REVISTA CIENTÍFICA
Ciencia y Método
ISSN 3103-1072

EDITORIAL GRUPO AEA

Artículo Científico

Modelo de Inteligencia artificial para el diagnóstico del dengue

Artificial intelligence model for dengue diagnosis

Arroyo-Ruales, Wilson Efren[1]
https://orcid.org/0009-0003-4853-7982
wilson.arroyo@espoch.edu.ec
Escuela Superior Politécnica de Chimborazo, Ecuador, Riobamba

Soria-Poma, Xavier [2]
https://orcid.org/0000-0003-2997-2439
xavier.soria@espoch.edu.ec
Escuela Superior Politécnica de Chimborazo, Ecuador, Riobamba.

Autor de correspondencia [1]

DOI / URL: https://doi.org/10.55813/gaea/rcym/v4/n1/126

Resumen: El diagnóstico diferencial del dengue frente a otras arbovirosis representa un desafío clínico significativo en regiones endémicas. Este estudio más allá de proponer un modelo de Inteligencia Artificial (AI) para el diagnóstico del Dengue, realiza un estudio y preparación del entorno de los conjuntos de datos de Dengue para entrenar moledos de AI, especilamente de machine learning. Este estudio evaluó la efectividad de técnicas de machine learning aprendizaje supervisado para predecir la infección por dengue utilizando datos clínicos y demográficos. Se evaluaron varios algoritmos de clasificación binaria tanto paramétricos como no paramétricos mediante un proceso de validación cruzada y métricas de desempeño ampliamente utilizadas como el accuracy o el F1-score. Se halló que la calidad del dato afecta el resultado del modelo ya que en el dataset balanceado y con datos mejor tratados, el modelo binario entrega mejores resultados que en el dataset desbalanceado o con ruido en sus registros. Se concluye que, revisado evidencias cuantitativas, se necesita realizar un estudio y experimentación más profunda de los dataset de Dengue para facilitar el proceso de entramiento de los modelos de machine learning.

Palabras clave: dengue; aprendizaje supervisado; regresión logística; boosting; Multi Layer Perceptron

Check for updates

Received: 25/Nov/2025
Accepted: 12/Dic/2025
Published: 05/Ene/2026

Cita: Arroyo-Ruales, W. E., & Soria-Poma, X. (2026). Modelo de Inteligencia artificial para el diagnóstico del dengue. *Revista Científica Ciencia Y Método*, *4*(1), 1-13. https://doi.org/10.55813/gaea/rcym/v4/n1/126

Revista Científica Ciencia y Método (RCyM)
https://revistacym.com
revistacym@editorialgrupo-aea.com
info@editoriagrupo-aea.com

Abstract:

The differential diagnosis of dengue from other arboviral diseases presents a significant clinical challenge in endemic regions. This study, beyond proposing an Artificial Intelligence (AI) model for dengue diagnosis, examines and prepares the environment of dengue datasets for training AI models, particularly machine learning models. This study evaluated the effectiveness of supervised machine learning techniques for predicting dengue infection using clinical and demographic data. Several parametric and non-parametric binary classification algorithms were evaluated through cross-validation and widely used performance metrics such as accuracy and F1-score. It was found that data quality affects model performance, as the binary model delivers better results in balanced datasets with better-processed data than in unbalanced datasets or those with noise in their records. Based on quantitative evidence, the study concludes that further study and experimentation with dengue datasets are needed to facilitate the training process for machine learning models.

Keywords: dengue; supervised learning; logistic regression; boosting; Multi-Layer Perceptor.

1. Introducción

El dengue es una enfermedad viral transmitida por mosquitos del género Aedes, principalmente Aedes aegypti, que constituye una de las mayores amenazas para la salud pública a nivel global (Bhatt et al., 2013; Brady & Hay, 2020). Según la Organización Mundial de la Salud (World Health Organization [WHO], 2025), hasta julio de 2025 se han reportado al menos 4 millones de casos de dengue, con alrededor de 3 000 muertes. En la región de las Américas, la Organización Panamericana de la Salud (Pan American Health Organization [PAHO], 2025) consolidó hasta noviembre de 2025 aproximadamente 4,2 millones de casos sospechosos, de los cuales fallecieron 2 099 personas. Estas cifras probablemente subestiman la verdadera magnitud del problema debido al subregistro y a las limitaciones diagnósticas.

