



Clasificación de los géneros de la subfamilia *Coronavirinae* basada en firmas genómicas

Abasolo-Cortés, Natalia., Gómez-Mora, Héctor E. y Pimentel-Ruíz, Carlos.

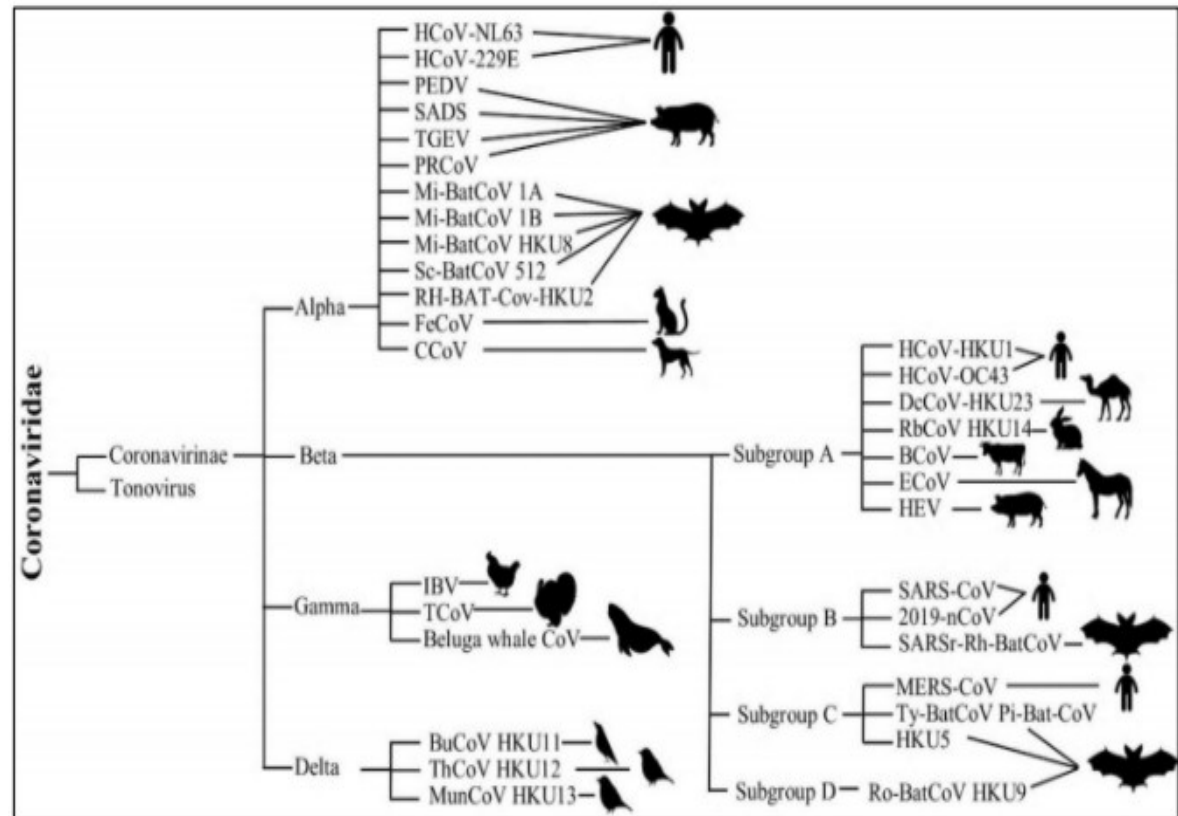
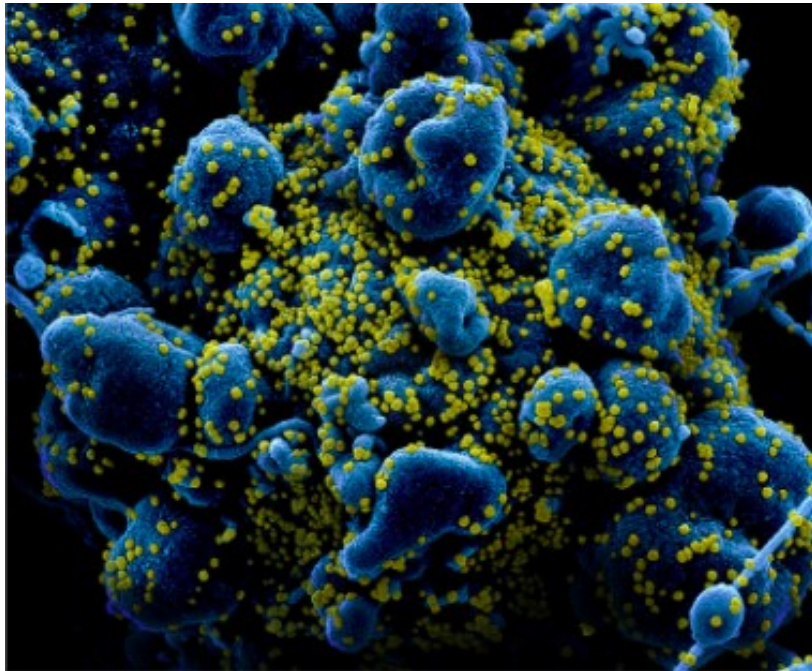
Universidad Nacional Autónoma de México
Facultad de Ciencias
Ciencias de la Computación
Genómica Computacional 7075

