A Hybrid Approach to Time Series Forecasting Based on the Support Vector machine algorithm and Feature Extraction Using Particle Swarm Optimization (PSO)

Zana Azeez Kakarash[a], Rizhin Nuree Othman[b], Nawroz Fadhil Ahmed[c]
[a] *Department of Engineering, Faculty of Engineering and Computer Science, Qaiwan International University, Sulaymaniyah, Iraq*
[b] *Department of Information Technology, Kurdistan Technical Institute, Sulaymaniyah, Kurdistan Region, Iraq*
[c] *College of Computer Science and Engineering, Lebanese French University, Erbil, Kurdistan Region, Iraq*

Abstract: In recent years, due to the non-linear nature, complexity, and irregularity of time series, especially in energy consumption and climate, studying this field has become very important. Therefore, this study aims to provide a high accuracy and efficiency hybrid approach to time series forecasting. The proposed model is called EMD-DWT-FPSO-SVR (Support Vector Regression- Feature selection with Particle Swarm Optimization-Discrete Wavelet Transform-Empirical Mode Decomposition). In the proposed hybrid approach, the first step is to decompose the signal into the IMF component using the MED algorithm. In the second step, each component is transformed into subsequences of approximation properties and details by converting the Wavelets. In the third step, the best feature is extracted by the PSO algorithm. The purpose of using the PSO algorithm is feature extraction and error minimization of the proposed approach. The fourth step, using time vector regression, has dealt with time series forecasting. Four data sets in two different fields have been used to evaluate the proposed method. The two datasets are Electric Charge of England and Poland, and the other two datasets are related to the temperature of Australia and Belgium. Evaluation criteria include MSE, RMSE, MAPE, and MAE. The evaluation results of the proposed method with other PCA feature extraction algorithms, and comparisons with methods and studies in this field, indicate the proper performance of the proposed approach.

Keywords: Time Series, Support Vector Regression, Feature Extraction, Particle Swarm Optimization

1. Introduction

Time series is a sequence of values (or observations) of an event in which values have a definite period, such as hourly, daily, weekly, monthly, yearly, and so on [1], and time series analysis deals with dynamic nature and real data [2]. Time series forecasting is commonly used in a wide range of fields for future decision making and planning [1]. One of its most widely used fields is energy. The need for energy and related services is increasing to meet human social and economic growth, welfare, and health [3] and depends on factors such as temperature, sunlight, wind, humidity, etc. [4]. Energy is very important in many areas of life. In addition, in recent decades, humans have been almost entirely dependent on energy, especially electrical energy. Although there have been many efforts to improve energy consumption in electronic devices, many emerging devices still rely on some kind of electric power [5]. The use of high-powered electrical devices and the development of technologies such as smart grids, electric cars, and renewable energy production have also expanded. All of these factors make it difficult to manage the power system [6], and energy companies are responsible to supply enough energy to consumers [5]. Moreover, ensuring energy supply and controlling it in climate variability, is one of the important energy challenges for a sustainable future [3]. Evidence demonstrates that energy consumption has a direct impact on the climate and has caused the production of greenhouse gases, which resulting in global warming [7]. Thus, electric charge, as a clean and efficient source, plays a significant role in human daily life and has recently increased dramatically, therefore, it has become a fundamental issue. In addition, electricity is more suitable and efficient than other traditional energy sources, such as natural gas, coal, and petroleum, to meet the needs of an environmentally friendly community. Therefore, before making a decision on electricity generation, it is necessary to predict electricity needs and loads [6].

Electric load forecasting helps suppliers to adjust supply and demand and ensure power grids in the event of electricity shortages. Load forecasting, to achieve different goals, is divided into several categories: short-term load forecasting (a few minutes to 1 day in advance) to adjust supply and demand, medium-term load forecasting (1 day to 1 year in advance) to definitive planning of the outage and maintenance and long-term load forecasting (more than 1 year in advance), for planning the development of electrical infrastructure [8]. Accurate short-term load forecasting is essential for effective performance in the electrical sector. Load forecasting in personal homes or buildings is challenging due to more fluctuations and uncertainties in load consumption [9]. Incorrect load forecasting causes significant financial loss. Therefore, correct forecasting using a good model leads to energy savings [6].

On the other hand, the dynamic and non-linear nature of the climate, due to variability in temperature and precipitation, has become an attractive field in time series forecasting [10]; because climate data is part of an uncertain time series [11] and socio-economic activities, and other human activities, in most countries, depend on climate parameters variabilities, such as temperature, humidity and wind speed, sunlight and wind direction [12]. Therefore, this study aims to forecast short-term load and air temperature using the proposed MED-DWT-FPSO-SVR method, which can be effective in both fields.

2. Related Works

The conducted studies are reviewed in three categories: forecasting electrical charge, climate (temperature), and optimization. The conducted works in the field of electric charge forecasting are:

In the article [8], the goal is short-term load forecasting. The present paper uses four machine learning methods, SVR[1], GBRT[2], FFNN[3] and LSTM[4], to forecast peak daily consumption and hourly energy in home buildings. In this research, the hourly load data set of England and London has been used; and also the electricity consumption of 15 houses was randomly selected and divided into 5 groups. This dataset fluctuates during peak hours and on weekends. The evaluation results show that the performance of the LSTM method is better in forecasting daily peak load consumption than other methods. Article [13], has dealt with providing a solution for renewable energy and reducing carbon emissions. This paper has assessed long-term load forecasting using the LSTM method and dynamic filtering of the potential highest electricity demand peaks, and to evaluate this method, a case study has been conducted on the UK building management system. The article [14] has dealt with seasonal load forecasting using the highest peak power consumption, by LSTM in Bangladesh. Article [15] investigates short-term load forecasting using LSTM and RNN methods for airline data. In this paper [16], the aim is short-term load forecasting using the hybrid method of signal decomposition algorithm to intrinsic components and support vector regression. The proposed method is used to analyze the intrinsic components as a noise reduction step in the training data, and then the SVR is used for forecasting. The evaluation is based on the Polish electric charge dataset, and the proposed method is EMD-SVR. The results show that the proposed algorithm of SVR and denoised-SVR algorithms have less error in forecasting electric charge. The paper [17] deals with load forecasting using a combination of support vector regression and LSTM. The data set used is collected from Iran's Power Generation and Distribution Company. In this research, the temperature features have also been used using the load features. Findings show there are good results in 24-hour forecasting.

Practices in the field of weather forecasting include:

The paper [10] analyzes the monthly temperature and precipitation rate by correlation and regression between indices using SPSS software in India and Uttarakhand. The research results can, as a forecasting tool, help to develop better methods of climate management in the region. According to the article [18], forecasting irregular time series has become one of the good research fields in recent years. Due to the complexity and irregularity of this type of time series, a hybrid model is needed for accurate forecasts. In this paper, a new hybrid model based on a Firefly Algorithm Optimized Neural Network called CEEMDAN – VMD – FABP is used. First, the signal is decomposed into primary components using variational mode decomposition (VMD), and Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). Then, the time series is forecasted by a firefly optimized backpropagation neural network (FABP). The dataset used in this article has used the temperature of Melbourne, Australia. The proposed model performs better than the VMD-FABP model and has a good outlook for forecasting irregular time series and lead to good results. The paper [19] has studied the time series forecasting for energy consumption. Temperature and humidity data sets were collected from sensors on a wireless network near an airport station in Belgium. In this paper, the following four models are used for time series forecasting: multiple linear regressions, support vector machine (SVM), random forest, and gradient boosting machines (GBM). According to the results, the performance of the GBM model has been better than the others. The paper [20] has dealt with available mixed time series forecasting in nature, such as climate with unpredictable and complex features, using a combined method called EMD-ARIMA-NN-FFOTR based on EMD and neural network. First, the intrinsic component is extracted using the EMD algorithm, then, using the run test, it is divided into three categories of high, low, and medium frequencies. Next, try to forecast the high-frequency signals using fuzzy first-order transition rules trained neural network (NN-FFOTR), and applied to the medium frequency of the ARIMA model. A simulated and unreal data set with time series, Lorenz-63 and Mackey-Glass systems were used to evaluate this approach. The results indicate that the hybrid method has good convergence speed and accuracy.

