

## Research Article

# Short-Term Infectious Diarrhea Prediction Using Weather and Search Data in Xiamen, China

Zhijin Wang<sup>1</sup>, Yaohui Huang<sup>2</sup>, Bingyan He<sup>1</sup>, Ting Luo<sup>2</sup>, Yongming Wang<sup>3</sup>,  
and Yonggang Fu<sup>1</sup>

<sup>1</sup>Computer Engineering College, Jimei University, Xiamen 361021, China

<sup>2</sup>Chengyi University College, Jimei University, Xiamen 361021, China

<sup>3</sup>China Electronics Technology Group Corporation, Shanghai 200001, China

Correspondence should be addressed to Yonggang Fu; yonggangfu@jmu.edu.cn

DWmHW \$%3bd^\$ " - Dm[eW # %3gYgef \$ " - 3UWbfW \$) 3gYgef \$ " - BgT[eZW % EwbfW TVd \$ " \$ "

Academic Editor: Chenxi Huang

Copyright © 2020 Zhijin Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Infectious diarrhea has high morbidity and mortality around the world. For this reason, diarrhea prediction has emerged as an important problem to prevent and control outbreaks. Numerous studies have built disease prediction models using large-scale data. However, these methods perform poorly on diarrhea data. To address this issue, this paper proposes a parsimonious model (PM), which takes historical outpatient visit counts, meteorological factors (MFs) and Baidu search indices (BSIs) as inputs to perform prediction. An experimental evaluation was done to compare the short-term prediction performance of ten algorithms for four groups of inputs, using data collected in Xiamen, China. Results show that the proposed method is effective in improving the prediction accuracy.

## 1. Introduction

To keep up with the pace of income growth, urbanization, and globalization, risk management of infectious diseases in public has become a critical task [1]. Infectious diarrhea (ID) [2] is one of the most common infectious diseases in the world, which infects more than 1 billion persons. It became the 37th legally notifiable disease in the China's National Notifiable Disease Reporting and Surveillance System, and any new case must be reported within 24 hours of diagnosis [3].

Early warning techniques [4] have been developed to monitor the status of infectious diseases and the demand for healthcare and health services. These techniques can support decision making for medical intervention strategies [5], by preinforming people, health service providers, and the government.

The problem of predicting upcoming diarrhea outpatient visits can be viewed as time series prediction problem. In past decades, numerous studies have used autoregression (AR), autoregressive integrated moving average

model (ARIMA), and machine learning methods to predict upcoming values based on past observations. The widely used machine learning methods are multiple linear regression (MLR), support vector regression (SVR), and random forest regression (RFR) [6–8]. A famous autoregression method is that of Box and Jenkins [9], which has been applied in many fields [10], such as for electricity load forecasting and stock price prediction. Another famous statistical method is spline interpolation [11], which learns and uses a cubic spline interpolation to predict future values. But the performance of these methods degrades when dealing with nonstationary and chaotic time series, such as those of diarrhea outpatients.

Recently, to alleviate the uncertainty of a time series, exogenous data have been collected and fused into machine learning methods to achieve better predictions [12–14]. Hereto, the methods using exogenous data are called NARX [15]. According to the structure of NARX, we categorize them into *wide models* and *deep models*.

A wide model commonly builds more than two components in a layer. To capture the sequential features of a





