Introducción

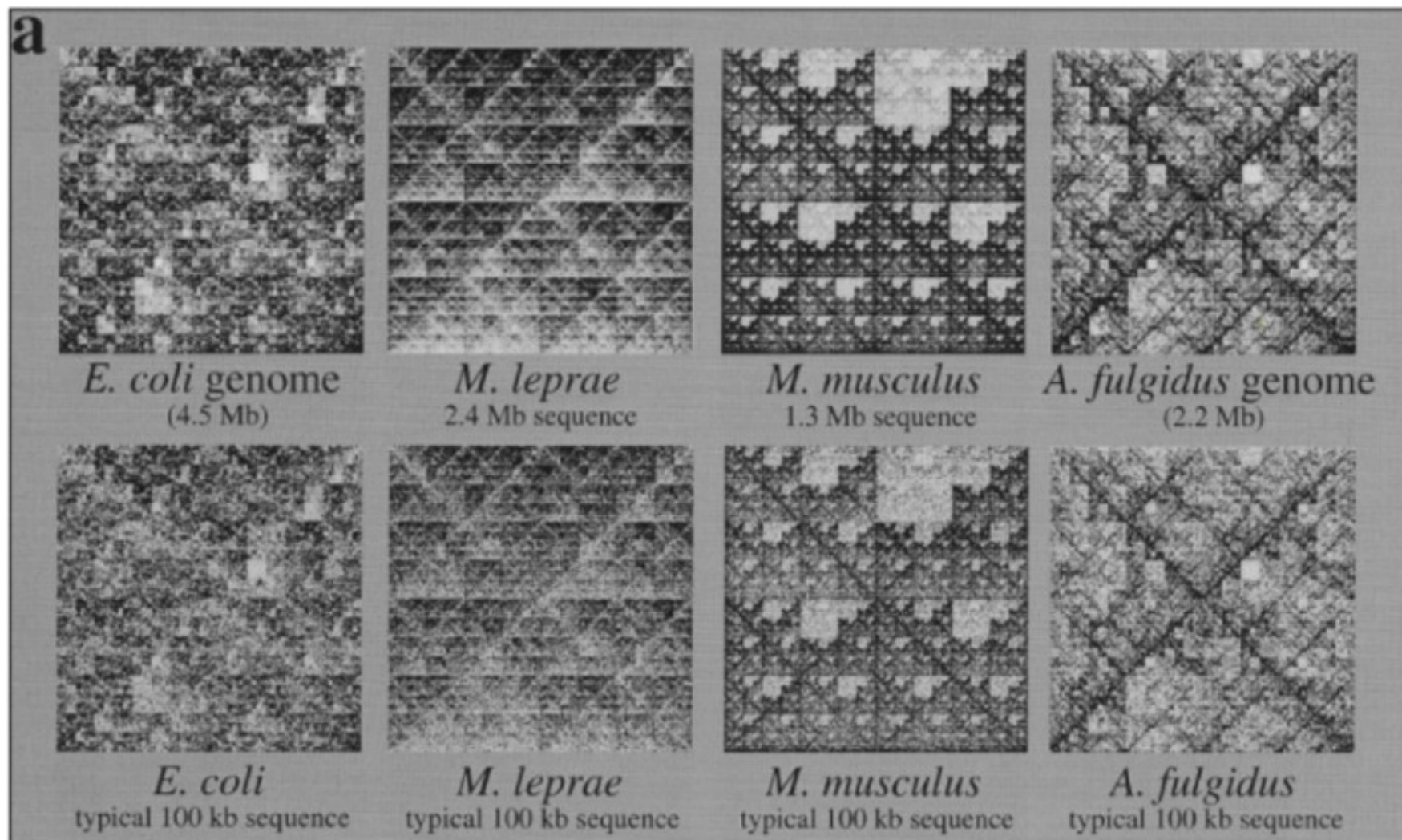
Grupos	Características	Tipo de ácido nucleico
ADN		
Grupo I	Doble hélice	Algunos bacteriófagos, adenovirus, herpesvirus, papovavirus (virus del papiloma humano o VPH)
Grupo II	Hélice simple	Otros bacteriófagos, parvovirus
ARN		
Grupo III	Doble hélice	Rotavirus
Grupo IV	Hélice simple, polaridad positiva	Coronavirus, virus de Zika, virus de la fiebre amarilla, del dengue, del Nilo occidental, de la hepatitis C, del mosaico del tabaco, de la polio, de la hepatitis A, de la hepatitis E, de la rubeola
Grupo V	Hélice simple, polaridad negativa	Virus de la gripe, de la rabia, de la parotiditis (paperas), del sarampión, de la fiebre hemorrágica del Ébola y de Marburgo
ADN o ARN, de transcripción inversa		
Grupo VI	Hélice simple-ARN	Retrovirus (virus de la inmunodeficiencia humana o VIH)
Grupo VII	Doble hélice-ADN	Hepadnavirus (virus de la hepatitis B)