Actualmente no se dispone de una cura específica para el dengue; el tratamiento se basa principalmente en una hidratación adecuada y en el uso de analgésicos para mitigar el dolor y la fiebre (WHO, 2025). La sintomatología incluye fiebre, mialgias, cefalea, náuseas y vómitos, entre otros signos y síntomas (Quinn et al., 2018; Peeling et al., 2010). Esta presentación clínica es muy similar a la de otras arbovirosis, como chikungunya y zika, lo que dificulta el diagnóstico diferencial y aumenta el riesgo de clasificaciones erróneas (Beltrán-Silva et al., 2018; Peeling et al., 2010).

El diagnóstico del dengue se apoya en pruebas directas orientadas a la detección del virus y pruebas indirectas centradas en la respuesta inmune del huésped. Entre las primeras se incluyen el aislamiento viral, las pruebas de reacción en cadena de la

polimerasa (PCR) y la detección del antígeno NS1; entre las segundas, la detección de anticuerpos IgM e IgG específicos (Peeling et al., 2010; WHO, 2016). Sin embargo, la disponibilidad de estos métodos puede ser limitada en contextos de escasos recursos, lo que abre espacio para enfoques complementarios basados en datos clínicos y epidemiológicos.

En este contexto, el aprendizaje automático se ha propuesto como una herramienta para mejorar la detección y clasificación de pacientes con sospecha de dengue (Bohm et al., 2024). El diagnóstico se plantea como un problema de clasificación binaria, donde los algoritmos aprenden a distinguir entre casos positivos y negativos a partir de un conjunto de variables clínicas, demográficas y, en algunos casos, epidemiológicas. Entre las técnicas más empleadas se encuentran la regresión logística (Lukman et al., 2024), los árboles de decisión, Random Forest (Hu et al., 2020; Boateng et al., 2020), los métodos de boosting como AdaBoost y XGBoost (Hatwell et al., 2020; Hu et al., 2020), KNN (Boateng et al., 2020) y las redes neuronales (Boateng et al., 2020; Weng & Szolovits, 2020).

El presente estudio emplea los datos de Bohm et al. (2024), que comprenden 20 000 registros, y de Neto et al. (2023), con 17 172 observaciones. Ambos conjuntos recogen información sobre síntomas clínicos, comorbilidades y características demográficas de los pacientes. Los síntomas incluyen fiebre, mialgia, cefalea, exantema, náuseas, artralgia, petequias, leucopenia y los días de evolución de la enfermedad. Las comorbilidades abarcan diabetes, hipertensión, enfermedades hepáticas, renales y hematológicas. Las variables sociodemográficas del paciente comprenden el sexo, la edad, el estado de gestación, la raza y la zona de residencia.

A fin de garantizar un entorno adecuado de análisis y compatibilidad con los algoritmos de aprendizaje supervisado, se implementó un proceso sistemático de preprocesamiento. En la primera etapa, las variables categóricas se codificaron a formato numérico binario (1 = presencia, 0 = ausencia) y se definió la variable objetivo como dengue = 1 y no dengue = 0. En esta fase inicial no se aplicó una curación exhaustiva de los datos, con el propósito de evaluar el impacto directo de la calidad original del registro sobre el desempeño de los modelos. Posteriormente, ambos conjuntos se dividieron aleatoriamente en una muestra de entrenamiento (90 %) y una muestra de prueba (10 %), siguiendo las recomendaciones de la literatura sobre muestreo en problemas de clasificación (Anderson, 2007; Lohr, 2021).

Con los datos preparados, se construyeron varios modelos de clasificación binaria, incluyendo métodos paramétricos y no paramétricos. El desempeño de cada modelo se evaluó mediante métricas ampliamente utilizadas en machine learning como accuracy, precisión, recall y F1-score (Grandini et al., 2020). Este enfoque permitió comparar la capacidad predictiva de los algoritmos y determinar cuáles presentan mayor robustez para la identificación de casos de dengue frente a casos no dengue.

