



RES
Revista de Ciencias Sociales

Depósito legal ppi 201502ZU4662
Esta publicación científica en formato
digital es continuidad de la revista impresa
Depósito Legal: pp 197402ZU789
● ISSN: 1315-9518 ● ISSN-E: 2477-9431

Universidad del Zulia. Revista de la Facultad de Ciencias Económicas y Sociales
Vol. XXIX, No. Especial 8 julio-diciembre 2023

Revista de Ciencias Sociales

Esta publicación científica en formato
digital es continuidad de la revista impresa
Depósito Legal: pp 197402ZU789
ISSN: 1315-9518

Modelo de ensamble para predecir el diagnóstico de pacientes con sospecha de Covid-19

Andrade-Girón, Daniel*
Sandivar-Rosas, Juana**
Carreño-Cisneros, Edgardo***

Resumen

La pandemia de Covid-19 ha afectado a millones de personas en todo el mundo, siendo calificada por la Organización Mundial de la Salud como una crisis de preocupación global. Esto ha generado la necesidad de realizar una predicción oportuna del diagnóstico de pacientes con alto riesgo de deterioro clínico en establecimientos médicos. El objetivo de este estudio es diseñar y comparar el rendimiento de modelos de aprendizaje automático basados en ensambles de máquinas para predecir pacientes con sospecha de Covid-19. La investigación sigue el paradigma positivista, enfoque cuantitativo, diseño observacional sin intervención, nivel predictivo. Para llevar a cabo este estudio, se recopilaron 1.000 registros de pacientes de dos establecimientos de salud en Perú. La construcción de los modelos se basó en algoritmos de ensamble, como Random Forest, Extra Trees, Gradient Boosting y AdaBoosting. Al comparar los modelos en términos de exactitud (accuracy), que mide el porcentaje de casos clasificados correctamente como pacientes con sospecha de Covid-19, se obtuvo un 97% de precisión para los modelos basados en Random Forest y Gradient Boosting. Además, el valor de Kappa de Cohen fue de 0.95, lo cual indica una concordancia muy buena entre el resultado de predicción del modelo y los datos reales.

Palabras clave: Modelo de ensamble; aprendizaje automático; riesgo clínico; clasificación; Covid-19.

* Doctor en Ingeniería. Magíster en Matemática. Docente Universitario en la Universidad Nacional José Faustino Sánchez Carrión, Huacho, Perú. E-mail: dandrade@unjfsc.edu.pe ORCID: <https://orcid.org/0000-0002-9746-3583>

** Doctora en Ingeniería. Magíster en Ingeniería de Sistemas. Docente Universitario en la Universidad Nacional Mayor de San Marcos, Lima, Perú. E-mail: jsandivarr@unmsm.edu.pe ORCID: <https://orcid.org/0000-0001-8494-2940>

*** Doctor en Educación. Magíster en Ingeniería de Sistemas. Docente Universitario en la Universidad Nacional José Faustino Sánchez Carrión, Huacho, Perú. E-mail: ecarreno@unjfsc.edu.pe ORCID: <https://orcid.org/0000-0001-7063-7072>

Ensemble model to predict the diagnosis of patients with suspected Covid-19

Abstract

The Covid-19 pandemic has affected millions of people around the world, being described by the World Health Organization as a crisis of global concern. This has generated the need to perform a timely prediction of the diagnosis of patients with high risk of clinical deterioration in medical establishments. The aim of this study is to design and compare the performance of machine-assembly-based machine learning models to predict patients with suspected Covid-19. The research follows the positivist paradigm, quantitative approach, observational design without intervention, predictive level. To carry out this study, 1,000 patient records from two health facilities in Peru were collected. The construction of the models was based on assembly algorithms, such as Random Forest, Extra Trees, Gradient Boosting and AdaBoosting. When comparing the models in terms of accuracy, which measures the percentage of cases correctly classified as patients with suspected Covid-19, a 97% accuracy was obtained for the models based on Random Forest and Gradient Boosting. In addition, Cohen's Kappa value was 0.95, which indicates a very good agreement between the model prediction result and the actual data.

Keywords: Assembly model; machine learning; clinical risk; classification; Covid-19.

Introducción

El coronavirus se ha convertido en una amenaza para la salud mundial debido a su rápida propagación geográfica en los últimos años (Sánchez-Duque, Arce-Villalobos y Rodríguez-Morales, 2020). El 31 de diciembre de 2019, las autoridades chinas notificaron a la Organización Mundial de la Salud (OMS) sobre un grupo de casos de neumonía de origen desconocido en Wuhan, China (Dunlop et al., 2020).

El 11 de marzo de 2020, la OMS declaró la pandemia de SARS-CoV-2, basándose en la información global disponible, la dinámica de propagación y su alta peligrosidad (Guan et al., 2020). Se mencionó que este virus era altamente letal y presentaba similitudes virológicas con el SARS-CoV-1 (HCoV-229E), que surgió en 2009, esta declaración marcaría un hito en la historia de la humanidad al cambiar su curso (De León et al., 2020; Luna-Nemecio, 2020; Marruecos, 2020).

La pandemia de SARS-CoV-2 ha provocado un gran número de muertes a nivel

mundial. Fundamentalmente, se debe a la falta de diagnóstico oportuno y atención inmediata a los pacientes vulnerables en el ámbito de la salud (Romero-Hernández et al., 2020). Las estrategias preventivas juegan un papel crucial en la reducción de la propagación pública del virus, al igual que el exitoso aislamiento de la enfermedad y la contención comunitaria (Sánchez-Duque et al., 2020).

