LSTM-Powered COVID-19 prediction in central Thailand incorporating meteorological and particulate matter data with a multi-feature selection approach

*Chanidapa Winalai, Suparinthon Anupong, **Charin Modchang**, Sudarat Chadsuthi*

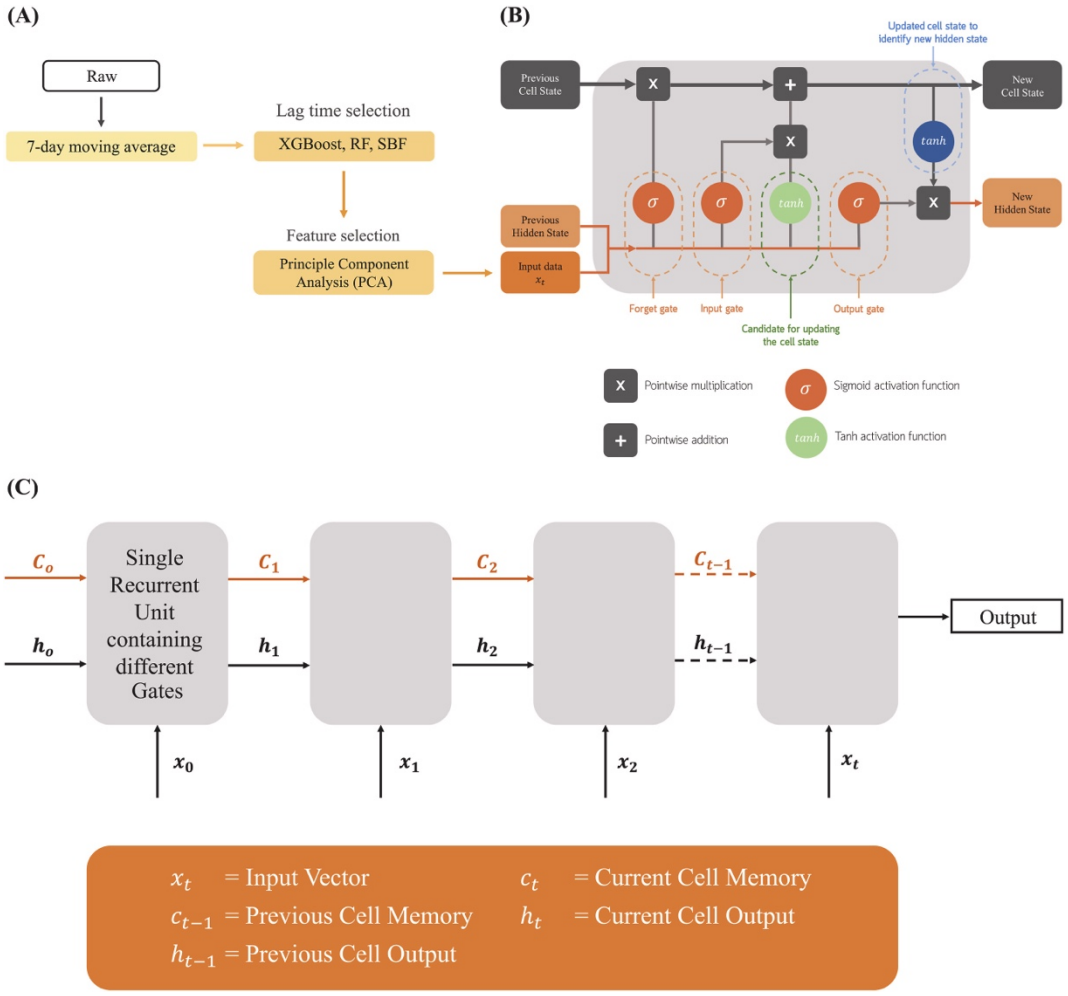
Rationale and objective: The COVID-19 pandemic has significantly impacted public health and necessitated urgent actions to mitigate its spread. Monitoring and predicting the outbreak's progression have become vital to devise effective strategies and allocate resources efficiently. This study presents a novel approach utilizing Multivariate Long Short-Term Memory (LSTM) to analyze and predict COVID-19 trends in Central Thailand, particularly emphasizing the multi-feature selection process. To ensure a comprehensive examination of predictive capabilities, we also conduct a comparative analysis to assess the efficacy of the LSTM model. This comparison involves evaluating its performance against both the Recurrent Neural Network (RNN) and the Generalized Linear Model (GLM). Through the integration of these methodologies and a prolonged data collection period, this study can provide significant insights into forecasting and comprehending the dynamics of COVID-19 within Central Thailand.

Summary: We propose a multi-feature selection technique to identify the most relevant and influential features that significantly impact the spread of COVID-19 in the region to enhance the model's performance. Our results highlight that relative humidity is the key factor driving COVID-19 transmission in Central Thailand. The proposed multi-feature selection technique significantly improves the model's accuracy, ensuring that only the most informative variables contribute to the predictions, avoiding the potential noise or redundancy from less relevant features. The proposed LSTM model demonstrates its capability to forecast COVID-19 cases, facilitating informed decision-making for public health authorities and policymakers.

Outcome: Our research on the analysis and prediction of COVID-19 cases in Central Thailand using multivariate long short-term memory (LSTM), incorporating meteorological and particulate matter factors, has yielded significant findings. The multi-feature selection process, which involved XGBoost, RF, SBF, PCA, and correlation coefficient analysis, was instrumental in identifying the most crucial input parameters for our predictive model. Among the selected input parameters, relative humidity emerged as a crucial predictor, strengthening our understanding of the virus's transmission dynamics. Our model can also demonstrate good performance in predicting COVID-19 cases.

Research grant and acknowledgement: 1. The Thailand Center of Excellence in Physics (ThEP). 2. The Centre of Excellence in Mathematics, Thailand.

Related SDGs goal: 3. Good health and well-being.



Graphical summary: Detailed Structure of an LSTM Memory Cell in an LSTM Recurrent Neural Network. (A) Data preprocessing, lag time selection, and feature selection, (B) An LSTM unit cell architecture, and (C) LSTM model.

Related publication:

Chanidapa Winalai, Suparinthon Anupong, **Charin Modchang**, Sudarat Chadsuthi. LSTM-Powered COVID-19 prediction in central Thailand incorporating meteorological and particulate matter data with a multi-feature selection approach. *Heliyon* 10 (2024) e30319. https://doi.org/10.1016/j.heliyon.2024.e30319