2. Materiales y métodos

Consideraciones generales sobre datos de salud y aprendizaje automático:

Desde la perspectiva del aprendizaje automático aplicado a enfermedades virales, la literatura especializada coincide en que la calidad del modelo depende en gran medida de la calidad del dato, más que de la sofisticación del algoritmo seleccionado (Maletzky et al., 2022; Bohm et al., 2024; Neto et al., 2023). Aspectos como el ruido en las etiquetas, los valores faltantes, la inconsistencia entre fuentes y el desbalance de clases pueden degradar significativamente el rendimiento de los modelos, aun cuando se utilicen técnicas avanzadas.

En el caso específico del dengue, diversos autores recomiendan combinar el criterio clínico, la inspección manual de los registros y procedimientos automatizados de limpieza, con el fin de reducir errores sistemáticos en los sistemas de vigilancia (Shi et al., 2021; Syed et al., 2023). En los datos de salud digitales, dimensiones como accesibilidad, completitud, consistencia, validez contextual y actualidad influyen directamente en los resultados analíticos (Syed et al., 2023; Hosseinzadeh et al., 2025). Las intervenciones más efectivas son aquellas que integran retroalimentación operacional, capacitación del personal y soluciones informáticas, articularndo la limpieza técnica con mejoras en los flujos de captura para prevenir la recurrencia de errores (Ahmed et al., 2023; Lighterness et al., 2024).

3. Resultados

3.1. Repositorio SINAN y datasets de dengue

El Sistema de Informação de Agravos de Notificação (SINAN) de Brasil (Sistema de Informação de Agravos de Notificação, 2025) dispone de un repositorio de datos con información detallada sobre síntomas, características demográficas y comorbilidades de los pacientes. La base incluye variables como fiebre, cefalea, mialgia, náuseas, vómito, rash cutáneo, dolor retroocular, artralgia y leucopenia, así como comorbilidades y variables sociodemográficas adicionales. Este nivel de granularidad es esencial para entrenar algoritmos capaces de distinguir entre casos de dengue y no dengue.

En este contexto, los conjuntos de datos derivados del SINAN y empleados previamente por Bohm et al. (2024) y Neto et al. (2023), denominados Dengue20K y Dengue17K, respectivamente, se consideraron la opción más apropiada para entrenar y evaluar modelos de IA orientados al diagnóstico del dengue. El conjunto Dengue20K contiene 20 000 registros individuales e incluye una variable dependiente binaria que clasifica cada caso como dengue o no dengue. La variable objetivo está completamente balanceada, con 10 000 casos de dengue y 10 000 de no dengue, e incorpora variables clínicas y sociodemográficas codificadas principalmente en formato binario.

El segundo conjunto, Dengue17K, está compuesto por 17 172 registros. En este caso, la variable dependiente original presenta tres categorías, que distinguen entre dengue, chikungunya y otras arbovirosis. Debido a la similitud clínica entre estas infecciones, existe el riesgo de que pacientes con chikungunya o zika hayan sido registrados como dengue, introduciendo ruido en las etiquetas (Neto et al., 2023; Bhatt et al., 2013; PAHO, 2025). Para el presente estudio, esta variable se recodificó como binaria: dengue = 1 y no dengue = 0, agrupando las otras arbovirosis en la segunda categoría.

3.2. Modelos de Machine Learning para el Diagnóstico de Dengue

Para el desarrollo de los modelos predictivos se utilizaron algoritmos de *machine learning* supervisado, dado que este enfoque permite modelar la complejidad de los patrones clínicos asociados al diagnóstico del dengue. La literatura en analítica clínica recomienda explorar varios modelos de clasificación y comparar su desempeño, a fin de seleccionar aquellos con mayor robustez y capacidad de generalización (Weng & Szolovits, 2020; Hu et al., 2020).

La regresión logística permite calcular la probabilidad de pertenecer a una clase mediante la función logística o sigmoide, que transforma una combinación lineal de las variables independientes en valores entre 0 y 1 (Lukman et al., 2024). KNN asigna la clase de un nuevo registro en función de la clase mayoritaria de sus vecinos más cercanos, según una medida de distancia (Boateng et al., 2020).