In this paper [5], the aim is to provide a solution to the problem of electric charge forecasting using conventional feedforward neural network optimized by a PSO algorithm. In this method, the PSO algorithm is used to optimize the neural network weight. The data set used is the electricity consumption of a company in Cyprus. The network is optimized, the average MAE and MSE error is reduced, and it converges faster, and the data training is faster using this method. According to the article [21], a large part of electricity consumption is used in the production sector. Electric charge forecasting can be useful in managing power consumption, scheduling the optimal generation, and planning for electricity maintenance. This can improve energy efficiency, and reduce production costs. In this paper, a short-term model of electrical charge forecasting is presented based on the GA-PSO-BPNN hybrid algorithm. The GA-PSO algorithm is used to optimize BPNN parameters. The data set used is the power consumption of a selected paper mill. According to the results, the proposed model can have good performance. The article [22] suggests that short-term electrical charge forecasting plays an essential role in the performance of power systems. This paper has presented a hybrid wavelet transform-based method, the GM algorithm, whose parameters are optimized by the PSO algorithm. The data set is the daily load of Iran and New York. In the proposed model, climate data including average temperature, average relative humidity, average wind speed, and load data of the past days are considered as model input. Wavelet transform is used to eliminate high frequency. The simulation results confirm the optimal performance of the proposed method compared to the previous one. According to the article [23], short-term electric charge forecasting plays an essential role in the energy management system. In this paper, there is a short-term load forecasting method, based on particle swarm optimization, and Deep belief Network DBN (PSO-DBN). The PSO

[1] Support Vector Regression (SVR)

[2] Gradient Boosting Regression Trees (GBRT)

[3] feedforward neural networks (FFNNs)

[4] Long-Short Term Memory Network (LSTM)

algorithm has been used for the initial network values and the reduction of iterations. The simulation shows the effectiveness of the proposed method. The article [24] has addressed short-term gas load forecasting using a hybrid model. The model is CF-SA-FFOA-SVM, a combination of Cross Factor (CF) algorithm, Simulated Annealing Algorithm (SA), Fruit Fly Optimization Algorithm (FFOA), and SVM. The optimization algorithm is used to find the best SVM parameters. The data set is collected from the produced temperature of China's municipal gas consumption. The results of the proposed model are compared with four other methods, such as BPNN, which indicate the proper performance of the proposed model. The paper [25] has investigated the short-term electric charge forecasting for a smart city using a hybrid method called LWSVR-MGOA, which is a combination of the Locally Weighted Support Vector Regression (LWSVR) algorithm and Modified Grasshopper Optimization Algorithm (MGOA). The MGOA algorithm is used to adjust the LWSVR parameters. The evaluation of the proposed method was performed on six real datasets of New York (hourly and daily) Victoria, EKPC, GEFCom, and ISO New England, which indicate the good performance of the proposed method.

Considering previous works, the PSO algorithm and other optimization algorithms can be used to optimize the parameters of models such as neural networks, SVR, and SVM, and there is no study on this algorithm in feature extraction in time series forecasting. This has led to a new and innovative aspect of the proposed EMD-DWT-FPSO-SVR method in this field. The proposed method is compared with other datasets in the comparison section with other tasks.

3. Proposed Research Method

Figure 1 shows the process of the proposed method called EMD-DWT-FPSO-SVR.

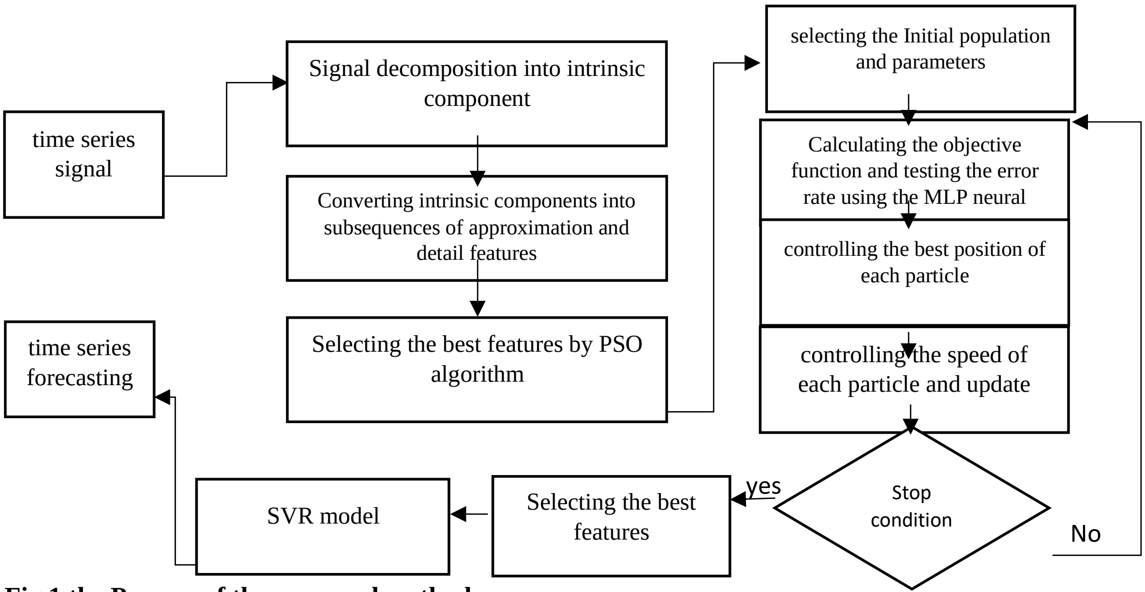

Fig.1 the Process of the proposed method

The first step is decomposing the signal into the IMF component using the MED algorithm. In the second stage, each component is transformed into a subsequence of approximation and detail features by converting the Haar wavelet. In the third step, the best feature is extracted by the PSO algorithm. A suitable objective function is selected for feature extraction by the PSO algorithm to minimize model error. Therefore, the error of this model should be calculated according to a classification algorithm, and then, in each iteration, should be by the PSO algorithm, and these steps continue until the stop condition. MLP neural network is considered to this end. The purpose of the PSO algorithm is to select the appropriate feature to minimize the error of the proposed method. The fourth step, using support vector regression tries to forecast time series.

3.1 EMD Algorithm (Empirical Mode Decomposition)

Component decomposition into intrinsic signals is an efficient nonlinear analytical method for time series data. Nonlinear and nonconstant time series can be divided into a group of mean and quasi-periodic signals, each of which is called the Intrinsic Mode Functions (IMF) [16]. In the EMD algorithm, the main signal is the sum of the IMFs plus remainder, as shown in Equation 1.

$$x(t) = \sum_{i=1}^{n} h_i(t) + r(t) \qquad (1)$$

In the above equation, $x(t)$is the main data, each of h_i^k indicate the i-th value of the IMF and $r(t)$is the remainder, and n is the number of IMFs.

