International Journal of Management, Economics and Social Sciences
Special Issue-International Conference on Medical and Health Informatics (ICMHI 2017)
2017, Vol. 6(S1), pp.166 – 184.
ISSN 2304 – 1366
http://www.ijmess.com

A Hybrid Time Series Model based on AR-EMD and Volatility for Medical Data Forecasting - A Case Study in the Emergency Department

[*]Liang-Ying Wei
Deng-Yang Huang
Shun-Chuan Ho
Jyh-Shyan Lin
Hao-En Chueh
Chin-Sung Liu
Tien-Hwa Ho

Dept. of Information Management, Yuanpei University of Medical Technology, Taiwan

Time series methods have been applied to forecast clinical data, such as daily patient number forecasting for emergency medical centers. However, the application of conventional time series models needs to meet the statistical assumptions, and not all models can be applied in all datasets. Most of the traditional time series models use a single variable for forecasting, but there are many noises involutedly in raw data that are caused by changes in weather conditions and environments for daily patient number forecasting. Time series models that use complicated raw data would reduce the forecasting performance. For solving the above problems, this paper develops a hybrid time series support vector regression (SVR) model based on empirical mode decomposition (EMD), AR (autoregressive) method, and volatility of data. The proposed model considers that EMD can decompose complicated raw data into highly correlations frequency components. Further, the volatility of data measures data change and implies the dataset's power. For the reason, this paper utilizes the data's volatility as an important variable to improve the proposed forecasting model. Then, this study utilizes SVR as a forecasting model that can overcome the limitations of statistical methods (data need to obey some mathematical distribution). In verification, this paper collects daily patient volumes in emergency departments as experimental datasets to evaluate the proposed model. Numerical results indicate that the proposed model outperforms the listing models.

Keywords: Support vector regression (SVR), empirical mode decomposition (EMD), clinical data forecasting

A hospital is the main feature of national health systems, and an essential part of the health system is the emergency department (ED). ED provides accurate health care and immediate medical help, which is relying on high technological devices and vocational practitioners. Because of gradual aging of the population, emergency services overcrowding problem is arising in many countries. If hospitals cannot make efficient

Manuscript received May 20, 2017; revised September 9, 2017; accepted October 5, 2017.
*Corresponding author: lywei@mail.ypu.edu.tw

allocation of ED resources, it would result in overcrowding problems. ED overcrowding must affect not only the quality of treatment and prognosis but also patient satisfaction. The crucial step for allocating ED resources is how to model and forecast the demand for EDs. Forecasting medical demand is a vital activity that guides decision-making in many tasks, such as expansions of beds, staff supplementation, and diversification of test equipment. Recently, several research works have concerned demand forecasting for resources, staff allocation, and emergency department flow measures in hospitals [4, 17, 22, 32]. Most researchers firstly consider conventional time series models as forecasting models to forecast medical resource demand, and many time series models have been proposed and applied to handle the different forecasting areas [5, 13, 21, 33]. Engle [13] proposed the ARCH (p) (Autoregressive Conditional Heteroscedasticity) model that has been used by many financial and economy analysts, and the GARCH [5] (Generalized ARCH) model is the generalized form of ARCH. Box and Jenkins [6] proposed the autoregressive moving average (ARMA) model, which combines a moving average process with a linear difference equation to obtain an autoregressive moving average model, and the ARMA model performs forecasting at linear stationary conditions. Models that describe such homogeneous non-stationary behavior can be obtained by supposing some suitable difference of the process to be stationary. Therefore, the autoregressive integrated moving average model (ARIMA) [6], with the assumption of linearity among variables, was proposed to handle the non-stationary behavior datasets. Besides, linguistic expressions are often used to describe daily observations. Hence, Song and Chissom [33] first proposed the original model of the fuzzy time series, and the following researcher, Chen [9], proposed a refined fuzzy time series model for enrollment forecasting. In focusing on establishing fuzzy relationships of fuzzy time series models, Yu [43] recommended that different weights should be set in various fuzzy relationships and proposed a weighted fuzzy time series method to forecast stock index. From the literatures above, AR (autoregressive) is a fundamental and important method in time series models. Not all models can be applied in all datasets; the reason is that application of conventional time series models need to meet statistical assumptions [38]. In addition, most of the traditional time series models use a single variable for forecasting. However, there are many noises involutedly in raw data that are caused by changes in surrounding conditions for patient volume forecasting. Conventional time series models get poor performance, because those models use complicated

raw data [39]. Further, volatility of data has been applied to time series forecasting [26, 34]. The data' s volatility measures data change and represents the power of the dataset. For the reasons, this paper reputes that the volatility of data plays an important role to affect data trends in the future. Thus, the volatility of data could be adopted as an important variable for the forecasting model.

