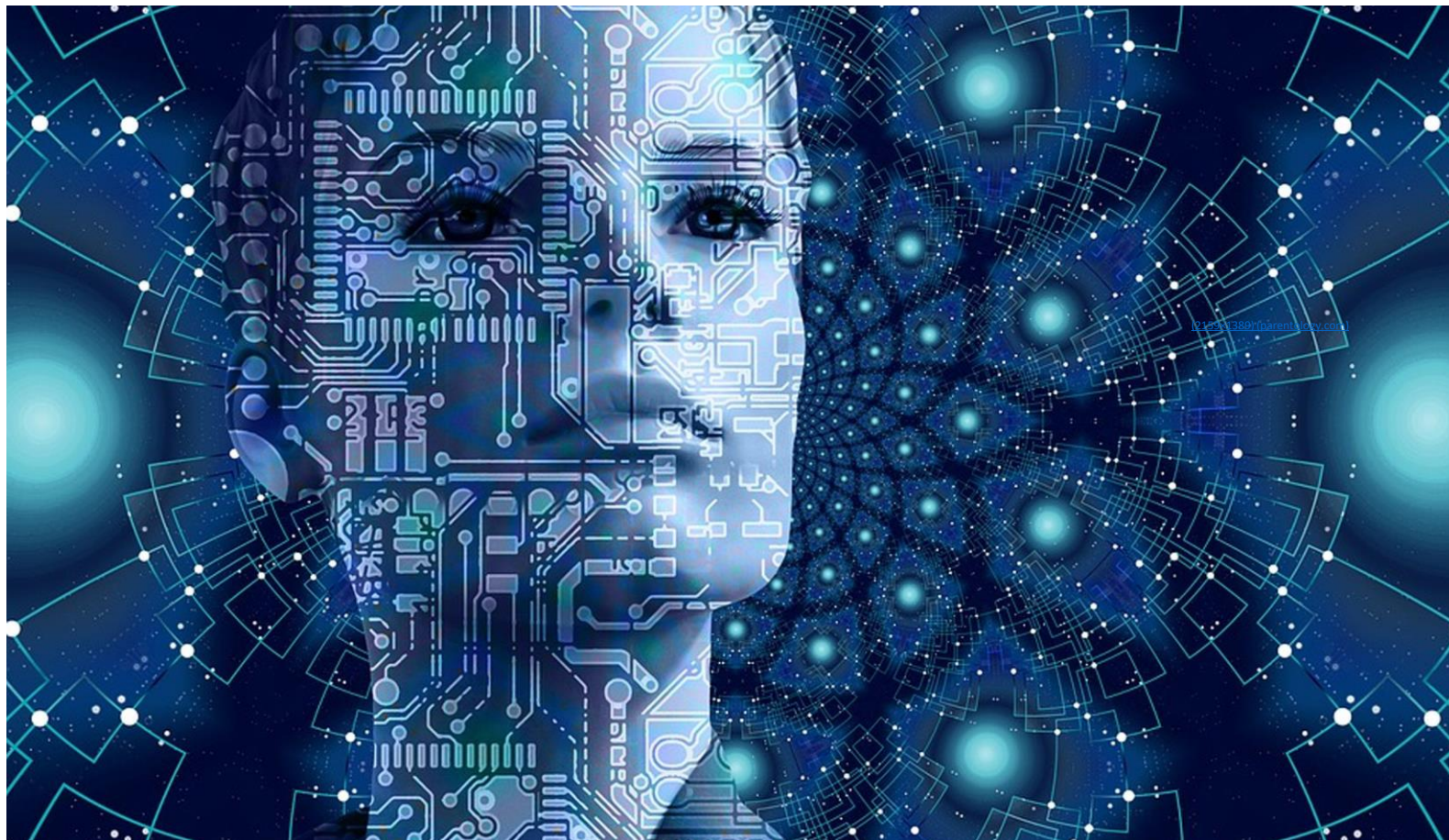


$$f_w(X) \approx Y$$

Créditos: Hoe Jiun Tian

<https://github.com/jiuntian/pytorch-mnist-example/blob/master/pytorch-mnist.ipynb>