3.2 Discrete Wavelet Transform (DWT)

Wavelet Transform is the decomposition of function based on wavelet functions. In discrete wavelet transform, the signals can be represented by approximation and detail [26]. The discrete form of the wavelet function is as equation 2.

$$\Psi_{j.k}(t) = \frac{1}{\sqrt{|s_0^j|}}\Psi(\frac{t - k\tau_0 s_0^j}{s_0^j}) \qquad (2)$$

Where $\Psi_{j.k}$ is the wavelet function, for certain values of k and j (integers), t, time (s> 1) is a constant of dilation parameter, τ_0 is a time transfer constant, and is dependent on s_0 .

3.3 Particle Swarm Optimization Algorithm (PSO)

The PSO algorithm is a universal optimization method, which can be used to deal with problems whose answer is a point or surface in n-dimensional space. In such a space, there are hypotheses and an initial velocity is assigned to the particles. Each particle has a position, which determines the particle coordinates in the multidimensional search space. As the particle moves over time, its position changes. $X_i(t)$ specifies the position of the i-th particle at time t. Moreover, each particle needs a velocity to move in space. $V_i(t)$ specifies the velocity of the i-th particle at time t. Accelerating the position of each particle, cause a new position for the particle. Equation 3 indicates the particle position update.

$$X_i(t+1) = X_i(t) + V_i(t) \qquad (3)$$

Whether the position of a particle in the search space is good or not is assessed by a fitness function. Particles are able to remember the best situation they have ever been in. The best individual experience of a particle, or the best position met by the $y_i$particle is called $pbest$[5] in some algorithms), and the particles can also know the best position met by the whole group, which is called $\widehat{y_i}$. (In some algorithms $\widehat{y_i}$ is also referred to as $gbest$[6].) The particle velocity vector, in the optimization process, reflects the empirical knowledge of the particle, its information, and the particle population information. Each particle considers two components to move in the search space: 1- Cognition Component, which is the best solution that a particle can get alone ($pbest$). 2- Social Component which is the best solution that is recognized by the whole group ($gbest$), was mentioned in the previous section.

According to $gbest$ and $pbest$, each particle uses equations 4 to 6 to determine the next position, and is updated:

$$V_{ij}(t+1) = WV_{ij}(t) + c_1 r_1 \big(p_{ij}(t) - X_{ij}\ (t)\big) + c_2 r_2 \big(p_{ij}(t) - X_{ij}\ (t)\big) \qquad (4)$$

$$X_{ij}(t+1) = X_{ij}(t) + V_{ij}(t+1) \qquad (5)$$

$$W(t+1)\ = w_{max} - (\frac{w_{max} - w_{min}}{t_{max}}).(t+1) \qquad (6)$$

In these equations, i is the particle subscript. The constants $c_1$and $c_2$in the above equations determine the personal and global learning parameters (impact rate) for $pbest$ and $gbest$. $r_1$and $r_2$are random numbers in the range [1 and 0]. $X_{ij}(t)$ and $V_{ij}(t)$ are the current positions and the velocity of the particles respectively. W is a parameter that controls the Particle motion inertia, which at the beginning of the algorithm decreases more rapidly, and after a while, decreases more slowly as we get closer to the answer [5].

For the simulation and steps of the PSO algorithm, first, the initial particle population is randomly generated, and the number of initial particles and other parameters is defined. The next step of evaluation calculates the particle's objective function (calculation of cost or viability), which the function considered in this study, is calculated by Equation 7.

[5] previous best position
[6] global best position

$$f = \min(\frac{1}{n}\sum_{i=1}^{N} e_i^2) \qquad (7)$$

The next step is recording the best position for each particle ($P_{i\,.best}$), and the best position among all the particles ($P_{g\,.best}$), and its updating, until the end of stop condition.

3.4 Support vector regression

The goal is the regression of the backup vector based on the SVM regression model, which is adapted from the regression tasks and classification. The support vector machine is a type of supervised learning system that is used both for grouping and estimating the data fitting function in regression problems so that the least error occurs in the data grouping or fitting function. This method is based on statistical learning theory, which takes advantage of the principle of structural error minimization and leads to a global optimum answer (Imani, 2019 et al., 2013). To implement support vector regression, the data is shown according to Equation 6.

$$SVR = \{x_i, t_i\} \quad \forall\ x_i \in R^m\ , t_i \in R \qquad (8)$$

Where x_i's are inputs that can have m dimensions and t_i is the target. From Equation 7 we can define SVR according to regression.

$$t_i \approx y_i = w^T x_i + b \quad \forall\ i = 1,2,\ldots\ldots,N \qquad (9)$$

Its penalty function is defined according to Equation 8.

$$L_\varepsilon(t_i, y_i) = \begin{cases} 0 & |t_i - y_i| \le \varepsilon \\ R^+ & other \end{cases} \qquad (10)$$

Where L_ε is the penalty function, and such that the desired output should be defined between the positive and negative interval of ε equal to $\aleph_i = |t_i - y_i| - \varepsilon$ that for y_i the desired output of the network, $\aleph_i$ the error due to the target and the output should be less than ε. Eventually, for the penalty function we will have:

$$L_\varepsilon(t_i, y_i) = \begin{cases} 0 & |t_i - y_i| \le \varepsilon \\ |t_i - y_i| - \varepsilon & other \end{cases} \qquad (11)$$

for all data with the operational risk, it should be minimized according to Equation 9. Therefore, in general, the target will be achieved according to Equation 10.

$$\min \quad \frac{1}{2}\ w^T w + C\sum_{i=1}^{N}(\aleph_i^+ + \aleph_i^-) \qquad (12)$$

In the above equation, the value of C is a constant number. Now if for the above equations we consider its dual form, for $-t_i + y_i + \varepsilon + \aleph_i^+ \ge 0$ the coefficient α_i^+, for $t_i - y_i + \varepsilon + \aleph_i^+ \ge 0$ the coefficient α_i^-, for $\aleph_i^+ \ge 0$ the coefficient $\alpha\mu_i^+$, and for $\aleph_i^- \ge 0$ the coefficient μ_i^- is replaced. If in the above equation the derivation operation is taken with respect to weight, bias, we will finally have equation 11 for the dual target function:

$$min \quad \frac{1}{2}\sum_i\sum_j(\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-)\, x_i^T x_i - \sum\nolimits_i(\alpha_i^+ - \alpha_i^-)\, t_i + \sum\nolimits_i(\alpha_i^+ - \alpha_i^-)\, \varepsilon \qquad (13)$$

$$s.t. \sum\nolimits_i(\alpha_i^+ - \alpha_i^-) = 0 \quad , \ 0 \le \alpha_i^+ \le C\ , \ 0 \le\ \alpha_i^- \le C \qquad (14)$$

The sum of the support vector can also be calculated using the value of α_i^+ and α_i^+, which the product of this value must be zero. Finally, if $\{x_i, t_i\}$is the input $\forall\ i = 1,2,\ldots\ldots,N$and the output is equal $y_i = w^T x_i + b$, we will have:

$$w = \sum_i(\alpha_i^+ - \alpha_i^-)x_i \qquad (15)$$

$$b = \frac{1}{|s|}\sum_{i\in s}[t_i - w^T x_i - Sing\ (\alpha_i^+ - \alpha_i^-)\varepsilon]\ (16)$$

In the support vector regression, a kernel function is used where if it is replaced in the relation $y = wx + b$ equal to $\sum_i(\alpha_i^+ - \alpha_i^-)\, x_i^T x_i + b$, Equation 14 will be formed.

$$y = \sum\nolimits_i (\alpha_i^+ - \alpha_i^-)\, K(x_i, x) + b \qquad (17)$$

Linear, sigmoid, polynomial, and radial (RBF) basis functions are the most common kernels. Due to the computational efficiency of RBF over the years, this kernel is known as one of the best kernels and in this study, this kernel that is shown in Equation 15 has been used as well.

