

Modelización de la evolución del Covid-19 Región de Centroamérica

Carlos González, Soizic Gibeaux, Roger Sánchez
Vicerrectoría de Investigación y Posgrados
Universidad Autónoma de Chiriquí
David, Panamá

carlos.gonzalez5@unachi.ac.pa, soizic.gibeaux@gmail.com, roger.sanchez@unachi.ac.pa

Resumen—La Inteligencia Artificial (IA) es una herramienta con un enorme potencial en la lucha contra la pandemia de COVID-19. Desde el inicio de la pandemia se ha utilizado esta herramienta en diferentes iniciativas de investigación epidemiológicas. En este artículo, se presenta un análisis estadístico comparativo de la evolución de los 100 primeros días de la pandemia en Centroamérica. El tratamiento de los datos se realiza con el uso de Pandas, un popular paquete Python utilizado para la ciencia de los datos, con estructuras eficaces, precisas y flexibles, que facilitan la manipulación y el análisis de la información.

Palabras claves—Aprendizaje Automático, Inteligencia Artificial, Covid-19, Modelización, Dataframes

I. INTRODUCTION

La epidemia del Covid-19, causada por el virus del SARS-CoV-2 identificado en diciembre de 2019 en China, fue declarada pandemia mundial por la OMS el 11 de marzo de 2020 [1]. La Inteligencia Artificial (AI) puede convertirse en una herramienta poderosa en la lucha contra la pandemia del Covid-19 [2], [3]. En la actualidad la AI puede definirse como Aprendizaje Automático (ML), el Procesamiento de lenguajes naturales (PNL), y aplicaciones de Visión por Computadora, para enseñar a las computadoras a utilizar modelos de Big Data el cual brinda el reconocimiento, la explicación y la predicción de ciertos patrones. Estas funciones pueden ser útiles para reconocer, predecir y tratar posibles contagios por Covid-19, y ayudar a gestionar los impactos socioeconómicos. Desde el inicio del brote de la pandemia, la AI ha revolucionado su uso, y la combinación con herramientas de análisis de datos[4]–[6].

En la lucha contra el coronavirus SARS-CoV-2, que ya ha infectado a más de 8 millones de personas y provocado la defunción de alrededor de 500,000 personas en todo el mundo; la AI y la Big Data podrían ser los mejores activos de la humanidad. El uso de plataformas tecnológicas se ha convertido en un eje fundamental en todas las áreas de la investigación. En algunos ámbitos relacionados con la medicina o la farmacología, las plataformas de cálculo numérico intensivo permiten, entre otras cosas, la ejecución de simulaciones que entran en la cadena de diseño de

medicamentos, el estudio y el análisis de los datos de las pruebas clínicas, el estudio de los datos relacionados con la propagación de virus y la validación de los modelos predictivos, entre otras aplicaciones.

En este estudio, se presenta una revisión resumida del estado actual de la AI en la lucha contra el Covid-19, introduciendo el potencial y la contribución de esta herramienta, así como las actuales limitaciones de dichas contribuciones en el flujo de la información. El documento tiene como objetivo obtener conclusiones rápidas sobre la evolución de la pandemia, en Centroamérica, donde se está expandiendo rápidamente y el volumen de este trabajo puede servir de insumo para respuestas rápidas en investigación, políticas y análisis médicos. El costo de la pandemia, en términos de vidas y daños económicos tiene un impacto considerable, creando una gran incertidumbre rodeada de estimaciones en cada uno de los países afectados. El ML, es una de las herramientas de análisis de datos más prometedoras que se ha desarrollado en la última década, para ayudar a reducir ciertas incertidumbres, en la búsqueda de validación de la información. Los datos fueron obtenidos de fuentes oficiales de cada uno de los países estudiados.

La AI puede ser usada para rastrear y predecir cómo, la epidemia Covid-19, se extenderá en el tiempo. Por ejemplo, tras una pandemia anterior, la del virus Zika- 2015, desarrollaron una red neural dinámica para predecir su evolución[7]. Sin embargo, los modelos de este tipo necesitarán ser *re-entrenados* usando los datos de la pandemia Covid-19. Esto parece estar sucediendo ahora. En la Universidad Carnegie Mellon, los algoritmos entrenados para predecir la gripe estacional, están siendo *re-entrenados* [8], para ser adaptados a la situación actual.