CORONAVIRUS (CoV):

- Virus con corona
- Proteínas en su superficie.
- Estrategia especializada de replicación.
- Genoma de gran tamaño.



Juego del Caos. Conjunto de reglas y parámetros que genera un fractal.



Machine Learning :

Redes Neuronales Convolucionales (CNN)

Conjunto de entidades denominadas neuronas artificiales interconectadas que emiten y reciben señales.

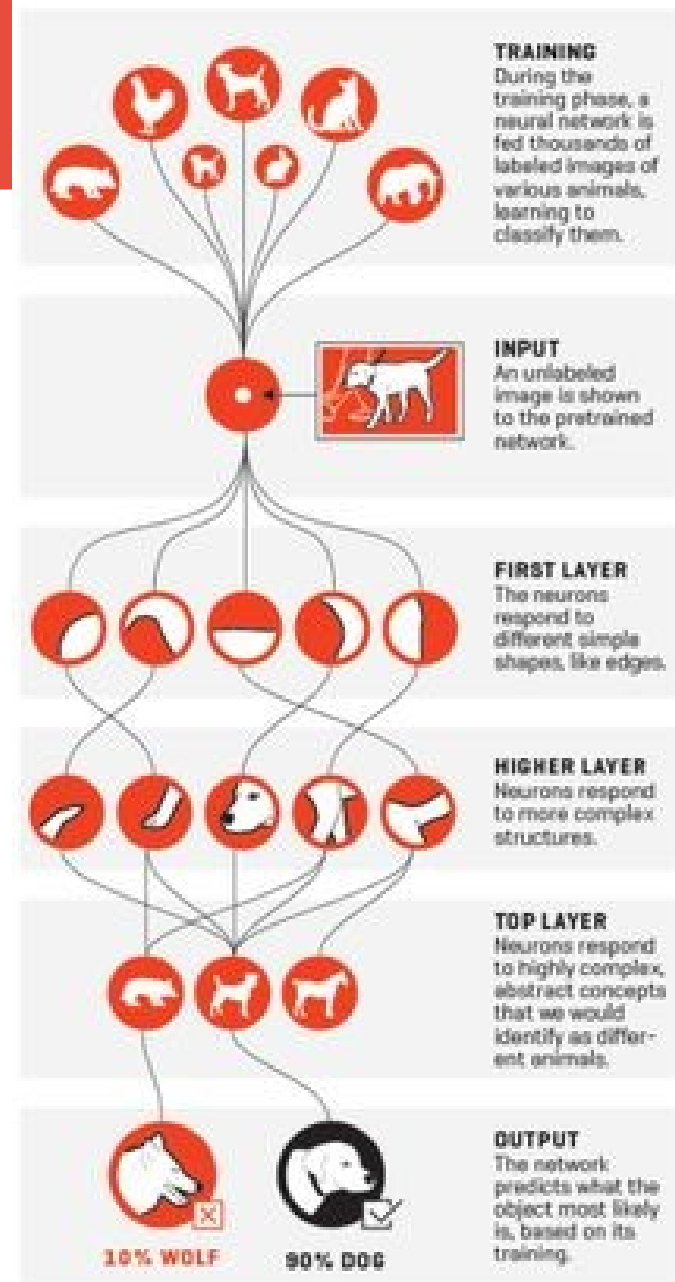
Esquema básico: Capa de entrada y de salida, capas ocultas.