Nowadays, information technology (IT) plays an important role to manage knowledge in healthcare environments [35, 40]. Although there are many challenges in creation, dissemination, and preservation of health care knowledge using advanced technologies, information technology is still the critical tool in the application field of medical care management. IT also has been applied to forecast health care demand, and because of continuous increases in the demand for emergency medical services, using IT for forecasting is becoming more and more important. Data mining method is one of IT and is a rapidly growing technology in the information processing application and has attracted great attention in knowledge discovery research fields. Knowledge discovery process consists of an iterative sequence of data integration, data selection, and data mining pattern recognition and knowledge presentation and has been applied to various disciplines, such as business and medical data forecasting [2, 3, 12]. Kim and Han [25] proposed a genetic algorithms approach to feature discretization and the determination of connection weights for artificial neural networks (ANNs) to predict the stock price index. Roh [31] integrated neural network and a time series model for forecasting the volatility of stock price index. Jones et al. [23] use artificial neural networks (ANNs) to forecast daily patient volumes in emergency departments. ANNs are promising forecasting methods and have been applied extensively in many domains. However, they would suffer from a network construction problem and the need of large training datasets. There are many research works that SVR overcomes the problem of deciding architecture and other shortcomings for ANN and has better prediction results in different domains [8, 19, 27].

Support vector regression (SVR) method [36], one of data mining methods, has been used in widespread applications, including medical data prediction, and performs well. Support vector machine (SVM) is based on structural risk minimization rather than empirical risk minimization, which is a maximum margin model proposed by Vapnik [36]. SVR has a global optimum and appears better forecasting accuracy for its structural risk minimization principle. It thinks both the capacity of regression and training error to avoid over-

fitting and under-fitting problems in the training step. Support vector regression is adapted from SVM for regression tasks. Unlike pattern recognition problems, where the desired outputs are discrete values (e.g., Boolean), the support vector regression (SVR) deals with ' real valued' functions. The principle of structural risk minimization makes SVM to have a stronger generalization ability, and generally, it performs better than the prior developed methodologies, such as neural networks and other conventional statistical models [29]. Thus, this work mainly focuses on the advantage of SVR and applies it to the proposed model to enhance forecasting performance.

In order to enhance prediction performances, except single forecasting algorithms, hybrid models are often utilized, and models that use empirical mode decomposition (EMD) have gained great attention [11]. EMD [20] is a useful method to deal with non-linear signal analysis (such as stock data) or other related fields [37, 42] and offers a new way to deal with nonlinear and non-stationary signals. EMD-based prediction methods have been used on wind speed prediction [1, 30], industry [14], tourism management [27], and financial time series forecasting [15]. Based on EMD, any complicated signal can be decomposed into a finite and often small number of intrinsic mode functions (IMFs), which have simpler frequency components and stronger correlations and are thus easier and more accurate to forecast.

From the mention above, there are some major drawbacks in those models: (1) some traditional time series models cannot be applied to the datasets that do not follow the statistical assumptions; and (2) most conventional time series models utilize late day data with noises as input variable in forecasting. However, there are noises, including in raw data, which are generated by environment and weather conditions. In order to overcome the drawbacks above, this paper considers that EMD can decompose the complicated raw data into simpler frequency components and highly correlating variables. Then, this study utilizes SVR as a forecasting model that can overcome the limitations of statistical methods (data need to obey some mathematical distribution).

Based on abovementioned concepts, the proposed model first tests the lag of AR model. Secondly, input variables of AR are decomposed by EMD into several IMFs and a residue. Thirdly, the AR method and IMFs and volatility of data are then modeled and forecasted by SVR. Thus, the proposed model could be expected to solve ED overcrowding problems by higher accurate forecasting results for daily patient volumes.

The rest of this paper is organized in the following. Sec. 2 describes related methodologies; Sec. 3 presents briefly the proposed model; Sec. 4 describes experiments and comparisons; and Sec. 5 illustrates some findings of this paper. Finally, conclusions of the study are made in Sec.6.

METHODOLOGY

This section reviews related methodologies of the autoregressive model, empirical mode decomposition, and support vector regression.

-Autoregressive Model

In time series forecast, predictions are practically obtained by forecasting a value at the next time period based on a specifc prediction algorithm. In addition, forecasting non-periodic short-term time series is much more difficult than that for long-term time series. The autoregressive moving average (ARMA) is a traditional method that is very suitable for forecasting regular periodic data, such as seasonal or cyclical time series [7].

Box and Jenkins [6] developed a general linear stochastic model by assuming that time series data can be generated by a linear aggregation of random shocks. In this study, we focus on the AR model, which is a model that includes one or more past values of the dependent variable among its explanatory variables, and the simplest AR(1) is defined as:

$$y_t = \phi_1 y_{t-1} \qquad (1)$$

When the random error and the constant term are taken into account, the modified AR(1) model becomes

$$y_t = \mu + \phi_1 y_{t-1} + u_t \qquad (2)$$

where ϕ_1 is the first-order autoregression coefficient and u_t is the white noise, viewed as a random error. An autoregressive model is simply a linear regression of the current value of the series against one or more prior values of the series. In the AR(1) model, it can be thought of as that for a given value y in time period t that has a relationship with time period t–1. If there is an autoregressive model of order p, an AR(p) model can be expressed as

