ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.2471

Comparative Analysis of SEIR/SIR/SIS Models and Machine Learning in the Prediction of COVID-19, Honduras

Brennedy Martinez Claros[1], Celene Yamileth Ventura[1], Henry Osorto Ruiz[1,2]
[1]Facultad de Postgrado, Universidad Tecnológica Centroamericana, San Pedro Sula, Honduras,
brennedy_martinez@unitec.edu, celene_ventura@unitec.edu, henry.osorto@unitec.edu.hn
[2]Universidad Nacional Autónoma de Honduras, UNAH, Honduras, *henry.osorto@unah.edu.hn*

Abstract– ***This study aims to compare traditional mathematical models used to analyze the spread of infectious diseases with machine learning methods, using COVID-19 as a case study. Classical models such as SIR, SEIR, and SIS help describe the progression of an epidemic; however, they typically rely on fixed parameters and struggle to adapt to real-time changes. In contrast, machine learning models including Polynomial Regression, SVM, and Random Forest are capable of processing large datasets, detecting more complex patterns, and adjusting their predictions as new information becomes available. For this analysis, historical data from the World Health Organization (WHO), adapted to national records, were used, and the performance of each model was evaluated using metrics such as RMSE and R^2. Overall, the results showed that machine learning models provided a better fit and greater adaptability, making them a valuable option for anticipating and controlling future epidemic outbreaks.***

Keywords-- Epidemiological models, Machine Learning, SIR, COVID-19, Python programming language.

Análisis comparativo de los modelos SEIR/SIR/SIS y el aprendizaje automático en la predicción de la COVID-19 en Honduras.

Brennedy Martinez Claros[1], Celene Yamileth Ventura[1], Henry Osorto Ruiz[1,2]
[1]Facultad de Postgrado, Universidad Tecnológica Centroamericana, San Pedro Sula, Honduras,
brennedy_martinez@unitec.edu, celene_ventura@unitec.edu, henry.osorto@unitec.edu.hn
[2]Universidad Nacional Autónoma de Honduras, UNAH, Honduras, *henry.osorto@unah.edu.hn*

Resumen– ***Este trabajo tiene como propósito comparar los modelos matemáticos tradicionales que se usan para estudiar la propagación de enfermedades con los métodos de aprendizaje automático, tomando como caso el COVID-19. Los modelos clásicos, como SIR, SEIR y SIS, ayudan a entender cómo avanza una epidemia, pero suelen trabajar con parámetros fijos y no logran adaptarse a los cambios reales que se dan con el tiempo. En cambio, los modelos de aprendizaje automático, como la Regresión Polinomial, SVM y Random Forest, pueden analizar grandes cantidades de datos, detectar patrones más complejos y ajustar sus predicciones conforme aparece nueva información. Para el análisis se usaron datos históricos de la Organización Mundial de la Salud (OMS), adaptados a registros nacionales, y se evaluó el rendimiento de cada modelo con métricas como RMSE y R^2. En general, los resultados mostraron que los modelos de aprendizaje automático se ajustaron mejor y ofrecieron mayor capacidad de adaptación, lo que los convierte en una opción valiosa para anticipar y controlar futuros brotes epidémicos.***

Palabras clave-- Modelos epidemiológicos, Aprendizaje automático, SIR, COVID-19, Lenguaje de programación Python.

I. Introducción

Este estudio pretende ser una herramienta valiosa para académicos, formuladores de políticas y líderes gubernamentales, proporcionando un análisis robusto sobre cómo optimizar los factores que influyen en la estabilidad gubernamental. Sus hallazgos pueden inspirar reformas significativas para lograr un desarrollo más estable y sostenible en la región. A medida que aumenta la necesidad de comprender e investigar enfermedades epidemiológicas contagiosas como el COVID-19 que en años recientes se extendió por todo el mundo de forma significativa, también crece la urgencia de contar con herramientas más precisas para anticipar su comportamiento. Los modelos matemáticos han contribuido enormemente a estimar los niveles de propagación y a describir la evolución de la enfermedad; sin embargo, presentan limitaciones, ya que suelen basarse en parámetros fijos que no siempre reflejan los cambios reales que ocurren en la población.

Además, la evidencia reciente subraya que la predicción de puntos de inflexión (pico y declive) en epidemias es particularmente difícil cuando se asume estabilidad paramétrica y cuando existen retardos en la observación y cambios de comportamiento inducidos por políticas. En este sentido, se ha documentado que el "turning point" de una epidemia en expansión puede no ser pronosticable con precisión en tiempo real, incluso con datos de alta frecuencia, debido a la sensibilidad del sistema a variaciones pequeñas en parámetros y a problemas de identificabilidad en etapas tempranas [14]. Este argumento refuerza la necesidad de enfoques que permitan parámetros dependientes del tiempo o estrategias de ajuste que incorporen información exógena (movilidad, vacunación, adherencia social).

Aunque la pandemia de COVID-19 ya quedó atrás, los virus siguen siendo una amenaza para la salud pública, y es importante seguir estudiando modelos que ayuden a entender cómo podrían comportarse futuras epidemias. Entre los más conocidos están los modelos compartimentales SIR (Susceptibles, Infectados, Recuperados), SEIR (Susceptibles, Expuestos, Infectados, Recuperados) y SIS (Susceptibles, Infectados, Susceptibles). Estos modelos han sido muy útiles para describir la forma en que se propagan las enfermedades y cómo cambian los contagios con el tiempo. Aun así, tienen una limitación clara: no pueden adaptarse por sí mismos a nuevas situaciones ni aprender de los datos que van surgiendo, lo que hace que sus predicciones pierdan precisión cuando las condiciones cambian o aparecen nuevos brotes.

El modelo SEIR es particularmente útil para enfermedades con un período de incubación, como el COVID-19, donde los individuos pasan por una fase expuesta antes de volverse infecciosos [1]. Mientras que, para otras enfermedades, los modelos SIR y SEIR también han sido empleados para comprender y predecir brotes [2], [3]. Estos modelos han ayudado a identificar estrategias efectivas de control y a evaluar la efectividad de intervenciones durante los diferentes brotes epidemiológicos que han existido [4].

A lo largo del tiempo, varios investigadores han tratado de encontrar mejores formas de entender y predecir cómo se

propagan las enfermedades infecciosas [5]. Esa falta de flexibilidad en los modelos clásicos ha hecho que se busquen alternativas más adaptables. Una de ellas es el aprendizaje automático (Machine Learning), que tiene la ventaja de poder analizar grandes cantidades de datos, detectar patrones difíciles de ver y ajustar sus predicciones conforme va recibiendo nueva información.

