Applied Soft Computing Journal 89 (2020) 106091

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

A granular deep learning approach for predicting energy consumption

Rabin K. Jana [a,*], Indranil Ghosh [b,1], Manas K. Sanyal [c]

[a] Operations & Quantitative Methods Area, Indian Institute of Management Raipur, CG 493661, India
[b] Department of Operations Management & IT, Calcutta Business School, WB 743503, India
[c] Department of Business Administration, University of Kalyani, WB 741235, India

ARTICLE INFO

Article history:
Received 22 June 2019
Received in revised form 23 October 2019
Accepted 9 January 2020
Available online 21 January 2020

Keywords:
Deep learning
Long short-term memory network
Predicting energy consumption
Maximal overlap discrete wavelet transformation

ABSTRACT

This paper proposes a granular deep learning approach consisting of maximal overlap discrete wavelet transformation (MODWT) and long short-term memory (LSTM) network for predicting the energy consumption of different sectors at macro levels. Input features are first evaluated using Boruta algorithm-based feature selection model. MODWT is then used to decompose the energy consumption time series to alienate the linear and nonlinear components. The LSTM network, a deep learning tool, is used to make predictions on individual sub-series at a granular level. The final prediction is obtained by aggregating the forecasts obtained on decomposed components. Statistical analyses rationalize the efficacy and superiority of the proposed hybrid framework over six other well-known prediction algorithms. Monthly data for residential, commercial, industrial and transportation sectors of the USA have been taken for analyses. It is observed that energy consumption in commercial and transportation sectors are easier to predict than residential and industrial sectors.

1. Introduction

Owing to a significant influence of energy consumption at various levels on overall economic growth, sustainability, and CO_2 emission, critical modeling of inherent patterns driving energy intake becomes very important [1–3]. Understanding the dynamics of energy consumption at micro and macro levels has drawn attentions of academicians and practitioners [4,5]. Due to its sensitiveness to ever-changing market demand and personal traits, evidence of nonlinear and nonparametric behaviors entrenched in its temporal dynamics is amply apparent. On the other hand, policymakers play a pivotal role in regulating energy consumption. Systematic inspection of the governing patterns for making closer estimations of future trends of energy consumption can assist policymakers in shaping regulations for sustainable needs. Hence, carrying out predictive analytics of energy consumption is extremely challenging and important.

Modeling energy consumption has been addressed in increasing amount in literature of late. Its undeniable impact on formulating energy policies, reducing carbon emissions, GDP, etc. are well-acknowledged in literature [6,7]. It is driven by factors like population growth, urbanization, renewable energy sources, interest rate, fuel poverty, GDP, and government policies [8]. The major predictive modeling literature of energy intake confine to forecasting at micro levels, such as households, buildings, ships, and mechanical processes. Household energy usage is modeled based on demographic characteristics, appliances, and topological features [9]. Energy usage at industrial operations, which constitutes a significant amount of global energy balance, is predominantly estimated by scale and nature of operations. On the contrary, analyses of energy usage at sectoral levels are largely dedicated to determining dynamic interplay with other sectors or to measure respective impact on environment and GDP mainly through econometric models. There exists a significant gap in research on robust predictive modeling energy consumption in different sectors. The present study aims to void the gap by proposing a deep learning approach for predicting monthly estimates of energy at sectoral levels.

Most of the predictive analytics research in micro-level energy demand utilizes multivariate frameworks comprising of independent features allied to nature of operations. As the present research deals with prediction of energy usage at macro levels, identifying exogenous influencing features can be quite vague. Instead, the study uses technical indicators-based time series framework for forecasting. Since we deal with monthly energy consumption records, technical indicators are estimated based on the same to form input constructs [10]. A detailed description of the input features is provided later. Justification and the final selection of the said features are made by employing Boruta feature selection algorithm [11].

* Corresponding author.
E-mail addresses: rkjana@iimraipur.ac.in (R.K. Jana), fri.indra@gmail.com (I. Ghosh), manas_sanyal@klyuniv.ac.in (M.K. Sanyal).

1 The second author is pursuing his Ph.D. at the Department of Business Administration, University of Kalyani, Kalyani, India.

https://doi.org/10.1016/j.asoc.2020.106091
1568-4946/

Unlike predictive modeling frameworks for determining energy demands at micro-units, there exists a dearth of efforts for building prediction models for forecasting sectoral level energy intake. Technical indicators-based forecasting models are common for equity markets [12,13]. There is hardly any research that used them for modeling energy usage and forecasting. In this study, technical indicators are used based on the key statistical properties of the dataset. They are evaluated separately for the final selection before proceeding with the one month ahead prediction. The usefulness of the approach is ascertained after extensive statistical tests and comparison with other advanced models.

This research contributes to the existing predictive analytics literature of energy consumption by proposing a neoteric hybrid deep learning approach. The proposed approach combines the MODWT and LSTM network. MODWT is used to decompose the original time series data into granular components by separating linear and non-linear components. By this decomposition process, it becomes easier to model inherent non-linearity of energy consumption data. MODWT based decomposition technique mimics the principle of granular computing. It decomposes the complex energy consumption datasets into smaller granules of approximately indistinguishable objects by eliminating linear and nonlinear components. LSTM, a deep learning technique and a variant of traditional recurrent neural network (RNN) technique, is used to discover the hidden patterns in the large-scale complex data sequences. Here, LSTM is applied to the decomposed subseries for pattern recognition and yielding forecasts separately. The final prediction is computed by adding the forecasts obtained from granular sub-components. It must be noted here that deep learning approaches are usually applicable for automatic feature extraction. However, in this work, the deep learning approach LSTM is used purely for discovering patterns and forecasting future figures. Usage of deep learning for pattern mining without invoking automatic feature extraction has been reported in the financial market forecasting literature [14,15]. The present study utilizes technical indicator-based approach for feature extraction. Hybridization of Boruta feature selection algorithm and MODWT with LSTM for predicting macro-level energy demand, therefore, contributes to the deep learning paradigm for energy modeling. A performance assessment of the proposed approach is carried out. Statistical comparisons with six distinct predictive modeling algorithms, namely, autoregressive integrated moving average (ARIMA) [16], decision trees (DT) [17], artificial neural network (ANN) [18], recurrent neural network (RNN) [19], multivariate adaptive regression splines (MARS) [20], and bagging [21] are made to justify the effectiveness and superiority of the proposed framework.