Los árboles de decisión (CART) generan reglas jerárquicas explícitas a partir de divisiones recursivas de las variables predictoras, lo que facilita la interpretación de la relación entre las variables independientes y la variable objetivo (Hu et al., 2020). Random Forest agrupa múltiples árboles construidos sobre subconjuntos aleatorios de datos y variables, reduciendo la varianza y el riesgo de sobreajuste en comparación con un único árbol (Hu et al., 2020; Boateng et al., 2020).

Los métodos de *boosting* buscan mejorar el rendimiento combinando múltiples clasificadores débiles. Gradient Boosting Machine (GBM) construye un modelo como suma secuencial de árboles de decisión, donde cada nuevo árbol se entrena para corregir los errores del ensamble anterior (Florek & Zagdański, 2023). AdaBoost ajusta iterativamente clasificadores débiles aumentando el peso de las observaciones mal clasificadas, permitiendo capturar relaciones no lineales en conjuntos altamente categóricos (Hatwell et al., 2020). XGBoost optimiza el *gradient boosting* mediante regularización explícita y técnicas de ingeniería computacional, produciendo modelos más robustos y eficientes (Hu et al., 2020).

Finalmente, la red neuronal artificial (RNA) de arquitectura multicapa (MLP) permite modelar interacciones no lineales y patrones de alta dimensionalidad. Este tipo de modelo puede capturar combinaciones complejas de variables que escapan a los enfoques basados únicamente en reglas o distancias (Weng & Szolovits, 2020; Hu et al., 2020).

3.3. Estudio de Datasets para el diagnóstico de dengue

El entorno de trabajo seleccionado fueron los notebooks de Google Colab (Google, 2024), que permiten desarrollar modelos de analítica avanzada utilizando el lenguaje Python (Python Software Foundation, 2024) y librerías especializadas como Scikit-learn (Pedregosa et al., 2011).

En los trabajos originales de Bohm et al. (2024) y Neto et al. (2023) no se documenta con detalle todo el proceso de preprocesamiento aplicado a los datos. En este estudio, una vez implementado el entorno de trabajo, se realizaron recodificaciones de variables categóricas, revisión de valores nulos o vacíos y análisis básico de registros atípicos. Las variables dicotómicas se transformaron a formato binario, y la variable CLASSI_FIN de Dengue17K se recodificó a dengue/no dengue, como se indicó anteriormente.

En Dengue17K no se detectaron valores nulos, pero se observó desbalance de clases, ya que la proporción de casos de dengue es sensiblemente menor que la clase no dengue. En Dengue20K, en cambio, la variable de clasificación ya se encontraba definida en dos categorías y balanceada, sin valores perdidos en los registros incluidos en el análisis.

3.4. Muestreo y tamaño de muestra

La selección de la muestra es un paso crítico para una modelización adecuada. La división 90 %/10 % entre entrenamiento y prueba maximiza la información disponible para el ajuste del modelo, lo cual es especialmente relevante en contextos clínicos donde cada registro aporta información valiosa, y mantiene al mismo tiempo un subconjunto independiente para la evaluación (Esteva et al., 2017).

Para verificar que el tamaño de muestra de entrenamiento sea igual o superior al mínimo teórico, se aplicó la fórmula de tamaño de muestra óptimo para muestreo aleatorio simple (Lohr, 2021). En Dengue20K, el tamaño óptimo estimado fue de 376 casos y la muestra de entrenamiento cuenta con 18 000 registros. En Dengue17K, el tamaño óptimo resultó de 368 casos y la muestra de entrenamiento tiene 15 455 registros. En ambos casos, la muestra real de entrenamiento supera ampliamente el tamaño mínimo requerido.

3.5. Métricas de evaluación

Una vez desarrollado cada modelo de clasificación binaria, se evaluó su capacidad para discriminar entre las clases dengue y no dengue. Las métricas seleccionadas miden tanto la coincidencia global entre predicciones y valores reales como la capacidad del modelo para reconocer correctamente la clase positiva.