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + u_t \qquad (3)$$

-Empirical Mode Decomposition

The empirical mode decomposition (EMD) technique, proposed by Huang et al. [20], is a form of an adaptive time series decomposition technique using the Hilbert-Huang transform (HHT) for nonlinear and non-stationary time series data. Recently, EMD has been applied to forecasting models in many research fields [18, 24, 41]. The basic principle of EMD is to decompose a time series into a sum of oscillatory functions—namely, intrinsic mode functions (IMFs). In the EMD, the IMFs must satisfy two conditions: (1) the number of extrema (sum of maxima and minima) and the number of zero crossing differ only by 1, and (2) the local average is 0. The condition that the local average is 0 implies that the envelope mean of the upper envelope and lower envelope is equal to 0. The first condition is similar to the traditional narrow band requirements for a stationary Gaussian process [20]. The second condition modifies the classical global requirement to a local one; it is necessary so that the instantaneous frequency will not have the unwanted fluctuations induced by asymmetric wave forms [20]. The detail algorithm for EMD is shown as follows [20]:

Step 1: Identify local extrema in the experimental data $\{x(t)\}$. All the local maxima are connected by a cubic spline line $U(t)$, which forms the upper envelope of the data. Repeat the same procedure for the local minima to produce the lower envelope $L(t)$. Both envelopes will cover all the data between them. The mean of the upper envelope and lower envelope $m_1(t)$ is given by:

$$m_1(t) = U(t) + L(t)/2 \quad (4)$$

Subtracting the running mean $m_1(t)$ from the original time series $x(t)$, we get the first component $h_1(t)$:

$$h_1(t) = x(t) - m_1(t) \quad (5)$$

The resulting component $h_1(t)$ is an IMF if it is symmetric and has all maxima positive and all minima negative. An additional condition of intermittence can be imposed here to sift out waveforms with a certain range of intermittence for physical consideration. If $h_1(t)$ is not an IMF, the sifting process has to be repeated as many times as it is required to reduce the extracted signal to an IMF. In the subsequent sifting process steps, $h_1(t)$ is treated as the data to repeat the steps mentioned above:

$$h_{11}(t) = h_1(t) - m_{11}(t) \quad (6)$$

Again, if the function $h_{11}(t)$ does not yet satisfy criteria for IMF, the sifting process continues up to k times until some acceptable tolerance is reached:

$$h_{1k}(t) = h_{1_{(k-1)}}(t) - m_{1k}(t) \quad (7)$$

Step 2: If the resulting time series is an IMF, it is designated as

$c_1 = h_{1k}(t)$.

The first IMF is then subtracted from the original data, and the difference r_1 given by

$$r_1(t) = x(t) - c_1(t) \quad (8)$$

which is the residue. The residue $r_1(t)$ is taken as if it were the original data, and we apply to it again the sifting process of Step 1.

Following the procedures above, we continue the process to find more intrinsic modes ci until the last one. The final residue will be a constant or a monotonic function that represents the general trend of the time series. Finally, we obtain

$$\begin{aligned} x(t) &= \sum_{i=1}^{n} c_i(t) + r_n \\ r_{i-1}(t) - c_i(t) &= r_i(t) \end{aligned} \quad (9)$$

where r_n is a residue.

Thus, residue $r_n(t)$ is the mean trend of x(t) $x(t)$. The IMFs $c_1(t), c_2(t), ..., c_n(t)$ include different frequency bands ranging from high to low. The frequency components contained in each frequency band are different, and they change with the variation of signal $x(t)$, while $r_n(t)$ represents the central tendency of signal $x(t)$.

-Support Vector Regression

SVR is a popular artificial intelligence algorithm, and many recent studies have also applied SVR to forecasting models in many different research areas [10, 28, 44]. In this section, the basic SVR concepts are briefly described, which can also be found in [36]. Given a training set (x_i, y_i), i = 1,2,. . .,m $(x_i, y_i), i = 1, 2, ..., m,$ where the input variable $x_i \in R^n$ is the n-dimensional vector, and the response variable $y_i \in R^n$ is the continuous value. SVR builds the linear regression function as the following form:

$$f(x, w) = w^T x + b \quad (10)$$

Based on the Vapnik' s linear ε-Insensitivity loss (error) function (Equation (11)), the linear regression $f(x,w)$ is estimated by simultaneously minimizing $\| w \|^2$ and the sum of the linear ε-Insensitivity losses (Equation (13)). The constant C, which influences a trade-off between an approximation error and the weights vector norm ||w||, is a design parameter chosen by the user.

$$| y - f(x,w) |_{\varepsilon} = \begin{Bmatrix} 0, if \mid y - f(x,w) \leq \varepsilon \\ | y - f(x,w) | - \varepsilon \ \ otherwise \end{Bmatrix} \quad (11)$$

$$R = \frac{1}{2} \| w \|^2 + C \left(\sum_{i=1}^{m} | y_i - f(x_i, w) |_{\varepsilon} \right) \quad (12)$$

Minimizing the risk R is equivalent to minimizing the following risk:

$$R_{w,\zeta\zeta^*} = \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{m} (\zeta + \zeta^*) \quad (13)$$

under constraints:

$$(w^T x_i + b) - y_i \leq \varepsilon + \zeta_i \quad (14)$$

$$y_i - (w^T x_i + b) \leq \varepsilon + \zeta_i^* \quad (15)$$

$$\zeta_i, \zeta_i^* \geq 0, \ \ i = 1, 2, ..., m \quad (16)$$

where ζ_i and ζ_i^* are slack variables: one for exceeding the target value by more than e, and the other for being more than ε below the target. As with procedures applied to SVM classifiers [36], this constrained optimization problem is solved by applying the Lagrangian theory and the Karush Kuhn-Tucker condition to obtain the optimal desired weights vector of the regression function.