La discusión contemporánea no plantea una sustitución total de los modelos mecanicistas por modelos puramente empíricos, sino una convergencia metodológica: extensiones compartimentales que incorporan vacunación, asintomáticos, reinfección y/o mutaciones, junto con procedimientos de estimación asistidos por aprendizaje automático. Por ejemplo, se han propuesto variantes extendidas del SIR/SEIR para modelar campañas de vacunación con calibración basada en redes neuronales, así como modelos que incorporan explícitamente múltiples mutaciones y políticas de intervención durante la fase de ajuste, con mejoras en la capacidad predictiva y en métricas epidemiológicas clave [6], [7]. Esta línea se articula con el objetivo del presente estudio: combinar predicción robusta con toma de decisiones en política pública.

En este contexto, existen técnicas de aprendizaje automático (Machine Learning), que son modelos que analizan grandes volúmenes de datos para predecir la propagación del virus, identificando los factores de riesgos y evaluando su efectividad al compararlos con los modelos tradiciones SEIR/SIR/SIS [8].

Finalmente, desde la perspectiva de política pública, la predicción es un insumo y no un fin. La literatura reciente enfatiza marcos "prediction-and-decision", donde un módulo epidemiológico alimenta un problema de optimización para seleccionar intervenciones bajo incertidumbre y restricciones operativas. Se han reportado enfoques integrados que combinan pronóstico epidemiológico con modelos de demanda y programación entera estocástica para decidir capacidad logística, así como herramientas que convierten modelos epidemiológicos y funciones de costo en problemas de optimización multiobjetivo (por ejemplo, minimización conjunta de mortalidad y recesión económica), utilizando algoritmos evolutivos y aprendizaje por refuerzo [30,31]. Esto alinea directamente el aporte del paper con decisiones de política: minimizar muertes y costo económico.

En este trabajo se busca comparar dos enfoques: los modelos epidemiológicos y los métodos de aprendizaje automático. La idea es analizar cuál de ellos logra adaptarse mejor a los cambios que se dieron durante el desarrollo del COVID-19 y cuál ofrece predicciones más precisas a partir de los datos históricos publicados por la Organización Mundial de la Salud (OMS). Con ello, se pretende reducir la distancia que todavía existe entre los modelos teóricos tradicionales y las nuevas herramientas de predicción que se apoyan directamente en los datos reales.

II. Revisión de Literatura

En síntesis, la literatura sobre los modelos epidemiológicos compartimentales son representaciones matemáticas que dividen a una población en diferentes grupos o compartimentos según su estado de salud [9]. Estos modelos son herramientas fundamentales para estudiar la propagación de enfermedades infecciosas, ya que permiten comprender cómo los individuos se trasladan entre distintos estados a lo largo del tiempo, lo que facilita la predicción de brotes y el diseño de estrategias de control [10].

Los modelos más conocidos son el SIR, el SEIR y el SIS. Modelo SIR (Susceptibles, Infectados, Recuperados): Este modelo divide a la población en tres grupos: los Susceptibles (S), que son aquellos individuos que aún no han contraído la enfermedad; los Infectados (I), que son los que han contraído la enfermedad y pueden transmitirla; y los Recuperados (R), que son los individuos que se han recuperado y han adquirido inmunidad [11].

Las transiciones entre estos estados están regidas por las tasas de transmisión (β) y recuperación (γ) [9]. Este modelo se aplica principalmente a enfermedades que no tienen reinfección. Modelo SEIR (Susceptibles, Expuestos, Infectados, Recuperados): A diferencia del modelo SIR, el modelo SEIR incorpora un compartimento adicional de Exposición (E), que incluye a los individuos que han estado en contacto con el virus pero que aún no son contagiosos [12]. Este modelo es especialmente útil para enfermedades que tienen un período de incubación, como el COVID-19 [13]. La dinámica de propagación es más compleja y permite modelar con mayor precisión el comportamiento de enfermedades con una fase de exposición. Modelo SIS (Susceptibles, Infectados, Susceptibles): Este modelo es apropiado para enfermedades en las que la inmunidad no es duradera, y los individuos pueden volver a ser susceptibles después de haberse recuperado [14]. Las transiciones entre los estados están determinadas por la tasa de infección y la tasa de recuperación, pero no se asume que los individuos adquieran inmunidad permanente. Estos modelos han sido fundamentales en la predicción y el control de brotes epidémicos, pero presentan limitaciones, principalmente debido a la necesidad de suponer que ciertos parámetros, como la tasa de transmisión o la tasa de recuperación, son constantes a lo largo del tiempo [15].

En aplicaciones recientes de COVID-19, la literatura ha tendido a ampliar el alcance de los modelos compartimentales para representar mecanismos ausentes en SIR/SEIR/SIS básicos, especialmente cuando se estudian escenarios de política. Por ejemplo, la incorporación de compartimentos de vacunación y asintomáticos (y la estimación conjunta de parámetros y condiciones iniciales) permite evaluar escenarios de despliegue vacunal y niveles de hesitación, cuestionando incluso nociones simplificadas de "inmunidad de rebaño" bajo variantes más contagiosas [29]. Asimismo, modelos diseñados para múltiples mutaciones e intervención durante el ajuste (en lugar de tratar la política como un shock exógeno posterior) tienden a capturar mejor la dinámica real observada,

especialmente cuando aparecen cepas con distinta transmisibilidad [7].

A. Limitaciones de los Modelos Epidemiológicos Tradicionales

A pesar de la amplia aplicación de los modelos compartimentales, estos tienen limitaciones significativas como ser:

Parámetros Fijos: Los modelos como el SIR, SEIR y SIS dependen de parámetros constantes que no siempre reflejan la variabilidad observada en situaciones de brotes [16]. La tasa de transmisión, por ejemplo, puede cambiar debido a factores como las intervenciones de salud pública, las mutaciones del virus o los cambios en el comportamiento de la población.

Falta de Adaptabilidad: Los modelos tradicionales no pueden adaptarse fácilmente a cambios en los datos o a nuevas variantes del virus. En situaciones como la pandemia del COVID-19, donde las tasas de transmisión varían significativamente con el tiempo, los modelos compartimentales tienden a ser imprecisos.

Tipos de Datos: Para calibrar correctamente los modelos, se necesita una gran cantidad de datos precisos, lo que no siempre es posible en brotes emergentes, donde los datos pueden ser escasos o poco confiables.

A estas limitaciones se suma un problema metodológico central: identificabilidad y sensibilidad. En etapas tempranas, distintos conjuntos de parámetros pueden explicar trayectorias similares, lo que dificulta inferencias confiables sobre β, γ o tasas de contacto cuando los datos están afectados por subregistro, cambios en testeo y rezagos. Este punto es consistente con evidencia que muestra que el "final" o el punto de giro de una epidemia en expansión puede no ser pronosticable con precisión, aun con modelos bien establecidos, debido a la dinámica no lineal y a la dependencia crítica de supuestos [16]. Adicionalmente, la aparición de variantes (p. ej., Ómicron) y la heterogeneidad entre países introducen cambios estructurales que vuelven insuficiente la calibración "estática". Estudios que analizan explícitamente periodos de transición por variantes han optado por estrategias multipropósito (ajuste de distribuciones, series de tiempo y modelos compartimentales), precisamente para reducir dependencia de un único modelo y mejorar robustez de predicción comparativa [17]. En términos de política pública, esto sugiere privilegiar enfoques de ensamble/híbridos y/o parámetros dependientes del tiempo.