Modelo supervisado: Entrenamiento previo a partir de un conjunto de datos específico, “conjunto de entrenamiento”, “conjunto de prueba”.

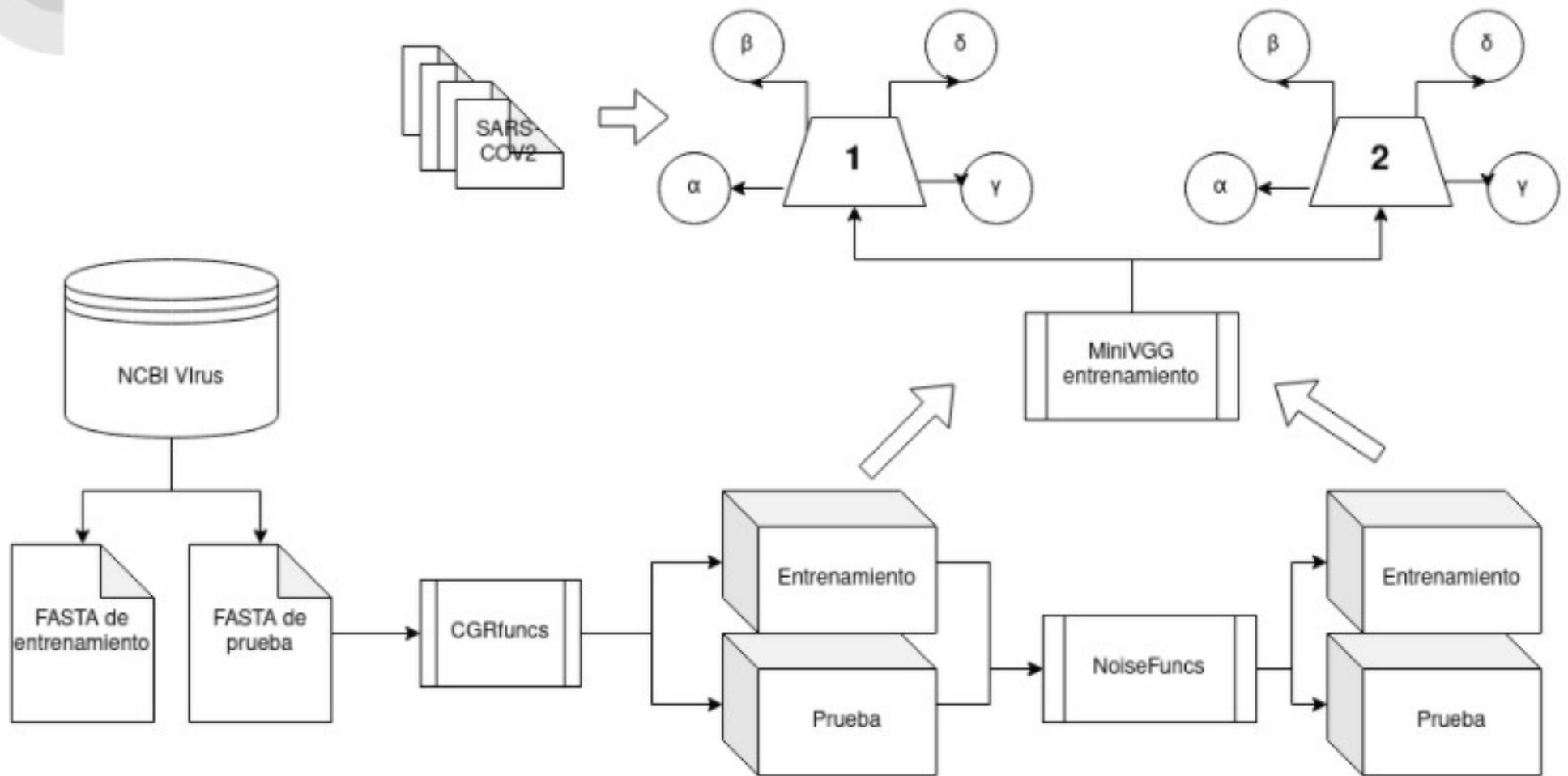
Las capas intermedias constituyen una secuencia de capas de convolución, activación y pooling, junto con capas tradicionales.