A continuación, se exponen diversas metodologías para reunir datos y entrenar los modelos sobre la actual pandemia en Centroamérica.

II. MATERIALES Y MÉTODOS

El estudio está enfocado en una investigación descriptiva [9], donde se busca realizar el diseño, la programación y la implementación de la plataforma para el análisis de la evolución del Covid-19 en Centroamérica. Una forma de

lidar con gran volumen de datos de forma flexible y dinámica es el uso de algoritmos de ML, a través de la modelización. Las principales plataformas de medios sociales como *Twitter*, *Instagram*, *Facebook* y comunicados oficiales, disponen de un gran volumen de datos sobre el Covid-19. En una nueva era de la AI, se utiliza cada vez más intensivamente para hacer modelización de contenidos, incluyendo la búsqueda de información actualizada.

Las variables de estudio son los casos confirmados, defunciones, recuperados y número de pruebas. Con estas variables, se busca determinar los comportamientos de acuerdo con la situación de cada país. El análisis está estructurado con una clasificación en bloque de los 100 primeros días de pandemia, el día uno, siendo el día de detección de la primera persona infectada en cada país. Por lo tanto, el día uno no es la misma fecha para cada país analizado.

	country	epidday	date	confirmed	deaths	recovered	tests \
0	Panamá	101	06/17/2020	22597	470	13774	97402
1	Panamá	100	06/16/2020	21962	457	13774	95299
2	Panamá	99	06/15/2020	21422	448	13766	93646

	confirmednorm	letalidad	recoverednorm
0	0.232000	0.020800	0.609550
1	0.230454	0.020809	0.627174
2	0.228755	0.020913	0.642610

	country	epidday	date	confirmed	deaths	recovered	tests \
781	Belice	3	03/25/2020	2			
782	Belice	2	03/24/2020	1			
783	Belice	1	03/23/2020	1			

Figura 1. Resumen de la estructura de datos.

La figura 1 muestra un resumen estructurado de algunos de los datos estudiados. Con frecuencia, cuando se trabaja en un modelo de aprendizaje profundo, se recuperan los datos de origen en formato *pandas.DataFrame* y se necesita convertirlo en un formato que *Tensorflow* pueda leer. Actualmente, *Tensorflow* proporciona conjuntos de datos que crean tuberías de datos para que la red se entrene. Convertir un *DataFrame* en un *tf.data.Dataset* puede ser un proceso tedioso sino se tienen clara las variables de estudio y la estructuración de los datos. Los *DataFrame* seleccionados tienen cuatro características generadas estáticamente y tres clases de objetivos para poder convertirlo en un conjunto de datos TensorFlow. El código está diseñado en Python y utiliza el modo de ejecución en un entorno Jupyter donde funcionará con gráficos dinámicos de acuerdo con los *DataFrames* en ejecución. Después de crear el *DataSet*, se pueden ejecutar funciones por lotes, aleatorias y otras para pre-procesar los datos antes del entrenamiento del modelo.

Con el procesamiento de grandes volúmenes de información, se pueden realizar análisis predictivos y cómputos científicos que están orientados a simplificar el desarrollo y la administración de los paquetes de software utilizados. Las diferentes versiones de paquetes utilizados, se administran mediante el sistema de administración del

software Anaconda, integrando el aprendizaje automático con *TensorFlow* y *SciPy*.

III. ANÁLISIS Y RESULTADOS

Esta sección, parte de la base real de los casos confirmados por país. El mayor número de casos reportados está en Panamá, seguido por Guatemala y Honduras, los primeros 100 días de la epidemia de estos países. Los casos reportados como positivos en Costa Rica, se mantienen estables en la duración de los 100 primeros días de la clasificación del estudio. Desde el día 60, El Salvador muestra un leve aumento. Los datos reportados por Nicaragua no son constantes con reportes semanales causando saltos de datos entre semanas. Actualmente, Belice tiene muy pocos casos reportados, lo que representa un limitado aporte de datos para realizar este análisis.

En los siguientes resultados presentados, Nicaragua y Belice no son considerados, ya que las cifras reportadas son datos preliminares y no permiten una evaluación precisa del modelo.