Aplicaciones



Control de Calidad en Manufacturación

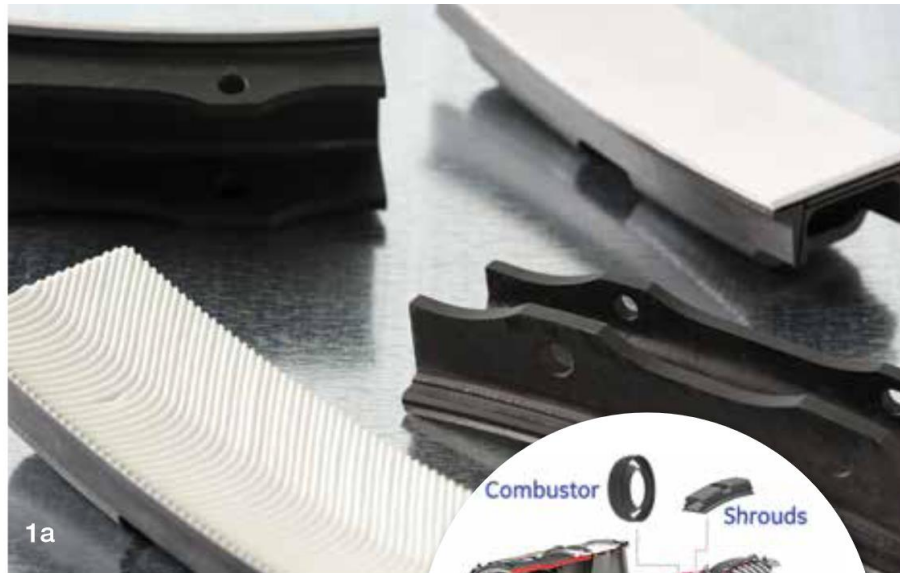
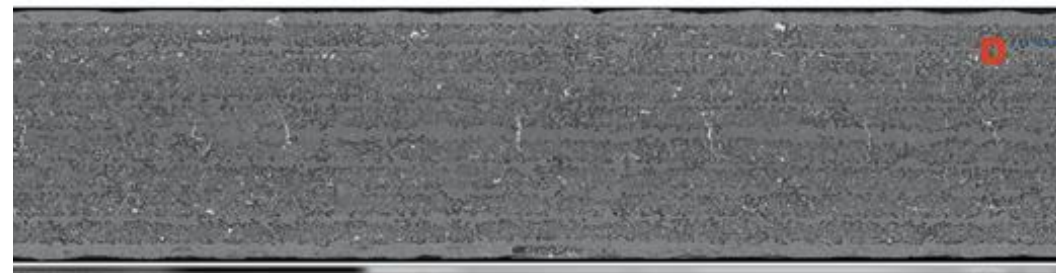


Figure 1a. CMC high-pressure turbine shrouds.

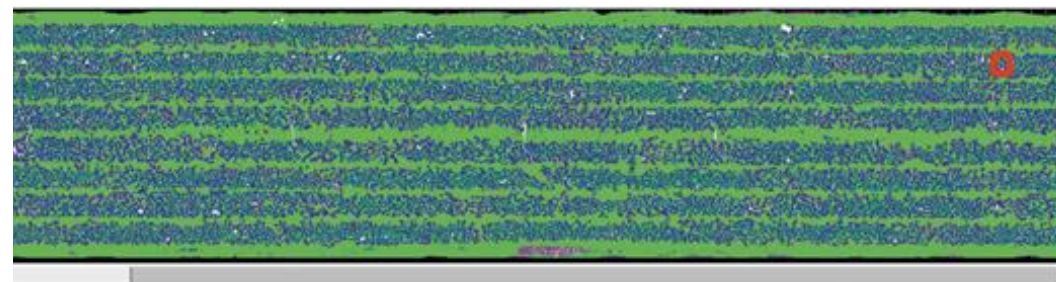
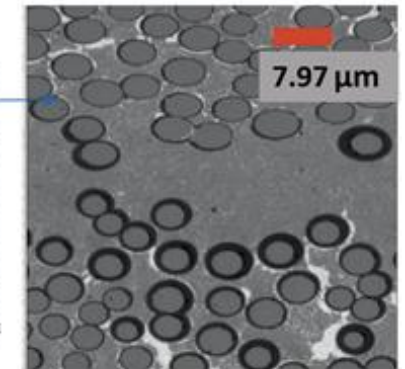
Figure 1b. Schematic drawing shows use of CMC components in jet engine.