Los sistemas de atención primaria desempeñan un papel fundamental como estrategia para proporcionar una respuesta inmediata durante una pandemia (Gómez-Gómez et al., 2020); en Reino Unido, la atención primaria abarca más del 95% de todas las actividades del sistema de salud (Dunlop et al., 2020).

En los establecimientos de salud a nivel mundial, se ha observado un amplio espectro clínico en los pacientes con Covid-19, que abarca desde una presentación asintomática hasta el desarrollo del síndrome de distrés respiratorio agudo (SDRA) e incluso la muerte (García, 2020). Esto ha generado un problema complejo en los diferentes establecimientos de

salud, puesto que muchos de ellos carecen de los equipos necesarios y del personal suficiente para brindar una atención oportuna y adecuada a los pacientes (Jiang et al., 2019).

La falta de un diagnóstico temprano y una intervención oportuna en el tratamiento aumentan el riesgo clínico para los pacientes con Covid-19 (Mojica-Crespo y Morales-Crespo, 2020). Como consecuencia, muchos países, incluyendo Perú que ocupa el primer lugar en la región, han experimentado un alto porcentaje de fallecimiento (Elguea, Prado y Barradas, 2019).

La optimización del tiempo en la detección de enfermedades es una estrategia fundamental en las actividades de los sistemas de salud. La identificación oportuna del riesgo clínico y el tratamiento adecuado del paciente permiten reducir los índices de mortalidad (Chen, Liu y Guo, 2020). En los servicios de emergencias médicas, es crucial clasificar de manera oportuna a los pacientes con Covid-19 según su nivel de gravedad clínica, puesto que muchos de ellos presentan alta comorbilidad y pluripatología, así como signos y síntomas específicos (Martínez y Del Pino, 2020; Lora et al., 2023).

En numerosos centros de salud a nivel mundial, se han empleado el triaje respiratorio y las escalas de evaluación de gravedad, junto con la valoración del riesgo de mortalidad, en pacientes con sospecha de infección (Chen et al., 2020). La Sociedad Torácica Británica sostiene que la implementación de la Escala Nacional de Alerta Temprana (NEWS, por sus siglas en inglés) en los servicios de emergencia facilita la detección de pacientes con infección por SARS-CoV-2, permitiendo tomar decisiones respecto al ingreso o la hospitalización (Manosalvas y Zamora, 2021).

En los centros de salud, es necesario automatizar el proceso de diagnóstico para proporcionar una clasificación oportuna del paciente de acuerdo con su estado clínico (Preciado, Valles y Lévano, 2021). Los métodos estadísticos tradicionales, como la clasificación basada en máxima verosimilitud, presentan limitaciones debido a las suposiciones de distribución normal y

las restricciones impuestas por los datos de entrada (Fuentes y Medina, 2021).

Se ha planteado como alternativa el diseño de modelos de aprendizaje automático que utiliza diversos algoritmos, como el de los k-vecinos más cercanos, los clasificadores de Bayes, los árboles de decisión, y la máquina de soporte vectorial, para realizar predicciones y clasificaciones (Tillaguango, 2021).

En ese sentido, se han realizado aplicaciones de gran relevancia en la clasificación de pacientes con sospecha de Covid-19 mediante técnicas de aprendizaje automático, que permiten identificar características en imágenes de rayos X de tórax (Zoabi, Deri-Rozov y Shomron, 2021; Dilmi, 2022; Kesav y Jubukumar, 2022; Luna y Carranza, 2023).

Dado un problema de clasificación o predicción, se pueden aprender muchas máquinas de aprendizaje automático con diferentes parámetros y, en general, se selecciona el modelo óptimo. Esta propuesta es ampliamente utilizada, pero adolece de dos problemas: Primero, se tiene que decidir cuál es el óptimo, la respuesta no es tan sencilla cuando más de un sistema candidato cumple los criterios (Xiao, Zhu y Liu, 2013). Es decir, el problema es complejo cuando el algoritmo de aprendizaje es propenso a diferentes óptimos locales, así mismo, cuando no se proporcionan suficientes datos de entrenamiento (Zachary y Papailiopoulos, 2018). En segundo lugar, se puede perder información potencialmente valiosa al descartar los modelos menos exitosos (Zachary y Papailiopoulos, 2018; Xue, 2019).

Una alternativa para abordar los desafíos de los óptimos locales y el sobre entrenamiento, es el enfoque del aprendizaje por ensamble. Este método se fundamenta en la noción de que los errores de un modelo pueden ser compensados por las predicciones acertadas de otros modelos (Xiao et al., 2013). En el aprendizaje por ensamble, se considera que un modelo es una combinación de sistemas de componentes que se unen para aumentar su capacidad de generalización en comparación con los modelos individuales que

lo conforman (Gerón, 2022; Pajankar y Joshi, 2022).

A pesar de los avances en la aplicación de la Inteligencia Artificial (IA) para hacer frente a los desafíos de la Covid-19, es crucial continuar implementando modelos eficientes para la detección, seguimiento y diagnóstico. Es necesario enfocar los esfuerzos en fortalecer las tecnologías existentes, especialmente en lo que respecta al diagnóstico diferencial temprano de Covid-19 a partir de datos clínicos (Comito y Pizzuti, 2022).

Con el fin de abordar estos vacíos de conocimiento, el objetivo de este estudio es desarrollar un modelo de ensamble para predecir el diagnóstico de pacientes con sospecha de Covid-19.