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \qquad (18)$$

In the above equation, σ is the width of RBF. Therefore, since the kernel function is nonlinear, we have Equation 16.

$$b = \frac{1}{|s|}\sum_{i \in s}[t_i - \sum\nolimits_i (\alpha_i^+ - \alpha_i^-)\, K(x_i, x) - Sing\,(\alpha_i^+ - \alpha_i^-)\varepsilon] \qquad (19)$$

4. Evaluation and Results

In this paper, the simulation environment using MATLAB R2018b software has been used for modeling, and the computer for simulation has an Intel i3-2350 series CPU and 4 GB of memory and 500 GB of hard drive.

4.1 Data Set

In different studies in the field of time series forecasting, depending on the proposed method and the purpose of the problem, different data sets belonging to various countries have been used. The data set used to evaluate the proposed method is presented in Table 1.

Table 1 the data set used to evaluate the proposed method

Title	Type	Number of sample data	The maximum value	The minimum value	The unit
London [28]	Daily electric charge of the London, collected until 2014	1000	1.04	0.06	KW/h
Poland [29]	Daily electric charge of Poland, until 1990	1400	1.34	0.61	MB
Australia [30]	Minimum daily temperature of Melbourne, Australia, until 1990	520	26.29	2.11	Santigrad
Belgium [31]	Data collected by the ZigBee Wireless Sensor Network from the 2017 Airport Meteorological Station	1002	23.78	18.39	Hourly

The first data set belongs to the United Kingdom and the City of London, which includes daily loads collected up to 2014. Figure 2 shows this dataset.

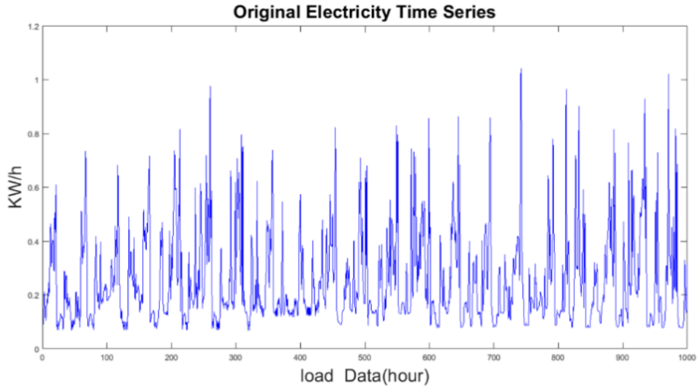

Fig. 2 the London daily load data set

The second data set belongs to Poland, which includes daily the daily hourly load collected up to 1990. Figure 3 shows this dataset.

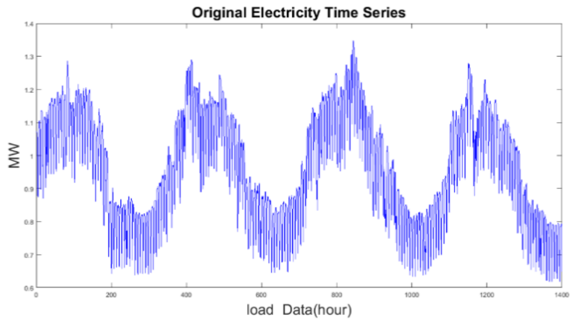

Fig. 3 the Poland daily load data set

The third dataset belongs to Australia and the city of Melbourne, which includes daily temperature of Meblevorn, which was collected from 1981 to 1990. Figure 4 shows this dataset.

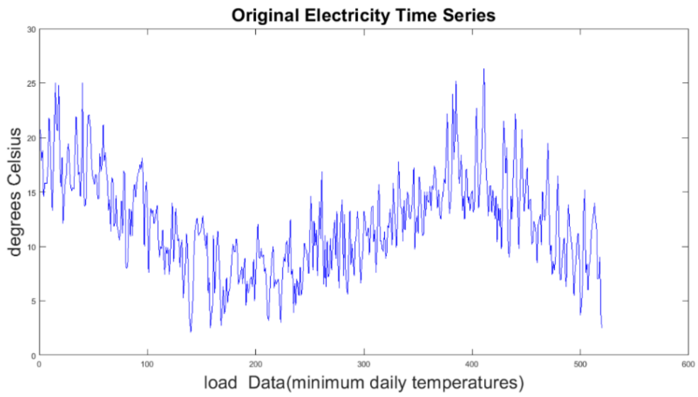

Fig. 4 Melbourne, Australia, Temperature Database

The fourth data set belongs to Belgium. This data set, including temperature-controlled by a ZigBee wireless sensor network. Each wireless node transmits temperature and humidity in about 3.3 minutes. This data set has been collected until 2017. Figure 5 shows this dataset.

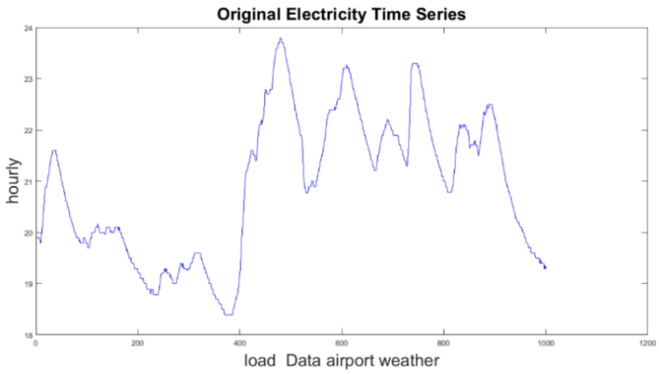

Fig. 5 Belgium Temperature data set

4.2 Evaluation criteria

All studies in the field of time series forecasting have used different criteria, and the criteria used in all articles have been used in this study too. The evaluation criteria of the proposed model are given in equations to 27 to 30.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_i - A_i)^2 \qquad (27)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - A_i)^2} \qquad (28)$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|(P_i - A_i)|}{P_i} \qquad (29)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(P_i - A_i)| \qquad (30)$$

In the above equations, $A_i$is the real value, $P_i$is the forecasted value, n is the number of samples.

4.3 Efficiency Results

To evaluate the proposed method, first, all the steps are described on the first data set, and then, in the end, the results of the proposed method are stated on the other data set. The first data set is the London daily electric load. First, the input signal is extracted using the EMD algorithm and its IMF components and is shown in Figure 6.

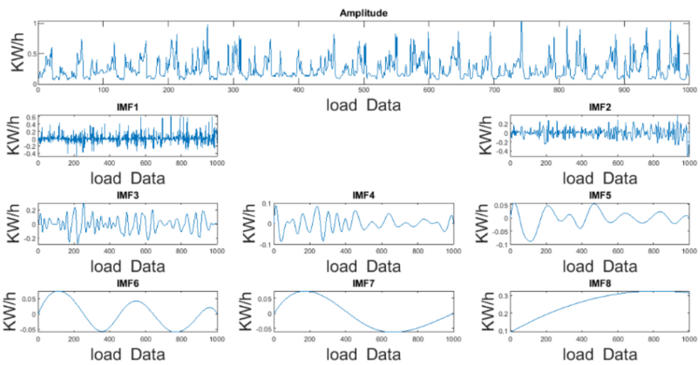

Fig.6 Signal decomposition using EMD algorithm to IMF components

Second, each decomposed component is transformed into a subsequence of approximations and details using depth 5 haar discrete wavelet transform, and examples of which are given in Figures 7, 8, and 9.

Fig. 7 subsequence of approximation and details of the first IMF

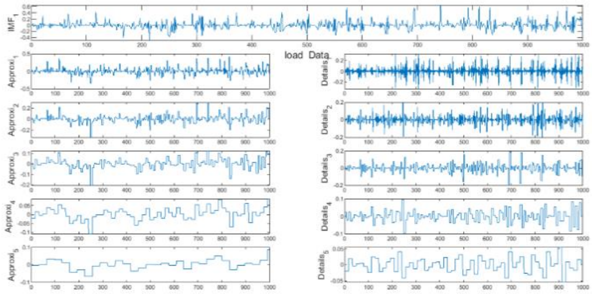

Fig. 7 subsequence of approximation and details of the first IMF

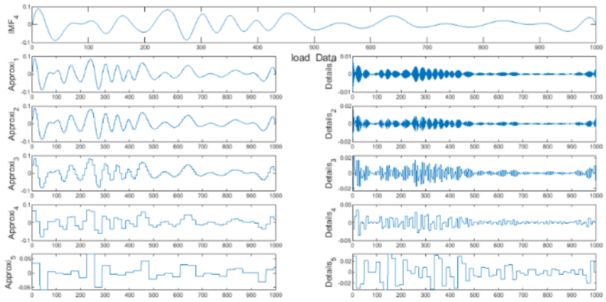

Fig. 8 subsequence of approximation and details of the fourth IMF

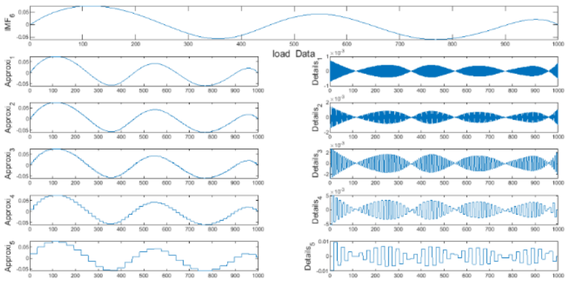

Fig. 9 subsequence of approximation and details of the sixth IMF

The output of the second stage contains 80 signal features from the subsequence of approximation and details. The third step is feature extraction using Particle Swarm Optimization(PSO). The required parameters of the PSO algorithm (initial particle population, selection of initial particles, and coefficients) are given in Table 2.

Table 2 the Parameters of PSO

Simulation and PSO parameters		
PSO algorithm	Initial population	30
	Iteration	10
	φ_1‘φ_2	2.05
	φ	$\varphi_1+\varphi_2$
	ρ	2/(φ-2+sqrt(φ^2-4*φ))
	w (inertia coefficient)	ρ

c_1 (personal learning coefficients)	$\rho*\varphi_1$
c_2(Global learning coefficients)	$\rho*\varphi_2$

Selecting the appropriate cost function is one of the most important parts of the optimization algorithm. The cost function for selecting a feature is to find the least possible error between the features and the purpose of the problem. So, by minimizing the model error (mean squared error), we can extract the best features which reduce problem error based on the order of its effect. Therefore, first, the error of this model must be calculated according to a classification algorithm, and then, each iteration is minimized by the PSO algorithm, and these steps continue until the stop condition. To this end, MLP neural network has been used to calculate the error. The cost function of this research is calculated according to Equation 7.

$$f = \min(\frac{1}{n}\sum_{i=1}^{N} e_i^2) \qquad (7)$$

Figure 10 shows the mean squared error minimized by the PSO algorithm for feature selection.

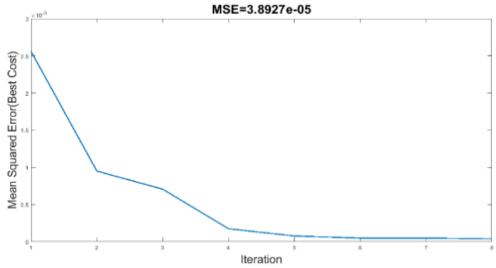

Figure 10 Mean square error with PSO algorithm for feature selection

According to Figure 10, the PSO algorithm shows the best cost function, and the least possible error of the proposed model, with 20 features equal to 3.8927e-05 in 10 iterations. Therefore, the output of the third step is the selection of 20 features of the signal, which is extracted by the PSO algorithm. The fourth step is time series forecasting by SVR. Figure 11 The red data is related to time series forecasting of electric charge and black color is related to the original signal.

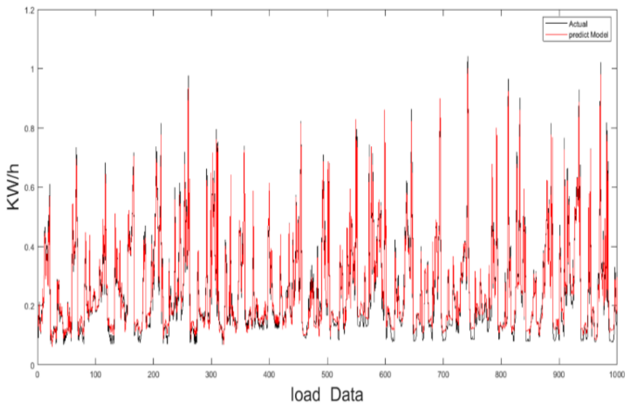

Fig. 11 Time series forecasting data, plotted on each other, related to the London data set

Figure 12 shows the load prediction signal output and original signal separately.

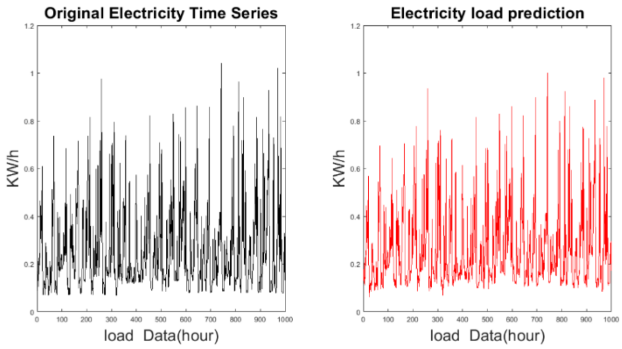

Fig. 12 Separated time series forecasting data for the London dataset

Table 3 shows the forecasted error value of the proposed method on the London dataset.

Table 3 forecasted error values of the proposed method with the London Electric Load dataset

Error	*MSE*	*RMSE*	*MAPE*	*MAE*
EMD-DWT-FPSO-SVR	6.*3166e*-04	0.0251	12.5721	0.0058

To show the efficiency of the PSO algorithm, its feature selection should be compared with the PCA and EMD-SVR algorithms (without feature selection), which Figure 13 shows the output of this comparison.

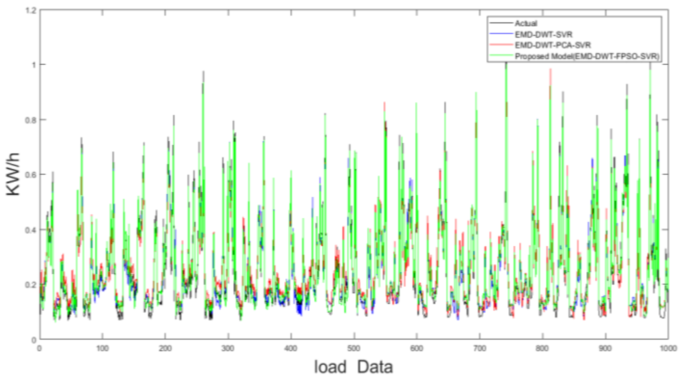

Fig.13 Output Comparison of the proposed model with other algorithms

Table 4 shows a comparison of the forecasted error values of the proposed method with other methods on the London dataset.

Table 4 Comparison of the forecasted error value of the proposed method with other methods on the London Electric Load Database

Error	*MSE*	*RMSE*	*MAPE*	*MAE*
EMD-DWT-SVR	8.6297e-04	0.0294	15.2977	0.0039
EMD-DWT-PCA-SVR	9.5408e-04	0.0309	16.3081	0.0117
EMD-DWT-FPSO-SVR	6.3166e-04	0.0251	12.5721	0.0058

Figure 14 shows the time series forecasting of the proposed method with other methods and Table 5 shows the comparison of the forecasted error value of the proposed method with other methods on the Polish data set.

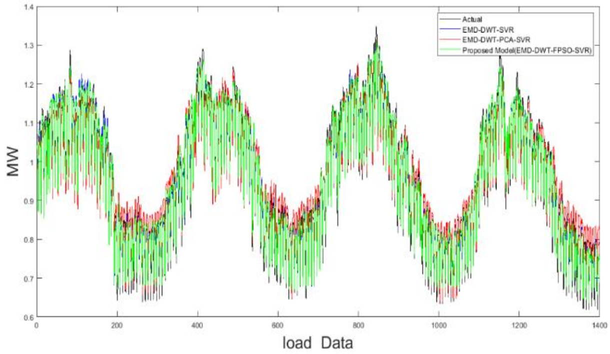

Fig. 14 Comparison of time series forecasting with the proposed method and other methods on the Polish electric load dataset

Table 5 Comparison of the forecasted error value of the proposed method with other methods on the Polish electric charge data set

Error	*MSE*	*RMSE*	*MAPE*	*MAE*
EMD-DWT-SVR	5.4363e-04	0.0233	2.0725	0.0039
EMD-DWT-PCA-SVR	8.9508e-04	0.0299	2.8708	0.0023
EMD-DWT-FPSO-SVR	4.5009e-04	0.0212	1.8877	0.0028

Figure 15 shows the time series forecasting of the proposed method with other methods and Table 6 shows the comparison of the forecasted error value of the proposed method with other methods on the Australian data set.

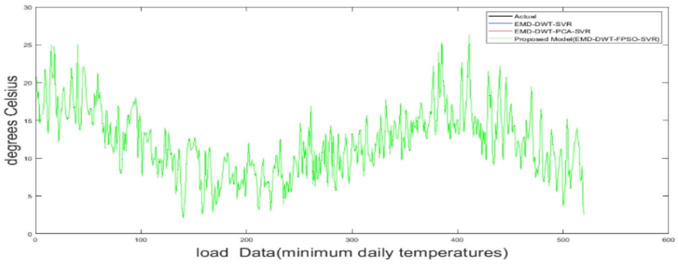

Fig. 15 Comparison of time series forecasting with the proposed method and other methods on the Australian temperature data set

Table 6 Comparison of the forecasted error value of the proposed method with other methods on the Australian temperature data set

Error	*MSE*	*RMSE*	*MAPE*	*MAE*
EMD-DWT-SVR	9.4926e-08	3.0810e-04	0.0025	5.7219e-05
EMD-DWT-PCA-SVR	1.2249e-07	3.4999e-04	0.0030	5.6486e-05
EMD-DWT-FPSO-SVR	9.4899e-08	3.0806e-04	0.0024	5.7159e-05

Figure 16 shows the time series forecasting of the proposed method with other methods and Table 7 shows the comparison of the forecasted error value of the proposed method with other methods on the Belgian data set.

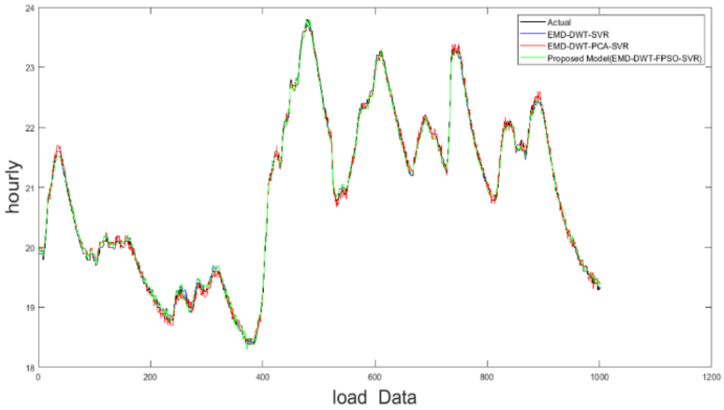

Fig. 16 Comparison of time series forecasting with the proposed method and other methods on the Belgium temperature data set

Table 7 Comparison of the forecasted error value of the proposed method with other methods on the Belgium temperature data set

Method	***MSE***	***RMSE***	***MAPE***	***MAE***
EMD-DWT-SVR	0.0055	0.0742	0.3336	9.7261e-04
EMD-DWT-PCA-SVR	0.0057	0.0756	0.3373	0.0033
EMD-DWT-FPSO-SVR	0.0047	0.0687	0.2988	4.9486e-04

Considering the comparison of the forecasted error value tables, it can be concluded that the proposed model, which is based on feature selection, has been able to reduce the error rate using the PSO algorithm. As can be seen, if the feature extraction algorithm does not work well, it can also increase the model error, which in some datasets, the application of the PCA algorithm has increased error. Moreover, according to the findings, this proposed model can be used in other fields for time series.

4.4 Comparison with rivals' work

In this part, the proposed method is compared with other similar work in this field. It should be noted that the comparison is based on common data sets and the same number of data. Chart 1 shows the forecasting error of the proposed model, and other similar articles to the London dataset, with the MAPE criterion.

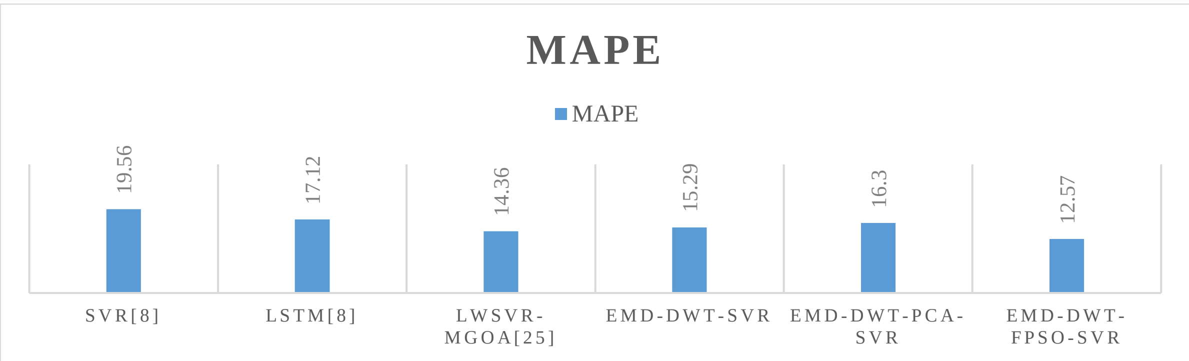

Chart 1 Comparison of the proposed method with other articles of London data set, with MAPE criteria

As can be seen, the result of the proposed method has been better than other ones, and the only method close to it is LWSVR-MGOA. It should also be noted that an efficient hybrid method may work better than a robust algorithm such as the LSTM.

Chart 2 shows the forecasting error of the proposed model, and other similar articles to the Polish dataset, with the MAPE criterion.

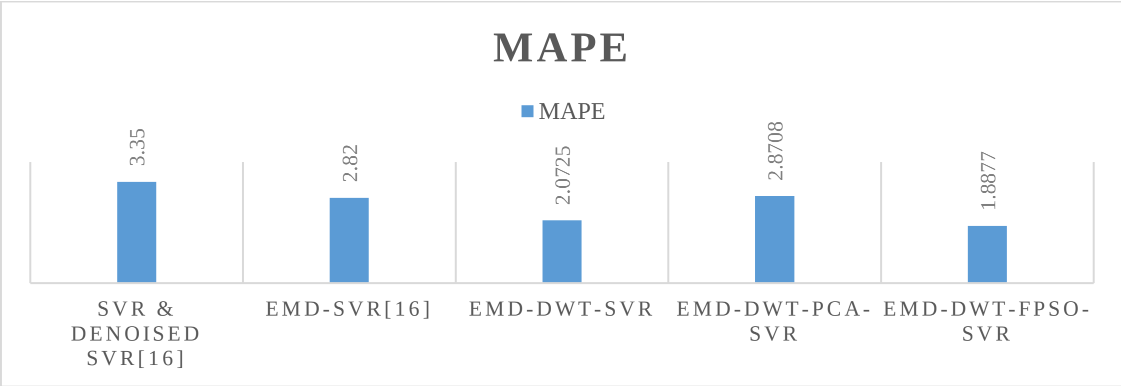

Chart 2 Comparison of the proposed method with other articles of Polish data set, with MAPE criteria

Chart 2 shows the forecasting error of the proposed model, and other similar articles to the Australian dataset, with the MAPE criterion.

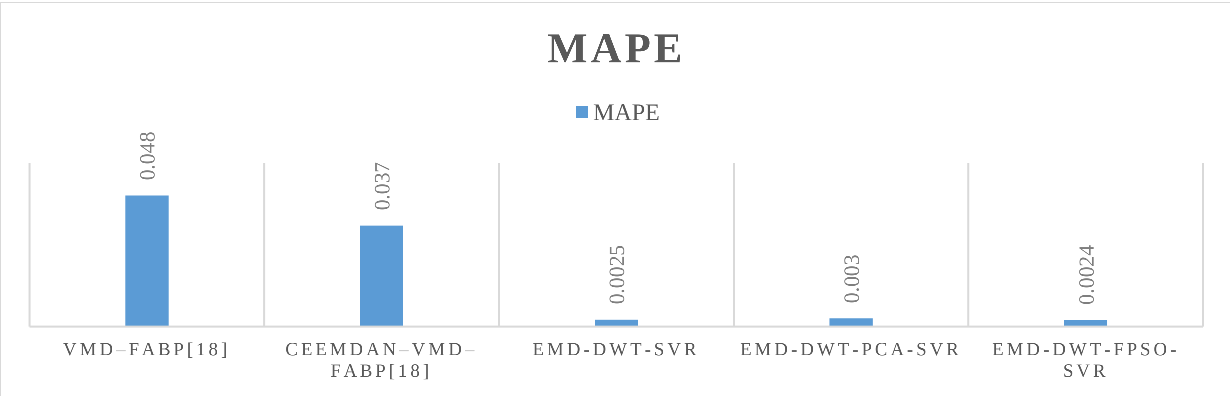

Chart 3 Comparison of the proposed method with other articles of Australian data set, with MAPE criteria

As can be seen, the result of the proposed method has been more significant than other ones.

Chart 4 shows the forecasting error of the proposed model, and other similar articles to the Belgian dataset, with the MAPE criterion.

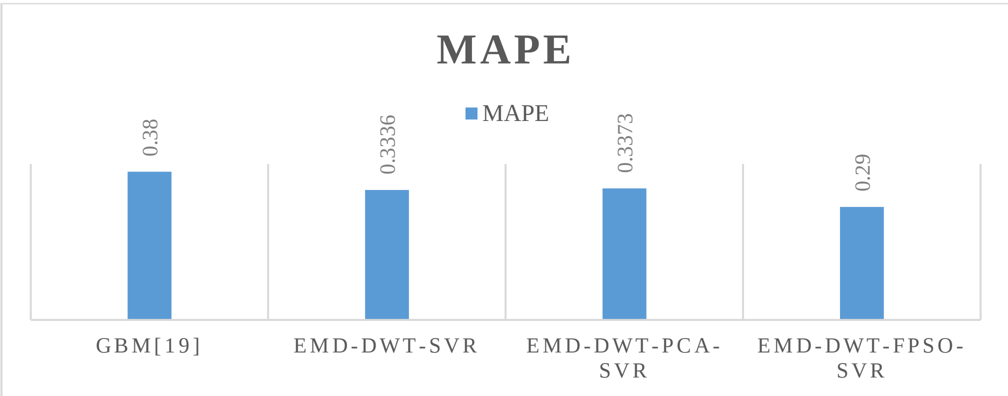

Chart 4 Comparison of the proposed method with other articles of Belgian data set, with MAPE criteria

According to the result of comparing the charts of the MAPE evaluation criterion used in most articles the proposed method in all comparisons has better results in time series forecasting, which indicates the proper performance of the proposed method.

5. Conclusion

Time series analysis deals with dynamic nature and real data and time series forecasting is commonly used in a wide range of fields for future decision making and planning. One of its most widely used fields is energy and climate. Due to the dynamic and non-linear nature of the climate and energy consumption, it has become an attractive field in time series forecasting. Therefore, this study aims to forecast short-term load and air temperature using the proposed MED-DWT-FPSO-SVR method, which can be effective in both fields. The first step is to decompose the signal into the IMF component using the MED algorithm. In the second stage, each component is transformed into a subsequence of approximation and detail features by converting the Haar wavelet. In the third step, the best feature is extracted by the PSO algorithm. The cost function for selecting a feature is to find the least possible error between the features and the purpose of the problem. The purpose of the PSO algorithm is to select the appropriate feature to minimize the error of the proposed method. The fourth step, using support vector regression tries to forecast time series. The results show that the hybrid model of the proposed method has better results than the compared methods, and the proposed method can be used for time series forecasting in both the field of the electric load and air temperature. Moreover, it should be noted that an efficient hybrid method may work better than a robust algorithm such as the LSTM method.

6. References

al series using clustering methods and support vector regression, *Springer, Artificial Intelligence Review*, vol. *52*, no. 1, pp: 743-773, 2019.

[2] Md. Karimuzzaman, Md. Moyazzem Hossain, Forecasting performance of nonlinear time-series models: an application to weather variable, *Springer, Modeling Earth Systems and Environment*, vol. *6*, no. 1, pp: 2451-2463, 2020.

[3] PH. A. Owusu, S. Asumadu-Sarkodie, A review of renewable energy sources, sustainability issues and climate change mitigation, *Cogent Engineering, Civil & Environmental Engineering*, vol. *3*, no. 1, pp: 1-14, 2016.

[4] M. Sabbir, A, Shourav. SH, Shahid. B, Singh. M. Mohsenipour, E. S. Chung, X. J. Wang, Potential Impact of Climate Change on Residential Energy Consumption in Dhaka City, *Springer, Environmental Modeling & Assessment* , vol. 23, no. 1, pp: 131-140, 2018.

[5] O. C. Ozerdem, E. O. Olaniyi, O. K. Oyedotun, Short term load forecasting using particle swarm optimization neural network. *Elsever*, 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW 2017, vol. *120*, no. 1, pp: 382-393, 2017.

[6] M. A. HAMMAD, B. JEREB, B. Rosi, D. DRAGAN, Methods and Models for Electric Load Forecasting: A Comprehensive Review, *Logistics & Sustainable Transport*, vol. *11*, no. 1, pp: 51-76, 2020.

[7] R. K. KAUFMANN, H. KAUPPI, J. H. STOCK, EMISSIONS, CONCENTRATIONS, & TEMPERATURE: A TIME SERIES ANALYSIS, *Springer, Climatic Change*, vol. 77, no. 1, pp: 249-278, 2006.

[8] A. M. Pirbazari, M. Farmanbar, A. Chakravorty, CH. Rong, Short-Term Load Forecasting Using Smart Meter Data: A Generalization Analysis, *MDPI, Processes*, vol. *8*, no. 4, pp: 484-504, 2020.

[9] S. Ali, H, Mansoor. I, Khan. N, Arshad. Short-Term Load Forecasting Using AMI Data, *Electrical Engineering and Systems Science, Signal Processing*, pp: 51-76, 2020.

[10] T. DIMRI, SH. AHMAD, M. SHARIF, Time series analysis of climate variables using seasonal ARIMA approach, *Springer, Journal of Earth System Science*, vol. *129*, no. 1, pp: 129-165, 2020.

[11] N. F. Mohd Radzuan, Z. Othman, A. Abu Bakar, Uncertain Time Series in Weather Prediction, *Elsever, The 4th International Conference on Electrical Engineering and Informatics (ICEEI 2013)*, vol. *11*, no. 1, pp: 557-564, 2013.

[12] Ch. Subaar, N. Apor, J. J. Fletcher, R. Galyuon, G. Eduse, V. W. Adayira, Time Series Analysis for Prediction of Meteorological Data from Wa, Upper West Region of Ghana, *Journal of Climatology & Weather Forecasting*, vol. *6*, no. 3, pp: 1-6, 2018.

[13] J. M. Jimenez, L. Stokes. CH. Moss, Q. Yang, V. N. Livina, Modelling energy demand response using long short-term memory neural networks, *Springer, Energy Efficiency*, vol. *13*, no. 3, pp: 1263-1280, 2020.

[14] H. Al-Shaikh, A. Rahman, Electric Load Forecasting with Hourly Precision Using Long Short-Term Memory Networks, IEEE, 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp: 1-6, 2019.

[15] J. Zheng, C. Xu, Z. Zhang, X. Li, Electric Load Forecasting in Smart Grids Using Long-Short-Term-Memory based Recurrent Neural Network, IEEE, 2017 51st Annual Conference on Information Sciences and Systems (CISS), pp: 1-6, 2017.

[16] Y. Yaslan, B. Bican, Empirical Mode Decomposition Based Denoising Method with Support Vector Regression for Time Series Prediction: A Case Study for Electricity Load Forecasting, *Measurement*, vol. 103, no. 1, pp: 1-26, 2017.

[17] M. Imani, Long Short-Term Memory Network and Support Vector Regression for Electrical Load Forecasting, *IEEE, In 2019 International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET)*, pp: 1-6, 2019.

[18] X. Xu, W. Ren, A Hybrid Model Based on a Two-Layer Decomposition Approach and an Optimized Neural Network for Chaotic Time Series Prediction, *MDPI, Symmetry*, vol. *11*, no. 5, pp: 610-627, 2019.

[19] L.M. Candanedo, V, Feldheim. D, Deramaix. Data driven prediction models of energy use of appliances in a low-energy house, *Elsever, Energy and Buildings*, vol. *140*, no. 1, pp: 81-97, 2017.

[20] L. H. Tang, Y. L. Bai, J. Yang, Y. N. Lu, A hybrid prediction method based on empirical mode decomposition and multiple model fusion for chaotic time series, *Elsever, Chaos, Solitons and Fractals Nonlinear Science, and Nonequilibrium and Complex Phenomena*, vol. 141, no. 1, pp: 1-12, 2020.

[21] Y. Hu, J, Li. M. Hong, J, Ren. R, Lin. Y. Liu, Short term electric load forecasting model and its verification for process industrial enterprises based on hybrid GA-PSO-BPNN algorithmdA case study of papermaking process, *Elsever, Energy*, vol. 170, no. 1, pp: 1215-1227, 2019.

[22] S. Bahrami, R. A. Hooshmand, M. Parastegari. Short term electric load forecasting by wavelet transform and grey model improved by PSO (particle swarm optimization) algorithm, Elsever, Energy, vol. 72, no. 1, pp: 434-442, 2014.

[23] Y. Shen, J. Zhang, J. Liu, P. Zhan, R. Chen. Y. Chen, Short-term load forecasting of power system based on similar day method and PSO-DBN, IEEE, 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2), pp: 1-6, 2018.

[24] H. Lu, M. Azimi, T. Iseley, Short term load forecasting of urban gas using a hybrid model based on improved fruit fly optimization algorithm and support vector machine, *Elsever, Energy Reports,* vol. *5*, no. 1, pp: 666-677, 2019.

[25] E.E. Elattar, N. A. Sabiha, M. Alsharef, M. K. Metwaly, Short term electric load forecasting using hybrid algorithm for smart cities, *Springer, Science+Business Media*, vol. *50*, no. 1, pp: 3379-3399, 2020.

[26] B. Cannas, A. Fanni, L. See, G. Sias, Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. *Elsevier, Physics and Chemistry of the Earth, Parts A/B/C*. Vol. 31, no. 1, pp: 1164-1171, 2006.

[27] N. I. Sapankevych, R. Sankar, Time Series Prediction Using Support Vector Machine A Survey, IEEE, COMPUTATIONAL INTELLIGENCE MAGAZINE, vol. *4*, no. 2, pp: 24-38, 2009.

[28] https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households.

[29] http://research.ics.aalto.fi/eiml/datasets.shtml.

[30] https://machinelearningmastery.com/time-series-datasets-for-machine-learning.

[31] https://www.kaggle.com/bulentsiyah/for-simple-exercises-time-series-forecasting.