Se utilizaron las siguientes métricas derivadas de la matriz de confusión (TP, FP, FN, TN): accuracy, precisión, recall (sensibilidad) y F1-score (Grandini et al., 2020). La combinación de estas métricas permite valorar no solo el porcentaje de aciertos globales, sino también el equilibrio entre la capacidad de detectar verdaderos positivos

y evitar falsos positivos en el contexto clínico de dengue. A partir de la matriz de confusión se tiene las siguientes métricas:

Accuracy: mide el porcentaje de predicciones correctas, es decir compara entre el total de registros todos los que se asignaron de manera correcta. Mientras más cercano a 100% es mejor.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: mide la proporción de predicciones positivas que el modelo acierta de forma correcta. Mientras más cercano a 100% es mejor.

$$Precision = \frac{TP}{TP + FP}$$

Recall o Sensitivity mide la capacidad del modelo para identificar correctamente los casos positivos. Mientras más cercano a 100% es mejor.

$$Recall = \frac{TP}{TP + FN}$$

F1-score combina precision y recall mediante su media armonica para equilibrar ambos componentes. Mientras más cercano a 100% es mejor.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

3.6. Resultados cuantitativos

Tanto para el conjunto Dengue17K como para Dengue20K se obtienen las métricas de desempeño de los diferentes modelos reimplementados.

Para Dengue17K las métricas indican que el mayor accuracy corresponde al XGBoost, y en general todos los modelos presentan este accuracy relativamente alto a excepción del KNN que tiene apenas 54.78%.

Tabla 1

Métricas para modelos con Dengue17K (%)

Modelo	Accuracy		Precision		Recall		F1-score	
	Acc. Art	Acc. Reimp	Prec. Art	Prec. Reimp	Recall Art	Recall Reimp	F1 Art.	F1 Reimp.
KNN	54.11	54.78	55.19	37.80	54.10	55.24	52.22	44.89
XGBoost	61.53	70.75	61.16	65.91	61.73	25.35	60.93	36.62
Adaboost	58.79	66.49	58.37	49.40	59.03	21.68	57.82	30.13
GBM	62.40	68.94	62.05	58.99	62.57	22.38	61.96	32.45
Rnd Forest	60.11	68.88	59.65	61.31	60.33	18.01	59.49	27.84
MLP	-	66.67	-	-	-	-	-	-

Nota: (Autores, 2026).

Las métricas del Multi Layer Perceptron (MLP) indican que este modelo tiene el peor desempeño, por lo que se indaga en la matriz de confusión, debido a que no hay predicciones para la categoría Dengue, es decir el modelo no clasifica para Dengue,

y con el antecedente de las métricas en cero y accuracy un poco alto, entonces el modelo clasifica todo hacia la categoría dominante, de esa manera maximiza el accuracy. Las incidencias en los datos de ingreso al modelo son la principal causa para que los modelos binarios presenten un pobre desempeño (Kordos et al 2016).

Por otro lado, también se tiene las métricas del conjunto Dengue20K, los modelos asociados a este conjunto presentan métricas de alto desempeño, y considerando que las métricas se probaron en una muestra que no se utilizó en la fase de entrenamiento, entonces se trata de modelos robustos. El modelo MLP es el de mejor desempeño ya que tiene las más altas métricas precision y recall, es decir es el que mejor discrimina a los casos de Dengue/No Dengue. El siguiente modelo en mejor desempeño es el Decision Tree ya que discrimina muy bien los Dengue, sin embargo, su desempeño para No Dengue es un poco menor comparado con la red neuronal. Bohm et al en su artículo no indican los valores de las métricas precision y recall, por lo que no se las incluye en la tabla de desempeño, sin embargo, estas métricas si se las calculó en la reimplementación.

Tabla 2

Métricas para modelos con Dengue17K (%)

Modelo	Accuracy		Precision		Recall		F1-score	
	Acc. Art	Acc. Reimp	Prec. Art	Prec. Reimp	Recall Art	Recall Reimp	F1 Art.	F1 Reimp.
Reg. Logístic	93.12	92.60	-	94.19	-	90.80	93.23	92.46
MLP	93.13	92.95	-	92.82	-	93.10	93.33	92.96
Decision Tree	92.52	92.25	-	96.38	-	87.80	92.83	91.89
KNN	92.23	88.75	-	87.66	-	90.20	92.51	88.91

Nota: (Autores, 2026).