B. Machine Learning en Epidemiología

El aprendizaje automático (Machine Learning) es una rama de la inteligencia artificial que utiliza algoritmos para analizar grandes volúmenes de datos y aprender patrones sin la necesidad de programación explícita [18], [19]. En el contexto epidemiológico, las técnicas de aprendizaje automático han mostrado un gran potencial para mejorar la predicción de brotes y la toma de decisiones en salud pública. A diferencia de los modelos compartimentales, que dependen de ecuaciones matemáticas predefinidas, el aprendizaje automático trabaja directamente con datos, lo que le permite adaptarse mejor a las características dinámicas de las epidemias [20].

En la literatura de alto impacto, el valor del aprendizaje automático no se limita a "ajustar curvas", sino a integrar señales heterogéneas y capturar no linealidades (movilidad, patrones de contacto, adherencia, vacunación) que afectan la transmisión. Un avance relevante es el uso de enfoques distribuidos para entornos con datos heterogéneos, como el aprendizaje federado/multitarea, que permite entrenar modelos sin centralizar información sensible y mejorar predicción de tasas de infección cuando existen diferencias regionales [21]. Esto es especialmente pertinente para sistemas de vigilancia con datos fragmentados y para el diseño de políticas territorialmente diferenciadas.

Otra vertiente complementaria es la incorporación de información social y comportamental mediante analítica de texto y sentimiento. Se ha argumentado que la minería de redes sociales puede aportar indicadores cuantificables de hesitación vacunal y actitudes hacia medidas de mitigación, mejorando la capacidad de ajustar modelos y refinar políticas en tiempo casi real [22]. Desde el enfoque de decisión pública, estas variables pueden tratarse como covariables que afectan $\beta(t)$(contactos efectivos) o como determinantes de parámetros de intervención (adherencia), elevando la validez externa del análisis.

Algunas de las aplicaciones de Machine Learning en Epidemiología son:

Regresión Polinomial: Los algoritmos de regresión pueden predecir la evolución de la epidemia a partir de variables como la movilidad de la población, las intervenciones de salud pública y los datos epidemiológicos [23].

Máquina de vectores y soporte (SVM): Es utilizada principalmente en problemas de clasificación la cual consiste en encontrar un vector (para el caso de dos dimensiones) y un hiperplano (en el caso de tener más de dos dimensiones) que logre maximizar la separación entre clases.

Random Forest: Este tipo de modelos permiten identificar los factores de riesgo más importantes para la propagación de enfermedades, lo que puede ayudar en la toma de decisiones para implementar medidas de control.

C. Ventajas del uso de Machine Learning

Mayor Flexibilidad: Se adapta rápidamente a nuevos datos y cambia con las condiciones del brote.

Capacidad de Manejar Grandes Volúmenes de Datos: El aprendizaje automático es capaz de procesar datos masivos de diversas fuentes, como informes de salud pública, datos de movilidad y redes sociales.

Modelado No Lineal: Permite captar relaciones no lineales entre los diferentes factores que afectan la propagación de la enfermedad, como la interacción de distintas variables o la aparición de nuevas variantes del virus.

No obstante, la literatura también subraya que el desempeño predictivo de ML depende críticamente de la calidad y estabilidad del proceso generador de datos; por ello, se

recomienda reportar validaciones fuera de muestra, análisis de sensibilidad y comparaciones con baseline mecanicistas. En contextos de cambios estructurales (variantes, vacunación, políticas), los modelos híbridos y los esquemas de recalibración periódica tienden a ser más defendibles científicamente que modelos puramente "caja negra", especialmente cuando el objetivo es informar política pública.

D. Integración de Modelos Tradicionales y Machine Learning

La combinación de modelos tradicionales de epidemiología y Machine Learning puede ofrecer un enfoque más robusto para predecir y controlar la propagación de enfermedades infecciosas [19]. Los modelos compartimentales pueden servir como una base sólida para describir la dinámica básica de la epidemia, mientras que los modelos de aprendizaje automático pueden proporcionar una mayor precisión al adaptarse a nuevos datos y condiciones cambiantes. Esta sinergia entre ambos enfoques puede mejorar la capacidad de respuesta frente a brotes epidémicos, al permitir una mejor predicción de la evolución de la enfermedad y optimizar las estrategias de intervención. Además, el uso de enfoques híbridos podría ayudar a identificar estrategias más eficaces para la prevención y control de enfermedades infecciosas, como se ha sugerido en estudios recientes [5], [8]. Los modelos epidemiológicos tradicionales, como SIR, SEIR y SIS, han sido fundamentales para la predicción de la propagación de enfermedades infecciosas. Sin embargo, debido a sus limitaciones, como la suposición de parámetros constantes y la falta de adaptabilidad a nuevos datos, han surgido enfoques más modernos basados en Machine Learning. Estos modelos permiten una mayor flexibilidad, la capacidad de manejar grandes volúmenes de datos y la detección de patrones complejos, especialmente en situaciones de brotes epidémicos como el COVID-19.

La integración se ha implementado, por ejemplo, usando redes neuronales para resolver/ajustar ecuaciones diferenciales y luego estimar parámetros que mejor explican curvas recientes, combinando aprendizaje supervisado y no supervisado sobre la estructura del modelo epidemiológico [29]. Alternativamente, metaheurísticas como PSO han sido utilizadas para ajustar parámetros de modelos tipo SIR y compararlos con aproximaciones de redes neuronales alimentadas por series diarias, reportando mejoras en métricas como R^2y reducciones en errores porcentuales [24]. Estas estrategias refuerzan el argumento de que la "hibridación" puede elevar precisión sin perder interpretabilidad.

Un paso adicional es formalizar la integración predicción → decisión. La literatura propone que los modelos epidemiológicos se conecten con funciones de costo y restricciones para derivar políticas óptimas, en lugar de limitarse a escenarios descriptivos. En esta línea, se han desarrollado marcos que combinan pronóstico epidemiológico (p. ej., SIR) con modelos econométricos/operativos y programación entera estocástica para decisiones bajo demanda incierta [30], y también herramientas generalizables que convierten modelos epidemiológicos en problemas de optimización multiobjetivo (minimizar muertes y recesión), resolubles con algoritmos evolutivos (NSGA-II) y aprendizaje por refuerzo profundo [25], [26]. Este es el fundamento metodológico directo para plantear en el paper un objeto de optimización centrado en minimización de muertes y costo económico.