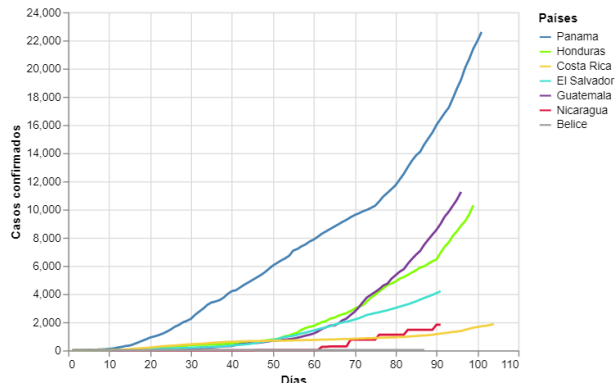


Figura 2. Casos confirmados por país los primeros 100 días de la pandemia en cada país.

En el caso de las defunciones se observa un aumento considerable en Guatemala desde el día 70. Panamá es el país con más defunciones reportadas de los países centroamericanos, con una constante de defunciones reportadas. Honduras al igual que Panamá mantienen un ritmo de crecimiento constante. Los casos reportados por El Salvador y Costa Rica son bajos, manteniendo un número inferior defunciones respecto de los demás países.

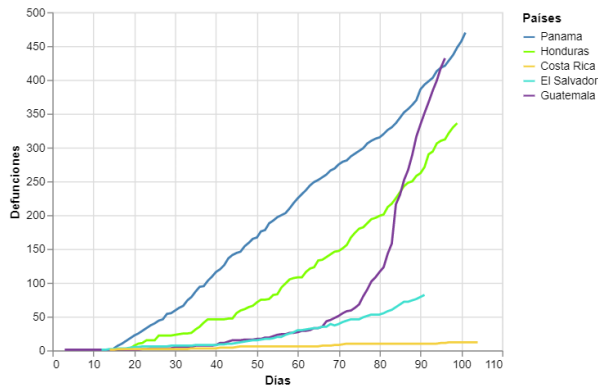


Figura 3. Defunciones por país los primeros 100 días de la pandemia.

En cuanto a los casos recuperados, Panamá tiene un alto número respecto de los demás países. Sin embargo, los datos proporcionados por el Ministerio de Salud no son constantes. Se pueden observar reportes diarios, donde se mantiene en diferentes días la misma cantidad de recuperados. Un punto importante a considerar es que se pueden remarcar los pocos casos recuperados en Guatemala y Honduras, dado el número de casos confirmados. Según datos proporcionados por El Salvador y Costa Rica, se observa una evolución estable según datos confirmados. Sin embargo, la tasa de letalidad permite medir el cociente de personas fallecidas entre las infectadas, de acuerdo con la clasificación de los *DataFrames*.

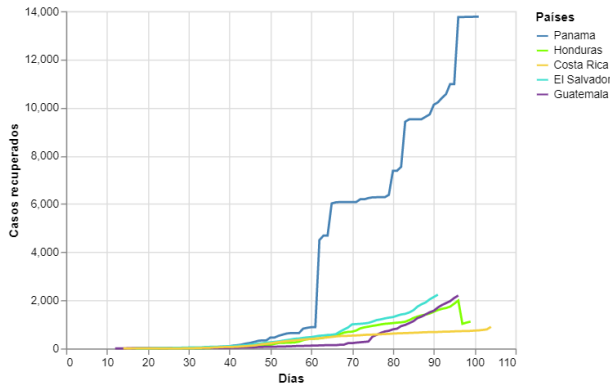


Figura 4. Recuperados por país los primeros 100 días de la pandemia.

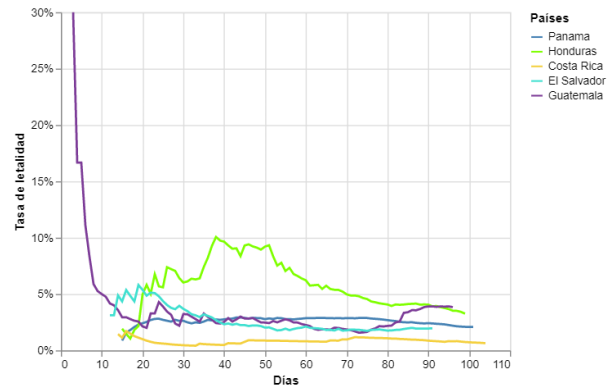


Figura 5. Tasa de letalidad por país los primeros 100 días de la pandemia.