The non-linear SVR maps the training samples from the input vectors into a higher-dimensional feature space via a mapping function Φ. By performing such a mapping method, in the feature space, the learning algorithm will be able to obtain a linear regression function by applying the linear regression SVR formulation. In the final expression for a predictor function, training data only appear in the form of scalar products $x_i^T x_j$, which are replaced by scalar products $\Phi^T(x_i)\Phi(x_j)$. The scalar product $\Phi^T(x_i)\Phi(x_j)$ is calculated directly by computing a kernel function, $K(x_i, x_j)$ — that is, $\Phi^T(x_i)\Phi(x_j) = K(x_i, x_j)$ UT(xi)U(xj)=K(xi,xj)—to avoid having to perform a mapping $\Phi(x)$. The most popular kernel function is Radial Basis Function (RBF), as shown in Equation (17).

$$K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2) \quad (17)$$

Proposed Model

In this section, the datasets utilized by the proposed model are illustrated in 3.1. The proposed algorithm of this study is presented in 3.2.

-Data Source

The datasets used in this study are extracted from a hospital information system (HIS) of a regional hospital. There are three diverse divisions (internal medicine, surgery, pediatrics) in the emergency department, and all daily patient volumes (all datasets) of the emergency department include patients of three different divisions. Patients visit the pediatrics division just in a few days, and patients visit the division of internal medicine and division of surgery every day. Therefore, this study only can collect daily patient volumes in the internal medical division and surgery as two sub-datasets. This study denotes all patient volumes of ED as dataset I and indicates patient volumes of internal medical division and surgery division as dataset II and dataset II, respectively. Each experimental dataset includes 731 observations from July 2010 to June 2012; training data are selected from July 2010 to December 2011, and remaining data (from January 2012 to June 2012) are testing data. The detailed definitions of these datasets are given in Table 1.

	Dataset I	**Dataset II**	**Dataset III**
Dataset description	All patient volumes of ED	Patient volumes of internal medical division	Patient volumes of surgical division
Max value	159	128	41
Min value	18	5	4
Mean	42.98	23.94	18.9
Standard deviation	12.8	10.95	5.52
The number of training data	549	549	549
The number of testing data	182	182	182

Table 1. Comparisons of Experimental Datasets

Proposed Algorithm

Based on the research concepts in section 1, this paper proposes a hybrid time series model; it considers EMD, AR, volatility of data, and SVR to forecast daily patient volumes in the emergency department. This study first tests the lag of AR by statistical analysis and then uses EMD to decompose input variables of AR. Finally, it utilizes SVR to forecast patient number, which combines AR method, IMFs, and data volatility. Then, the overall flowchart of the proposed model is shown as Figure 1.