III. Materiales y Métodos

En esta investigación se lleva a cabo un análisis comparativo entre los modelos epidemiológicos compartimentales tradicionales (SIR, SEIR y SIS) y modelos de aprendizaje automático para la predicción de la propagación del COVID-19. Además, esta investigación es tipo longitudinal retrospectiva debido que estudia la epidemia como cambio a lo largo del tiempo en el 2020. La investigación tiene un enfoque cuantitativo dado que se basa en la cuantificación de datos numéricos y se centra en medir el rendimiento de los modelos en función de la calidad y la exactitud de sus predicciones. La metodología se estructura en las siguientes fases:

A. Selección y Obtención de Datos

Se utilizarán datos históricos sobre la propagación del COVID-19, obtenidos de fuentes oficiales como la Organización Mundial de la Salud (OMS) y Our Word in data (OWD) entre otras bases de datos epidemiológicas de Honduras. Los datos se deben de preprocesar para eliminar valores atípicos, manejar datos faltantes y normalizar las variables relevantes para su posterior análisis [27].

A partir de esta información se reunieron los registros correspondientes a Honduras, con el propósito de analizar la evolución de la enfermedad en el país y utilizar dichos datos en los modelos planteados más adelante. Con esta información se organizó una base de datos enfocada exclusivamente en los registros diarios de Honduras, lo que permitió estudiar de manera más detallada el comportamiento de la pandemia en el territorio nacional.

Para adaptar la información tanto a los modelos epidemiológicos como a los métodos de aprendizaje automático, fue necesario incorporar variables adicionales, entre ellas Sigma (tasa de exposición), Gamma (tasa de recuperación), Beta_rate (tasa de transmisión) y Reproduction_rate.

También se incluyeron los compartimentos Exposed y Recovered. Estas variables se obtuvieron a partir del análisis realizado con los modelos SIR, SEIR y SIS explicados previamente, logrando así conformar una base de datos más completa y estructurada, lista para aplicar los enfoques matemáticos tradicionales y los modelos de aprendizaje automático.

B. Implementación de Modelos Epidemiológicos

Se implementarán los modelos compartimentales SIR, SEIR y SIS utilizando ecuaciones diferenciales que describen la dinámica de transmisión de las enfermedades. De acuerdo con [23] el Modelo SIR divide a la población en tres grandes grupos

Susceptibles(S), Infectados(I), Recuperados(R) los cuales se explican a partir de las siguientes ecuaciones diferenciales:

$$\frac{dS}{dt} = \frac{-\beta SI}{N} \quad (1)$$

$$\frac{dI}{dt} = \frac{-\beta SI}{N} - \gamma I \quad (2)$$

$$\frac{dR}{dt} = \gamma I \quad (3)$$

Tomando en cuenta que cuando se realice la simulación se trabaja con un número exacto de población (N) que debe ser igual al número de personas susceptibles, Infectados, Recuperados [28]. A continuación, se describe el nombre y el valor de cada uno de los parámetros que interviene en el cálculo de este modelo:

TABLA I
VALORES DE PARÁMETROS PARA EL MODELO SIR

Parámetro	Descripción	Valor
S_0	Susceptibles iniciales	9442000
I_0	Infectadas iniciales	3
R_0	Recuperados iniciales	0
N	Población total (2020)	9442000
t	Tiempo (días)	300

Para el modelo SEIR la población se separa en los mismos grupos solo agregando una más, siendo esta el grupo de Expuestos (E), los cuales se explicar a partir de las siguientes ecuaciones diferenciales:

$$\frac{dS}{dt} = \frac{-BSI}{N} \quad (4)$$

$$\frac{dE}{dt} = \frac{BSI}{N} - \sigma E \quad (5)$$

$$\frac{dI}{dt} = \sigma E - \gamma I \quad (6)$$

$$\frac{dR}{dt} = \gamma I - \varepsilon R - vR \quad (7)$$

La población para este modelo se calcula de acuerdo con la siguiente ecuación:

$$N = S+E+I+R \quad (8)$$

A continuación, se describe el nombre y el valor de cada uno de los parámetros que interviene en el cálculo de este modelo:

TABLA II
VALORES DE LOS PARÁMETROS PARA EL MODELO SEIR

Parámetro	Descripción	Valor
S_0	Susceptibles iniciales	9442000
E_0	Expuestos iniciales	3
I_0	Infectadas iniciales	1
R_0	Recuperados iniciales	0
N	Población total (2020)	9442000
t	Tiempo (días)	300

Para finalizar los modelos compartimentales o modelos tradicionales, el modelo SIS es un modelo matemático que comparte muchas similitudes como el modelo SIR, con la diferencia que el grupo de individuos recuperados pasan a ser susceptibles con las características que en este modelo no se consideran las muertes debido al tipo de enfermedad que se modela ya que un individuo fallecido no puede ser susceptible para el cual [29] describe el modelo SIS mediante las siguientes ecuaciones:

$$\frac{dS}{dt} = -\beta IS + \gamma I \quad (9)$$

$$\frac{dI}{dt} = \beta IS - \gamma I \quad (10)$$

Para cada modelo se ajustarán los parámetros de transmisión (β), recuperación (γ) y exposición (σ en el caso del modelo SEIR) con base en los datos recopilados. Además, se evaluará la capacidad de predicción de cada modelo mediante simulaciones para diferentes periodos de tiempo.

C. Implementación de Modelos de Aprendizaje Automático

En esta investigación se eligieron tres modelos de aprendizaje automático: Máquinas de Vectores de Soporte (SVM), Bosques Aleatorios (Random Forest) y Regresión Polinomial. La elección se basó en trabajos anteriores que mostraron buenos resultados en la predicción de casos de COVID-19 usando este tipo de enfoques. Los modelos SVM y Random Forest son adecuados porque pueden reconocer patrones no lineales y cambios en el tiempo dentro de los datos, logrando una buena precisión incluso cuando la cantidad de información no es muy grande [30].

Por otro lado, se incluyó la Regresión Polinomial como un modelo más simple que permite representar la tendencia general de los contagios. Este método ayuda a analizar la forma de la curva epidémica y a describir cómo cambian los casos, muertes y recuperaciones a lo largo del tiempo. En conjunto, estos modelos ofrecen un equilibrio entre precisión y facilidad de interpretación, lo que permite obtener resultados confiables sin necesidad de estructuras muy complejas [31].

Los modelos de machine Learning tienen la facilidad de aprender de un volumen grande de datos y a partir de ahí realizar tareas de clasificación o predicción lo que es una opción viable que puede reemplazar las técnicas básicas de los modelos tradicionales [23].

Para los modelos de *Machine Learning*, se utilizará la Regresión Lineal, un método que permite estimar el valor de una variable dependiente a partir de una o más variables independientes, las cuales describen la relación o correlación existente con la variable a predecir [32]. Este modelo se representa de la siguiente manera:

$$y = b_0 + b_1x_1 + b_2x_2 + b_nx_n \qquad (11)$$

Donde b_0 es el interceptó y cada X_n son los factores que intervienen dentro del modelo y que para tener una mayor aproximación se realizara una transformación polinomial en los datos cambiando los valores de X por:

$$\varphi(x) = [1, x, x^2, .., x^d] \qquad (12)$$

Para el modelo Maquinas de Vectores de Soporte (SVM) consiste en encontrar un vector en el caso que sea dos dimensiones o un hiperplano que logre separar los grupos de manera más optima, donde para lograr la mejor predicción en este modelo se busca entrenar de manera general con la mayoría de los datos para poder mejorar la predicción.