La tasa de letalidad es un análisis muy importante en este estudio, ya que permite visualizar si las medidas de confinamientos impuestas, por los gobiernos son efectivas en cada una de las poblaciones en estudio. Al inicio de la pandemia, los datos reportados por Guatemala reflejan un mismo número de defunciones, por lo que la tasa de letalidad aparece elevada. El dato interesante se observa en los últimos días de la pandemia donde, a pesar de las medidas tomadas por los gobiernos, en los últimos días del bloque de clasificación, Guatemala y Honduras mantienen los porcentajes más elevados. En Panamá, la tasa de letalidad se mantiene estable con un promedio del 2%. Costa Rica, mantiene la tasa de letalidad más baja con un promedio de 0.8%.

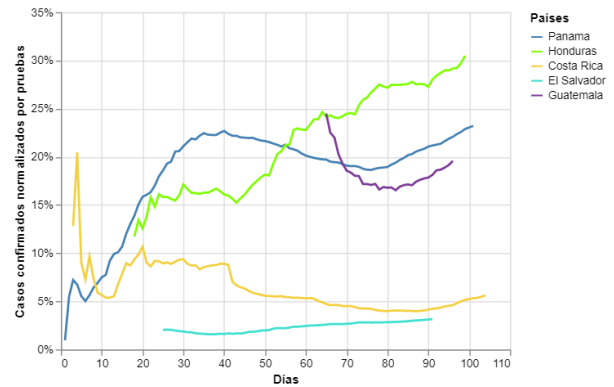


Figura 6. Confirmados normalizados por país los primeros 100 días de la pandemia en Centroamérica.

Un análisis de suma importancia es la cantidad de pruebas aplicadas por casos confirmados, donde se puede observar el número de casos positivos. En muchos estudios solo muestran los casos confirmados vs las defunciones o personas recuperadas. En este estudio, se hace énfasis en las pruebas realizadas ya que permite visualizar la efectividad de los gobiernos en la búsqueda de los posibles contagios. Por falta de datos de las fuentes oficiales, en países como Honduras y El Salvador no se puede visualizar el comportamiento de las pruebas realizadas, por el número de casos confirmados los primeros 100 días. Los porcentajes más elevados, de casos positivos, por pruebas realizadas de

acuerdo con los casos confirmados son Honduras y Panamá. El porcentaje de casos positivos, por pruebas realizadas en Guatemala, comienzan el día 65 con un alto porcentaje de pruebas que, decae gradualmente y en los últimos días comienzan a repuntar. Dado el bajo número de casos reportados por El Salvador y Costa Rica, el porcentaje de casos positivos por pruebas realizadas es más bajo.

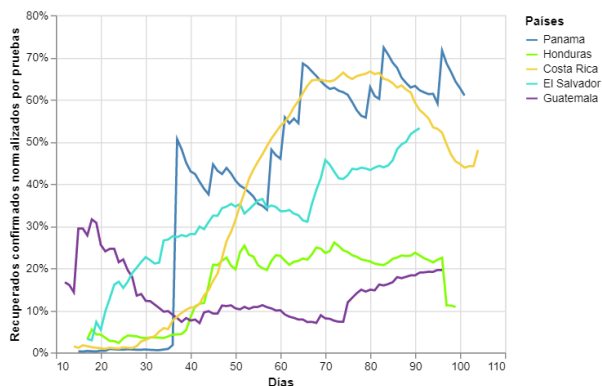


Figura 7. Recuperados por casos confirmados

Panamá presenta el mejor porcentaje de recuperados, en relación con los casos confirmados. En los últimos días, Costa Rica quien mantenía buenos resultados en los indicadores estudiados anteriormente, parece disminuir el porcentaje de recuperados. El Salvador aumenta su porcentaje de recuperados gradualmente, respecto de los demás países centroamericanos. Los porcentajes de Guatemala y Honduras están por debajo de un 20%, observando un descenso pronunciado los últimos días en Honduras.

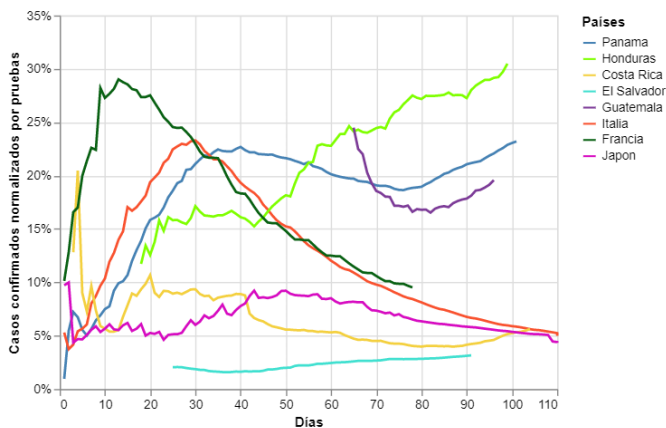


Figura 8. Confirmados normalizados por total de pruebas por país los primeros 100 días de la epidemia en Centroamérica, Italia, Francia y Japón.

Hay un análisis que permite visualizar la evolución de la pandemia, los primeros 100 días, entre los países de Centroamericanos con respecto de dos países europeos, en este caso particular Italia y Francia. Además, se agrega Japón para tener una visión más amplia de la evolución de la pandemia. El porcentaje de casos positivos, por pruebas realizadas desde el día 70 en Italia y Francia, está por debajo del 10%. Sin embargo, en Centroamérica, solo Costa Rica y El Salvador están en ese porcentaje según cifras oficiales. Guatemala, Honduras y Panamá mantienen porcentajes de alrededor 20% o más de casos positivos por pruebas realizadas, mantenido una constante al alza de los casos positivos.

CONCLUSIONES

La inteligencia artificial está jugando un rol sumamente importante en la lucha contra el Covid-19. Este estudio ofrece una versión resumida de la situación actual de la epidemia en Centroamérica. La falta de información por parte de algunas entidades gubernamentales dificulta el análisis comparativo de la evolución del Covid-19. En el futuro se plantea ampliar el estudio evolutivo y realizar comparaciones con países donde se ha regresado a la nueva normalidad. De igual manera, se espera que el mismo contribuya con la toma de decisiones, por parte de las autoridades y la implementación de políticas públicas, frente a esta y otras posibles emergencias en salud.

REFERENCIAS

- [1] M. P. Cruz, E. Santos, M. V. Cervantes, y M. L. Juárez, «COVID-19, una emergencia de salud pública mundial», *Rev. Clínica Esp.*, 2020.
- [2] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, y M. Luengo-Oroz, *Mapping the Landscape of Artificial Intelligence Applications against COVID-19*. 2020.
- [3] G. Petropoulos, «Artificial intelligence in the fight against COVID-19», *Bruegel 23 March*, 2020.
- [4] Z. Hu, Q. Ge, L. Jin, y M. Xiong, «Artificial intelligence forecasting of covid-19 in china», *ArXiv Prepr. ArXiv200207112*, 2020.
- [5] A. Alimadadi, S. Aryal, I. Manandhar, P. B. Munroe, B. Joe, y X. Cheng, *Artificial intelligence and machine learning to fight COVID-19*. American Physiological Society Bethesda, MD, 2020.
- [6] R. Vaishya, M. Javaid, I. H. Khan, y A. Haleem, «Artificial Intelligence (AI) applications for COVID-19 pandemic», *Diabetes Metab. Syndr. Clin. Res. Rev.*, 2020.
- [7] M. Akhtar, M. U. G. Kraemer, y L. M. Gardner, «A dynamic neural network model for predicting risk of Zika in real time», *BMC Med.*, vol. 17, n.º 1, p. 171, sep. 2019, doi: 10.1186/s12916-019-1389-3.
- [8] W. Naudé, «Artificial intelligence vs COVID-19: limitations, constraints and pitfalls», *Ai Soc.*, p. 1, 2020.
- [9] R. Sautu *et al.*, *METODOLOGIA DE LA INVESTIGACION I*. México DF: INTERAMERICANA EDITORES, SA DE CV, 2014.