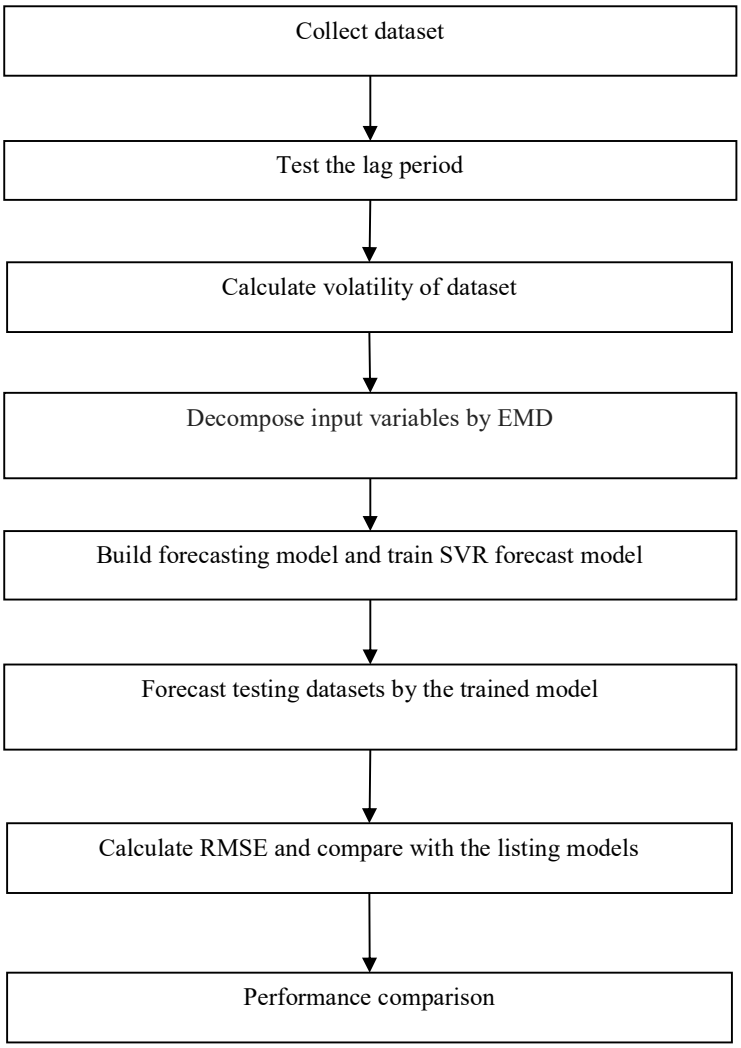

Figure 1. Flowchart of the Proposed Procedure

This section uses practically collected data (Dataset I) as the example step by step to show the core

concept of the proposed algorithm as follows.

Step1: Collect dataset

In this step, this study collects all patient volumes of ED (Dataset I) from July 2010 to June 2012 to illustrate the proposed model. The training dataset is selected from July 2010 to December 2011, and the remainder dataset (from January 2012 to June 2012) is used for testing

Step 2: Test the lag period

In this step, this paper uses E-Views software package to fit the AR model for different lags and orders of patient volumes (PV). In dataset I, five linear regression variables, i.e., from PV (t − 1) to PV (t − 5), are selected to be estimated and tested. If the p-value is less than the significant level of 0.05, then reject the null hypothesis. Take Dataset I as an example; Figure 2 shows that the p-value (0.0000) for PV (t− 1)) is less than the significant level of 0.05 among five variables, from PV (t -1) to PV (t − 5). Further, the variable PV (t -1) is not equal to 0. Therefore, the order of AR is 1.

Dependent Variable: PV
Method: Least Squares
Date: 06/17/17 Time: 20:44
Sample (adjusted): 6 731
Included observations: 726 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	34.94163	2.930068	11.92519	0.0000
PV(-1)	0.233577	0.037218	6.275975	0.0000
PV(-2)	0.023926	0.038066	0.628543	0.5298
PV(-3)	-0.024174	0.038080	-0.634816	0.5258
PV(-4)	0.000135	0.038079	0.003537	0.9972
PV(-5)	-0.046642	0.037120	-1.256506	0.2093

R-squared	0.059936	Mean dependent var	42.96281
Adjusted R-squared	0.053408	S.D. dependent var	12.79601
S.E. of regression	12.44962	Akaike info criterion	7.889487
Sum squared resid	111595.0	Schwarz criterion	7.927401
Log likelihood	-2857.884	Hannan-Quinn criter.	7.904118
F-statistic	9.181080	Durbin-Watson stat	1.984936
Prob(F-statistic)	0.000000		

Figure 2. Testing the Lag Period of PV in Dataset I

Step 3: Calculate volatility of dataset

In order to calculate the volatility of patient volumes (PV), this paper defines Vo(*t*) as one period of volatility of PV, and formula of Vo(*t*) is presented in equation (18). Then, this study uses equation (18) to calculate Vo(*t*) values for the proposed model.

The partial instances of Vo(*t*) during 2010/7/1 to 2010/7/31 are shown in Table 2

$$Vo(t) = PV(t) - PV(t-1) \quad (18)$$

Step 4: Decompose input variables by EMD

In step 2, this study tests the lag period of PV, and the results show that the order of AR is 1. Therefore, this paper uses the EMD to decompose the input variable (PV (t)) into a finite set of IMFs (the residual $r_{n+1}(t)$ is also considered as an IMF) to obtain interpretable information of (PV (t)). In this study, the PV (t) in Dataset I is decomposed into ten IMFs and one residue, which exhibits a stable and regular variation. This means that the interruption and coupling between the different characteristics information embedded in the original data have been weakened to an extent. Therefore, the forecasting model is easier to build.

Step 5: Build forecasting model and train SVR forecast model

In this step, this paper utilizes PV(*t*) and 11 IMFs (generated in Step 4) and Vo(*t*) (generated in Step 3) as input variables and uses PV(*t+1*)(PV value in next day) as the output variable. Then, this paper applies the SVR with the ε-insensitive loss function (ε-SVR) to build the forecasting model. Then, this study sets the radial basis function as the type of kernel function, degree in kernel function=3, the ε in loss function of ε-SVR=0.1, and gamma in kernel function=0. For each training step, this paper sets ε=0.001 for tolerance as the stopping criterion.

Step 6: Forecast testing datasets by the trained model

The ε-SVR parameters of the forecasting models are determined when the stopping criterion is reached from step 5; then, the training forecasting model is used to forecast PV(t+1) for the target testing datasets.

Step 7: Calculate RMSE and compare with the listing models

Calculate RMSE values in testing datasets by Equation (19). Then, the RMSE is taken as an evaluation criterion to compare with the listing models.