The remaining part of the paper is arranged as follows: In Section 2, recent allied research in this domain is summarized. The trend of the existing literature, gaps, and motivations for the current work are discussed. The detailed description and important statistical properties of the dataset are outlined in Section 3. The research methodology, operations of Boruta algorithm, MODWT, LSTM, performance measures, and statistical tests are enunciated in Section 4. The results of predictive performance and comparative statistical analysis are discussed in detail in Section 5. Finally, Section 6 concludes the paper by highlighting the overall findings, discussing limitations and further research scopes.

2. Literature review

Due to its significant influence on economic health, carbon emission, sustainable development, and nuclear energy development, researchers attempted to figure out the key drivers of temporal dynamics of energy usage at micro and macro levels [12,13]. The empirical studies identified human capital income, globalization, and climatic impact as the primary drivers [22]. Identification of explanatory variables enabled researchers in conducting a predictive exercise of energy consumption of different units [23] and framing policies accordingly to regulate the same [7,16].

A multivariate predictive modeling framework using support vector machine (SVM) is proposed for electricity price forecasting on daily Phelix index [24]. A technique combining support vector regression (SVR), ANN and ridge regression (RR) is proposed for forecasting energy consumption. The technique yielded superior forecast [25]. A framework is designed based on SVR for predicting household electricity consumption in China [9]. The algorithmic model used radial basis function (RBF) as kernel function for SVR and resulted in a robust one month ahead forecast. Multiple regression is used to estimate heating energy intake of residential buildings in Palestine and found a significant influence of physical dwelling characteristics, socio-econometric parameters and heating configurations on overall energy usage [26].

Regression-based predictive modeling is carried out of energy intensity and its components in Iran using a wavelet-based neural network [27]. The findings indicated a significant influence of capital–output ratio, capital–labor ratio, and urbanization on target variables. A logarithmic mean Divisia index is used to comprehend the impact of industry delocalization on global energy usage across the globe [28]. The study revealed a significant influence of delocalization on increase in energy usage. A hybrid multi-objective genetic programming (GP) and linear regression (LR) approach are designed for predictive analysis of spectral energy demands [29]. The model outperformed several other energy demand models. Another effective approach is proposed for improvement of energy efficiency in Android app by predicting energy consumption using LR model in application program interface (API)-level perspective [30]. Mathematical models are developed to estimate the demand for electricity and natural gas in the household of California in short and long run [31]. The study critically identified the influence of various assets in consumption of electricity and natural gas. Predictive modeling of hourly energy usage in buildings is successfully conducted [32]. Results suggested that the energy usage prediction for educational buildings could further be improved by monitoring energy behavior changes during semesters.

A two-stage hybrid predictive model is proposed for forecasting energy consumption in China [33]. The model uses group method of data handling based autoregressive tool for estimating the trend of time series corresponding to energy consumption at the initial stage. The remaining nonlinear components are then predicted based on adaptive boosting ensemble model utilizing back propagation neural network (BPNN), RBF network, SVR, and GP. The obtained results justified the efficiency of the proposed model for predicting energy consumption in China. Multilevel regression and adaptive elastic net regularization techniques are used for automatically deducing the importance of features and making predictions of residential energy usage [34]. The findings inferred significant impact of household attributes on governing the dynamics of energy intake. The efficacy of deep recurrent neural network models is also demonstrated in short-range building energy forecasting [5,25].

A review of the existing literature reveals that majority of the predictive modeling approaches are used for predicting the energy consumption at the micro-level by incorporating the necessary operational and allied process parameters. Conventional machine learning algorithms dominate other statistical and parametric methods in pattern mining and forecasting future energy consumption figures. This shows that there exists a significant research gap linked to predictive analytics of energy consumption

Table 1
Summary statistics.

	Residential	Commercial	Industrial	Transportation
Minimum	923.2	685.3	2124.0	1374.0
Maximum	2778.0	1867.0	3076.0	2545.0
Mean	1559.0	1242.0	2652.0	2014.0
Median	1488.0	1288.0	2667.0	2035.0
Variance	150636.7	75730.09	35061.2	79063.53
SD	388.12	275.19	187.25	281.18
Skewness	0.67	−0.20	−0.26	−0.12
Kurtosis	−0.08	−0.95	−0.30	−1.26
Jarque–Bera	38.97***	22.17***	7.66*	35.02***
Shapiro–Wilk	0.96***	0.97***	0.99***	0.95***
Frosini Test	0.76***	0.83***	0.35*	1.05***
Weisberg–Bingham Test	0.96***	0.97***	0.99**	0.95***
ADF Test	−1.12#	−0.95#	−2.53#	−0.90#
Ljung–Box Test	379.06***	3107.03***	2284.40***	4099.01***
Terasvirta's NN Test	10.23***	32.42***	2.96#	23.45***
White's NN Test	10.91***	34.52***	2.87#	23.0***
Hurst Exponent	0.75	0.75	0.76	0.72

*** Significant at 1%, * Significant at 10% levels of significance, # Not significant, SD: standard deviation, ADF: Augmented Dickey Fuller, NN: Neural Network.

at the macro level. Deep learning algorithms have been found to generate commendable forecasts for historical financial time series [35,36]. Energy consumption data are instances of time series. Therefore, the present research leverages the effectiveness of technical indicator-based forecasting frameworks for energy consumption. To critically model the volatile dynamics, MODWT decomposition is used in conjunction with LSTM deep learning framework utilizing technical indicators as explanatory features.

3. Data description

Monthly total energy consumption of residential, commercial, industrial, and transportation sectors from January 1973 to December 2018 is compiled from the official web portal of the USA energy administration. Figs. 1–4 portray the temporal evolutionary patterns of energy consumption for the considered sectors. The horizontal axis represents sample observations, and the vertical axis represents the net energy intake.

Table 1 reports the summary statistics and outcome of several conventional statistical tests to ascertain the behavioral pattern of the dataset.