4. Discusión

Los resultados obtenidos permiten analizar el efecto del proceso de depuración y curación sobre la estabilidad y la capacidad predictiva de los modelos entrenados. En el conjunto Dengue17K se observa un mayor nivel de ruido estructural, evidenciado por la presencia de registros duplicados entre las categorías analizadas, posiblemente relacionado con la similitud clínica entre dengue, chikungunya, zika y otras arbovirosis. Esta similitud sintomática favorece errores de clasificación ya desde la etapa de registro y codificación de los casos (Neto et al., 2023; Bhatt et al., 2013; PAHO, 2025).

Esta calidad subóptima de los datos se refleja directamente en las métricas de desempeño. Aunque el modelo XGBoost alcanza un accuracy de 70.75 %, su precisión desciende a 65.91 % y el recall se sitúa en 25.35 %. En general, los modelos presentan una baja capacidad para identificar correctamente los casos positivos de dengue, dado que el recall, que mide la proporción de casos de dengue correctamente identificados, se mantiene bajo; en el mejor escenario, el modelo KNN solo alcanza un recall de 55.24 %. En el caso de la red neuronal MLP, la clasificación de la clase dengue es prácticamente nula, ya que el modelo tiende a maximizar el accuracy

clasificando la mayoría de las observaciones en la clase más numerosa, un comportamiento ya descrito en la literatura cuando existen ruido de etiquetado y desbalance entre clases en problemas de arbovirosis (Bohm et al.; Neto et al., 2023).

El hecho de que todos los modelos entrenados con Dengue17K presenten valores de desempeño consistentemente bajos sugiere que el problema no se resuelve únicamente mediante el ajuste de parámetros o hiperparámetros, sino que está asociado a la calidad intrínseca del dataset. Como se ha señalado, las etiquetas de la variable dependiente pueden estar mal asignadas debido a la ambigüedad diagnóstica y a la ausencia de una depuración sistemática de los registros. Además, el conjunto Dengue17K no fue originalmente diseñado para el entrenamiento de redes neuronales, sino para modelos de boosting, lo que refuerza la necesidad de procesos de curación y reestructuración de los datos antes de aplicar enfoques de inteligencia artificial más complejos.

Por otro lado, el conjunto Dengue20K de entrada está totalmente balanceado con 50% de casos para Dengue, y 50% para No Dengue, con menor ruido. Las métricas globales de todos los modelos entrenados se mantuvieron estables y bastante cercanas entre si. El hecho que los modelos se hayan testeado en una muestra independiente distinta a la utilizada en el entrenamiento del modelo garantiza la robustez de los modelos.

En ese sentido el modelo con mejor performance es el de red neuronal, ya que logra tener el mejor accuracy, asi como los mejores valores de precision y recall simultanemanente. Este dataset tambien consideraba entrenar una red neuronal, recordando que estos modelos son sensibles a los datos atípicos y es requisito que los datos sean estandarizados y uniformes, razón por la cual el dataset tiene una mejor normalización. Sin embargo, se necesita realizar una experimentación más profunda analizando con procedimientos de limpieza de datos.

Nuestras primeras observaciones sugieren que, en los datasets de dengue considerados en este documento, la falta de limpieza puede incrementar el nivel de ruido e inconsistencia del conjunto de datos, aumentar la sensibilidad de los algoritmos al ruido presente durante el entrenamiento y afectar la estabilidad de la distribución entre clases.

5. Conclusiones

En los dos conjuntos Dengue17K y Dengue20K se comparó el desempeño de varios modelos de clasificación binaria, incluyendo una red neuronal multicapa. Los resultados muestran que, en términos globales, el comportamiento de los modelos es marcadamente distinto entre ambos datasets. Mientras que en Dengue20K los clasificadores alcanzan métricas de desempeño elevadas y relativamente estables, en Dengue17K el rendimiento es considerablemente más discreto, especialmente en la identificación de los casos positivos de dengue.

En particular, en el conjunto Dengue17K la red neuronal se comporta como el peor clasificador, lo que puede estar asociado al desbalance de clases y a la presencia de ruido en las etiquetas. Bajo estas condiciones, la red tiende a optimizar el accuracy global favoreciendo a la clase mayoritaria, en detrimento de la sensibilidad para la clase minoritaria (dengue), fenómeno ampliamente descrito en la literatura de *machine learning* aplicado a problemas clínicos con desbalance de clases. En cambio, en Dengue20K los modelos presentan métricas con valores altos y similares (accuracy, precisión, *recall* y F1-score), lo que evidencia una mayor robustez y capacidad de generalización, favorecida por el hecho de que tanto la muestra de entrenamiento como la de prueba se encuentran bien balanceadas.