Date	PV(t+1)	PV(*t*)	Vo(*t*)
2010/7/1	36		
2010/7/2	40	36	
2010/7/3	43	40	4
2010/7/4	72	43	3
2010/7/5	38	72	29
2010/7/6	29	38	-34
2010/7/7	37	29	-9
2010/7/8	37	37	8
2010/7/9	27	37	0
2010/7/10	38	27	-10
2010/7/11	63	38	11
2010/7/12	45	63	25
2010/7/13	30	45	-18
2010/7/14	46	30	-15
2010/7/15	42	46	16
2010/7/16	36	42	-4
2010/7/17	59	36	-6
2010/7/18	62	59	23
2010/7/19	44	62	3
2010/7/20	47	44	-18
2010/7/21	40	47	3
2010/7/22	44	40	-7
2010/7/23	30	44	4
2010/7/24	42	30	-14
2010/7/25	47	42	12
2010/7/26	36	47	5
2010/7/27	29	36	-11
2010/7/28	29	29	-7
2010/7/29	46	29	0
2010/7/30	48	46	17
2010/7/31	44	48	2

Table 2 Partial Instances of Patient Volumes

$$\text{RMSE} = \sqrt{\sum_{t=1}^{n} | actual(t) - forecast(t) |^2 \Big/ n} \qquad (19)$$

where *actual(t)* denotes the real PV value, *forecast(t)* denotes the predicting PV value, and n · is the number of data.

Step 8: Performance comparison

RMSE values in testing datasets are calculated by equation (19). Then, the RMSE is taken as the evaluation criterion to compare with the listing models.

Experiments and Comparisons

This section provides accuracy evaluations and comparisons, and the RMSE is taken as the evaluation criterion. To verify the proposed model, the Datasets I, II, and III are used as the experiment datasets. In each dataset, data from July 2010 to December 2011 are used for training, and those from January 2012 to June 2012 are selected for testing. Further, this paper compares the performances of the proposed model with the traditional time series model (AR (1) model [13]), fuzzy time series model (Chen's model [9], Yu's model [43]), and ε-SVR [36] model. The AR-EMD-SVR model (exclude Vo(t) as input variable) is also a comparison model. This study tests the lag period of PV in Datasets II and II, and the orders of AR for two datasets are all 1. The numbers of decomposed IMFs in Datasets II and III are all 9.

The performances of the listing models above used to forecast PV are compared to the proposed model. The forecasting performances of the AR (1) model, Chen model, Yu model, ε-SVR model, AR-EMD-SVR model, and the proposed model are listed in the Table 3. From Table 3, it is clear that the proposed model surpasses the other five models in each dataset. These PV forecasting performance evaluations demonstrate the outstanding performance of the proposed model.

Models	Dataset		
	I	II	III
Chen's model	17.33	26.4	6.12
Yu's model	14.65	12.06	6.63
AR(1)	12.4	10.31	5.43
ε-SVR	13.02	11.05	5.45
AR-EMD-SVR	9.79	8.13	4.44
Proposed model	8.86 [a]	6.83 [a]	3.62 [a]

[a] The best performance among six models

Table 3. The Results of Different Models for Testing Data in RMSE

This paper uses a nonparametric statistical method, the Friedman test [16], to verify that the proposed model is superior to the other methods in three datasets. Using the data from Table 3, chi-square test is

used to test the hypothesis H_0: equal performance in three datasets. The results (p=0.012) with regard to this hypothesis, which rejects H_0 = 0, are listed in Table 4. Table 5 shows the mean rank by Friedman test, demonstrating that the proposed model (mean rank = 1) outperforms the other models. Based on Tables 4-5, the difference in performance is significant.

Parameter	
n	3
Chi-Square	14.619
df	5
Asymp. Sig.	0.012

Note: n is the number of data points; df denotes degrees of freedom.

Table 4. Results of Friedman Test

Models	Mean Rank
Chen	5.67
Yu	5.33
AR(1)	3.00
SVR	4.00
AR-EMD-SVR	2.00
Proposed model	1.00

Table 5. Mean Rank of Friedman Test

FINDINGS

A new model, based on AR-EMD and volatility of data joining to fusion SVR procedure, has been proposed to forecast patient volumes in Taiwan; furthermore, the proposed model is compared with five forecasting models— Chen's model Yu's model, AR(1) model, AR-EMD-SVR model, and ε-SVR model—to evaluate the performance of the proposed model. After verification and comparison, the proposed method outperforms the listing methods. From the experimental results, there are three findings in this paper as follows:

(1) The superiority of the hybrid model:

According to Table 3, it is evident that the hybrid models (AR-EMD-SVR model and proposed model) are superior to the single methods (Chen model, Yu model, AR model, SVR model) in terms of RMSE. The main reason is that the hybrid models take into account the AR-EMD method with SVR learning for PV forecasting, integrating the advantage of SVR, which has ability to avoid over-fitting and under-fitting problems.