1. Fundamentación teórica

Los métodos de aprendizaje automático se vienen aplicando en la ciencia de la salud, en particular su aplicación en pandemia ha estado centrado en diferentes ámbitos. Ahamad et al. (2020), presentó un estudio para el diagnóstico de Covid-19 que implementa seis modelos predictivos utilizando seis clasificadores diferentes basados en 14 características clínicas. Esta investigación emplea algoritmos de clasificación de aprendizaje automático, incluidos el clasificador de Bayes, la regresión logística, el clasificador perezoso, el meta clasificador, la Clasificación por Regresión (CR), el aprendizaje de reglas y el árbol de decisiones.

Patiño et al. (2021), llevaron a cabo un estudio con el objetivo de determinar la efectividad de las pruebas rápidas en la detección del Covid-19 mediante el uso de máquinas de aprendizaje. Para ello, se aplicó una metodología que incluyó la creación de modelos de regresión lineal y *Random Forest*.

Comito et al. (2022), realizaron una revisión sistemática y exhaustiva sobre la aplicación de tecnologías de aprendizaje automático y aprendizaje profundo para el pronóstico, detección y diagnóstico del Covid-19. El estudio examinó y revisó

diversos algoritmos utilizados para predecir y diagnosticar Covid-19, brindando una descripción detallada de las técnicas de inteligencia artificial empleadas. Cada estudio analizado fue evaluado en términos de su enfoque lógico, métodos utilizados, tipos y tamaños de datos analizados, métodos de validación, aplicaciones objetivo y resultados obtenidos. Se hizo especial énfasis en la predicción de pacientes positivos y la predicción de riesgo.

Vaishya et al. (2020), identificaron siete aplicaciones significativas de IA para la pandemia de Covid-19. Esta tecnología juega un papel importante para detectar el grupo de casos y predecir dónde afectará este virus en el futuro mediante la recopilación y el análisis de todos los datos anteriores (Sreepadmanabh, Sahu y Chande, 2020). Moulaei et al. (2022), en el estudio realizado, encontraron que el uso de máquina de aprendizaje permite obtener un nivel de precisión razonable en la predicción de la mortalidad por Covid-19. Específicamente, los modelos predictivos basados en máquina de aprendizaje, especialmente el algoritmo de *Random Forest*, tienen el potencial de identificar a los pacientes con mayor riesgo de mortalidad y proporcionar información relevante para que los médicos realicen las intervenciones adecuadas.

Por su parte, Xiong et al. (2022) establecieron tres modelos para predecir los resultados del tratamiento utilizando 23 características, que incluían una clínica, una Tomografía Computarizada (TC) de tórax y 21 características de laboratorio. Entre los tres modelos, *Random Forest* demostró el mejor rendimiento general con un AUC más alto de 0,970, en comparación con SVM de 0,948 y LR de 0,928. Además, *Random Forest* logró una sensibilidad del 96,7%, una especificidad del 69,5% y una precisión del 84,5%. SVM tuvo una sensibilidad del 93,9%, una especificidad del 79,0% y una precisión del 88,5%. Así mismo, LR alcanzó una sensibilidad del 92,3%, una especificidad del 72,3% y una precisión del 85,2%.

Dado que no existe investigación orientada a la clasificación de pacientes

mediante el rango clínico aplicando algoritmo de ensamble de máquina de aprendizaje (Comito et al., 2022), en los últimos años se ha manifestado un creciente interés, por parte de investigadores, en la definición de métodos que combinan hipótesis denominados multi clasificadores (Corso et al., 2018).

El término ensamblado de máquina de aprendizaje, se refiere a un conjunto de metodologías que permiten combinar múltiples modelos de clasificación (Yaman, Rattay y Subasi, 2021). El objetivo es crear un meta-algoritmo que mejore el rendimiento en clasificación de los modelos individuales que lo componen (Rückert, Papenberg y Tracht, 2021).

Los expertos en el campo coinciden en que combinar diferentes modelos en un ensamblado puede ser beneficioso por tres razones principales: Primero, desde una perspectiva estadística, cualquier algoritmo de aprendizaje busca el mejor modelo en un espacio de hipótesis (Pu y Shamir, 2023). Sin embargo, debido a la incertidumbre asociada al proceso de selección del modelo y sus parámetros, siempre existe el riesgo de elegir una hipótesis incorrecta debido a un conjunto de entrenamiento limitado. Mediante la creación de un ensamblado, es posible reducir ese riesgo al combinar varios modelos efectivos disponibles (Khoshgoftaar, Van y Napolitano, 2011).

La segunda razón, se relaciona con la computación, puesto que ciertos algoritmos, como los árboles de decisión, pueden generar funciones de costo altamente no convexas, lo que puede llevar a que los métodos utilizados para resolverlos queden atrapados en óptimos locales (Yaman y Subasi, 2019). Al ensamblar diferentes hipótesis, utilizando diferentes enfoques en cada una de ellas, se incrementa la probabilidad de aproximar de manera más precisa la hipótesis verdadera. Esto se debe a que se exploran diferentes soluciones y se reduce la dependencia de un único óptimo local (Abdul et al., 2023).

La tercera razón es de naturaleza representacional. En ocasiones, la función

objetivo que se desea aproximar, f^* , no puede ser representada por ninguno de los modelos en el espacio de hipótesis (Gerón, 2022). Sin embargo, al combinar diferentes métodos que sí se encuentran en el espacio de hipótesis disponible, se puede ampliar y lograr una mejor aproximación de la hipótesis. Al incorporar una variedad de enfoques, se incrementa la capacidad del ensamblado para capturar patrones complejos y adaptarse a la función objetivo de manera más efectiva (Xiao et al., 2013).

Los métodos de ensamble pueden presentar dos tipos de arquitecturas: Homogénea o híbrida (Yaman y Subasi, 2019). En el primer caso, se utiliza un solo algoritmo de aprendizaje; mientras que en el segundo caso se permite la combinación de diferentes algoritmos. Entre los métodos de ensamble homogéneos que se hará uso en el presente estudio se encuentran *Boosting*, *Bagging* y *Random Forest*, los cuales utilizan árboles de decisión como clasificador base (Zachary y Papailiopoulos, 2018).

2. Metodología

La presente investigación sigue el paradigma positivista, enfoque cuantitativo, diseño observacional sin intervención, nivel predictivo. Para el desarrollo del modelo de aprendizaje automático se aplicó la metodología utilizada en el diseño de máquina de aprendizaje automático (Gerón, 2022).

Para desarrollar la presente investigación se recolectaron 1.000 Registros de pacientes con diagnóstico de infección por SARS-CoV-2 que ingresaron por el servicio de emergencia en centros de salud del Perú. Las variables que se consideraron fueron: Edad, género, peso, talla, frecuencia respiratoria, saturación de oxígeno, tensión arterial sistólica, frecuencia cardíaca y temperatura. Dado que el estudio es observacional sin intervención, no se solicitó consentimiento informado y se ha mantenido la confidencialidad de los datos.

2.1. Preprocesamiento de los datos

El preprocesamiento es una fase fundamental que se realiza antes de proceder al análisis. Con ello se garantiza que los datos sean adecuados para el entrenamiento y prueba del modelo. En la presente investigación se analizaron los datos con el paquete Pandas del lenguaje de programación *Python*. Se validaron los datos y registros con respecto al riesgo clínico a fin de garantizar la adecuada clasificación del modelo; se imputaron los datos faltantes; se estandarizaron las características; así como se solucionó el problema de desequilibrio en los datos. Después de analizar, reducir las características y equilibrar, se consideró una muestra de 700 registros.

2.2. Selección de algoritmo de aprendizaje

Existen diversas herramientas de aprendizaje automático. Para este estudio se eligió cuatro algoritmos de aprendizaje automático en conjunto, que incluyen *Random Forest*, *Gradient Boosting* (GB), *Adaptive Gradient Boosting* (AGB) y XGB.

2.3. Implementación del modelo y entrenamiento

Se ha desarrollado el programa haciendo uso del paquete *Scikit-learn* de *Python*. Se ha utilizado un conjunto de entrenamiento, que es el 70% de la muestra, y un conjunto de validación, que es el 30%, para lo cual se diseñó un programa orientado a mejorar los resultados. De esta manera, en cada entrenamiento se minimiza el error, lo que garantiza una mejora incremental en la eficiencia del modelo.

2.4. Evaluación y validación

Se ha comprobado que el modelo realiza una clasificación eficiente, para lo cual

se ha calculado la matriz de confusión, y se ha optimizado el rendimiento del modelo, realizando un ajuste a la validación cruzada. Se evaluó el rendimiento de los modelos de aprendizaje automático en conjunto utilizando la métrica Kappa de Kohen y la matriz de confusión (Tran y Kim, 2023).

3. Resultados y discusión

En este estudio, se implementaron modelos de ensamble de máquinas de aprendizaje para la clasificación de pacientes con sospecha de Covid-19 y se optimizaron los parámetros de cada modelo para comparar su rendimiento. El modelo que demostró el mejor rendimiento de clasificación fue *RandomForest* y *GradientBoosting*.

En la fase de recolección de datos, se identificaron 1.000 historias clínicas de pacientes con el diagnóstico de Covid-19 en dos centros de salud en Perú. Se tomaron registros de pacientes que cumplían los criterios de validación en términos de riesgo clínico. Después de realizar el preprocesamiento de datos y la ingeniería de atributos, que incluyó la imputación de datos, la selección de características y el balanceo de los datos, se obtuvo una muestra de 700 registros. La edad promedio en esta muestra fue de 59 años, con una desviación estándar de 15.90.

Se llevó a cabo un proceso de entrenamiento utilizando el 70% de las muestras tomadas aleatoriamente, y se realizó la prueba con el 30% restante. Con el objetivo de asegurar que el clasificador no dependa de la cantidad de características de los datos para su funcionamiento, es decir, con el fin de lograr la validez externa del modelo, se optimizaron los hiperparámetros, mediante la aplicación del método *GridSearchCV()* del paquete *sklearn*, para el algoritmo *RandomForest* los hiperparámetros optimizados fueron: {'max_depth': 10, 'n_estimators': 50}, como consecuencia, la evaluación reportó los siguientes resultados, que se observan en la Tabla 1.

Tabla 1
Resultados de evaluación con el Algoritmo *RandomForest*

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
Leve	0,93	1,00	0,97	57
Moderado	1,00	0,93	0,96	100
Severo	0,95	1,00	0,97	53
<i>accuracy</i>			0,97	210
<i>macro avg</i>	0,96	0,98	0,97	210
<i>weighted avg</i>	0,97	0,97	0,97	210

Fuente: Elaboración propia, 2023.

Los resultados muestran que el modelo basado en el algoritmo *RandomForest* tuvo una tasa de acierto total de 97%, una precisión en la clasificación de los pacientes en función al riesgo clínico de 96% para el macro promedio, y 97% para el promedio ponderado. Asimismo, para predecir un paciente con rango clínico severo tuvo una precisión de 95% y una *recall* (sensibilidad) de 100%; para predecir un paciente con rango clínico moderado tuvo una precisión de 100% y una *recall* (sensibilidad) de 93%; para predecir un paciente con rango clínico leve tuvo una precisión de 93% y una *recall* (sensibilidad) de 100%.

El valor *f1-score*, asume que importa de igual forma la precisión y la sensibilidad y fue calculado sobre la base de la media armónica; para el modelo en evaluación se ha obtenido para un paciente con rango clínico severo 97%, para un paciente con rango clínico moderado 96%, y para un rango clínico leve 97% (ver Tabla 1).

De igual manera, al entrenar el modelo para el algoritmo *ExtraTrees* con los hiperparámetros optimizados {'max_depth': 10, 'n_estimators': 100}, se obtuvieron los siguientes resultados que se aprecian en la Tabla 2.

Tabla 2
Resultados de evaluación con el *ExtraTrees*

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
Leve	0,92	1,00	0,96	57
Moderado	1,00	0,92	0,96	100
Severo	0,95	1,00	0,97	53
<i>accuracy</i>			0,96	210
<i>macro avg</i>	0,96	0,97	0,96	210
<i>weighted avg</i>	0,96	0,96	0,96	210

Fuente: Elaboración propia, 2023.

Los resultados de la matriz de confusión muestran que el modelo basado en el algoritmo *ExtraTrees* tuvo una tasa de acierto total de 96%; una precisión en la clasificación de los pacientes en función al riesgo clínico de 96% para el macro promedio, y 96% para el promedio ponderado. Igualmente, para predecir un paciente con rango clínico severo, tuvo una

precisión de 95% y una *recall* (sensibilidad) de 100%; para predecir un paciente con rango clínico moderado, presentó una precisión de 100% y una *recall* (sensibilidad) de 92%; para predecir un paciente con rango clínico leve, tuvo una precisión de 92% y una *recall* (sensibilidad) de 100%.

El valor *f1-score*, donde se asume

que importa de igual forma la precisión y la sensibilidad y fue calculado sobre la base de la media armónica; para el modelo en evaluación se ha obtenido para un paciente con rango clínico severo 97%, para un paciente con rango clínico moderado 96%, y para un rango clínico leve 96% (ver Tabla 2).

De igual forma, al entrenar el modelo para el algoritmo *GradientBoosting* con los hiperparámetros optimizados {'max_depth': 10, 'n_estimators': 150}, se obtuvieron los siguientes resultados que se muestran en la Tabla 3.

Tabla 3
Resultados de evaluación con el algoritmo *GradientBoosting*

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
Leve	0,95	1,00	0,97	57
Moderado	1,00	0,93	0,96	100
Severo	0,93	1,00	0,96	53
<i>accuracy</i>			0,97	210
<i>macro avg</i>	0,96	0,98	0,97	210
<i>weighted avg</i>	0,97	0,97	0,97	210

Fuente: Elaboración propia, 2023.

Los resultados de la matriz de confusión muestran que el modelo basado en el algoritmo *GradientBoosting* radial tuvo una tasa de acierto total de 97%, una precisión en la clasificación de los pacientes en función al riesgo clínico de 96% para el macro promedio, y 97% para el promedio ponderado. Asimismo, para predecir un paciente con rango clínico severo, tuvo una precisión de 93% y una *recall* (sensibilidad) de 100%. Para predecir un paciente con rango clínico moderado presentó una precisión de 100% y una *recall* (sensibilidad) de 93%; y para predecir un paciente con rango clínico leve, tuvo una precisión de 95% y una *recall* (sensibilidad) de 100%.

El valor *f1-score*, donde se asume que importa de igual forma la precisión y la sensibilidad y fue calculado sobre la base de la media armónica, para el modelo en evaluación se ha obtenido para un paciente con rango clínico severo 96%, para un paciente con rango clínico moderado 96%, y para un rango clínico leve 97% (ver Tabla 3).

Asimismo, al entrenar el modelo para el algoritmo *GradientBoosting* con los hiperparámetros optimizados {'learning_rate': 0.5, 'n_estimators': 200}, se obtuvieron los siguientes resultados que se observan en la Tabla 4.

Tabla 4
Resultados de evaluación con el algoritmo *Adaboost*

	<i>Precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
Leve	0,90	1,00	0,95	57
Moderado	0,97	0,86	0,91	100
Severo	0,86	0,94	0,90	53
<i>Accuracy</i>			0,92	210
<i>macro avg</i>	0,91	0,93	0,92	210
<i>weighted avg</i>	0,92	0,92	0,92	210

Fuente: Elaboración propia, 2023.

Los resultados de la matriz de confusión muestran que el modelo basado en el algoritmo *AdaBoost* tuvo una tasa de acierto total de 92%, una precisión en la clasificación de los pacientes en función al riesgo clínico de 91% para el macro promedio, y 92% para el promedio ponderado. Asimismo, para predecir un paciente con rango clínico severo, tuvo una precisión de 86% y una *recall* (sensibilidad) de 94%; para predecir un paciente con rango clínico moderado, presentó una precisión de 97% y una *recall* (sensibilidad) de 86%; para predecir un paciente con rango clínico leve, tuvo una precisión de 90% y una *recall* (sensibilidad) de 100%.

El valor *f1-score*, donde se asume

que importa de igual forma la precisión y la sensibilidad y fue calculado sobre la base de la media armónica, para el modelo en evaluación se ha obtenido para un paciente con rango clínico severo 90%, para un paciente con rango clínico moderado 91%, y para un rango clínico leve 95% (ver Tabla 4).

Para realizar una comparación adecuada, se evaluó la precisión del modelo y se calculó el índice de Kappa de Cohen para medir la concordancia entre las categorizaciones obtenidas y las etiquetas de referencia, estas medidas permiten evaluar la precisión y concordancia del modelo en su capacidad de clasificar correctamente los casos (ver Tabla 5).

Tabla 5
Precisión y Kappa de Cohen para los algoritmos para evaluar el rendimiento de los modelos ensamble

Algoritmo	Accuracy	Kappa de Cohen	IC-95%
<i>RandomForest</i>	97%	0.95	0.93-0.96
<i>GradientBooting</i>	97%	0.95	0.93-0.96
<i>ExtraTrees</i>	96%	0.94	0.92-0.95
<i>AdaBoost</i>	92%	0.87	0.85-0.89

Fuente: Elaboración propia, 2023.

Dado que en el modelo basado en ensamble de máquinas de aprendizaje aplicado en el sistema de salud, interesa mantener más bajos los falsos negativos (error de tipo II), puesto que, sería perjudicial para el paciente darle un diagnóstico negativo de Covid-19, cuando en realidad sí la tiene, se debe optar por un mayor valor de sensibilidad, donde la cantidad de falsos negativos es menor entre los modelos implementados, los modelos que presentan mejores resultados con esta consideración son: *RandomForest* y *GradientBoosting* con un *Accuracy* de 97% y Kappa de Cohen de 0.95 mostrando muy buena concordancia. Los resultados obtenidos superan en precisión, sensibilidad y especificidad a los resultados de las investigaciones orientadas al mismo propósito.

Zoabi et al. (2021), desarrollaron modelo de predicción de pacientes con sospecha de Covid-19 utilizando ocho características binarias:

Sexo, edad ≥ 60 años, contacto conocido con una persona infectada y la aparición de cinco síntomas clínicos iniciales, el modelo predijo con 0,90 AU-ROC (área bajo la curva característica operativa del receptor) con un IC del 95%: 0,892–0,905. Utilizando las predicciones del conjunto de pruebas, los resultados que presentaron son: 87,30% de sensibilidad y 71,98% de especificidad. El modelo de ensamble desarrollado en el presente estudio, evidencia mejor resultado con un *Accuracy* de 97% y una Kappa de Cohen de 0.95, para los modelos basados en *RandomForest* y *GradientBoosting*.

Conclusiones

En este estudio, se han desarrollado y comparado modelos basados en ensamble de aprendizaje automático para clasificar a

pacientes con sospecha de Covid-19 según su rango clínico. El objetivo de este estudio fue diseñar y comparar el rendimiento de modelos de aprendizaje automático basados en ensambles de máquinas para predecir pacientes con sospecha de Covid-19, para reducir las consecuencias clínicas y el coste sanitario. Una clasificación oportuna de pacientes con Covid-19 permitirá mejorar la gestión de estrategias de vigilancia, prevención y control hospitalario.

Se han entrenado modelos utilizando los algoritmos *RandomForest*, *GradientBoosting*, *Extratrees* y *AdaBoost*. De estos modelos, *RandomForest* y *GradientBoosting* reportaron una precisión (*accuracy*) del 97% y un coeficiente kappa de Cohen de 0.95, lo que muestra una concordancia muy buena en comparación con los modelos basados en *Extratrees* y *AdaBoost*. El modelo que presenta la precisión más baja, con un 92% de *accuracy* y un coeficiente kappa de Cohen de 0.82, corresponde al algoritmo *AdaBoost*.

Una ventaja de los modelos basado en ensamble de aprendizaje automático es su consistencia, puesto que evita en la medida de lo posible los óptimos locales y el problema de sobreajuste. Sin embargo, requiere una mayor cantidad de datos disponibles para lograr la validez externa. Se recomienda llevar a cabo más investigaciones con grandes volúmenes de datos y comparar los resultados con algoritmos de aprendizaje profundo.

Referencias bibliográficas

- Abdul, N. H., Wan, W. F., Wah, Y. B., Azlin, S., Shaadam, N., y Indratno, S. W. (2023). Comparison of ensemble hybrid sampling with bagging and boosting machine learning approach for imbalanced data. *Indonesian Journal of Electrical Engineering and Computer Science*, 29(1), 598-608. <http://doi.org/10.11591/ijeecs.v29.i1.pp598-608>
- Ahamad, M. M., Aktar, S., Rashed-Al-Mahfuz, M., Uddin, S., Liò, P., Xu, H., Summers, M. A., Quinn, J. M. W., y Moni, M. A. (2020). A machine learning model to identify early stage symptoms of SARS-Cov-2 infected patients. *Expert Systems with Applications*, 160, 113661. <https://doi.org/10.1016/j.eswa.2020.113661>
- Chen, Y., Liu, Q., y Guo, D. (2020). Emerging coronaviruses: genome structure, replication, and pathogenesis. *Journal of Medical Virology*, 92(4), 418-423. <https://doi.org/10.1002/jmv.25681>
- Comito, C., y Pizzuti, C. (2022). Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review. *Artificial Intelligence in Medicine*, 128, 102286. <https://doi.org/10.1016/j.artmed.2022.102286>
- Corso, C. L., Maldonado, C., Martínez, G., Casatti, M., y Mallo, A. (2018). Modelo de ensamble homogéneo basado en un proceso de reducción de datos simultaneo dirigido a la resolución de problemas de clasificación supervisada. *Revista de la Facultad de Ciencias Exactas, Físicas y Naturales*, 5(S-2), 51-55. <https://revistas.unc.edu.ar/index.php/FCEfyN/article/view/22897>
- De León, J., Pareja, A., Aguilar, P., Enriquez, Y., Quiroz, C., y Valencia, E. (2020). SARS-CoV-2 y sistema inmune: Una batalla de titanes. *Horizonte Médico*, 20(2), e1209. <https://doi.org/10.24265/horizmed.2020.v20n2.12>
- Dilmi, S. (2022). Automatic COVID-19 diagnosis using deep learning features and support vector machines based on Bayesian optimization. *ICAASE 2022 - 5th Edition of the International Conference on Advanced Aspects of Software Engineering, Proceedings*, Constantine, Algeria, 1-8. <https://doi.org/10.1109/ICAASE56196.2022.9931584>
- Dunlop, C., Howe, A., Li, D., y Allen, L. N. (2020). The coronavirus outbreak:

