Comparative Analysis of Statistical, Machine Learning, and Deep Learning Time Series Models for Forecasting Dengue, Zika, and Chikungunya Incidence in Cuban Provinces

Yanet Garcia Serrano, Jennifer de la Caridad Sanchez Santana

Université de La Havane

11 septembre 2026

Arboviral diseases, specifically dengue, Zika, and chikungunya, represent a critical public health challenge in Cuba due to their high morbidity and the direct influence of climatic variables on their spread. Despite entomological surveillance efforts, the non-linear nature of these phenomena demands advanced predictive methods. This research aims to determine the most accurate model for forecasting the incidence of these pathologies across Cuban provinces through a comparative study of time series algorithms. The methodology integrates epidemiological and meteorological data (processed using Thiessen Polygons), applying statistical models (SARIMAX), machine learning (SVR, Random Forest, XGBoost), and deep learning (LSTM). The effect of time lags is examined using Pearson correlation hypothesis tests and the cross-correlation function (CCF). The findings identify the optimal architecture for each territory, providing a scientific methodological basis to strengthen early warning systems and decision-making within the national health system.