(2) EMD strength:

From Table 3, it is shown that the performances of the proposed model and AR-EMD-SVR are better than the ε-SVR model. It is evident that EMD methods could decompose the noise raw data into simpler frequency components and highly correlating input variables and reduce forecasting error more effectively.

(3) Volatility of data for forecasting performance:

Table 3 reveals that the proposed model including the volatility variable performs better than the AR-EMD-SVR model, which does not consider volatility as input variable. This implies that volatility is an important variable for patient volumes forecasting to enhance forecasting performance.

CONCLUSION

The number of ED visits increased in recent years, and shifts in the supply of and demand for emergency department resources make the efficient allocation of ED resources increasingly important. For clinical research fields, scholars have proposed many models to forecast daily patient numbers in the emergency department. However, there are still drawbacks in previous forecasting models: (1) some datasets do not follow the statistical assumptions, and several traditional time series models can not be applied to those datasets; and (2) late day data are an input variable in most conventional time series models, but there are noises including in raw data.

To solve the shortage of previous conventional models, this paper has presented a new model by integrating AR, EMD (EMD can decompose raw data with noises into IMFs that have stronger correlations), and SVR (SVR can overcome the restriction that datasets should follow the statistical assumptions) and also considers the causality of volatility with patient volumes forecasting to improve model performance.

Experimental results showed that the proposed model appears to reliably and accurately forecast the number of patients admitted to the ED per day. The reason why the proposed model outperforms the listing models is that EMD, which can fully capture the local fluctuations of data, can be used as a preprocessor to denoise the original signal to grasp the general tendency of the dataset, which has simpler frequency components and high correlations. Further, the proposed model utilizes data volatility, which would affect the data trend in the next day as input variable, and premeditates the fluctuation of data to enhance forecasting performance

By this preprocessing, this study can not only advance the simplification of SVR modeling but also be more precise than listing models based on RMSE. Therefore, the proposed method is very suitable for prediction with nonlinear and strong noise data and is an efficient method for patient volumes forecasting. Moreover, the results of this paper are useful and viable for decision-makers and future scholars. We deem that hospital managers can utilize this forecasting model to discover useful knowledge with benefits in the medical research field.

In future work, a more practical patient volumes dataset can be collected as the experimental dataset to verify the robustness of the proposed model. Other traditional time series could be combined in the proposed model to enhance forecasting performance, and a rule-based data mining method could be utilized to generate useful decision rules for decision-makers.

REFERENCES

[1] An X., Jiang D., Zhao M., Liu C., (2012) Short-term prediction of wind power using EMD and chaotic theory, Commun. Nonlinear Sci. Numer. Simul. 17 1036-1042,

[2] Aneeshkumar A.S., Venkateswaran C. J., (2015) Reverse Sequential Covering Algorithm for Medical Data Mining, Procedia Computer Science, 47, 109-117,

[3] Arslan A. K., Colak C., Sarihan M. E. (2016) Different medical data mining approaches based prediction of ischemic stroke, Computer Methods and Programs in Biomedicine, 130, 87-92,

[4] Asplin BR, Flottemesch TJ, Gordon BR. (2006) Developing models for patient flow and daily surge capacity research. Acad Emerg Med; 13: 1109-1113.

[5] Bollerslev T., (1986) Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics. 31 307-327.

[6] Box G., Jenkins G., (1976) Time series analysis: Forecasting and control, San Francisco: Holden-Day.

[7] Chang B.R., (2008) Resolving the forecasting problems of overshoot and volatility clustering using ANFIS coupling nonlinear heteroscedasticity with quantum tuning. Fuzzy Sets and Systems, 159(23) 3183-3200.

[8] Chen B., Chang M., Lin C., (2001) Machines: A Study on EUNITE Competition 2001, in: IEEE Trans. Power Syst. , pp. 1821-1830.

[9] Chen S. M., (1996) Forecasting enrollments based on fuzzy time series, Fuzzy Sets Systems, 81, 311-319.

[10] Chen Y., Xu P., Chu Y., Li W., Wang K. (2017) Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings, Applied Energy, 195, 659-670

[11] Chen K. L., Yeh C. C., Lu T., (2012) Forecasting the output of Taiwan' s integrated circuit (IC) industry using empirical mode decomposition and support vector machines, Int. J. Phys. Sci. 3 (78) 5460-5467.

[12] Cheng C. H. , Wei L.Y., (2014) A novel time series model based on empirical mode decomposition for forecasting TAIEX, Economic Modelling, 136-141

[13] Engle R. F., (1982) Autoregressive conditional heteroscedasticity with estimator of the variance of United Kingdom inflation. Econometrica. 50(4) 987-1008.

[14] Feng F., Zhu D., Jiang P., Jiang H., (2010) GA-EMD-SVR condition prediction for a certain diesel engine, 2010 Progn. Syst. Heal. Manag. Conf. PHM ' 10,

[15] Fu C., (2010) Forecasting exchange rate with EMD-based Support Vector Regression, in: 2010 Int. Conf. Manag. Serv. Sci. MASS,

[16] Friedman M. (1937), The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Stat. Assoc.; 675-701.