- The central role of primary care in emergency preparedness and response. *BJGP Open*, 4(1). <https://doi.org/10.3399/bjgpopen20X101041>
- Elguea, P. A., Prado, O. A., y Barradas, J. (2019). Implementación de una escala de gravedad para la activación del equipo de respuesta rápida: NEWS 2. *Medicina Critica*, 33(2), 98-103. <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=87296>
- Fuentes, M. D., y Medina, W. D. (2021). *Diseño de un modelo predictivo-asistencial de pacientes infectados por Covid-19, mediante un modelo supervisado de Machine Learning basado en criterios de derivación hospitalaria o ambulatoria* [Tesis de pregrado, Universidad de Guayaquil]. <http://repositorio.ug.edu.ec/handle/redug/52650>
- García, L. F. (2020). Immune response, inflammation, and the clinical spectrum of COVID-19. *Frontiers in Immunology*, 11, 1441. <https://doi.org/10.3389/fimmu.2020.01441>
- Gerón, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Inc.
- Gómez-Gómez, A., Gómez-Escanamé M. J., Sánchez-Ramos L., y Noyola, D. E. (2020). Diagnóstico y tratamiento temprano de neumonía ante la pandemia por COVID-19 en San Luis Potosí. ¿Es posible implementar una estrategia para lograrlo? *NCT Neumología y Cirugía de Tórax*, 79(4), 214-220. <https://doi.org/10.35366/97962>
- Guan, W.-J., Ni, Z.-Y., Hu, Y., Lian, W.-H., Ou, C.-Q., He, J.-X., Liu, L., Shan, H., Lei, C.-L., Hui, D. S. C., Du, B., Li, L.-J., Zeng, Guang., Yuen, K.-Yung., Chen, R.-C., Tang, C.-L., Wang, T., Chen, P.Y., Xiang, J.,...Zhong, N.-S. (2020). Clinical characteristics of coronavirus disease 2019 in China. *New England Journal of Medicine*, 382(18), 1708-1720. <https://doi.org/10.1056/NEJMoa2002032>
- Jiang, F., Deng, L., Zhang, L., Cai, Y., Cheung, C. W., y Xia, Z. (2019). Review of the clinical characteristics of coronavirus disease 2019 (COVID-19). *Journal of General Internal Medicine*, 35, 1545-1549. <https://doi.org/10.1007/s11606-020-05762-w>
- Kesav, N., y Jubukumar, M. G. (2022). A deep learning approach with Bayesian optimized Kernel support vector machine for Covid-19 diagnosis. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 11(3), 623-637. <https://doi.org/10.1080/21681163.2022.2099299>
- Khoshgoftaar, T. M., Van, H., y Napolitano, A. (2011). Comparing boosting and bagging techniques with noisy and imbalanced data. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 3, 552-568. <https://doi.org/10.1109/TSMCA.2010.2084081>
- Lora, M., Barrantes, S. A. M., Peñalver, M., y Rodríguez, J. L. (2023). Apoyo familiar y calidad de vida percibida por pacientes con diabetes mellitus-2 en tiempos Covid-19. *Revista de Ciencias Sociales (Ve)*, XXIX(1), 357-368. <https://doi.org/10.31876/rcs.v29i1.39756>
- Luna, B. D., y Carranza, L. E. (2023). Reconocimiento de la presencia de sars-cov-2 en pulmones a través de imágenes de radiodiagnóstico haciendo uso de Machine Learning con Python. *LATAM Revista Latinoamericana de Ciencias Sociales*

- y *Humanidades*, 4(1), 788-806. <https://doi.org/10.56712/latam.v4i1.296>
- Luna-Nemecio, J. (2020). Determinaciones socioambientales del COVID-19 y vulnerabilidad económica, espacial y sanitario-institucional. *Revista de Ciencias Sociales (Ve)*, XXVI(2), 21-26. <https://doi.org/10.31876/rcs.v26i2.32419>
- Manosalvas, M. G., y Zamora, S. C. (2021). Validación de la escala News 2 para predecir deterioro clínico en pacientes adultos mayores hospitalizados con neumonía por SARS CoV-2/ COVID-19, Hospital de Especialidades Eugenio Espejo, marzo 2020 a marzo 2021 [Tesis de especialista, Pontificia Universidad Católica del Ecuador]. <http://repositorio.puce.edu.ec:80/handle/22000/20775>
- Marruecos, M. E. (2020). Incidencia del Covid-19 en el derecho humano fundamental a la libertad ambulatoria. *Revista de Ciencias Sociales (Ve)*, XXVI(2), 13-15. <https://doi.org/10.31876/rcs.v26i2.32418>
- Martínez, J. R., y Del Pino, R. (2020). *Manual práctico de enfermería comunitaria*. Elsevier.
- Mojica-Crespo, R., y Morales-Crespo, M. M. (2020). Pandemia COVID-19, la nueva emergencia sanitaria de preocupación internacional: Una revisión. *Medicina de Familia. SEMERGEN*, 46(S-1), 65-77. <https://doi.org/10.1016/j.semerg.2020.05.010>
- Moulaei, K., Shanbehzadeh, M., Mohammadi-Taghiabad, Z., y Kazemi-Arpanahi, H. (2022). Comparing machine learning algorithms for predicting COVID-19 mortality. *BMC Medical Informatics and Decision Making*, 22(1), 2. <https://doi.org/10.1186/s12911-021-01742-0>
- Pajankar, A., y Joshi, A. (2022). *Hands-on machine learning with Python: Implement neural network solutions with Scikit-learn and PyTorch*. Apress.
- Patiño, D., Munive, C., Cevallos-Torres, L., y Botto-Tobar, M. (2021). Predicting the effectiveness of rapid tests performed to patients with COVID-19 through Linear Regression and Random Forest. *Ecuadorian Science Journal*, 5(2), 31-43. <https://doi.org/10.46480/esj.5.2.108>
- Preciado, A. J., Valles, M. A., y Lévano, D. (2021). Importancia del uso de sistemas de información en la automatización de historiales clínicos, una revisión sistemática. *Revista Cubana de Informática Médica*, 13(1), e417. <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=105413>
- Pu, L., y Shamir, R. (2023). 4CAC: 4-class classification of metagenome assemblies using machine learning and assembly graphs. *bioRxiv*, 2023-01. <https://doi.org/10.1101/2023.01.20.524935>
- Romero-Hernández, S., Zamarrón, E. I., Saavedra, J., Pérez, O. R., Guerrero, M. A., Soriano, R., Sánchez-Díaz, S., y Deloya, E. (2020). *Protocolo de atención para COVID-19 (SARS-CoV-2) de la Sociedad Mexicana de Medicina de Emergencias*. Sociedad Mexicana de Medicina de Emergencias.
- Rückert, P., Papenberg, B., y Tracht, K. (2021). Classification of assembly operations using machine learning algorithms based on visual sensor data. *Procedia CIRP*, 97, 110-116. <https://doi.org/10.1016/j.procir.2020.05.211>
- Sánchez-Duque, J. A., Arce-Villalobos, L. R., y Rodríguez-Morales, A. J. (2020). Enfermedad por coronavirus 2019 (COVID-19) en América Latina: papel de la atención primaria en la preparación y respuesta. *Atención*

- Primaria, 52(6), 369-372. <https://doi.org/10.1016/j.aprim.2020.04.001>
- Sreepadmanabh, M., Sahu, A. K., y Chande, A. (2020). COVID-19: Advances in diagnostic tools, treatment strategies, and vaccine development. *Journal of Biosciences*, 45(1), 148. <https://doi.org/10.1007/s12038-020-00114-6>
- Tillaguango, J. R. (2021). Revisión sistemática de literatura: Análisis de viabilidad para la detección y diagnóstico de Covid-19, aplicando modelos de Inteligencia Artificial (IA). *CEDAMAZ*, 11(2), 142-151. <https://doi.org/10.54753/cedamaz.v11i2.1183>
- Tran, V.-L., y Kim, J.-K. (2023). Ensemble machine learning-based models for estimating the transfer length of strands in PSC beams. *Expert Systems with Applications*, 221, 119768. <https://doi.org/10.1016/j.eswa.2023.119768>
- Vaishya, R., Javaid, M., Haleem, I., y Haleem, A. (2020). Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Review*, 14(4), 337-339. <https://doi.org/10.1016/j.dsx.2020.04.012>
- Xiao, T., Zhu, J., y Liu, T. (2013). Bagging and Boosting statistical machine translation systems. *Artificial Intelligence*, 195, 496-527. <https://doi.org/10.1016/j.artint.2012.11.005>
- Xiong, Y., Ma, Y., Ruan, L., Li, D., Lu, C., y Huang, L. (2022). Comparing different machine learning techniques for predicting COVID-19 severity. *Infectious Diseases of Poverty*, 11, 19. <https://doi.org/10.1186/s40249-022-00946-4>
- Xue, Y. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, 1168(2), 022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>
- Yaman, E., y Subasi, A. (2019). Comparison of bagging and boosting ensemble machine learning methods for automated EMG signal classification. *BioMed Research International*, 2019, 9152506. <https://doi.org/10.1155/2019/9152506>
- Yaman, M. A., Rattay, F., y Subasi, A. (2021). Comparison of bagging and boosting ensemble machine learning methods for face recognition. *Procedia Computer Science*, 194, 202-209. <https://doi.org/10.1016/j.procs.2021.10.074>
- Zachary, C., y Papailiopoulos, D. (2018). Stability and generalization of learning algorithms that converge to global optima. *Proceedings of the 35th International Conference on Machine Learning*, PMLR, 80, 745-754. <https://proceedings.mlr.press/v80/charles18a.html>
- Zoabi, Y., Deri-Rozov, S., y Shomron, N. (2021). Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *Digital Medicine*, 4, 3. <https://doi.org/10.1038/s41746-020-00372-6>