Table 1 shows that the Jarque–Bera, Shapiro–Wilk, Frosini, and Weisberg–Bingham (WB) test statistics values are highly significant. This implies that energy consumption for the chosen sectors displays non-parametric behavior. Insignificant ADF test statistic value suggests that no series is stationary in nature. The Ljung–Box test at lag 10 implies a strong presence of serial autocorrelation structure entrenched in the monthly energy consumption time series. The Terasvirta's and White's neural network tests indicate the presence of nonlinear behavior in Residential, Commercial and Transportation energy consumption series, while Industrial energy consumption series do not clearly exhibit non-linear traits. Hurst exponent (H) is a measure to characterize fractal properties in a time sequence. The value of H equal to 0.5 indicates the existence of pure random walk. Values of H greater than 0.5 imply the presence of persistent trend while values less than 0.5 indicate existence of antipersistent trend. As the estimated values of H for the respective series are substantially greater than 0.5, dominance of persistent trend in governing the evolutionary patterns of energy intakes can be concluded. Persistent trend infers long memory structure embedded in the temporal dynamics of the considered time series which in turn suggests the dependence of future movements on the historical information. Since the presence of prominent serial autocorrelation and long memory structure is apparent, the use of technical indicator-based prediction framework is more appropriate as they can effectively capture the previous information of varying time steps. Table 2 outlines the technical indicator-based independent features which are used in this work.

Technical indicators are widely used in identifying patterns and forecasting future figures. Particularly, in stock market modeling, usage of these features is extremely useful as stock indices often exhibit autocorrelation and persistent behavior. Like stock markets, energy consumption series under consideration display long memory dependence. Therefore, it is appropriate to use these features for estimating forecasts based on historical information. This paper uses 26 features that are critically evaluated using Boruta algorithm for the final judgment on the fitment for the use in the proposed prediction approach.

4. Research methodology

In this section, the proposed research framework is elucidated in detail. There are three key components of the framework. They are feature selection, time series decomposition, and deep learning. The numerical indices used for evaluating the predictive performance of hybrid granular framework and statistical tests used for the comparison are also presented.

4.1. Boruta feature selection algorithm

Boruta is an ensemble-based feature selection algorithm that mimics the working principle of random forest (RF) with additional mechanisms for obtaining superior results [11]. RF is an ensemble approach for classification having a series of unbiased constituent classification trees constructed based on bootstrapped samples for estimating the target. It uses multiple voting schemes for drawing final predictive decisions while evaluating the importance of a feature based on loss of prediction accuracy due to induced random perturbation of feature values. However, the methodology suffers from some statistical drawbacks [37]. To thwart the challenges, Boruta algorithm incorporates increased level of randomness in the existing information system for selecting appropriate features. The steps of Boruta algorithm are outlined below:
Step 1: Create copies of original features by extending the information system, known as shadow attributes.
Step 2: Perturb the values of shadow attributes to reduce the correlation with the target variable(s).
Step 3: Apply the RF model on the new extended information system to compute the Z score of all attributes including the shadow ones.
Step 4: Select the Z score with the maximum value among the shadow features (MZSF) and mark other attributes having higher Z score than MZSF.
Step 5: Attributes yet to be assessed for importance are then arranged to execute the two-sided test of equality with MZSF.
Step 6: Features having Z score significantly higher than MZSF are marked as important features, while features with lower Z score than MZSF are deemed as unimportant features.
Step 7: Remove the shadow attributes.
Step 8: Repeat steps 1 to 7 to a pre-specified number of runs.

Next, we mention the granular time-series decomposition model by applying MODWT.

4.2. MODWT based wavelet decomposition

Wavelet transformation allows efficient inspection of time-localized information, keeping the conservation of entropy of any signal [38,39]. It enables multi-resolution analysis by transforming original signal into multiple horizons [40]. Time series

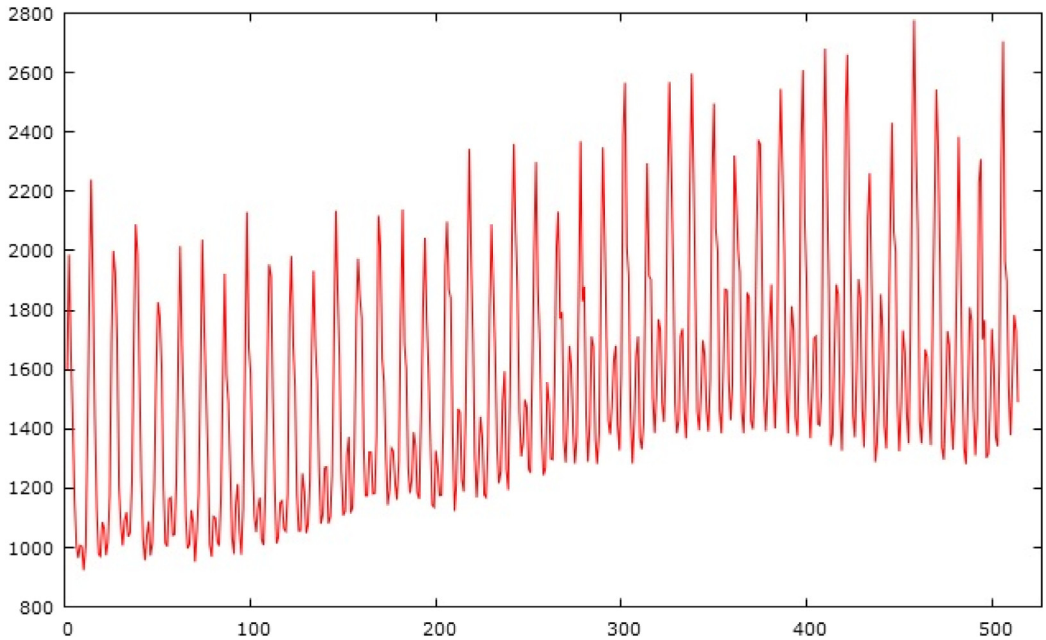

Fig. 1. Evolutionary pattern of the residential energy consumption.