coating steps to apply thin coatings on

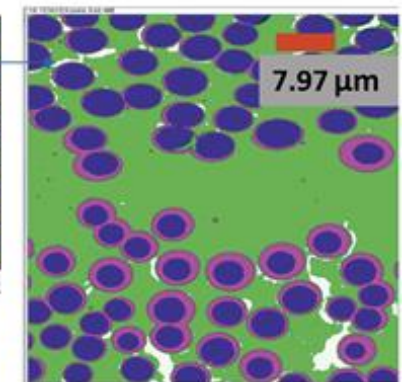
Jim Steibe, "Ceramic matrix composites taking flight at GE Aviation", American Ceramic Society Bulletin, Vol. 98, No. 3, 2019



Raw Data

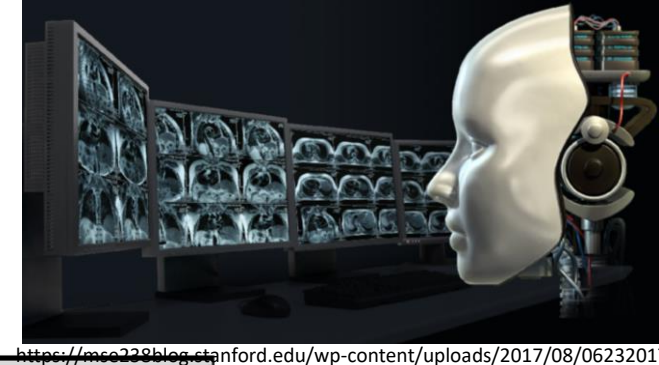


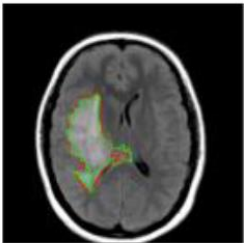
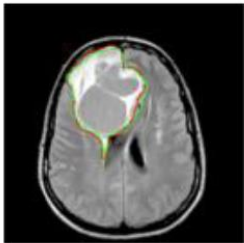
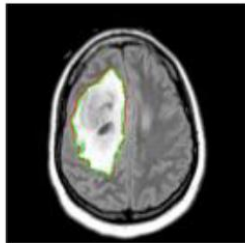
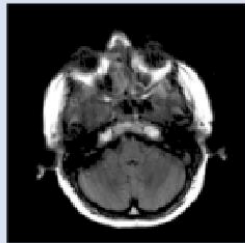
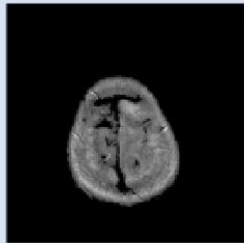

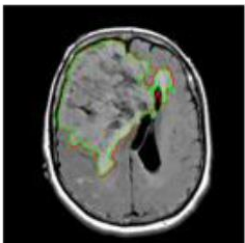
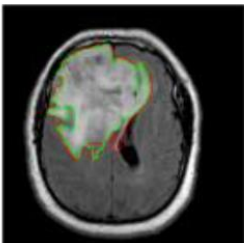
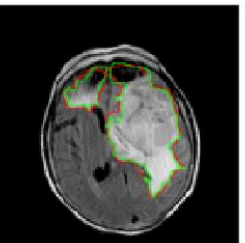
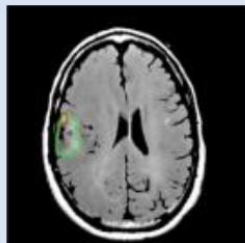
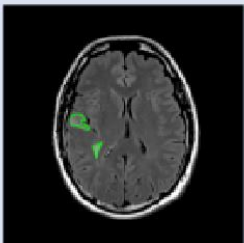
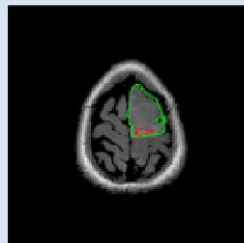
Semantic Segmentation