OBJETIVO: Generar una red neuronal convolucional que permita identificar a qué género pertenecen algunas secuencias no clasificadas

HOW NEURAL NETWORKS RECOGNIZE A DOG IN A PHOTO



Metodología



Chaos Game Representation (CGR)

Algoritmo

Dado un 4-gono asociado a A,C,T, G y una secuencia w

Se inicializa un conjunto S .

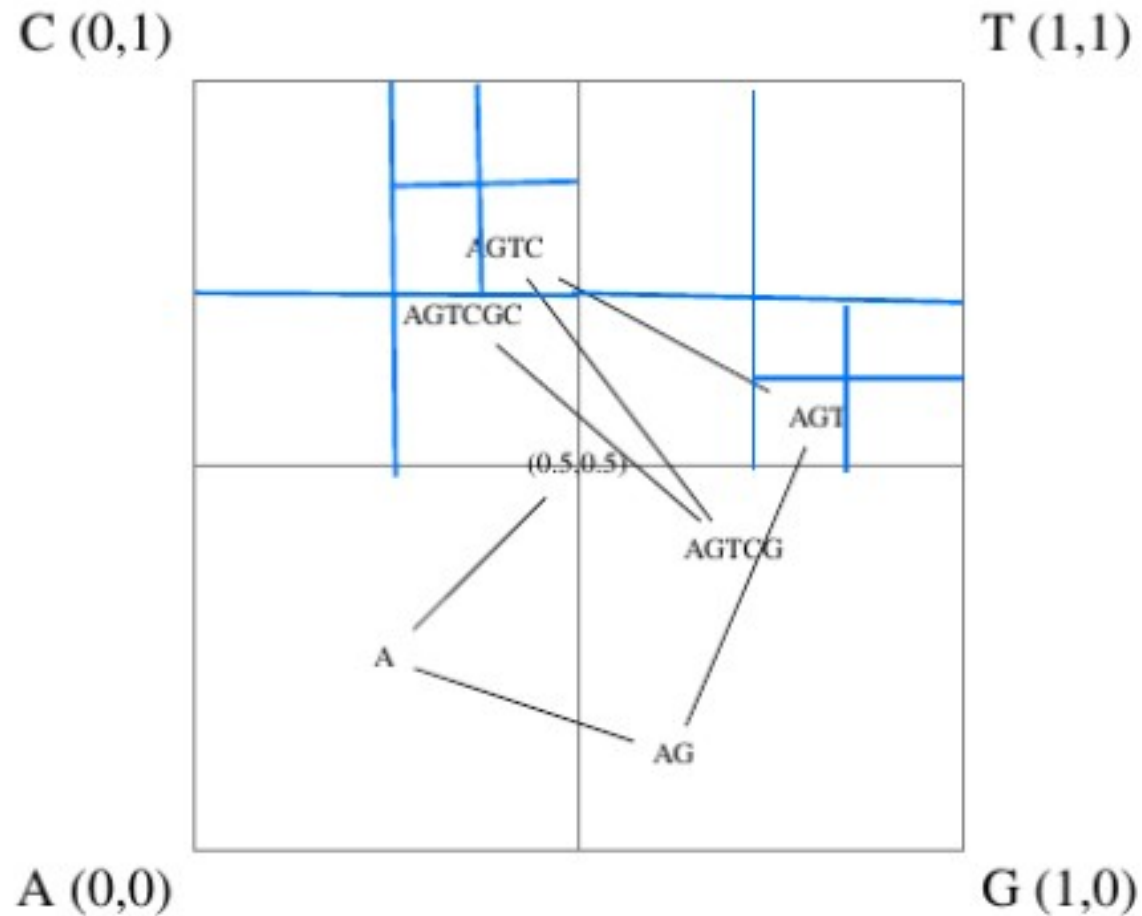
Se lee el primer símbolo y se grafica el punto a la mitad de la distancia entre el centro y el vértice correspondiente al símbolo.

Se agrega el punto a S .

Por cada símbolo leído se grafica el punto a la mitad de la distancia entre el último punto agregado y el vértice correspondiente al símbolo.