En este escenario, la red neuronal entrenada con Dengue20K emerge como el modelo "ganador", ya que alcanza los valores más altos de accuracy y F1-score, lo que indica un mejor equilibrio entre la correcta clasificación de casos positivos y negativos. La brecha observada entre los resultados de ambos datasets sugiere que la calidad y estructura de los datos condicionan de manera decisiva el desempeño de los modelos, más allá de la técnica utilizada. Por ello, se recomienda realizar un estudio y una experimentación más profunda sobre la construcción, depuración y balanceo de estos conjuntos de datos, con el fin de facilitar el proceso de entrenamiento de los modelos de *machine learning* para el diagnóstico de dengue y mejorar, en particular, las métricas obtenidas con Dengue17K, en concordancia con lo señalado en trabajos previos sobre arbovirosis y calidad del dato en salud.

CONFLICTO DE INTERESES

"Los autores declaran no tener ningún conflicto de intereses".

Referencias Bibliográficas

Ahmed, A., Xi, R., Hou, M., Shah, S. A., & Hameed, S. (2023). Harnessing big data analytics for healthcare: A comprehensive review of frameworks, implications, applications, and impacts. *IEEE Access, 11*, 112891–112928. https://doi.org/10.1109/ACCESS.2023.3323574

Anderson, R. (2007). *The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation*. Oxford University Press. https://global.oup.com/academic/product/the-credit-scoring-toolkit-9780199226405

Beltrán-Silva, S. L., Chacón-Hernández, S. S., Moreno-Palacios, E., & Pereyra-Molina, J. Á. (2018). Clinical and differential diagnosis: Dengue, chikungunya and Zika. *Revista Médica del Hospital General de México, 81*(4), 218–227. https://doi.org/10.1016/j.hgmx.2016.10.003

Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., et al. (2013). The

global distribution and burden of dengue. *Nature, 496*(7446), 504–507. https://doi.org/10.1038/nature12060

Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020). Basic tenets of classification algorithms: k-nearest-neighbor, support vector machine, random forest and neural network – A review. *Journal of Data Analysis and Information Processing, 8*(4), 341–357. https://doi.org/10.4236/jdaip.2020.84020

Bohm, B. C., Borges, F. E., Silva, S. C., Soares, A. T., Ferreira, D. D., Belo, V. S., Lignon, J. S., & Bruhn, F. R. P. (2024). Utilization of machine learning for dengue case screening. *BMC Public Health, 24*, 1573. https://doi.org/10.1186/s12889-024-19083-8

Brady, O. J., & Hay, S. I. (2020). The global expansion of dengue: How *Aedes aegypti* mosquitoes enabled the first pandemic arbovirus. *The Lancet Infectious Diseases, 20*(1), e42–e51. https://doi.org/10.1016/S1473-3099(19)30446-X

Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115–118. https://doi.org/10.1038/nature21056

Florek, P., & Zagdański, A. (2023). Benchmarking state-of-the-art gradient boosting algorithms for classification. *arXiv*. https://doi.org/10.48550/arXiv.2305.17094

Google. (2024). *Google Colaboratory* [Computer software]. https://colab.research.google.com

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: An overview. *arXiv*. https://doi.org/10.48550/arXiv.2008.05756

Hatwell, J., Gaber, M. M., & Azad, R. M. A. (2020). Ada-WHIPS: Explaining AdaBoost classification with applications in the health sciences. *BMC Medical Informatics and Decision Making, 20*, 250. https://doi.org/10.1186/s12911-020-01201-2

Hosseinzadeh, E., Afkanpour, M., Momeni, M., et al. (2025). Data quality assessment in healthcare: Dimensions, methods and tools – A systematic review. *BMC Medical Informatics and Decision Making, 25*, 296. https://doi.org/10.1186/s12911-025-03136-y