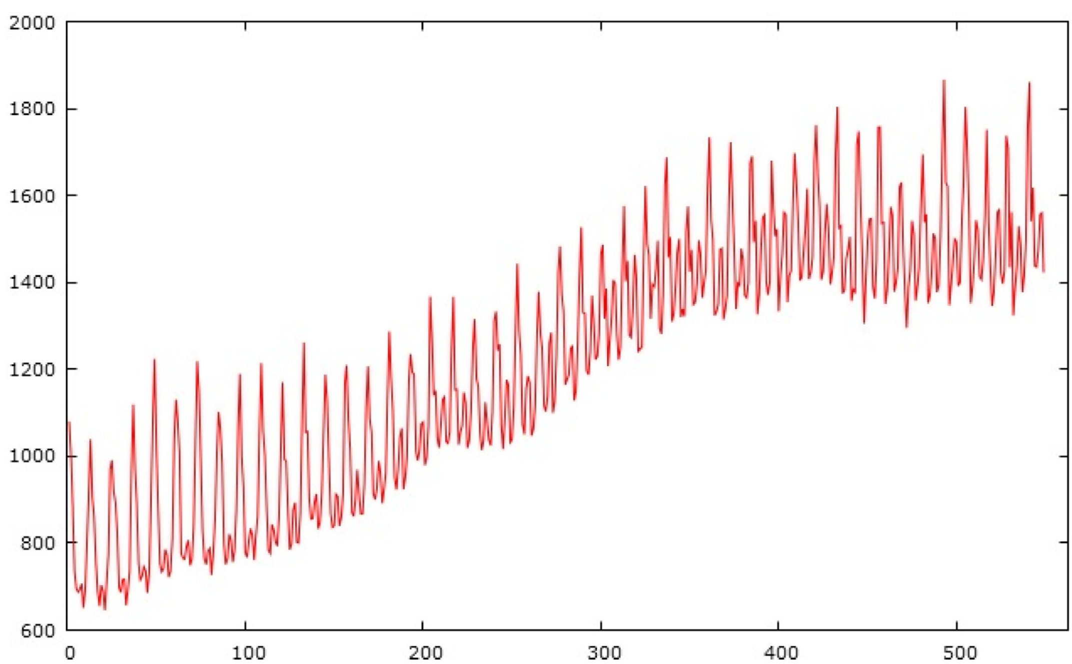

Fig. 2. Evolutionary pattern of commercial energy consumption.

time and frequency space are decomposed simultaneously by representing the series as projections of father (ϕ) and mother (ψ) wavelets. $\phi(t)$ represents the approximation with unit mean, and $\psi(t)$ represents the detail having unit energy and zero mean. Mathematically, it can be expressed as:

$$\int \phi(t)dt = 1 \tag{1}$$

$$\int \psi(t)dt = 0 \tag{2}$$

Father wavelets represent long scale smooth components, and mother wavelets represent deviations from the smooth components. Father and mother wavelets are expressed using the following equations:

$$\phi_{(j,k)} = 2^{-j/2}\phi\left(\frac{t-2^jk}{2^j}\right) \tag{3}$$

$$\psi_{(j,k)} = 2^{-j/2}\psi\left(\frac{t-2^jk}{2^j}\right) \tag{4}$$

The smooth and detailed coefficients are estimated from an original signal $f(t)$ as follows:

$$s_{(j,k)} = \int f(t)\phi_{(j,k)} \tag{5}$$

$$d_{(j,k)} = \int f(t)\psi_{(j,k)} \tag{6}$$

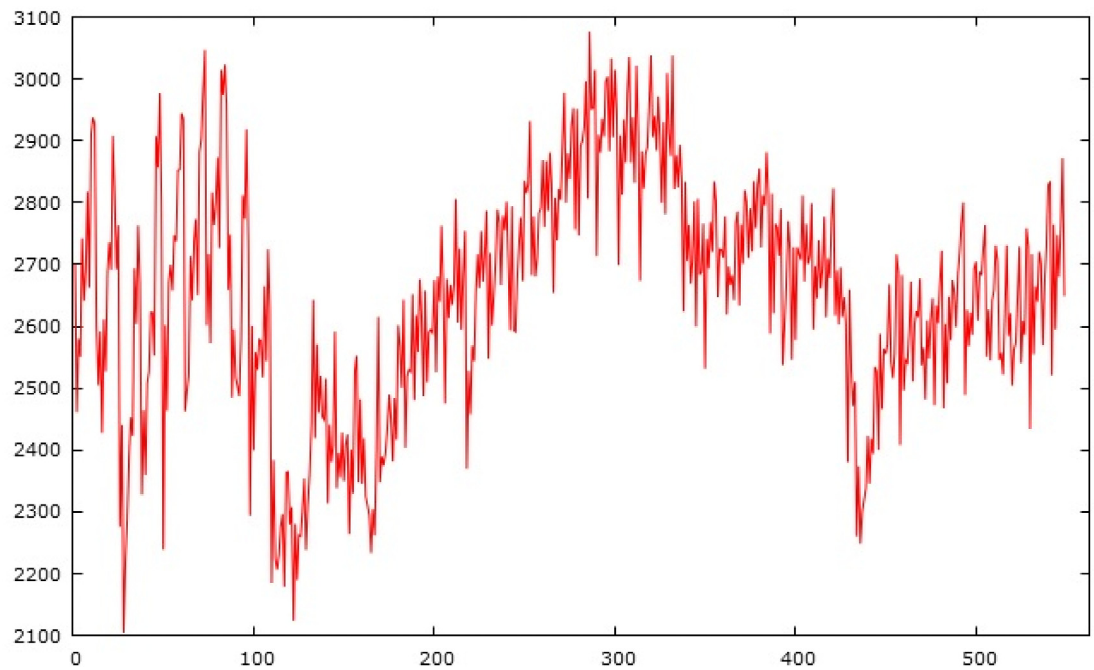

Fig. 3. Evolutionary pattern of industrial energy consumption.

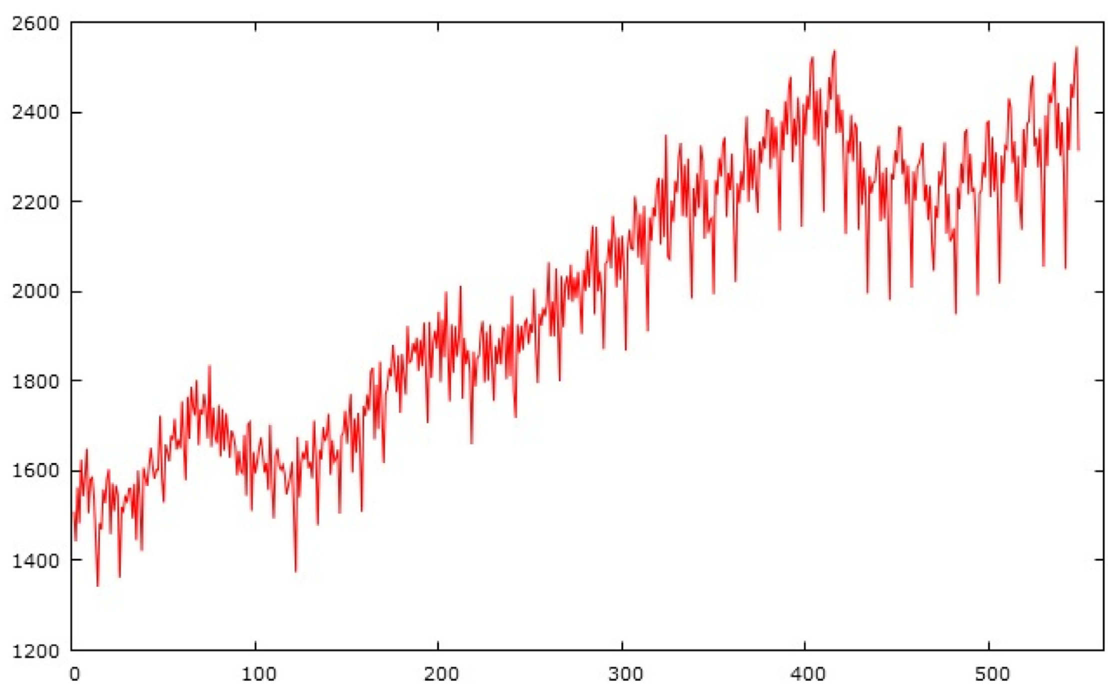

Fig. 4. Evolutionary pattern of transportation energy consumption.

Hence, the original function can be expressed as a linear combination of wavelet functions as:

$$f\,(t) = \sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\psi_{J,k}(t) + \cdots + \sum_k d_{j,k}\psi_{j,k}\,(t) + \cdots + \sum_k d_{1,k}\psi_{1,k}(t) \tag{7}$$

Eq. (7) can be rewritten as:

$$f\,(t) = S_J + D_J + D_{J-1} + \cdots + D_j + \cdots + D_1 \tag{8}$$

where

$$S_J = \sum_k s_{J,k}\phi_{J,k}\,(t)\,, j = 1, \ldots, J \tag{9}$$

$$D_J = \sum_k d_{J,k}\psi_{J,k}\,(t)\,, j = 1, \ldots, J \tag{10}$$

The orthogonal components $\left[S_J, D_J, D_{J-1}, \ldots, D_1\right]$ are resultant of the multiscale decomposition of $f\,(t)$. The wavelet detail at the Jth level is denoted by D_J. The sum of deviations at the respective detail scales is defined by S_J.

Table 2
Summary of input features.

Sl. No.	Feature	Formula
1.	One month back energy consumption (LAG1)	$LAG1 = EC_{i-1}$ where EC_i denotes total energy intake of present month (i)
2.	Two months back energy consumption (LAG2)	$LAG2 = EC_{i-2}$
3.	Three months back energy consumption (LAG3)	$LAG3 = EC_{i-3}$
4.	Four months back energy consumption (LAG4)	$LAG4 = EC_{i-4}$
5.	Four months back energy consumption (LAG5)	$LAG5 = EC_{i-5}$
6.	5-month moving average (MA5)	$MA5 = \frac{\sum_{i=j-4}^{j} EC_i}{5}$
7.	10-month moving average (MA10)	$MA10 = \frac{\sum_{i=j-9}^{j} EC_i}{10}$
8.	20-month moving average (MA10)	$MA20 = \frac{\sum_{i=j-19}^{j} EC_i}{20}$
9.	5-month bias (B5)	$B5 = \frac{EC_i - MA5}{MA5}$
10.	10-month bias (B10)	$B10 = \frac{EC_i - MA10}{MA10}$
11.	20-month bias (B20)	$B20 = \frac{EC_i - MA20}{MA20}$
12.	5-month momentum (MTM5)	$MTM5 = EC_i - EC_{i-5}$
13.	10-month momentum (MTM10)	$MTM10 = EC_i - EC_{i-10}$
14.	20-month momentum (MTM20)	$MTM20 = EC_i - EC_{i-20}$
15.	5-month rate of change (ROC5)	$ROC5 = \frac{EC_i - EC_{i-5}}{EC_{i-5}}$
16.	10-month rate of change (ROC10)	$ROC10 = \frac{EC_i - EC_{i-10}}{EC_{i-10}}$
17.	20-month rate of change (ROC10)	$ROC20 = \frac{EC_i - EC_{i-20}}{EC_{i-20}}$
18.	5-month exponential moving average (EMA5)	$EMA5 = \frac{2}{5+1} \times EC_5 + \frac{5-1}{5+1} \times EMA4$, where $EMA1 = EC_1$
19.	10-month exponential moving average (EMA10)	$EMA10 = \frac{2}{10+1} \times EC_9 + \frac{10-1}{10+1} \times EMA9$
20.	20-month exponential moving average (EMA20)	$EMA20 = \frac{2}{20+1} \times EC_{19} + \frac{20-1}{20+1} \times EMA19$
21.	Upper Bollinger band (UB)	$UB = MA20 + (20 \times \sigma_{20})$ where σ_{20} denotes standard deviation of energy consumption of the previous 20 months
22.	Lower Bollinger band (LB)	$LB = MA20 - (20 \times \sigma_{20})$
23.	Moving Average Convergence Divergence (MACD)	$MACD = 2 \times (DIF - DEA)$; $DIF = EMA12 - EMA26$; $DEA = EMA(DIF)$
24.	5-month relative strength index (RSI5)	$RSI5 = \frac{UC5_i}{UC5_i + DC5_i} \times 100$ where $UC5_i = \sum_{t=1}^{5} max\{0, EC_{i-t+1} - EC_{i-t}\}$ and $DC5_i = \sum_{t=1}^{5} max\{0, EC_{i-t} - EC_{i-t+1}\}$
25.	10-month relative strength index (RSI10)	$RSI10 = \frac{UC10_i}{UC10_i + DC10_i} \times 100$
26.	20-month relative strength index (RSI20)	$RSI20 = \frac{UC20_i}{UC20_i + DC20_i} \times 100$

MODWT is used here for estimating the respective scale and wavelet coefficients. The process of decomposition is carried out using the Daubechies least asymmetric filter of length eight (la8). It belongs to the family of orthogonal wavelets having several advantages over Haar wavelet filters manifested in terms of smoother operations [38] and generation of uncorrelated coefficients [41].

LSTM algorithm is used for making predictions on granular decomposed sub-series obtained by MODWT. Obtained forecasts on individual components are added to generate the final forecast reflecting the estimation of total energy consumption of the next month for the respective sectors. The basic principle of LSTM is described next.

4.3. Deep learning using LSTM

LSTM [42] is a deep learning approach that has garnered high level of attention among the researchers in carrying out predictive modeling of complex pattern recognition tasks [32,33]. LSTM is a variant of conventional RNN that can effectively tackle the 'vanishing gradient' problem and generate superior predictions [43]. It consists of a series of memory cells for holding records of states and three units of gates (input, forget, and output gates) to monitor the movements of information flow. The memory cells can contain large scale past historical information. Forget gates are responsible for deciding the portion of cell state to be kept or to be replaced with the latest information. Input gates gradually attempt to learn the condition under which any information should be kept or updated in cell state. Output gates are responsible for propagation of information toward the forward direction in LSTM. Through proper controlling of information passage and memory structure, LSTM attempts to extract long-range dependence entrenched in time series data.

Subject to permissions of input gates, input figures are restored in the state of the cell. The input (i_t) and candidate (C_t) values of the respective memory units are calculated as follows:

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \quad (11)$$

$$C_t = sigm\left(W_C x_t + U_C h_{t-1} + b_C\right), \quad (12)$$

where W, U, b account for the weight matrices and bias units of respective layers, and *sigm* denotes Sigmoid activation function.

Forget gate values ($f(t)$) are determined as follows:

$$f(t) = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \quad (13)$$

The new state ($\tilde{C}_t$) of a memory cell is updated as follows:

$$\tilde{C}_t = i_t \times C_t + f_t \times C_{t-1} \quad (14)$$

Values of output gates (o_t) are updated as

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + V_o C_t + b_o\right) \quad (15)$$

The outcome of the cell (h_t) is computed as:

$$h_t = o_t \times sigm\left(C_t\right) \quad (16)$$

For estimation of the near-optimal weights to construct the model, back-propagation through time (BPTT) algorithm has been used. BPTT is an extension of the standard back-propagation algorithm capable of unrolling the recurrence structure of LSTM as additional layers. BPTT learns the parameters of the unfolded network by well-defined series of computational steps linking the temporal dependence. Sigmoid functions are chosen as activation functions. Two LSTM blocks of 50 neurons each are considered for modeling. The conventional mean squared residual is chosen as the loss function, and the learning is conducted using the '*adam*' optimizer [44] in BPTT framework. The LSTM is used on individual decomposed components obtained through MODWT for capturing the inherent patterns and making predictions. The final prediction is made by taking the sum of obtained predictions on decomposed components by LSTM. This integrated framework consisting of MODWT and LSTM is denoted as M-LSTM for forecasting. The model is simulated using Python Programming language in the *'Tensorflow'* platform. The proposed research methodology is depicted in Fig. 5.

4.4. Effectiveness measurement

For measuring the effectiveness of the proposed M-LSTM approach, the following indices are used:

Root Mean Squared Error (RMSE)

The RMSE measures the standard deviation of a prediction model. It is estimated as follows:

$$RMSE=\sqrt{\frac{\sum_{t=1}^{N}\left(\hat{Y}_t-Y_t\right)^2}{N}} \tag{17}$$

where $\hat{Y}_t$ *and* Y_t represent predicted and observed figures at time step t. A lower RMSE value indicates a better efficiency of the model.

Nash-Sutcliffe Efficiency (NSE)

NSE is a normalized statistic that measures the relative strength of the residual variance arising from a predictive model compared to the data variance of the original dataset [45]. It is defined as follows:

$$NSE=1-\frac{\sum_{t=1}^{N}\left\{\hat{Y}_t-Y_t\right\}^2}{\sum_{i=1}^{N}\left\{Y_t-\overline{Y_t}\right\}^2}, \tag{18}$$

NSE values vary between $-\infty$ to 1. NSE value corresponding to a prediction model closer to 1 indicates a superior prediction.

Index of Agreement (IA)

IA measures the magnitude of the error resulting from the prediction model [46]. It is defined as follows:

$$IA=1-\frac{\sum_{t=1}^{N}(\hat{Y}_t-Y_t)^2}{\sum_{i=1}^{N}\left\{\left|\hat{Y}_t-\overline{Y_t}\right|+|Y_t-\overline{Y_t}|\right\}^2} \tag{19}$$

Similar to NSE, IA values should be close to 1 for a top-quality prediction.

Theil Index (TI)

The TI index is defined as follows [47]:

$$TI=\frac{\left[\frac{1}{N}\sum_{t=1}^{N}\left(\hat{Y}_t-Y_t\right)^2\right]^{1/2}}{\left[\frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y}_t\right)^2\right]^{1/2}+\left[\frac{1}{N}\sum_{i=1}^{N}(Y_t)^2\right]^{1/2}} \tag{20}$$

TI values should ideally be closer to 0 for an effective prediction tool.

Directional Predictive Accuracy (DA)

The DA is defined as follows [48]:

$$DA=\frac{1}{N}\sum_{t=1}^{N}D_t,\ D_t=\begin{cases}1,\ (Y_{t+1}-Y_t)\left(\hat{Y}_{t+1}-Y_t\right)\geq 0\\ 0,\ Otherwise\end{cases} \tag{21}$$

The closer the value of DA to 1, the higher the accuracy of the model of directional prediction, whereas values closer to 0 implies lower the directional predictive accuracy.

Diebold–Mariano (DM) Test

DM is a statistical test to assess the comparative differences of multiple forecasting models in terms of predictive accuracy [49]. The MSE measure is used as the loss function in the DM test. It tests the null hypothesis that the MSE of a model under test is higher than the benchmark one. The MSE is calculated as follows:

$$MSE=\frac{1}{N}\sum_{t=1}^{N}\left(\hat{Y}_t-Y_t\right)^2 \tag{22}$$

The DM statistic (DMS) is calculated as follows:

$$DMS=\frac{\overline{D}}{\sqrt{V_{\overline{D}}/M}} \tag{23}$$

where

$$\overline{D}=\frac{1}{M}\sum_{t=1}^{M}d_t \tag{24}$$

$$d_t=\frac{1}{M}\sum_{t=1}^{M}\left(\left(Y_t-\hat{Y}_t^{test}\right)^2-\left(Y_t-\hat{Y}_t^{bench}\right)^2\right) \tag{25}$$

$$V_{\overline{D}}=\gamma_0+2\sum_{q=1}^{\infty}\gamma_q \tag{26}$$

and

$$\gamma_q=cov\left(d_t,d_{t-q}\right), \tag{27}$$

$\hat{Y}_t^{test}$ and $\hat{Y}_t^{bench}$ represent predicted figures obtained by tested and benchmark models.

5. Results and discussions

This section presents the findings of feature selection, time series decompositions, predictive exercises, and statistical analyses sequentially.

5.1. Feature selection

Initially, 26 features are considered for the final selection of the explanatory variables in the proposed approach. Boruta algorithm is applied using 100 runs. Beyond 100 iterations, no significant changes in the findings were observed. Figs. 6–9 report the findings of Boruta based feature selection algorithm.

The above figures display box plots of importance scores of individual factors obtained through the Boruta algorithm. Factors having green box plots are the significant ones as they exhibit better ability to predict than the shadow attributes. For the residential and commercial sectors, all the independent features have emerged as significant. Hence, all the 26 features are used for forecasting the future figures for these two sectors. However, for industrial and transportation sectors, few factors are marked in yellow and red colors. Features marked in red are the factors

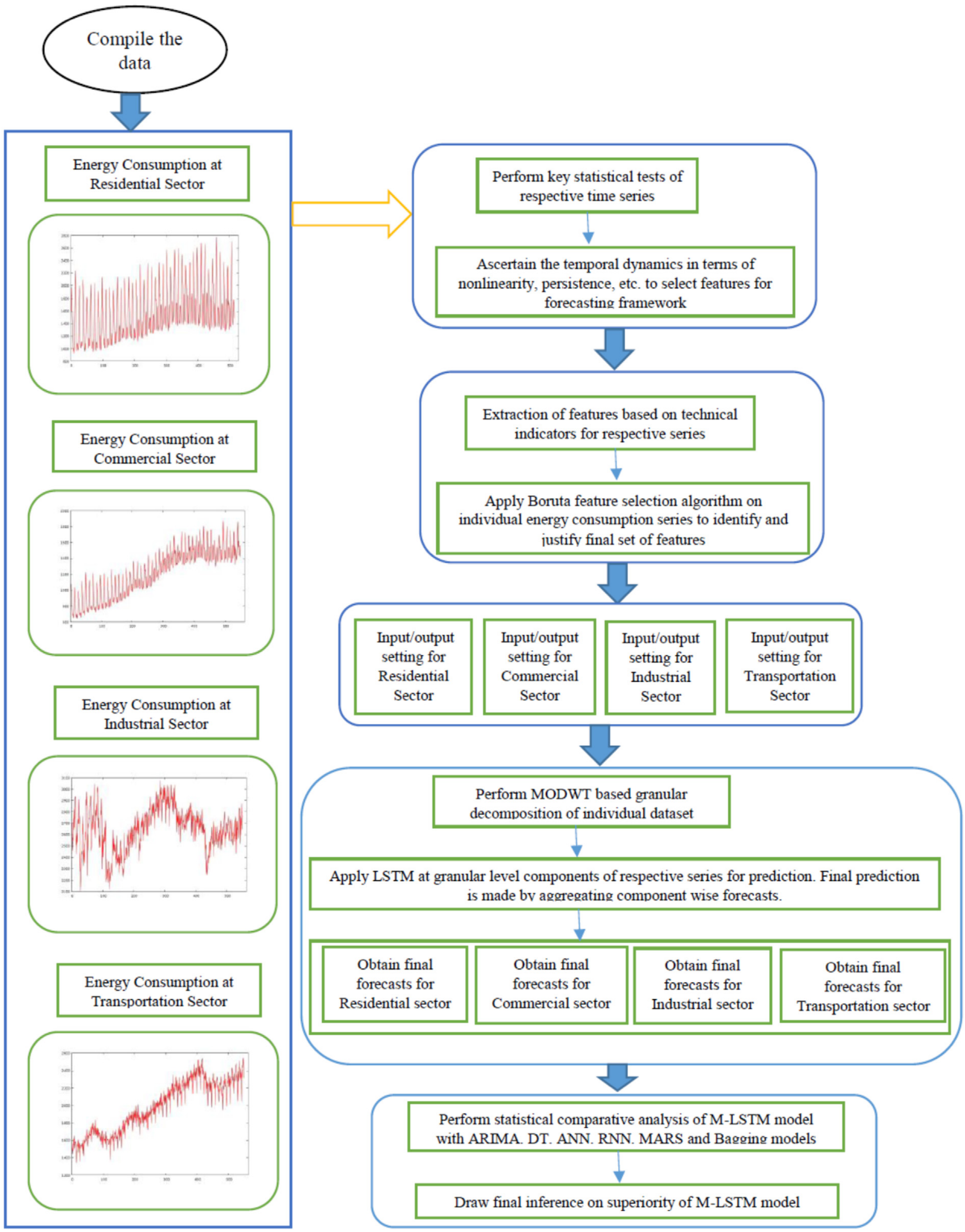

Fig. 5. Flowchart of the proposed research framework.