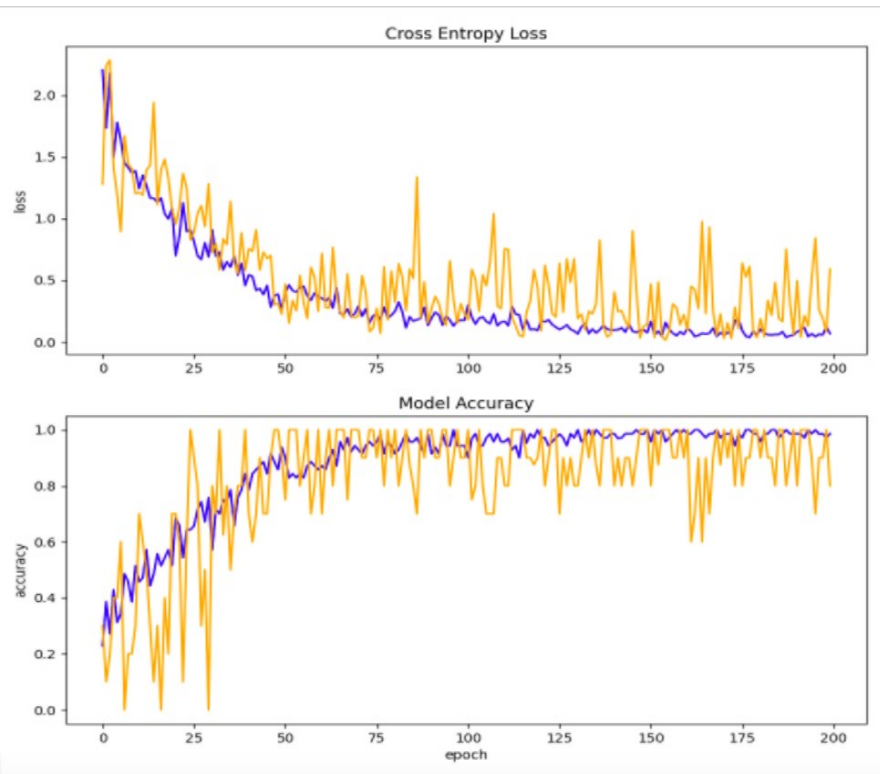
Repetir hasta que la secuencia haya sido procesada en su totalidad.

Chaos Game Representation (CGR)



CGR para la
secuencia AGTCGC.