Hu, L., Chen, J., Vaughan, J., & Yang, H. (2020). Supervised machine learning techniques: An overview with applications to banking. *arXiv*. https://doi.org/10.48550/arXiv.2008.04059

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., et al. (2016). Jupyter Notebooks – A publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), *Positioning and power in academic publishing: Players, agents and agendas* (pp. 87–90). IOS Press. https://doi.org/10.3233/978-1-61499-649-1-87

Kordos, M., & Rusiecki, A. (2016). Reducing noise impact on MLP training. *Soft Computing, 20*(1), 49–65. https://doi.org/10.1007/s00500-015-1690-9

Lighterness, A., Adcock, M., Scanlon, L. A., & Price, G. (2024). Data quality–driven improvement in health care: Systematic literature review. *Journal of Medical Internet Research, 26*, e57615. https://doi.org/10.2196/57615

Lohr, S. L. (2021). *Sampling: Design and analysis* (3rd ed.). CRC Press. https://doi.org/10.1201/9780429298899

Lukman, A. F., Mohammed, S., Olaluwoye, O., & Farghali, R. A. (2024). Handling multicollinearity and outliers in logistic regression using the robust Kibria–Lukman estimator. *Axioms, 13*(1), 19. https://doi.org/10.3390/axioms13010019

Maletzky, A., Böck, C., Tschoellitsch, T., Roland, T., Ludwig, H., Thumfart, S., Giretzlehner, M., Hochreiter, S., & Meier, J. (2022). Lifting hospital electronic health record data treasures: Challenges and opportunities. *JMIR Medical Informatics, 10*(10), e38557. https://doi.org/10.2196/38557

Neto, S. R. S., Oliveira, T. T., & Neto, L. M. (2023). Binary models for arboviruses classification using machine learning: A benchmarking evaluation. In *Proceedings of the 56th Hawaii International Conference on System Sciences* (pp. 2834–2843). https://doi.org/10.24251/HICSS.2023.348

Pan American Health Organization. (2025). *Dengue: Datos y análisis*. Pan American Health Organization. https://www.paho.org/es/temas/dengue/datos

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. https://www.jmlr.org/papers/v12/pedregosa11a.html

Peeling, R. W., Artsob, H., Pelegrino, J., et al. (2010). Evaluation of diagnostic tests: Dengue. *Nature Reviews Microbiology, 8*(12 Suppl), S30–S38. https://doi.org/10.1038/nrmicro2459

Python Software Foundation. (2024). *Python* (Version 3.x) [Computer software]. https://www.python.org

Quinn, E., Cheong, A., Calvert, J., Higgins, G., Hahesy, T., & Carr, J. (2018). Clinical features and laboratory findings of travelers returning to South Australia with dengue virus infection. *Tropical Medicine and Infectious Disease, 3*(1), 6. https://doi.org/10.3390/tropicalmed3010006

Shi, X., Prins, C., Van Pottelbergh, G., Mamouris, P., Vaes, B., & De Moor, B. (2021). An automated data cleaning method for electronic health records by incorporating clinical knowledge. *BMC Medical Informatics and Decision Making, 21*, 267. https://doi.org/10.1186/s12911-021-01630-7

Sistema de Informação de Agravos de Notificação. (2025). *Dados epidemiológicos – SINAN*. Ministério da Saúde, Brasil. https://portalsinan.saude.gov.br/dados-epidemiologicos-sinan

Syed, R., Eden, R., Makasi, T., Chukwudi, I., Mamudu, A., et al. (2023). Digital health data quality issues: Systematic review. *Journal of Medical Internet Research, 25*, e42615. https://doi.org/10.2196/42615

Weng, W. H., & Szolovits, P. (2020). Machine learning for clinical predictive analytics. In S. R. Steinhubl, P. W. Zimlichman, & B. D. Topol (Eds.), *Data science for healthcare: Methodologies and applications* (pp. 199–217). Springer. https://doi.org/10.1007/978-3-030-47994-7_12

World Health Organization. (2016). *Laboratory testing for Zika virus infection: Interim guidance* (WHO/ZIKV/LAB/16.1). World Health Organization. https://apps.who.int/iris/handle/10665/204671

World Health Organization. (2025). *Dengue and severe dengue* [Fact sheet]. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue