Article

Forecasting Weekly Influenza Outpatient Visits Using a Two-Dimensional Hierarchical Decision Tree Scheme

Tian-Shyug Lee [1,2], I-Fei Chen [3], Ting-Jen Chang [1,]* and Chi-Jie Lu [1,4,]*

[1]   Graduate Institute of Business Administration, Fu Jen Catholic University, New Taipei City 242062, Taiwan; 036665@mail.fju.edu.tw

[2]   Artificial Intelligence Development Center, Fu Jen Catholic University, New Taipei City 242062, Taiwan

[3]   Department of Management Sciences, Tamkang University, New Taipei City 251301, Taiwan; enfa@mail.tku.edu.tw

[4]   Department of Information Management, Fu Jen Catholic University, New Taipei City 242062, Taiwan

*   Correspondence: metheus.c@gmail.com (T.-J.C.); 059099@mail.fju.edu.tw (C.-J.L.);
    Tel.: +886-2-2905-2973 (C.-J.L.)

check for updates

Abstract: Influenza is a serious public health issue, as it can cause acute suffering and even death, social disruption, and economic loss. Effective forecasting of influenza outpatient visits is beneficial to anticipate and prevent medical resource shortages. This study uses regional data on influenza outpatient visits to propose a two-dimensional hierarchical decision tree scheme for forecasting influenza outpatient visits. The Taiwan weekly influenza outpatient visit data were collected from the national infectious disease statistics system and used for an empirical example. The 788 data points start in the first week of 2005 and end in the second week of 2020. The empirical results revealed that the proposed forecasting scheme outperformed five competing models and was able to forecast one to four weeks of anticipated influenza outpatient visits. The scheme may be an effective and promising alternative for forecasting one to four steps (weeks) ahead of nationwide influenza outpatient visits in Taiwan. Our results also suggest that, for forecasting nationwide influenza outpatient visits in Taiwan, one- and two-time lag information and regional information from the Taipei, North, and South regions are significant.

Keywords: public health; influenza outpatient visits; hierarchical structure; forecasting; decision tree

1. Introduction

Influenza, caused by the influenza virus, is a serious public health issue worldwide, as it is an important cause of individual morbidity and mortality in the population, and existing vaccines are not completely protective [1–3]. Influenza pandemics do not only cause a sudden increase in deaths, but also social disruption, and economic loss [4,5]. These effects, although in a more local and lower scale, can result from an epidemic episode even affecting a single country or continent. This has a direct impact on the outpatient departments of hospitals through significant overcrowding [6]. Accurate forecasting of influenza outpatient visits is beneficial for the reasonable planning and allocation of healthcare resources to meet medical demands or to anticipate potential shortages in medical resources.

Many existing studies use mathematical models, statistical methods, and machine learning to forecast influenza outpatient visits. For example, Towers and Chowell [7] use the susceptible-exposed-infectious-recovered mathematical model to estimate the latent period of influenza; Dugas et al. [8] employed a statistical model of generalized linear autoregressive moving average combined with negative binomial distribution and Google flu trends information to predict influenza

Int. J. Environ. Res. Public Health **2020**, *17*, 4743

2 of 14

cases; Nsoesie et al. [9] combined an individual-based mathematical model and optimization approach for influenza prediction; while Osthus et al. [10] used a deterministic mathematical model of the susceptible-infectious-recovered model to forecast seasonal influenza using multiple sources of uncertainty.

Volkova et al. [11] proposed a machine learning model of long short-term memory (LSTM) trained by combining data of influenza-like illness with social media signals; its performance was better than when only using historical data of influenza-like illness for influenza-like illness prediction. Basile et al. [1] used a functional generalized least squares regression model to predict epidemic intensity levels; Venna et al. [12] utilized a LSTM neural network model to use the spatiotemporal and environmental factors for influenza prediction. Machine learning does not require strong model assumptions and has advantages in that it is able to build a relatively effective forecasting model. It can also explain non-linear relationships and extract features, and is suitable for analyzing complex, high dimensional data [13]. Thus, machine learning has been adapted in this study to develop an influenza outpatient visit forecasting model.

The centers for disease control and prevention of a country usually have a duty to track influenza activity, and to monitor and collect data on influenza outpatient visits. The national infectious disease statistics system collects the data, which contain both national and regional statistics. The data on influenza outpatient visits therefore have a hierarchical data structure. The hierarchical data structure is different from a single independent autocorrelation time series, as it is a lag series. Individual time series are grouped by their proximity, and aggregated levels are the sum of the corresponding subsets [14,15]. For example, if we have the regional numbers of influenza outpatient visits, the region is probably a city, county, or state, and the aggregated levels mean that the nationwide numbers are the sum of all of the regional numbers. The national time series collection often has some aggregation structure: An obvious example is the national collection of disease statistics, which records the numbers nationally and covers each city or state. This aggregation is not only a consequence of the data collection method but also a requirement for ensuring privacy.

The existing influenza forecasting studies usually use a nationwide lag time series for multivariate data analysis. The nationwide lag time series can carry information or patterns depending merely on time. In order to collect additional information, the researcher introduced other independent time series, such as Google flu trends, Twitter, Wikipedia, meteorological data, metapopulational data, and laboratory data into the model [1,6,11,12,16,17]. These unstructured variables make it difficult to capture interaction information between variables. The structure patterns of hierarchical influenza data are useful for predicting aggregated nationwide values by capturing the data interactions between each region. The existing studies have supported the efficacy of building forecasting model based on hierarchical data structure [18,19]. For example, Collins [20] presented that the proposed hierarchical forecasting model can generate more accurate forecasts. Dunn et al. [21] and Dangerfield and Morris [22] stated that building a forecasting model using hierarchical data from a lower level can generate better performance.

As regions and time lags often provide useful spatial and temporal information for modeling predictions, this study was based on time lags and region information on influenza outpatient visits to propose a two-dimensional hierarchical influenza forecasting model using machine learning methods. Decision tree methods are one of the most important and frequently used machine learning methods because they have the advantage of being easy to understand and to explain the built model, easy to extract hidden patterns of data, and are suitable for processing unrelated features of data. Random forest (RF) and extreme gradient boosting (XGBoost) are two major decision trees as they can measure the importance of variables and handle high dimensional data [23]. Thus, RF and XGBoost are used in the proposed two-dimensional hierarchical influenza outpatient visit forecasting model to provide effective forecasting results.

In influenza forecasting, the most important forecasting horizon is one to four weeks ahead. Reich et al. [24] indicate one- and two-weeks-ahead forecasting of influenza are referred to as nowcast,

which means the short-time forecasting is approximate to the current time. Similarly, three- and four-weeks-ahead forecasting of influenza is called a forecast, which means long-time forecasting or estimates about events in the future. We considered one- to four-weeks-ahead forecasting at the same time to include both views of nowcast and forecast. In this study, we used a multi-stage forecasting strategy to predict influenza outpatient visits one to four steps ahead. To evaluate the reliability of the proposed two-dimensional hierarchical decision tree scheme for influenza outpatient visit forecasting, popular statistical methods for medical data forecasting, including autoregressive integrated moving average (ARIMA) and multivariate adaptive regression splines (MARS), were used in this study for comparison [25,26].

2. Methods

2.1. Random Forest

RF is a combination classification method of a supervised machine learning algorithm based on a decision tree, and aims to build up a collection of decision trees [27]. A popular aggregation function is considered as the unweighted majority of class votes. RF changes constructed methods of classification or regression trees. RF uses the best value among a subset of predictors randomly chosen at that node to split each node. In the process, RF uses a different bootstrap sample of the data to construct each tree, and uses the bagging program to select multiple random variable samples as the training data set. The bagging program indicates replacing the meta-algorithm with random sampling. Then, under the training process, a tree classifier corresponding to the selected sample is constructed, and a large number of the trees from the selected samples are used to produce an RF.

The typically tree-based classifier of RF is a classification and regression tree. RF uses the metric called out of bag error to measure the performance, which is the average of the rate of error in each weak learner. Each tree of the RF depends on the sampling random vector, and all trees in the forest have the same distribution. When the trees are developed to its maximum expansion, RF will use a particular method to explore each individual tree and select the most important variable for splitting randomly [28]. The common stopping criteria of RF for the split are that the number of samples in the leaf node is less than the given threshold, and the Gini index in the sample set is less than the given threshold.

2.2. XGBoost

XGBoost is an advanced method based on the tree boosting algorithm. It can perform massive parallel-boosted tree calculations. It optimizes the loss function of the traditional gradient boosting decision tree model, and provides regularization, a new sparsity-aware algorithm, a distributed weighted quantile sketch algorithm, and optimizes the cut-point search algorithm. The algorithm optimizes the value of the objective function and applies the feature vectors for prediction. The gradient enhancement builds an enhancement tree to adaptively obtain feature scores, indicating the difference in importance of each feature in the training model. The enhanced tree is used for key decision-making, in which some tree features are more, and the score is higher. XGBoost calculates the importance of features by gain, frequency, and coverage [29].

The XGBoost algorithm uses the regular term in the cost function to effectively prevent overfitting and to perform a second-order Taylor expansion of the cost function, and uses both first-order and second-order derivatives, while applying the loss function to calculate the pseudo-residuals of the first and second derivatives for generating learning. Its *t*-th loss function can be expressed as [25]:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{1}$$

XGBoost considers the inefficiency of the greedy method, so it uses an approximate algorithm to find the best segmentation point. It can use the sparse value to specify the default direction of the branch for a missing value or specified value, improving the efficiency of the algorithm [30]. When the gain of new branches is less than zero or the tree reaches the maximum depth, the XGBoost algorithm will be stopped.

2.3. Autoregressive Integrated Moving Average

The autoregressive integrated moving average (ARIMA) is constructed on historical data and attempts to identify patterns in the data. An ARIMA model contains three components: Autoregressive (AR), moving average (MA), and an integrated component. An ARIMA model is a generalization of an ARMA model [31]. In simple terms, the ARIMA model of nonstationary random process can be described as ARIMA (*p,d,q*) by the order of AR, the difference (integrated component), and MA, and its mathematical model can be denoted as [32]:

$$\left(1 - \sum_{i=1}^{p} \alpha_i B^i\right)(1 - B)^d Y(t) = \beta_0 + \left(1 - \sum_{i=1}^{q} \beta_i B^i\right)\varepsilon(t) \tag{2}$$

where B is the backshift operator; $\{\alpha_i\}$ are the AR coefficients; $\{\beta_i\}$ are the MA coefficients; $\varepsilon(t)$ are the white Gaussian noise process with mean zero and variance σ^2; β_0 is the deterministic trend term, and there is an implied polynomial of order d in the forecast function when $d > 0$. Therefore, $\left(1 - \sum_{i=1}^{p} \alpha_i B^i\right)$ and $\left(1 - \sum_{i=1}^{q} \beta_i B^i\right)$ are polynomials of order p and q, respectively. The Box-Jenkins methods of model identification for a non-stationary time series can be abstracted in four steps [33]. The first step is selecting the appropriate conversion of the observed time series. The most basic transformations are variance-stable transformations and differential operations. The second step is calculating the ACF (Auto-correlation) and PACF (Partial Auto-correlation Function) of the samples in the observed time series to determine the necessity and degree of difference. In the third step, we can identify the orders of p and q for the ARIMA model depending on calculating the sample ACF and PACF of the correctly transformed time series. In the last, if $d > 0$, we test the deterministic trend term β_0 to determine the necessity of including β_0 in the model.

2.4. MARS

MARS is a new technology of nonparametric statistics using a multivariable nonlinear algorithm to aim at finding optimal variable transformations and interactions. It is a flexible process to model relationships that are approximately additive or involve interactions with fewer variables. The model is a sum of several basic equations or the basis function (BF), and then MARS combines them for prediction. Under the modeling process, MARS is dependent on a divide-and-conquer strategy. The training data sets are divided into separate regions and each region has its own regression equation. In different intervals of the independent variable space, MARS uses the separate linear regression slopes in distinct intervals to approximate the non-linearity of a model [34]. We can use the following mathematical formula to express it:

$$\hat{f}(x) = \alpha_0 \sum_{i=1}^{m} \alpha_i \prod_{k=1}^{kn} [p_{kn}(x(k, n) - t_{kn})] \tag{3}$$

where α_0 and α_i are parameters, m is the number of basic functions, k_n is the number of knots, p_{kn} is given as values of either 1 or −1, indicating the right/left sense of the associated step function, $x(k, n)$ is the label of the independent variable, and t_{kn} means the knot location. We can consider that α_0 and α_i give a value similar to regression coefficients, mainly the number of BF determined by the evaluation criteria.

We can infer that the data points within the expectation interval show a linear relationship, and then use a linear function (basic equation) in each interval to describe the relationship between the data in the interval. Among the intervals, the node or knot is used as a separation point. This means that these nodes are not only the inflection point where the slope of the equation changes at each interval, but also where the data behavior changes. After accumulating the basic equations of these intervals, we can obtain a more flexible nonlinear model to solve complex problems that are multiple and nonlinear. In other words, there are three steps for MARS procedure. In the first step, the algorithm is employed to pick all probable basic functions and their corresponding knots. The second step is in order to produce the best combination of existing knots. MARS is dependent on the general cross-validation (GCV) criterion using a backward algorithm to delete all basic functions in the order of the least contribution. GCV is a form of regularization, which can balance goodness-of-fit against model complexity. When the lowest GCV value is reached, the best combination is chosen.

2.5. Model Implementation

The methods used in this study have been implemented in R of version 3.61 (R core team, Vienna, Austria). The algorithm of the methods is based on the relevant R package. There are five R packages used in our study; in particular, the *e1071* R package of version 1.7-3 [35] is used to construct the RF model and the *tuneRF* function of *e1071* R package is used to tune and obtain the best model. Then, we implemented MARS using the *earth* R package of version 5.1.2 [36]. The default setting of this package was used to tune and then build the best additive MARS model.

Because we are trying to determine the importance of variables from RF, the *RandomForest* R package of version 4.6-14 [37] was used to extract the important variables from the best model of RF. In the next section, we find the optimal orders for ARIMA using the *auto.arima* function of the *forecast* R package of version 8.11 [38] to provide a method to compute the best orders for the model automatically, and then we use the order to build the ARIMA model. Finally, the XGBoost model was constructed using the *xgboost* version 0.90.0.2 of the R package [39] for tuning and modeling the best XGBoost model.

2.6. Proposed Influenza Outpatient Visit Forecasting Scheme

In this study, two decision tree techniques, RF and XGBoost, were integrated into a two-dimensional hierarchical data scheme to build an effective scheme for predicting influenza outpatient visits in Taiwan. A flowchart of the proposed scheme is shown in Figure 1.

The first step in the proposed scheme was collecting the data. The second step was to construct a two-dimensional hierarchical structure based on nationwide and regional data. The two-dimensional hierarchical structure considered the regional values, and used the time lag of each region to generate potential predictor variables for the RF and XGBoost methods. In this structure, the total number of outpatient visits (nationwide) was equal to the sum of the outpatient visits of all regions, so we were able to obtain a data matrix of size *n* by *p* that constructed the number of influenza outpatient visits in *n* regional and its *p* time-lag. That is, the number of influenza outpatient visits in the region *n* and time-lag *p* of the two-dimensional hierarchical structure can be expressed as $X_{n,t-p}$.

For example, as depicted in Figure 1, $X_{1,t-1}$ represents the number of influenza outpatient visits in the region 1 and lag 1 week; $X_{3,t-2}$ stands for the number of influenza outpatient visits in the region 3 and lag 2 weeks, and so on.
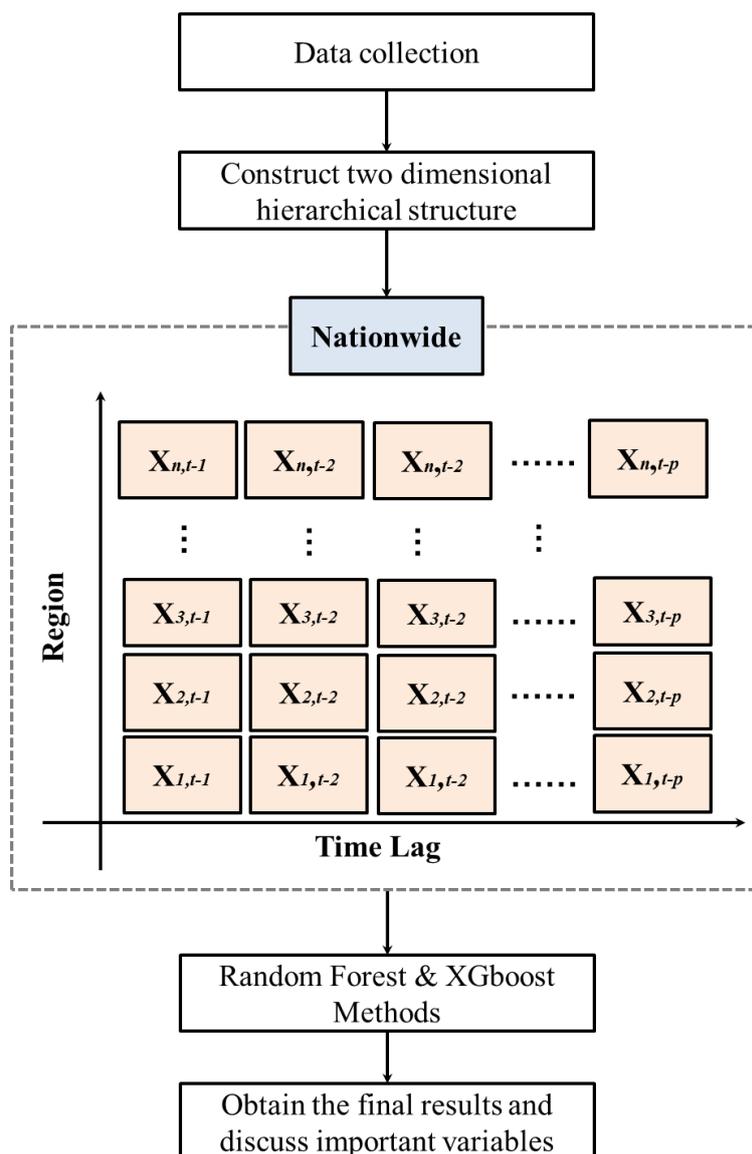
Figure 1. The flowchart of the proposed influenza outpatient visits forecasting scheme.

As the Taiwan influenza data contains six regions, namely, the Taipei, North, Central, South, Kaoping, and East regions, $n = 6$ was used in this study. The notations $\{T, N, C, S, K, E\}$ are, respectively, corresponding to the Taipei, North, Central, South, Kaoping, and East regions. Moreover, since the time-lag information of one to four weeks was the one most used and produced reasonable information for influenza outpatient visit forecasting [24], $p = 4$ was considered in this search. Therefore, we could generate 24 predictor variables, $X_{n,t-p}$, $n = \{T, N, C, S, K, E\}$, $p = 1, 2, 3, 4$. The target variable was the total (nationwide) influenza outpatient visits at time t (Y_t).

In the third step, we used RF and XGBoost decision tree methods to predict nationwide influenza outpatient visits using the generated predictor variables. In the two models constructed, the 24 hierarchical variables were directly used as predictors for RF and XGBoost. The model used RF and was called a two-dimensional hierarchical random forest (TDHS-RF) scheme. Similarly, the model utilizing XGBoost as the prediction method was called TDHS-XGB. Meanwhile, the MARS was also used in the proposed two-dimensional hierarchical scheme to evaluate the advantage of the decision tree, which was termed TDHS-MARS. Moreover, the RF, XGBoost, MARS, and ARIMA methods, without incorporating a two-dimensional hierarchical-based scheme, were used as benchmarking methods for evaluating the proposed influenza outpatient visit prediction scheme.

Int. J. Environ. Res. Public Health **2020**, *17*, 4743

7 of 14

In order to evaluate the performance of the proposed scheme and the competing methods, five commonly used performance metrics including mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute scaled error (MASE), and root mean square percentage error (RMSPE) were used in this study. The definitions of these metrics are summarized in Table 1. These metrics were used to measure the deviations between the actual and predicted values. The deviation was smaller, and the accuracy was higher.

Table 1. The equations of the five performance metrics.

Metrics	Calculation *
MAE	$\text{MAE} = \frac{1}{m} \sum\limits_{j=1}^{m} \left| F_j - A_j \right|$
RMSE	$\text{RMSE} = \sqrt{\frac{1}{m} \sum\limits_{j=1}^{m} \left(F_j - A_j \right)^2}$
MAPE	$\text{MAPE} = \frac{1}{m} \sum\limits_{j=1}^{m} \left| \frac{F_j - A_j}{A_j} \right| \times 100$
MASE	$\text{MASE} = \dfrac{\frac{1}{m} \sum_{j=1}^{m} \left| F_j - A_j \right|}{\frac{1}{z-1} \sum_{t=2}^{n} \left| A_t - A_{t-1} \right|}$
RMSPE	$\text{RMSPE} = \sqrt{\frac{1}{m} \sum\limits_{j=1}^{m} \left(\frac{F_j - A_j}{A_j} \right)^2}$

* Note that F and A represent the predicted and actual value; m is the total number of testing set; z is the total number of training set.

In the last step, after obtaining the final forecasting results of each model, we compared the prediction errors of all the models and obtained the final results for predicting influenza outpatient visits in Taiwan. The meanings of the identified important variables were also discussed.

3. Empirical Study

In this study, we used weekly influenza outpatient visit data collected from Taiwan's national infectious disease statistics system (https://nidss.cdc.gov.tw/en/). The data range was from the first week of 2005 to the second week of 2020, and included 788 data points. The dataset contained nationwide outpatient visits for influenza, and six regional outpatient visits for influenza. The regions were Taipei, North, Central, South, East, and Kaoping, and the nationwide values were the sum of the six regional values. Figures 2–4 depict Taiwan's total national influenza outpatient visits from the first week of 2005 to the second week of 2020.

The sliding window approach was used in this study to split the data into training and testing data sets based on their time period. After data cleansing, the first 630 data points (80% of the total sample points, from the first week of 2005 to the first week of 2017) were selected as the training samples, while the remaining 158 data points (20% of the total sample points, from the second week of 2017 to the second week of 2020) were employed as the testing sample for measuring the out-of-sample forecasting ability of the proposed scheme.

The proposed TDHS-RF and TDHS-XGB models were used in one to four step (week) forecasts of nationwide influenza outpatient visits in Taiwan. First, Table 2 presents the results of one-step (week)-ahead outpatient visit forecasting of the proposed TDHS-RF and TDHS-XGB models, and the competing models including TDHS-MARS, RF, XGBoost, MARS, and ARIMA. As Table 2 depicts, the performance of the proposed TDHS-RF and TDHS-XG models are better than those of the five comparison models, and the TDHS-XG model can provide the best forecasting accuracy of all the models for one-week-ahead nationwide influenza outpatient visit forecasting. It can also be observed from Table 2 that both the TDHS-RF and TDHS-XGB models outperformed the RF and XGBoost models, which did not use two-dimensional hierarchical structure data. This indicates that the proposed data scheme can be used to improve forecasting accuracy. In addition, Table 2 shows that all

machine learning-based methods, including TDHS-RF, TDHS-XGB, RF, and XGBoost, outperformed the statistical-based methods, including TDHS-MARS, MARS, and ARIMA.
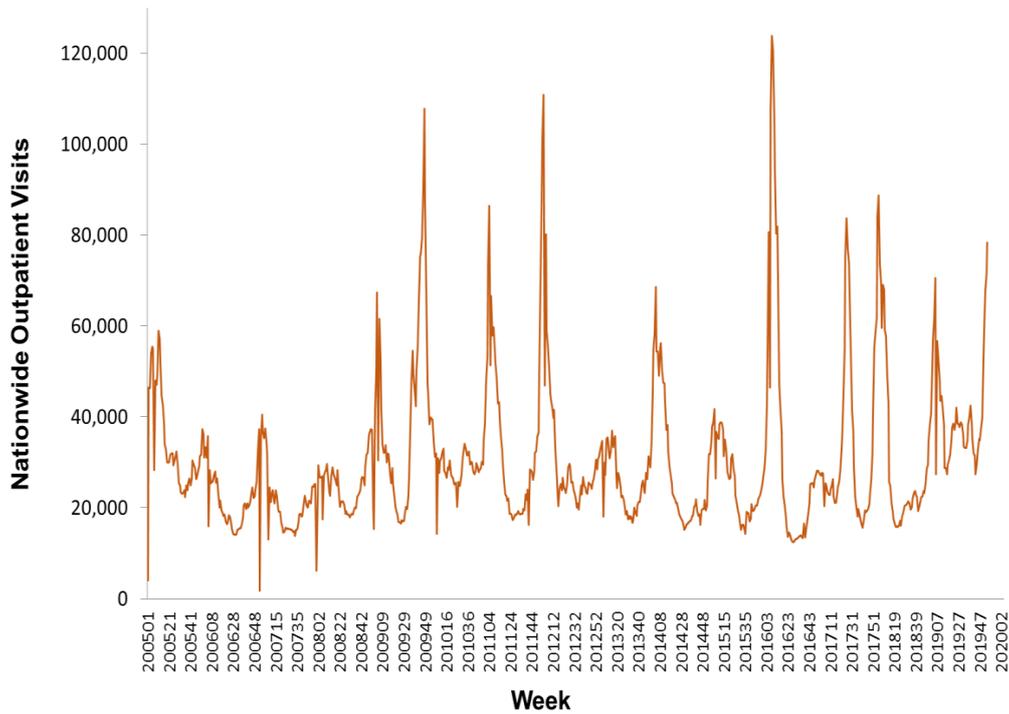

Figure 2. The nationwide outpatient visits for influenza in Taiwan from the first week of 2005 to the second week of 2020.
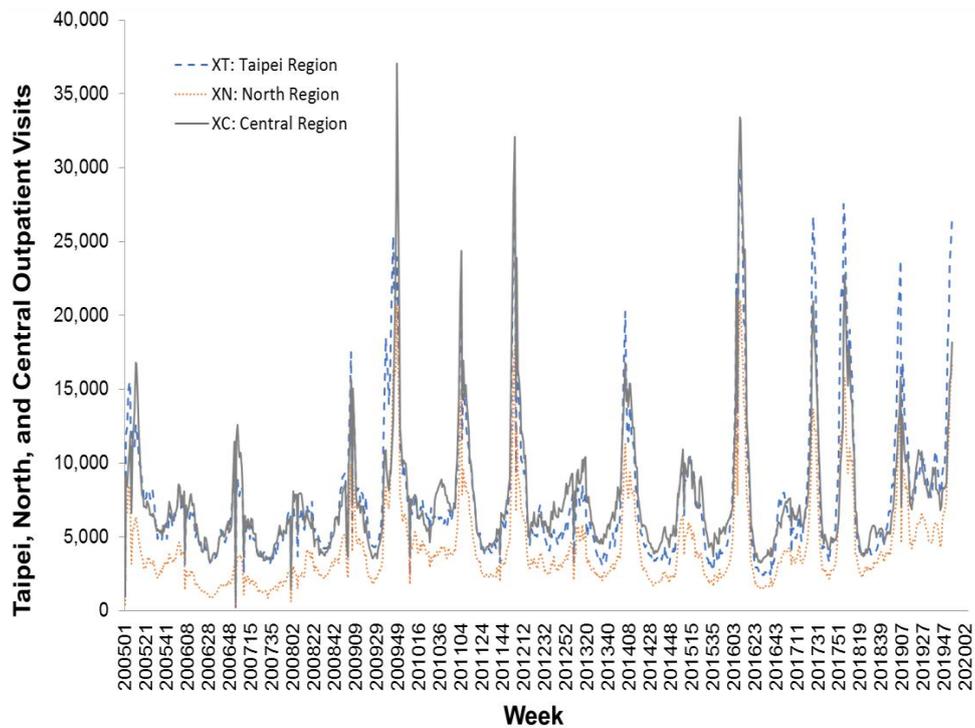

Figure 3. The influenza outpatient visits for Taipei, North, and Central regions in Taiwan from the first week of 2005 to the second week of 2020.
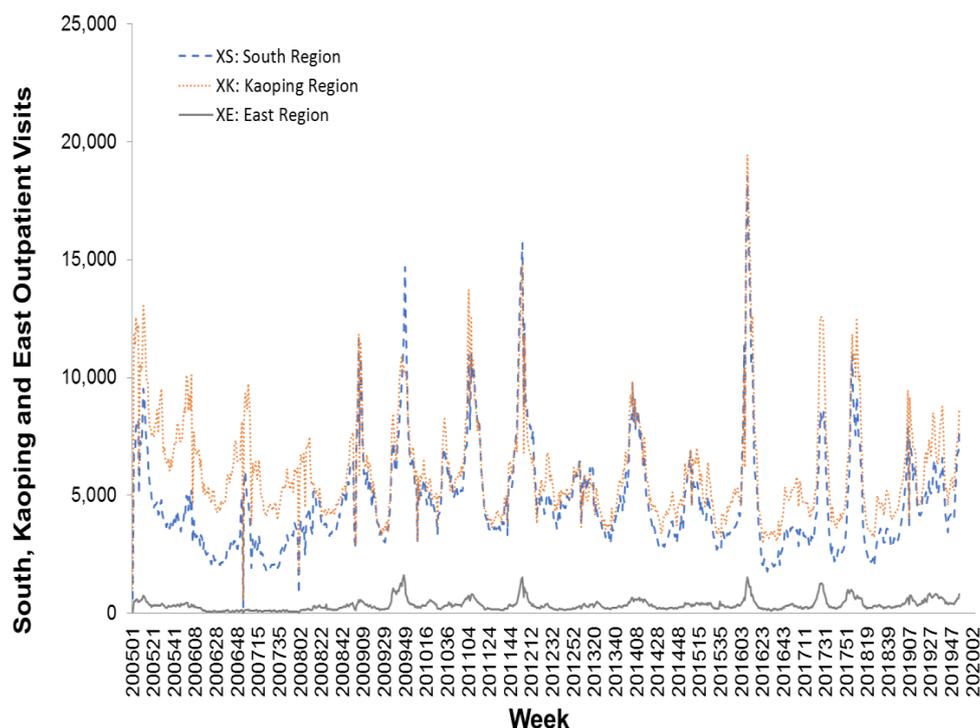
Figure 4. The influenza outpatient visits for South, East, and Kaoping regions in Taiwan from the first week of 2005 to the second week of 2020.

Table 2. Forecasting results of one step (week) ahead outpatient visits.

Method	MAE	RMSE	MAPE	MASE	RMSPE
TDHS-RF	4693	8702	0.126	1.167	0.188
TDHS-XG	4409	8050	0.113	1.096	0.168
TDHS-MARS	6415	14,000	0.159	1.595	0.226
RF	5040	9179	0.133	1.253	0.194
XGB	4432	8029	0.117	1.102	0.173
MARS	5158	9861	0.133	1.282	0.201
ARIMA	13,986	20,774	0.417	3.477	0.512

Note that TDHS–RF means two-dimensional hierarchical random forest; TDHS-XGB means two-dimensional hierarchical extreme gradient boosting; TDHS-MARS means two-dimensional hierarchical multivariate adaptive regression splines; RF means random forest; XGB means extreme gradient boosting; MARS means multivariate adaptive regression splines; ARIMA means autoregressive integrated moving average.

Table 3 shows the forecasting results of two- to four-steps (weeks)-ahead of outpatient visits of the proposed TDHS-RF and TDHS-XG models, and the five comparison methods. Table 3 shows that the proposed TDHS-RF and TDHS-XGB models outperformed the five comparison methods in all two- to four-step (week) forecasts. The machine learning forecasting models can generate better forecasting results than the statistical methods including ARIMA and MARS. Thus, it can be concluded that the proposed TDHS-RF and TDHS-XG models are effective and promising alternatives for forecasting one to four steps (weeks) ahead for nationwide influenza outpatient visits in Taiwan.

Table 3. Forecasting results of two to four steps (weeks) ahead of outpatient visits.

Step (Weeks) Ahead	Method	MAE	RMSE	MAPE	MASE	RMSPE
	TDHS-RF	6451	10,614	0.176	1.599	0.240
	TDHS-XGB	6439	11,018	0.163	1.596	0.225
	TDHS-MARS	8077	14,819	0.205	2.002	0.268
2	RF	6858	11,342	0.186	1.700	0.264
	XGB	6489	11,009	0.168	1.608	0.231
	MARS	7071	11,689	0.185	1.753	0.247
	ARIMA	14,037	20,818	0.418	3.479	0.513
	TDHS-RF	8222	13,098	0.227	2.033	0.316
	TDHS-XGB	8329	13,764	0.211	2.059	0.291
	TDHS-MARS	9671	16,012	0.252	2.391	0.328
3	RF	8691	13,857	0.241	2.149	0.346
	XGBoost	8564	14,119	0.222	2.118	0.312
	MARS	8780	14,054	0.234	2.171	0.323
	ARIMA	14,082	20,862	0.419	3.482	0.514
	TDHS-RF	9994	15,609	0.280	2.463	0.401
	TDHS-XGB	10,178	16,550	0.264	2.508	0.375
	TDHS-MARS	11,114	17,746	0.295	2.739	0.404
4	RF	10,478	16,595	0.295	2.582	0.435
	XGB	10,433	17,027	0.274	2.571	0.401
	MARS	10,462	16,654	0.285	2.578	0.414
	ARIMA	14,118	20,903	0.419	3.479	0.514

Table 4 shows the relative importance value of each predictor variable in decreasing order based on the importance identification results of the proposed TDHS-RF and TDHS-XGB models. It is the average value of the variable importance values of one- to four-step forecasting using the TDHS-RF and TDHS-XGB models. From Table 4, it can be found that $X_{T,t-1}$ is the most important variable. It means that the number of influenza outpatient visits in the Taipei region and lag one week is the most important reference information to plan and allocate healthcare resources for influenza prevention.

Table 4. The relative importance value of each predictor variable based on the proposed schemes.

Variable	Relative Importance	Variable	Relative Importance
$X_{T,t-1}$	19.06%	$X_{C,t-4}$	0.88%
$X_{S,t-1}$	17.45%	$X_{C,t-3}$	0.85%
$X_{N,t-1}$	15.43%	$X_{T,t-4}$	0.81%
$X_{E,t-1}$	9.86%	$X_{S,t-3}$	0.74%
$X_{C,t-1}$	8.72%	$X_{K,t-2}$	0.74%
$X_{T,t-2}$	7.32%	$X_{N,t-3}$	0.65%
$X_{E,t-2}$	3.84%	$X_{K,t-3}$	0.65%
$X_{S,t-2}$	3.26%	$X_{S,t-4}$	0.64%
$X_{N,t-2}$	2.79%	$X_{N,t-4}$	0.57%
$X_{C,t-2}$	1.89%	$X_{E,t-3}$	0.53%
$X_{K,t-1}$	1.33%	$X_{E,t-4}$	0.52%
$X_{T,t-3}$	0.92%	$X_{K,t-4}$	0.52%
		Total	100.00%

4. Discussion

Figure 5 shows the relative importance value and cumulative importance value of each variable based on Table 4. For identifying important variables, after discussing with three experts in public health, the first three variables such as $X_{T,t-1}$, $X_{S,t-1}$, and $X_{N,t-1}$ are inferred to be the most important variables and can be considered as a crucial sign for influenza prevention. From Figure 5, we can find the cumulative importance of the first three important variables is 52% which is around 60%

and these three variables account for only 12.5% of all 24 variables. According to the first three important variables, we can infer that when the number of influenza outpatient visits in the previous one week are bumping up in the Taipei, South, and North regions, it is highly probable that within the next four weeks, there will be a nationwide pandemic. From an administrative area perspective, it seems reasonable because the Taipei region includes Taipei City, New Taipei City, Keelung City, Yilan County, Kinmen County, and Lianjiang County. The North region includes Hsinchu City, Hsinchu County, Taoyuan City, and Miaoli County; and the South region includes Tainan City, Chiayi City, Chiayi County, and Yunlin County. These regions cover the most important areas of economic activity in Taiwan.
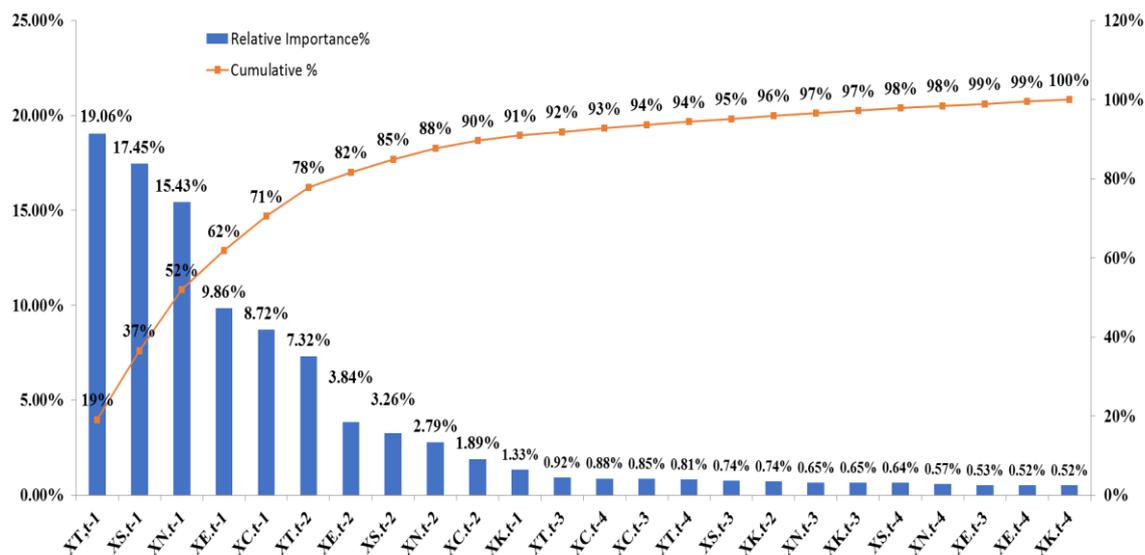

Figure 5. The relative and cumulated importance values of each predictor variables.

For using more important variables for influenza prevention, the experts suggested that the first ten variables including $X_{T,t-1}$, $X_{S,t-1}$, $X_{N,t-1}$, $X_{E,t-1}$, $X_{C,t-1}$, $X_{T,t-2}$, $X_{E,t-2}$, $X_{S,t-2}$, $X_{N,t-2}$, and $X_{C,t-2}$ could be considered since these then variables account for only 41.6% of all 24 variables. The cumulative importance value the first ten important variables is 90% which is also a common concept in the statistical analysis [40]. The first ten important variables depict the geographical area of Taipei, South, North, Central, and the East regions, and its influenza outpatient visits in the previous one and two weeks are important, and the variables for lag three to four weeks are not selected. The result presents that the high-frequency variable in this study is less of a contribution to predict influenza outpatient visits nationwide. The first ten variables can be used to predict the trend of influenza outpatient visits one week to four weeks ahead.

5. Limitations and Future Research

This study only used weekly influenza outpatient visits of Taiwan to evaluate the performance of the proposed scheme and to find important variables as a meaningful sign for influenza prevention. The findings of this study could not be directly extended to other countries' outpatient visits of influenza is one of the limitations. The other limitations of this study are that outpatient visits of influenza-like illness, emergency department visits of influenza, and influenza-like illness were not considered in this study.

The LSTM, one of the state-of-the-art machine learning methods, is an attractive method and has been used for influenza forecasting study in recent years [11,12]. But, the main disadvantage of LSTM is that it has a complicated training model which can easily get overfitting and require a long training time [41,42]. Meanwhile, LSTM cannot be used to select important predictors. On the contrary, RF and XGboost have the advantages of being easy to build and explain forecasting models, easy to avoid

overfitting, and can select important predictors for further analysis. Moreover, RF and XGBoost are both promising methods and have been successfully used for different applications in recent years [43]. Thus, as the main aim of this study was to propose a two-dimensional hierarchical decision tree scheme for forecasting influenza outpatient visits and identifying important predictor variables, using LSTM in the proposed scheme and comparing the performance between the proposed scheme and the LSTM may not useful for our research purpose, although it might be one of our research limitations.

According to the limitations, using the proposed two-dimensional hierarchical decision tree forecasting scheme for other countries' influenza outpatient visit data and emergency department visits of influenza could be one of the future research directions. Moreover, as using the state-of-the-art machine learning methods with feature selection mechanism in the proposed scheme might generate a more promising influenza outpatient visit forecasting scheme, it would be considered as one of our future research subjects. In addition, utilizing more hierarchical variables such as population and economy indexes in the proposed scheme to provide a deeper insight for influenza prevention could be also be taken into account for our future study.

6. Conclusions

This study was based on spatial and temporal information on influenza outpatient visits to propose a two-dimensional hierarchical decision tree scheme for forecasting nationwide influenza outpatient visits in Taiwan. In the proposed scheme, first, a data matrix of predictor variables was constructed by the number of influenza outpatient visits in six regions and its four time-lags. The twenty-four predictor variables, including spatial and temporal information, were then used as input variables of RF and XGBboost algorithms to respectively construct the TDHS-RF and TDHS-XGB models. Finally, the proposed TDHS-RF and TDHS-XGB models were utilized to forecast one- to four-steps (weeks)-ahead nationwide influenza outpatient visits in Taiwan and identify important variables.

The empirical results demonstrated that the proposed TDHS-RF and TDHS-XGB models were effective and promising alternatives for forecasting one to four steps (weeks) ahead for nationwide influenza outpatient visits in Taiwan. Our results also suggest that, for forecasting nationwide influenza outpatient visits in Taiwan, one- and two-time lag information and region information from the Taipei, North, and South regions were most significant. If the numbers of influenza outpatients are significantly increased in those regions, officials should pay attention as there is a high probability of a nationwide influenza pandemic occurring within one to four weeks. This important spatial and temporal information can provide useful information to prepare and design reasonable planning and allocation of healthcare resources to prevent a potential influenza pandemic.

References

1. Basile, L.; Oviedo de la Fuente, M.; Torner, N.; Martı́nez, A.; Jane, M. Real-time predictive seasonal influenza model in Catalonia, Spain. *PLoS ONE* **2018**, *13*, e0193651. [CrossRef] [PubMed]
2. Liebowitz, D.; Gottlieb, K.; Kolhatkar, N.S.; Garg, S.J.; Asher, J.M.; Nazareno, J.; Tucker, S.N. Efficacy, immunogenicity, and safety of an oral influenza vaccine: A placebo-controlled and active-controlled phase 2 human challenge study. *Lancet Infect. Dis.* **2020**, *20*, 435–444. [CrossRef]
3. Al-qaness, M.A.A.; Ewees, A.A.; Fan, H.; Abd Elaziz, M. Optimized Forecasting Method for Weekly Influenza Confirmed Cases. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3510. [CrossRef] [PubMed]

4. Molinari, N.-A.M.; Ortega-Sanchez, I.R.; Messonnier, M.L.; Thompson, W.W.; Wortley, P.M.; Weintraub, E.; Bridges, C.B. The annual impact of seasonal influenza in the US: Measuring disease burden and costs. *Vaccine* **2007**, *25*, 5086–5096. [CrossRef]

5. Lu, J.; Meyer, S. Forecasting Flu Activity in the United States: Benchmarking an Endemic-Epidemic Beta Model. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1381. [CrossRef] [PubMed]

6. Araz, O.M.; Bentley, D.; Muelleman, R.L. Using Google Flu Trends data in forecasting influenza-like illness related ED visits in Omaha, Nebraska. *Am. J. Emerg. Med.* **2014**, *32*, 1016–1023. [CrossRef] [PubMed]

7. Towers, S.; Chowell, G. Impact of weekday social contact patterns on the modeling of influenza transmission, and determination of the influenza latent period. *J. Theor. Biol.* **2012**, *312*, 87–95. [CrossRef] [PubMed]

8. Dugas, A.F.; Jalalpour, M.; Gel, Y.; Levin, S.; Torcaso, F.; Igusa, T.; Rothman, R.E. Influenza forecasting with Google Flu Trends. *PLoS ONE* **2013**, *8*, e56176. [CrossRef]

9. Nsoesie, E.O.; Marathe, M.; Brownstein, J.S. Forecasting peaks of seasonal influenza epidemics. Edition 1. *PLOS Curr. Outbreaks* **2013**. [CrossRef]

10. Osthus, D.; Hickmann, K.S.; Caragea, P.C.; Higdon, D.; Valle, S.Y.D. Forecasting seasonal influenza with a state-space SIR model. *Ann. Appl. Stat.* **2017**, *11*, 202–224. [CrossRef]

11. Volkova, S.; Ayton, E.; Porterfield, K.; Corley, C.D. Forecasting influenza-like illness dynamics for military populations using neural networks and social media. *PLoS ONE* **2017**, *12*, e0188941. [CrossRef] [PubMed]

12. Venna, S.R.; Tavanaei, A.; Gottumukkala, R.N.; Raghavan, V.V.; Maida, A.S.; Nichols, S. A novel data-driven model for real-time influenza forecasting. *IEEE Access* **2018**, *7*, 7691–7701. [CrossRef]

13. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [CrossRef] [PubMed]

14. Cuenca, E.; Sallaberry, A.; Ying Wang, F.; Poncelet, P. MultiStream: A multiresolution streamgraph approach to explore hierarchical time series. *IEEE Trans. Vis. Comput. Graph.* **2018**, *24*, 3160–3173. [CrossRef]

15. Hyndman, R.J.; Lee, A.J.; Wang, E. Fast computation of reconciled forecasts for hierarchical and grouped time series. *Comput. Stat. Data Anal.* **2016**, *97*, 16–32. [CrossRef]

16. Pei, S.; Kandula, S.; Yang, W.; Shaman, J. Forecasting the spatial transmission of influenza in the United States. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 2752–2757. [CrossRef] [PubMed]

17. Wang, Y.; Xu, K.; Kang, Y.; Wang, H.; Wang, F.; Avram, A. Regional Influenza Prediction with Sampling Twitter Data and PDE Model. *Int. J. Environ. Res. Public Health* **2020**, *17*, 678. [CrossRef] [PubMed]

18. Tiao, G.C.; Guttman, I. Forecasting contemporal aggregates of multiple time series. *J. Econom.* **1980**, *12*, 219–230. [CrossRef]

19. Kohn, R. When is an aggregate of a time series efficiently forecast by its past? *J. Econom.* **1982**, *18*, 337–349. [CrossRef]

20. Collins, D.W. Predicting earnings with sub-entity data: Some further evidence. *J. Account. Res.* **1976**, *14*, 163–177. [CrossRef]

21. Dunn, D.M.; Williams, W.H.; DeChaine, T.L. Aggregate versus subaggregate models in local area forecasting. *J. Am. Stat. Assoc.* **1976**, *71*, 68–71. [CrossRef]

22. Dangerfield, B.J.; Morris, J.S. Top–down or bottom–up: Aggregate versus disaggregate extrapolations. *Int. J. Forecast.* **1992**, *8*, 233–241. [CrossRef]

23. Venkatesh, B.; Anuradha, J. A hybrid feature selection approach for handling high-dimensional data. Lecture notes in Networks and Systems. In *Innovations in Computer Science and Engineering*; Springer: Singapore, 2019; pp. 365–373.

24. Reich, N.G.; Brooks, L.C.; Fox, S.J.; Kandula, S.; McGowan, C.J.; Moore, E.; Moore, E.; Osthus, D.; Ray, E.L.; Tushar, A.; et al. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 3146–3154. [CrossRef]

25. Sharafi, M.; Ghaem, H.; Tabatabaee, H.R.; Faramarzi, H. Forecasting the number of zoonotic cutaneous leishmaniasis cases in south of Fars province, Iran using seasonal ARIMA time series method. *Asian Pac. J. Trop. Med.* **2017**, *10*, 79–86. [CrossRef]

26. Cong, J.; Ren, M.; Xie, S.; Wang, P. Predicting Seasonal Influenza Based on SARIMA Model, in Mainland China from 2005 to 2018. *Int. J. Environ. Res. Public Health* **2019**, *16*, 4760. [CrossRef] [PubMed]

27. Probst, P.; Wright, M.N.; Boulesteix, A.-L. Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, e1301. [CrossRef]

28. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

Int. J. Environ. Res. Public Health **2020**, *17*, 4743

14 of 14

29. Suo, G.; Song, L.; Dou, Y.; Cui, Z. Multi-dimensional short-term load Forecasting based on XGBoost and fireworks algorithm. In Proceedings of the 2019 18th International Symposium on Distributed Computing and Applications for Business Engineering and Science (*DCABES*), Wuhan, China, 8–10 November 2019.

30. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

31. Newbold, P. ARIMA model building and the time series analysis approach to forecasting. *J. Forecast.* **1983**, *2*, 23–35. [CrossRef]

32. Chen, P.; Pedersen, T.; Bak-Jensen, B.; Chen, Z. ARIMA-Based Time Series Model of Stochastic Wind Power Generation. *IEEE Trans. Power Syst.* **2010**, *25*, 667–676. [CrossRef]

33. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. *Time Series Analysis Forecasting and Control*, 3rd ed.; PrenticeHall: Englewood Cliffs, NJ, USA, 1994.

34. Friedman, J.H. Multivariate adaptive regression splines. *Ann. Stat.* **1991**, *19*, 1–67. [CrossRef]

35. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F. E1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TUWien, 2017. R Package Version, 1.7–3. Available online: https://www.rdocumentation.org/packages/e1071 (accessed on 18 January 2020).

36. Milborrow, S.; Hastie, T.; Tibshirani, R.; Miller, A.; Lumley, T. Earth: Multivariate Adaptive Regression Splines. R Package Version 5.1.2. Available online: https://www.rdocumentation.org/packages/earth (accessed on 18 January 2020).

37. Liaw, A.; Wiener, M. RandomForest: Breiman and Cutler's Random Forests for Classification and Regression. R Package Version, 4.6.14. Available online: https://www.rdocumentation.org/packages/randomForest (accessed on 18 January 2020).

38. Hyndman, R.J.; Khandakar, Y. *Automatic Time Series Forecasting: The Forecast Package for R.*; Monash University, Department of Econometrics and Business Statistics: Clayton, Australia, 2018.

39. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y. Xgboost: Extreme gradient boosting. R Package Version 0.90.0.2. Available online: https://www.rdocumentation.org/packages/xgboost (accessed on 18 January 2020).

40. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*, 6th ed.; Pearson: London, UK, 2014.

41. Niu, D.; Liu, Y.; Cai, T.; Zheng, X.; Liu, T.; Zhou, S. A Novel Distributed Duration-Aware LSTM for Large Scale Sequential Data Analysis. In *CCF Conference on Big Data*; Jin, H., Lin, X., Cheng, X., Shi, X., Xiao, N., Huang, Y., Eds.; Springer: Singapore, 2019; Volume 1120.

42. Lecuyer, G.; Ragot, M.; Martin, N.; Launay, L.; Jannin, P. Assisted phase and step annotation for surgical videos. *Int. J. Comput. Assist. Radiol. Surg.* **2020**, *15*, 673–680. [CrossRef] [PubMed]

43. Boutaba, R.; Salahuddin, M.A.; Limam, N.; Ayoubi, S.; Shahriar, N.; Estrada-Solano, F.; Caicedo, O.M.A. Comprehensive survey on machine learning for networking: Evolution, applications and research opportunities. *J. Internet Serv. Appl.* **2018**, *9*, 16. [CrossRef]