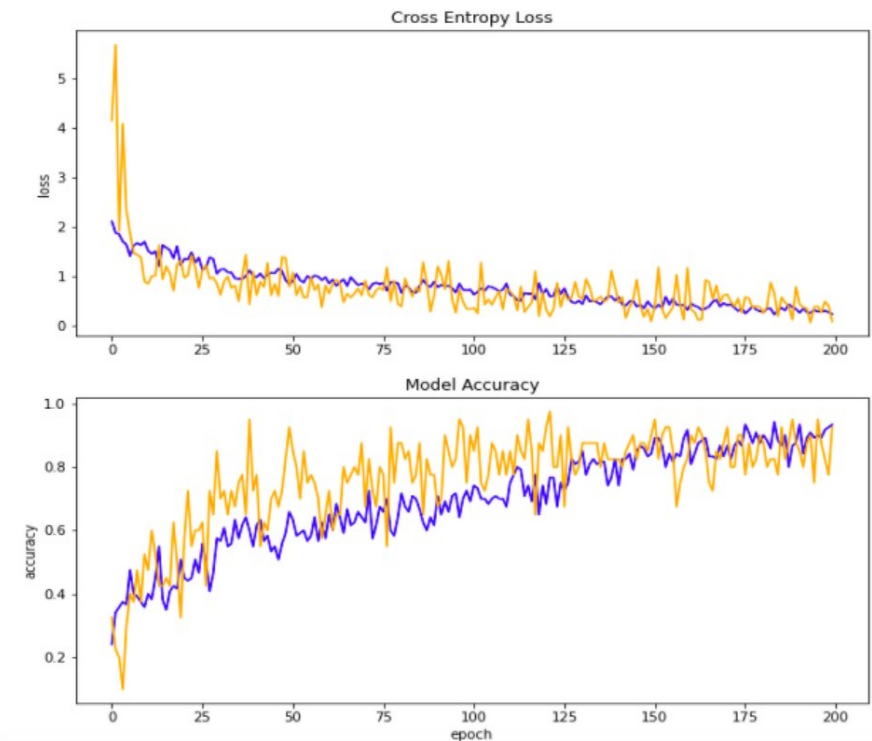
Resultados y discusión

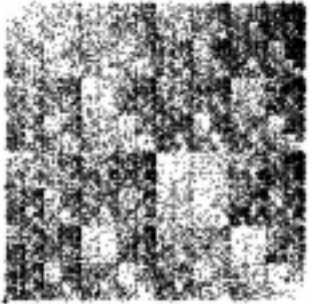


Modelo entrenado sin ruido

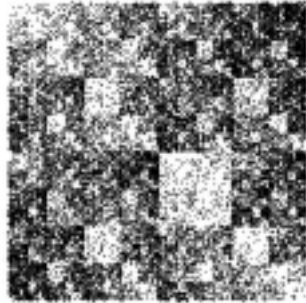
- Conjunto de entrenamiento
- Conjunto de validación

Modelo entrenado con ruido

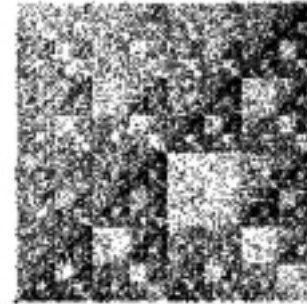


α β γ δ 

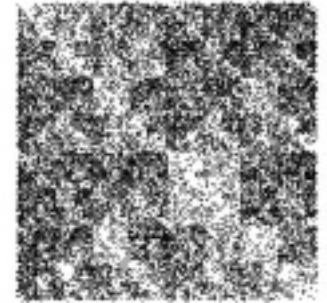
Coronavirus
Humano 229E
(HCoV-229E)



Coronavirus tipo 2
del síndrome
respiratorio agudo
grave (SARS-CoV-2)



Coronavirus de
beluga SW1
(beluga whale
coronavirus SW1)



Coronavirus de
diarrea porcina
(PDCoV)

ID	Alpha	Beta	Delta	Gamma
HCoV-229E	0.99971253	0.00009756162	0.0000712035	0.00011876108
SARS-CoV2	0.99954563	0.00010292028	0.00014498606	0.00020643209
Porcine Diarrhea Cov	0.99967992	0.00010776492	0.00011975131	0.00009254125
Beluga Whale Coronavirus	0.99952006	0.00019956121	0.00006335269	0.00021700891

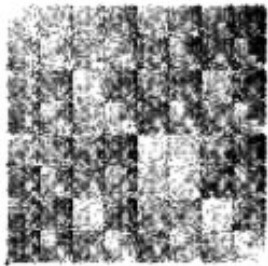
Modelo entrenado sin ruido

ID	Alpha	Beta	Delta	Gamma
HCoV-229E	0.9089777	0.0019587	0.00656711	0.08249648
SARS-CoV2	0.89561015	0.00243213	0.01359664	0.08836111
Porcine Diarrhea Cov	0.8996626	0.00152908	0.0211022	0.07770605
Beluga Whale Coronavirus	0.91656125	0.00301626	0.00623145	0.07419103