Por último, el modelo de Random Forest es una combinación de árboles predictivos lo que hace es promediar la colección de árboles que se genera tomando que cada árbol dependa de un valor de la muestra de manera independiente.

El elemento común en todos estos procedimientos de Random Forest que para el k-ésimo árbol se genera un vector aleatorio θk independiente de los últimos vectores aleatorios θ1, …, θk-1 pero con la misma distribución; y un árbol se desarrolla usando el conjunto de entrenamiento y de θk, lo que resulta en un clasificador donde h(x, θk) es un vector de entrada [33].

En la etapa de entrenamiento, el algoritmo intenta optimizar los parámetros de las funciones de Split a partir de las muestras de entrenamiento.

$$\theta_k = argmax_\theta \qquad (13)$$

El desarrollo de esta investigación se hizo a través del software Python donde se simulará cada uno de los modelos para ejemplificar mejor mediante gráficos dinámicos. Dentro del programa Python se emplean diferentes librerías las cuales se describen de la siguiente manera [33] :

1. *np para numpy*: sirve para el procesamiento numérico y de matrices

2. *pd de pandas*: funciona para cargar el data set y poder crear el dataframe

3. *sns para seaborn:* que se lo emplea para manejar un mejor entorno gráfico.
4. *odeint:* Una función que define el sistema de ecuaciones diferenciales (la derivada de las variables respecto al tiempo).

5. *TimeSeriesSplit:* se utiliza para realizar validación cruzada en series temporales.

D. Entrenamiento, validación y evaluación de los modelos

Durante esta fase de la investigación se trabajó con la base de datos previamente procesada, construida a partir de los registros diarios históricos de casos de COVID-19 mencionados anteriormente. A partir de esta información se derivaron variables epidemiológicas inspiradas en los modelos SIS, SIR y SEIR, que describen la evolución de la población entre los grupos de personas susceptibles, infectadas y recuperadas a lo largo del tiempo.

Para capturar la progresión temporal de los contagios se incorporaron variables autorregresivas como Infected_t-1, Infected_t-2 y Infected_t-3 que representan el número de personas infectadas en los días previos. Estas variables permiten que el modelo aprenda la tendencia de crecimiento o disminución de los contagios y anticipe su evolución. Este enfoque busca reflejar la naturaleza dinámica de la propagación del virus, de forma similar a investigaciones previas que aplicaron técnicas de aprendizaje automático para analizar y predecir la evolución temporal de la pandemia [30].

El conjunto de datos se dividió en dos partes: los primeros 200 días se utilizaron para el entrenamiento de los modelos y el resto para la fase de prueba, manteniendo la independencia temporal entre ambas etapas. Los modelos SVM y Regresión Polinomial se entrenaron con datos escalados mediante StandardScaler, lo que permitió normalizar las magnitudes y mejorar la estabilidad numérica. En el caso del SVM, se empleó un kernel radial (RBF) con los parámetros C = 100 y ε = 0.1, configurado para capturar relaciones no lineales entre los valores diarios de infectados, el uso de este tipo de kernel ha demostrado mejorar la precisión en la predicción de contagios al representar patrones no lineales en los datos epidemiológicos [30].

El modelo Random Forest se configuró con 100 árboles de decisión y una semilla aleatoria (random_state = 42) para asegurar la reproducibilidad de los resultados. La validación se realizó comparando las predicciones con los valores reales

mediante el Error Cuadrático Medio (RMSE) y el coeficiente de determinación (R^2), métricas comúnmente utilizadas para medir la precisión de modelos predictivos en salud pública. En general, la Regresión Polinomial obtuvo el mejor desempeño, seguida del SVM, mientras que Random Forest tendió a suavizar los valores extremos, en concordancia con su naturaleza promediadora.

IV. Resultados

En la Figura 1 se muestran las simulaciones del modelo SEIR (izquierda) y SIR (derecha), comparando los valores reales (línea azul) con las predicciones del modelo (línea verde). En la parte inferior se presenta el desempeño del modelo SIS(izquierda) y la comparación directa entre las predicciones SIR y SIS (derecha). Los ejes representan el número de infectados (eje Y) y los días transcurridos (eje X). Se observa que el modelo SIR logra un mejor ajuste general a los datos reales que el SEIR y SIS.

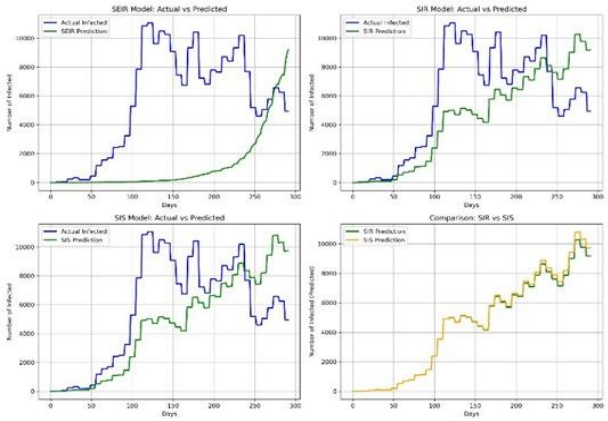

Fig. 1 Comparación de los modelos epidemiológicos tradicionales (SEIR, SIR y SIS) para la predicción de casos de infección por COVID-19 en Honduras (2020).

En la Figura 2 se muestran los resultados de la comparación de los valores reales de casos infectados (línea azul) con las predicciones generadas por tres modelos: Random Forest (arriba), Support Vector Machine – SVM (centro) y Regresión Polinómica (abajo). Cada modelo fue entrenado con datos hasta el día 200 y realizó predicciones a partir de ese punto (línea discontinua vertical). Los ejes representan el número de infectados (eje Y) y los días transcurridos (eje X). Se observa que la regresión polinómica logra el mejor ajuste general a los datos reales, seguida por SVM y Random Forest.

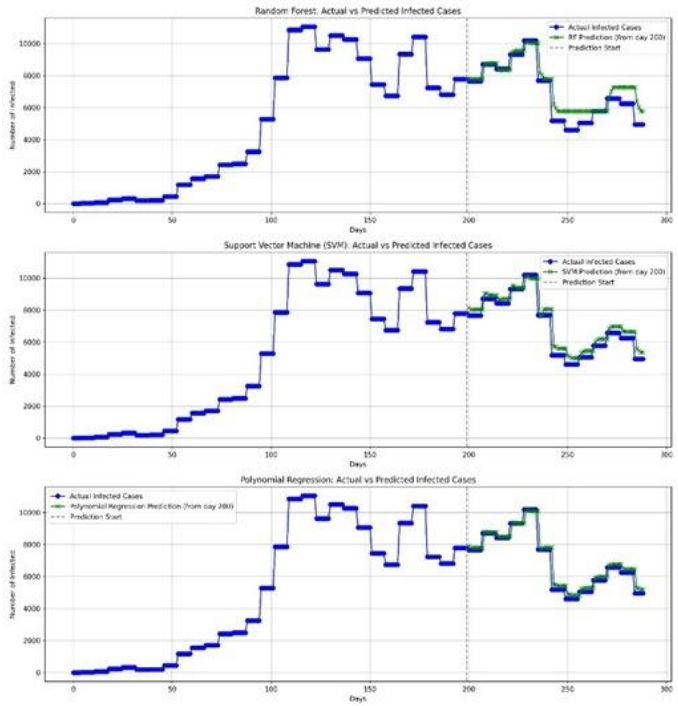

Fig. 2 Resultados de los modelos de aprendizaje automático en la predicción de casos de COVID-19 en Honduras (2020).

V. Discusión

La utilización de modelos epidemiológicos tradicionales como el SIR, SEIR y SIS ha proporcionado resultados diversos desde el punto de vista de la fiabilidad para predecir los casos de infección. Por ejemplo, el modelo SIR no refleja de manera íntegra la conducta epidemiológica descrita, especialmente para las fases iniciales del crecimiento. Tal como se expone en [34], este modelo no toma en consideración las interacciones locales entre la población, ni los casos de personas infectadas que han muerto, etc. Esta conducta se puede explicar, entonces, por la simplificación de la cual adolece el modelo, el cual no tiene en cuenta mecanismos importantes como la reinfección, la evolución temporal de la tasa de transmisión debida a las medidas de control social, fluctuaciones de la movilidad de la población, etc.

Resultados similares fueron constatados, ya que diversos estudios han observado que el modelo SIR tiende a subestimar los picos de contagio en contextos donde la transmisión varía de manera temporal o geográfica [1, 4]. El modelo SEIR, por su parte, presentó limitaciones aún más notorias, coincidiendo con algunas investigaciones donde señalaron que su desempeño disminuye al enfrentar datos reales con alta variabilidad o ruido [35]. En esta investigación, la aparición de valores negativos en la población expuesta refleja la rigidez del modelo para adaptarse a escenarios con cambios abruptos en las tasas de contagio o en el comportamiento social. Esto sugiere la necesidad de aplicar enfoques híbridos o versiones estocásticas

de los modelos SEIR que incorporen incertidumbre en los parámetros, tal como proponen [21] y [14].

Los resultados para el modelo SIS fueron intermedios entre los obtenidos para los modelos SIR y SEIR. Aunque este modelo logró captar mejor la tendencia general del número de infectados, se presenciaron grandes oscilaciones en los periodos de reinfección, lo que podría extender la hipótesis de que tener en cuenta una inmunidad no duradera es una aproximación. observada durante la pandemia esto coincide con lo afirmado en estudios previos, los cuales indican que el modelo SIS resulta adecuado para enfermedades en las que los individuos recuperados vuelven a ser susceptibles. Sin embargo, también señalan que, en muchos casos, este modelo tiende a ajustar de manera inadecuada la tasa de recuperación. De hecho, en los resultados presentados, las oscilaciones en los datos y la falta de información sobre la reinfección pudieron afectar la estabilidad del modelo [12].

A diferencia de los modelos tradicionalmente utilizados para dicho análisis, los modelos que se basan en el aprendizaje automático mostraron cualidades distintas, y también un rendimiento notablemente superior. La regresión polinómica, en particular, logró un excelente ajuste para la tendencia real de la cual estábamos dispuestos, como lo muestran los elevados valores de R^2 y el menor RMSE, lo cual pone de manifiesto la capacidad del modelo para recoger las variaciones no lineales que son propias de los datos epidemiológicos que se han estudiado. Resultados similares fueron obtenidos por [24]; para quienes los modelos polinómicos ofrecían una mejor predicción con respecto a los tradicionales modelos lineales en el modelado de curvas epidémicas de COVID- 19 para Ecuador. El modelo *Support* Vector Regression (SVR) tampoco presentó un desempeño deficiente, lo cual coincide con lo reportado en estudios que evidencian el potencial del SVR para identificar picos epidémicos a corto plazo [18]. No obstante, la ligera subestimación de los picos, evidenciada en este estudio, puede tener su origen en una calibración conservadora de los hiperparámetros (C y gamma), y, por tanto, provoca que el modelo sea poco sensible a situaciones extremas.

El modelo Random Forest, por su parte, mostró una tendencia a suavizar los comportamientos abruptos, donde se logra destacar que el promediado interno de los árboles reduce la varianza, pero también la capacidad del modelo para detectar cambios repentinos [36]. Aun así, su desempeño fue razonable y sugiere que este tipo de modelos puede ser útil en escenarios con datos incompletos o ruidosos, donde otros enfoques más sensibles podrían sobreajustarse.

Desde una posición crítica, se debe aceptar que este análisis poseía ciertas limitaciones. Primeramente, los datos utilizados son datos oficiales que pueden estar expuestos a un subregistro o a retrasos en las notificaciones, lo que, como consecuencia, incorpora sesgos en la estimación de los parámetros y, por lo tanto, impacta también en la validación de los modelos. Segundo, la comparación entre estos modelos epidemiológicos y de machine learning devienen de supuestos diferentes, de manera tal que, mientras que los primeros tenían el objetivo de interpretar la dinámica biológica del contagio, los segundos se hallan predispuestos a dar prioridad a la mayor precisión predictiva, por lo que esta comparación entre ambos modelos se hace difícilmente superponible. Por último, se puede indicar que no se probaron técnicas para ajustar hiperparámetros más exhaustivas y la selección de variables explicativas no fue evaluada para comprobar su influencia sobre la estabilidad de las soluciones.

Resumidamente, puesto que quedan muchos del análisis conceptual de la difusión de las enfermedades las fuentes epidemiológicas tradicionales no dejan de tener valor, los sistemas de Machine Learning sí que demuestran ser modelos predictivos de mejor desempeño, siempre que contemos con un conjunto de datos amplios y de calidad. Pero las predicciones sólidas requieren de una buena calibración, del control del sesgo presente en la información y de la capacidad para poner los resultados en su contexto. Trabajos futuros deberían evaluar planteamientos híbridos que incorporen a los modelos compartimentales y su interpretabilidad junto a las técnicas de Machine Learning debido a su capacidad de adaptarse, tal como ha sido planteado en trabajos previos [15, 17], con el fin de ayudar a la toma de decisión en la salud pública cuando se enfrentan a nuevas epidemias.

VI. Conclusiones

Se llevaron a cabo las prácticas de los modelos compartimentales SEIR, SIR y SIS, confidentes de la estimación dinámica de la tasa de transmisión $\beta(t)$ recabada con datos concretos de la difusión del COVID-19 en Honduras. Como se observa en la Figura 1, el modelo SIR reflejó parcialmente la tendencia del conjunto de infectados, principalmente durante la fase de crecimiento, pero mostró limitaciones por ocultar las crestas y sobrestimar las caídas, lo que puede explicarse por su estructura simplista y por pasar por alto fenómenos como la reinfección, la variabilidad en la movilidad social o intervenciones sanitarias que tuvieron lugar en el período analizado.

Por otro lado, el modelo SEIR arrojó el peor desempeño, con predicciones que se alejaron de los datos empíricos hacia el final del periodo analizado. A pesar del objetivo de restringir el valor máximo de $\beta(t)$, el modelo mostró desajustes como valores negativos en la población expuesta, problemas estructurales del modelo que se manifiestan ante datos ruidosos y dinámicas epidemiológicas complejas. En este comportamiento puede encontrarse una dificultad del modelo para representar con precisión la fase de la exposición en contextos de elevada variabilidad temporal en los datos.

Por su parte, el modelo SIS presentó un ajuste intermedio entre SIR y SEIR. Este modelo reprodujo de manera razonable la tendencia general de los casos infectados y mostró una mejor aproximación en los periodos de reinfección, lo cual es consistente con su supuesto de inmunidad no permanente. Sin embargo, también evidenció fluctuaciones marcadas en los picos de contagio, lo que sugiere que la tasa de recuperación

utilizada pudo haber sido insuficiente para representar correctamente los periodos de disminución de casos.

Por último, una comparación de los modelos SIR y SIS que se realizó (panel inferior derecho de la Figura 1) es indicativa de que presentaban comportamientos similares a la hora de hacer la predicción global, aunque el modelo SIR tendía a dar una suavidad a los picos de contagio, mientras que el modelo SIS siempre daba una respuesta más oscilante en respuesta de las abruptas variaciones de datos. En resumen, los resultados mostraban, que si bien los tres modelos podrían permitir tener una visión más o menos ampliada de la evolución de la epidemia, el modelo SIR podía ser el que daba un mejor ajuste a la tendencia del crecimiento y decrecimiento de los casos en la manera que lo describía, mientras que el modelo SIS aportaba datos complementarios respecto a otros posibles escenarios de reinfección y variabilidad temporal.

Los resultados que se han generado a partir de los modelos de aprendizaje automático demostraron un amplio ajuste general a los datos reales, en comparación con los modelos epidemiológicos que se habían utilizado con anterioridad. La regresión polinómica de segundo grado logró el mejor desempeño, ya que fue capaz de conseguir la mínima predicción de error cuadrático medio (RMSE = 492.07) y el mayor coeficiente de determinación (R^2 = 0.9222). Esto, se traduce como una buena capacidad para poder captar las variaciones no lineales en la serie de casos infectados. Como podemos observar en la Figura 2 (panel inferior), el modelo reproduce adecuadamente las fases de ascenso y descenso de la curva epidémica, manteniendo una elevada correlación con los datos reales, incluso en las zonas de cambio más acentuado. La tendencia observada se corresponde con la que pudieron evidenciar [26], donde las regresiones polinómicas logran describir adecuadamente los patrones no lineales en series epidemiológicas, siempre que los datos sean suficientemente densos y continuos.

El modelo Support Vector Regression (SVR) mostró un comportamiento competitivo (RMSE = 579.60, R^2 = 0.8920), si bien se observó cierta subestimación de los picos más significativos, que se produjeron sobre todo durante períodos de rápido crecimiento de los casos. Este resultado se puede explicar mediante el hecho de que, hasta cierto punto, el modelo SVR, que aproxima sus márgenes de error sobre el hiperplano de regresión al máximo, va a tender a la estabilidad en vez de ser extremadamente sensible a la variabilidad. Su rendimiento fue, sin embargo, muy estable y robusto ante posibles outliers, lo que se ha encontrado en otras pruebas [18] en las que indicaban la utilidad del SVR en las predicciones a corto plazo de las epidemias de COVID-19, aunque con una limitada detección de picos súbitos.

Respecto del modelo Random Forest, los resultados, aunque mostraran un RMSE = 764.11 (r modificado = 0.8123), deberían señalarse como valores aceptables en relación con los que ya se habían presentado en los modelos precedentes. En la Figura 2 (panel superior), se hace notar que el modelo commence a moderar los comportamientos abruptos que muestran los datos, especialmente los de después del día 200, lo cual podría interpretarse como un efecto de un exceso de promediado que podría estar propiciado por la naturaleza del modelo propio de Random Forest, dependiente de árboles de decisión. Este asunto fue destacado por [36], los cuales concluyeron que el modelo Random Forest podría ser considerado una opción válida en contextos donde hay ruido, pero también que reduce la sensibilidad precisamente en las situaciones que requieren que se detecten cambios locales en series con temporalidad elevada. No obstante, lo anterior, este modelo demostró ser el más estable en términos de la consistencia entre los momentos de entrenamiento y de validación, reforzando de esta manera su aplicación en contextos en los cuales la robustez se prioriza más que la precisión puntual.

En general, los tres modelos muestran que las aplicaciones de los algoritmos de machine learning como herramientas para la predicción de series temporales epidemiológicas son prometedores sobre todo cuando hay que incorporar relaciones no lineales y/o comportamientos de corta duración. Sin embargo, el desempeño final está muy condicionado por el adecuado ajuste de hiperparámetros y por el diseño de los datos de entradas. Un punto a tener en cuenta es que estos modelos no incorporan explícitamente mecanismos biológicos o sociales de transmisión, por lo que su posibilidad explicativa es limitada comparándolo con los modelos epidemiológicos clásicos; por tanto, el uso de enfoques híbridos donde los algoritmos de machine learning complementarían a los modelos compartimentales, dados los sistemas predictivos más adaptativos y útiles para la gestión de emergencias sanitarias, como proponen [15] y [17].

REFERENCES

[1] D. Ortega-Lenis, D. Arango-Londoño, E. Muñoz, and others, "Predicciones de un modelo SEIR para casos de COVID-19 en Cali, Colombia," *Revista de Salud Pública (Bogotá)*, vol. 22, no. 2, pp. 132–137, 2020.

[2] J. Wei, X. Yu, J. Liu, and others, "SIR Modelo de Propagación Basado en Autómatas Celulares y el Sistema de Entrega Basado en el Modelo de Mochila para Erradicar el Ébola," *Investigación Clínica*, vol. 60, p. 1358, 2019.

[3] F. J. Ariza Hernandes and M. P. Árciga Alejandre, "16250588_TM2016_1," Universidad Autónoma de Guerrero, 2018.

[4] S. I. Polo-Triana, Y. A. Ramírez-Sierra, J. E. Arias-Osorio, and others, "Métodos de aprendizaje automático para predecir el comportamiento epidemiológico de enfermedades arbovirales: revisión estructurada de literatura," *Revista de Salud*, vol. 55, no. 1, 2022.

[5] W. Belloso, F. Pasterán, V. Burgos, and others, *Inteligencia artificial en enfermedades infecciosas*.

[6] M. Angeli, G. Neofotistos, M. Mattheakis, and E. Kaxiras, "Modeling the effect of the vaccination campaign on the COVID-19 pandemic," *Chaos, Solitons & Fractals*, vol. 154, p. 111621, 2022, doi: 10.1016/j.chaos.2021.111621.

[7] T. Lazebnik and G. Blumrosen, "Advanced multi-mutation with intervention policies pandemic model," *IEEE Access*, 2022, doi: 10.1109/ACCESS.2022.3149956.

[8] E. Bausili Llamas, "Machine Learning para el Diagnóstico de COVID19," Universidad Autónoma de Madrid, 2021.
[9] Y. Quintana Cruz, *Matemáticas y epidemiología: modelos y conclusiones*.
[10] C. E. Álvarez Cabrera, E. J. Andrade, and V. Gauthier Umaña, "Modelos epidemiológicos en redes: una presentación introductoria," *Boletín de Matemáticas*, vol. 22, no. 1, pp. 21–37, 2015.
[11] J. N. Tavares, "Modelo SIR em epidemiologia," *Revista Ciência Elementar*, vol. 5, no. 2, 2017.
[12] S. Hyun Ho, *Modelos epidemiológicos basados en ecuaciones diferenciales*.
[13] D. Aleja, R. Criado, and M. Romsnce, *Modelo SEIR para el Coronavirus COVID-19*. España.
[14] M. Casals, K. Guzmán, and J. A. Caylà, "Modelos matemáticos utilizados en el estudio de las enfermedades transmisibles," *Revista Española de Salud Pública*, vol. 83, no. 5, pp. 689–695, 2009.
[15] B. J. A. Betancourt, E. Ortiz Hernández, A. González Mora, and others, "Enfoque de los sistemas complejos en la Epidemiología," *Revista Archivo Médico de Camagüey*, vol. 13, 2009.
[16] M. Castro, S. Ares, J. A. Cuesta, and others, "The turning point and end of an expanding epidemic cannot be precisely forecast," *Proceedings of the National Academy of Sciences*, vol. 117, no. 42, pp. 26190–26196, 2020.
[17] S. K. Yadav, V. Kumar, and Y. Akhter, "Modeling global COVID-19 dissemination data after the emergence of Omicron variant using multipronged approaches," *Current Microbiology*, 2022, doi: 10.1007/s00284-022-02985-4.
[18] L. J. Sandoval, "Algoritmos de aprendizaje automático para análisis y predicción de datos," *Revista Tecnológica*, no. 11, pp. 36–40, 2018.
[19] K. Quintero Acuña, "Aplicación de Machine Learning a un modelo tradicional de prevención y detección de fraude en entidad financiera proyectado a periodos trimestrales," Universidad de La Salle, Colombia.
[20] J. Garcia Piñeiro, "Inteligencia Artificial aplicada a la Epidemiología," Universidad de Salamanca, España, 2022.
[21] M. Kumaresan, M. Senthilkumar, and N. Muthukumar, "Analysis of mobility based COVID-19 epidemic model using federated multitask learning," *Mathematical Biosciences and Engineering*, 2022, doi: 10.3934/mbe.2022466.
[22] T. Daghriri, M. Proctor, and S. Matthews, "Evolution of select epidemiological modeling and the rise of population sentiment analysis: A literature review and COVID-19 sentiment illustration," *International Journal of Environmental Research and Public Health*, vol. 19, no. 6, p. 3230, 2022, doi: 10.3390/ijerph19063230.
[23] S. Prado Medina and S. A. Quintero Rodriguez, "Exploración del uso de técnicas de machine learning para obtener proyecciones del comportamiento de la pandemia Covid 19," Universidad Militar Nueva Granada, Colombia, 2021.
[24] R. Zreiq, S. Boubaker, S. Kamel, M. Alzain, and F. D. Algahtani, "Analysis and modeling of COVID-19 epidemic dynamics in Saudi Arabia using SIR-PSO and machine learning approaches," *Journal of Infection in Developing Countries*, 2022, doi: 10.3855/jidc.15004.
[25] H. Jia, S. Shen, J. A. Ramírez-García, and C. Shi, "Partner with a thirdparty delivery service or not? A prediction-and-decision tool for restaurants facing takeout demand surges during a pandemic," *Service Science*, 2022, doi: 10.1287/serv.2021.0294.
[26] C. Colas *et al.,* "EpidemiOptim: A toolbox for the optimization of control policies in epidemiological models," *Journal of Artificial Intelligence Research*, vol. 71, 2021, doi: 10.1613/JAIR.1.12588.
[27] J. M. Barros García and C. M. Guillén Juan, *Documentar las pruebas de limpieza: uso de bases de datos*.
[28] F. Bill and M. Gates, *Modelos SIR y SIRS — Documentación del modelo de VIH.* [Online]. Available: https://docs.idmod.org/projects/emod-hiv/en/2.20_a/model-sir.html
[29] A. S. Ríos Gutiérrez, "Modelos Epidemiológicos Estocásticos y su Inferencias: Casos SIS y SEIR," Universidad Nacional de Colombia.
[30] V. Chaurasia and S. Pal, "Application of machine learning time series analysis for prediction COVID-19 pandemic," *Research on Biomedical Engineering*, vol. 38, no. 1, pp. 35–47, 2022.
[31] G. E. Chanchí Golondrino, W. Y. Campo Muñoz, and L. M. Sierra Martinez, "Aplicación de la regresión polinomial para la caracterización de la curva del COVID-19, mediante técnicas de machine learning," *Investigación e Innovación en Ingeniería*, vol. 8, no. 2, pp. 87–105, 2020.
[32] D. Patiño Pérez, C. Munive Mora, L. Cevallos-Torres, and others, "Predicción de la Efectividad de las Pruebas Rápidas Realizadas a Pacientes con COVID-19 mediante Regresión Lineal y Random Forest," *Ecuadorian Science Journal*, vol. 5, no. 2, pp. 31–43, 2021.
[33] R. F. Medina Merino and C. I. Ñique Chacón, "Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python," *Interfases*, no. 10, pp. 165–189, 2017.
[34] M. Culqui Sanchez, J. Nasimba Quinatoa, and E. Chiquinga Calderón, "Aplicación del modelo matemático SEIR en la pandemia por Covid19, relevancia en salud pública," *Vive. Revista de Investigación en Salud*, vol. 3, no. 9, 2020.
[35] E. K. Grillo Ardila, J. Santaella-Tenorio, R. Guerrero, and others, "Mathematical model and COVID-19," *Colombia Médica*, vol. 51, no. 2, e4277, 2020.
[36] D. Patiño Pérez, R. Silva Bustillo, C. Munive Mora, and others, "Predicción de Covid19 con el uso de Algoritmo de Bosques Aleatorios y Redes Neuronales Artificiales," vol. 4, no. 2, 2020.