[17] Georgio G., Guttmann A., Doan Q. H., (2017) Emergency Department Flow Measures for Adult and Pediatric Patients in British Columbia and Ontario: A Retrospective, Repeated Cross-Sectional Study, The Journal of Emergency Medicine, In press

[18] He K., Yu L., Tang L. (2015) Electricity price forecasting with a BED (Bivariate EMD Denoising) methodology, Energy, 91, 601-609,

[19] Hong W.C., Pai P.F., (2006) Predicting engine reliability by support vector machines, Int. J. Adv. Manufact. Technol. 28 (1-2) 154-161.

[20] Huang N.E., Shen Z., Long S.R., Wu M.C., Shih H.H., Zheng Q., (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis, in: Proceedings of the royal society of London series a-mathematical physical and engineering sciences, series A, 454 903-995.

[21] Huarng K. H., (2001) Effective lengths of intervals to improve forecasting in fuzzy time series, Fuzzy Sets and Systems. 123 155-162.

[22] Hwang S.W., Lee H.J., (2008) Development of a revisit prediction model for the outpatient in a hospital. J Korean Soc Med Inform; 14: 137-145.

[23] Jones S. S., Thomas A., Evans R. S., Welch S. J., Haug P. J., Snow G. L., (2008) Forecasting Daily Patient Volumes in the Emergency Department, ACADEMIC EMERGENCY MEDICINE; 15:159– 170

[24] Jun W., Lingyu T., Yuyan L., Peng G. (2017) A weighted EMD-based prediction model based on TOPSIS and feed forward neural network for noised time series, Knowledge-Based Systems, 132, (15) 167-178,

[25] Kim K., Han I., (2000) Genetic algorithms approach to feature discretization in artificial neural networks for prediction of stock index. Expert System with Applications. 19 125-132.

[26] Kim M. J., & Min S. H., & Han I., (2006) An evolutionary approach to the combination of multiple classifiers to predict a stock price index, Expert Systems with Applications 31 241-247

[27] Lai M. C., Yeh C. C., (2013) A hybrid model by empirical mode decomposition and support vector regression for tourist arrivals forecasting, J. Test. Eval. 41.

[28] Liu Y., Wang R.X., (2016) Study on network traffic forecast model of SVR optimized by GAFSA,Chaos, Solitons & Fractals, 89, 153-159

[29] Mohammadi K., Shamshirband S., Tong C.W., Arif M., Petkovic D., Sudheer C., (2015) A new hybrid support vector machine-wavelet transform approach for estimation of horizontal global solar radiation, Energy Convers. Manag. 92 162-171,

[30] Ren Y., Suganthan P.N., Srikanth N., (2014) A novel empirical mode decomposition with support vector regression for wind speed forecasting, IEEE Trans. Neural Networks Learn. Syst. 1-6.

[31] Roh T. H., (2007) Forecasting the volatility of stock price index. Expert Systems with Applications. 33 916-922.

[32] Schweigler L.M., Desmond J.S., McCarthy M.L., Bukowski K.J., Ionides E.L., Younger JG. (2009) Forecasting models of emergency department crowding. Acad Emerg Med; 16: 301-308.

[33] Song Q., Chissom B.S., (1993) Forecasting enrollments with fuzzy time series Part I, Fuzzy Sets and Systems. 54 1-10.

[34] Tanaka-Yamawaki, M. & Tokuoka, S. (2007). Adaptive use of technical indicators for the prediction of intra-day stock prices, Physica A 383 125-133

[35] Turan A. H., Palvia P. C. (2014) Critical information technology issues in Turkish healthcare, Information & Management, 51, (1), 57-68,

[36] Vapnik, V. (1995). The nature of statistical learning theory. New York: Springer-Verlag.

[37] Vincent H.T., Hu S.-L.J., Hou Z., (1999) Damage detection using empirical mode decomposition method and a comparison with wavelet analysis, in: Proceedings of the second international workshop on structural health monitoring, Stanford 891-900.

[38] Wei L.Y., (2013) A GA-weighted ANFIS model based on multiple stock market volatility causality for TAIEX forecasting , Applied Soft Computing, 13 911-920

[39] Wei L.Y., (2016)A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting, Applied Soft Computing, 368-376

[40] Yang J.J., Li J., Mulder J., Wang Y., Pan H. (2015) Emerging information technologies for enhanced healthcare, Computers in Industry, 69, 3-11,

[41] Yi S., Guo K., Chen Z., (2016) Forecasting China's Service Outsourcing Development with an EMD-VAR-SVR Ensemble Method, Procedia Computer Science, 91, 392-401,

[42] Yu D.J., Cheng J.S., Yang Y., (2005) Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, Mech. Syst. Signal Process. 19 (2) 259-270.

[43] Yu H.K., (2005) Weighted fuzzy time series models for TAIEX forecasting, Physica A. 349 609-624.

[44] Zhang Z., Gao G., Tian Y., Yue J., (2016) Two-phase multi-kernel LP-SVR for feature sparsification and forecasting, Neurocomputing, 214, (19), 594-606.