

A review on statistical and machine learning approaches to forecasting the occurrence of covid-19 positive cases

ABSTRACT

The World Health Organization (WHO) classified the 2019 new corona virus a worldwide pandemic on March 11th, 2020. Corona virus, also known as COVID-19, initially appeared in Wuhan, Hubei province, China, around December 2019 and quickly spread around the world. Many research efforts in the battle against the pandemic have been made, a lot of prediction models based on mathematical models, infectious disease models, and machine learning models have been developed. Previous work shows that the LSTM algorithm is the most used deep learning technique in forecasting various infectious disease such as Dengue, Malaria and recent Covid-19 pandemic. Previous study shows that it is important to conduct comprehensive studies on infectious disease especially Covid-19 due to its fast infection rate worldwide. Thus, this paper summarizes datasets, method and hyperparameters setting used to design experiments and models for prediction diseases outbreaks. At the same time, several limitations have been identified and need to be considered in building a robust LSTM model to learn time series data related to the occurrence of Covid-19 positive cases and death cases. These limitations include model design based on assumption, restricted to short time-series data, exclusion of impact changes factors such as time changes, spatial influence, climate factors, small sample dimension, depended on historical data and finally changes of future policies based on assumption.