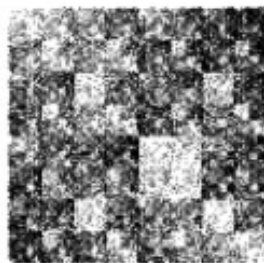
Modelo entrenado con ruido

Noticias de último momento

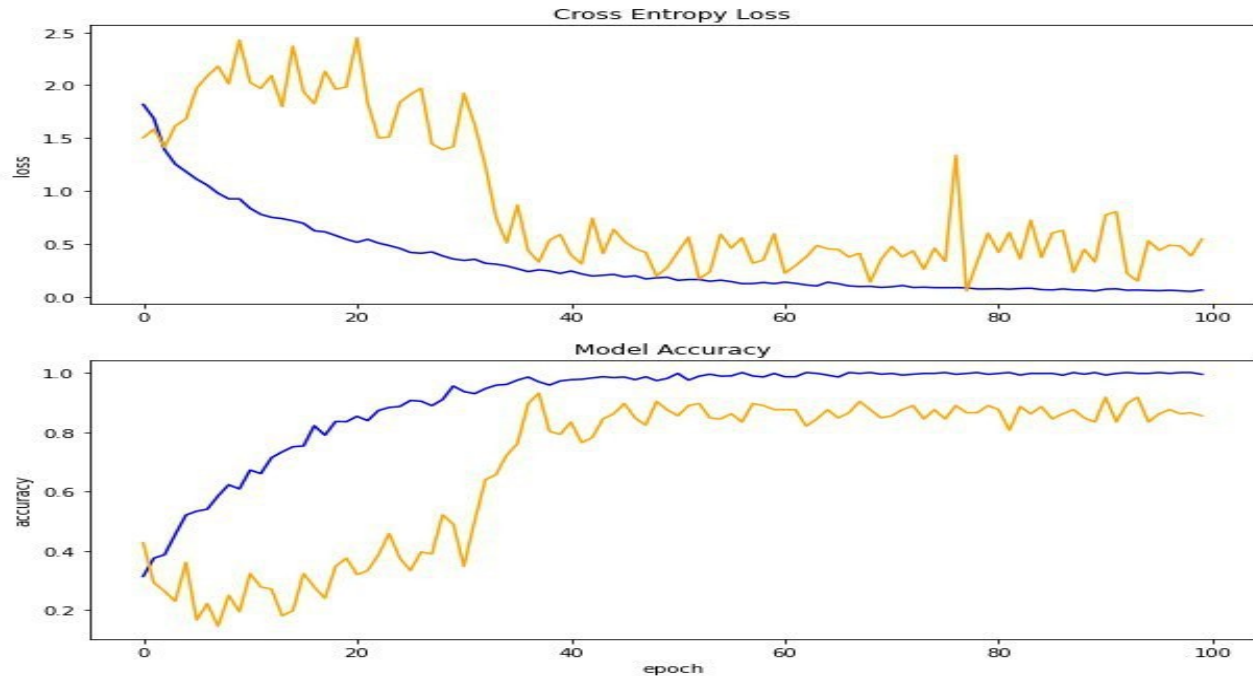
α β



Coronavirus
Humano 229E
(HCoV-229E)



Coronavirus tipo 2
del síndrome
respiratorio agudo
grave (SARS-CoV-2)



ID	Alpha	Beta	Delta	Gamma
HCoV-229E	0.99971253	0.00009756162	0.0000712035	0.00011876108
SARS-CoV2	0.99954563	0.00010292028	0.00014498606	0.00020643209

ID	Alpha	Beta	Gamma	Delta
HCoV-229E	0.5886072	0.03078082	0.07725197	0.30336004
SARS-CoV2	0.5335983	0.02950229	0.13653311	0.30036628

Conjunto de validación

Conjunto de entrenamiento

Referencias

- Deschavanne, P., Giron, A., Vilain, J., Vaury, A., Fertil, B., 2000. Genomic signature is preserved in short DNA fragments. IEEE International Symposium on Bioinformatics and Biomedical Engineering (BIBE'00), pp. 161–167.
- Fehr, A. and Perlman, S. 2015. Coronaviruses: An Overview of Their Replication and Pathogenesis. Coronaviruses, pp.1-23
- Gómez-Mora, H. E., Abasolo-Cortés, N. y Pimentel-Ruíz, C. 2020. https://github.com/cpimentelGH/genomica_computacional_proyecto_final.git
- Goodfellow, et. al. 2016. Deep Learning. 7: Regularization For Deep Learning. MIT Press, Disponible en: <https://www.deeplearningbook.org/>
- Habibi, A. H. y Jahani, H. E. 2017. Convolutional Neural Networks. In: Guide to Convolutional Neural Networks. Springer, Cham.
- Pattanayak, S. (2017) Introduction to Deep-Learning Concepts and TensorFlow. In: Pro Deep Learning with TensorFlow. Apress, Berkeley, CA.
- Russel, S. y Norvig, P. 2010. Artificial Intelligence: A Modern Aproach. Prentice Hall, pp. 729-737.
- Sun, Y., Xue, B., Zhang, M. y Yen, G.G. 2020. Evolving Deep Convolutional Neural Networks for Image Classification, in IEEE Transactions on Evolutionary Computation, vol. 24, no. 2, pp. 394-407.
- Wang, Y., Hill, K., Singh, S., & Kari, L. 2005. The spectrum of genomic signatures: from dinucleotides to chaos game representation. Gene, 346, 173–185. doi:10.1016/j.gene.2004.10.021
- Zhao, L., Jha, B., Wu, A., Elliott, R., Ziebuhr, J., Gorbalenya, A., Silverman, R. and Weiss, S., 2012. Antagonism of the Interferon-Induced OAS-RNase L Pathway by Murine Coronavirus ns2 Protein Is Required for Virus Replication and Liver Pathology. Cell Host & Microbe, 11(6), pp.607-616.