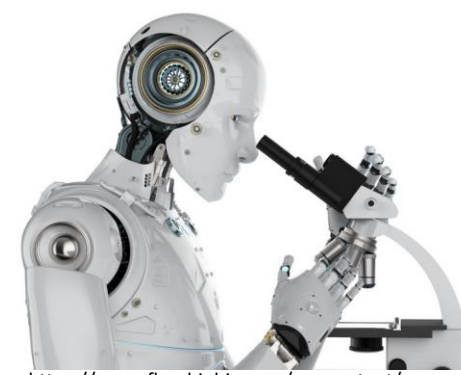
Kareem S. Aggour, et al, "Artificial intelligence/machine learning in manufacturing and inspection: A GE perspective", MRS Bulletin, 44(7), 545-558. doi:10.1557/mrs.2019.157

Radiología: Segmentación de tumor cerebral



Tumor Present			Tumor Absent		
CS_6669 Slice 13 357, 6563, 3956, *	DU_5855 Slice 16 5886, 4043, 5114, *	DU_A5TW Slice 20 5886, 3101, 5114, *	CS_5396 Slice 7 657, 653, 6531, *	HT_7856 Slice 29 657, 653, 6613, *	HT_A616 Slice 7 657, 653, 6531, *
					
Large Area			Small Area		
DU_7018 Slice 21 7313, 4626, 8872, *	HT_7694 Slice 12 8584, 4043, 5114, *	HT_7882 Slice 18 6670, 3578, 525, *	DU_7306 Slice 25 657, 2863, 2863, *	FG_7637 Slice 30 657, 653, 6613, *	DU_8166 Slice 29 657, 653, 6613, *
					

Patología Molecular: Cuantificar de células del sistema inmunológico



<https://www.flagshipbio.com/wp-content/uploads/Pathology-AI-poster-web-1024x768.png>

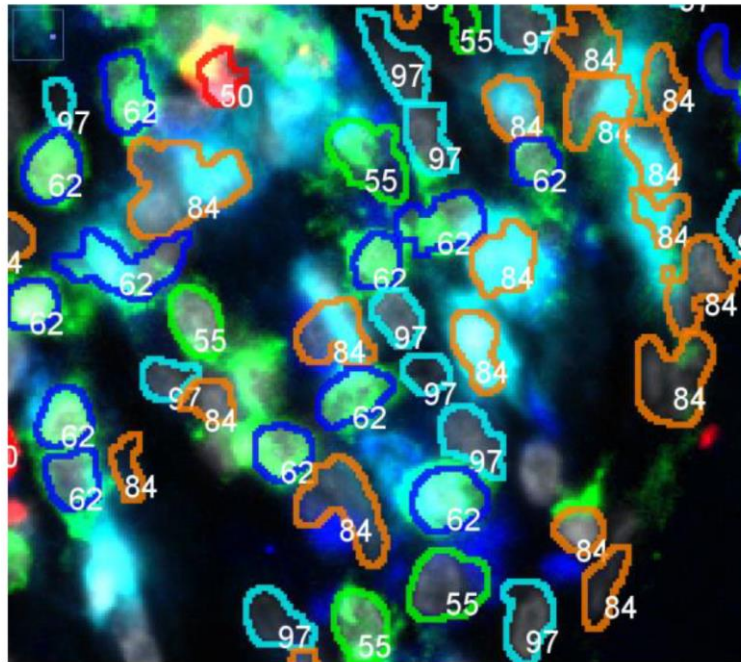


Fig. 4: Color composite image: CD20 (B-Cell), CD3 (T-Cell), CD68 (Macrophage), Claudin1 and DAPI (Cell Nuclei), in Red, Green, Blue, Cyan and Gray respectively. The segmented stroma nuclei is overlaid with the symbol identifier for each cell. We observe that the different cell phenotype markers are associated with distinct symbols, but are same for the cells with the same marker. The symbols generated correspond to Experiment 1 in Table 1 and Fig. 3(a).

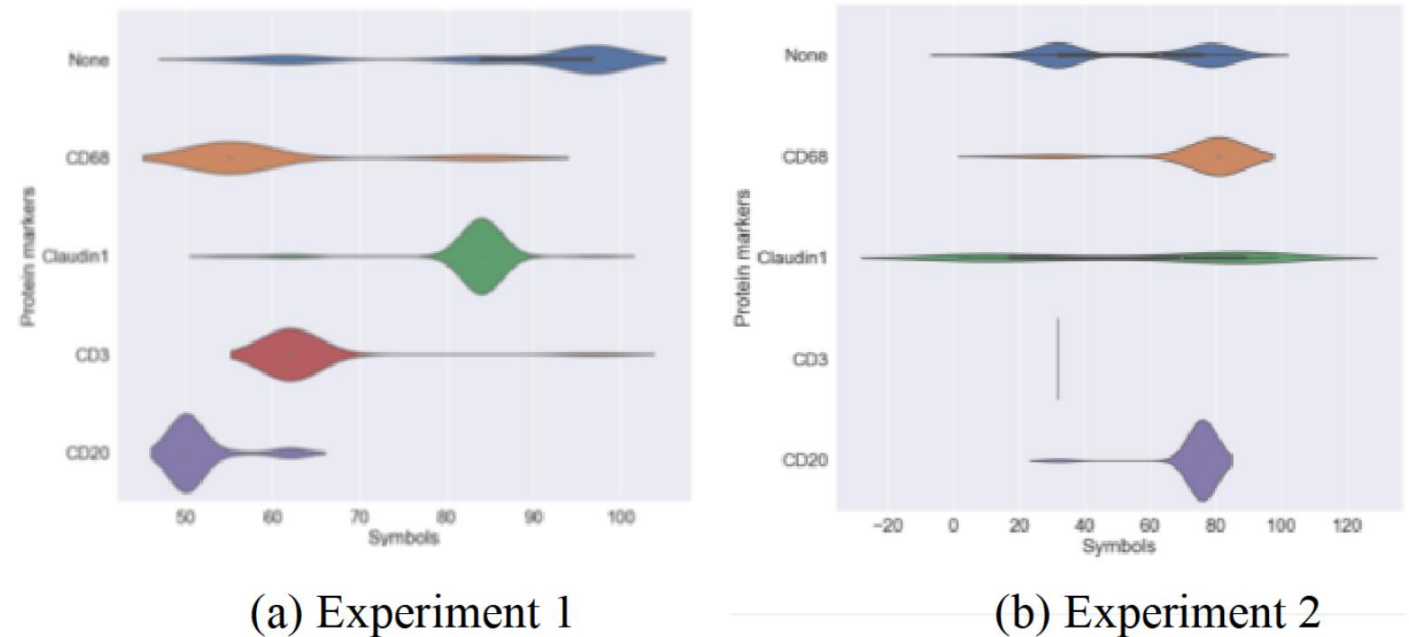


Fig. 3: Violin plots of the distributions of symbols from 1-100 with respect to the protein markers in (a) Experiment 1 and (b) Experiment 2

Conclusión

- Actualmente es el mejor momento de la historia de la humanidad para ser matemático.

Agradecimientos

- Mat. Rafael Davis Velati
- Dr. Homero Ríos Figueroa
- M.I.A. Patricia González Gaspar
- Profesores de la Facultad de Matemáticas
- Compañeros de generación