that have less explanatory power than the shadow features, and hence they can be omitted from the final list. It is observed that RSI20 features for both series have turned out to be insignificant. Factors in yellow are tentative factors. ROC10 and MTM10 are tentative factors in the industrial dataset, while RSI10 and EMA5 act as tentative features for the transportation data. As adding more features may result in overfitting problem, the present study has discarded them from the final feature list. Alternatively, one may increase the number of runs in Boruta algorithm for checking whether they remain at the tentative state or not at the

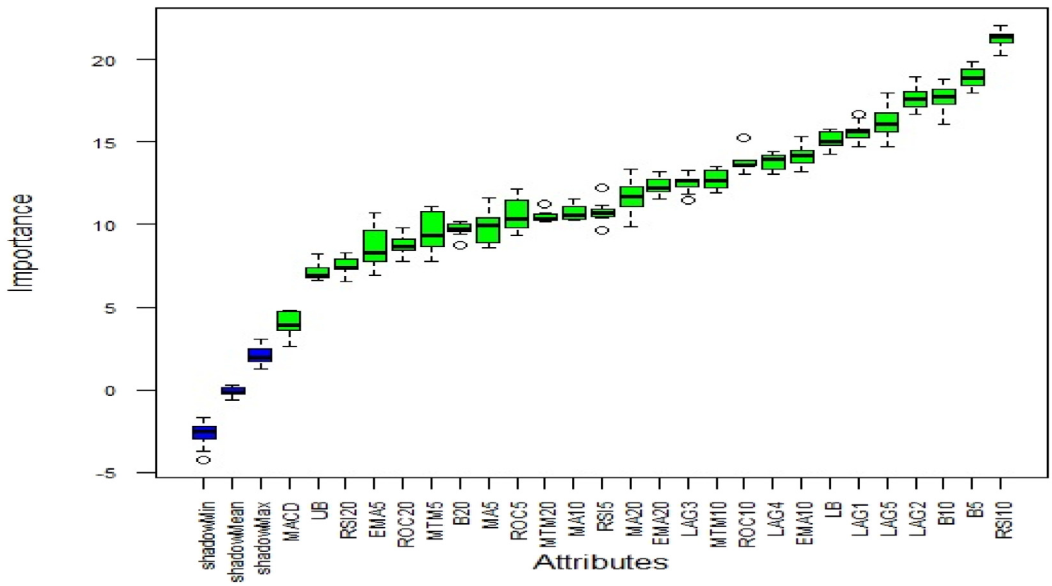

Fig. 6. The outcome of the feature inspection process on residential energy consumption series.

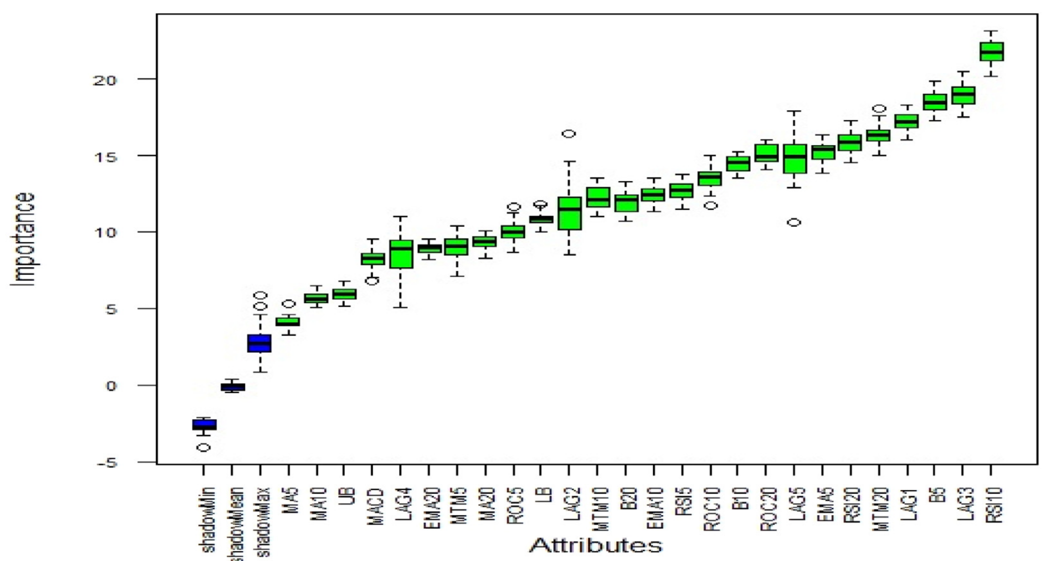

Fig. 7. The outcome of the feature inspection process on commercial energy consumption series.

expense of computational resources and time in order to take the final call on tentative factors.

5.2. Time series decomposition

As stated earlier, the paper uses the MODWT model using "la8" filter for generating granular decomposed components. Six levels of decomposition are used. Figs. 10–13 display the resultant decomposed components of the energy consumption series. In Figs. 10–13, the horizontal axis represents observations and the vertical axis represents the absolute values of respective decomposed components.

Respective independent features are also decomposed into six levels, and component-wise target and input variables are arranged for predictive modeling exercise.

5.3. Predictive analytics

After identification of the independent features and granular decomposition, LSTM is applied to individual decomposed components for carrying out predictive analytics. Component-wise obtained forecasts are sum together to generate the final estimation of energy consumption at the monthly level for the chosen sectors. To conduct experimental trials, process parameters of LSTM are varied for 20 trials. The original monthly time series dataset is portioned randomly into training (85%) and test (15%) datasets. All the variables are rescaled between 0 to 1 for computing the performance measures. Table 3 presents the values of performance indicators to judge the efficiency of M-LSTM model.

From Table 3, we observe that values of NSE and IA are greater than 0.9 on both training and test cases for all the four sectors. The values of DA are greater than 0.9, as well. The values of RMSE and TI are estimated to be quite low. Therefore, it can be concluded that the performance of M-LSTM in estimating the next month's forecasts of energy usage for the residential,

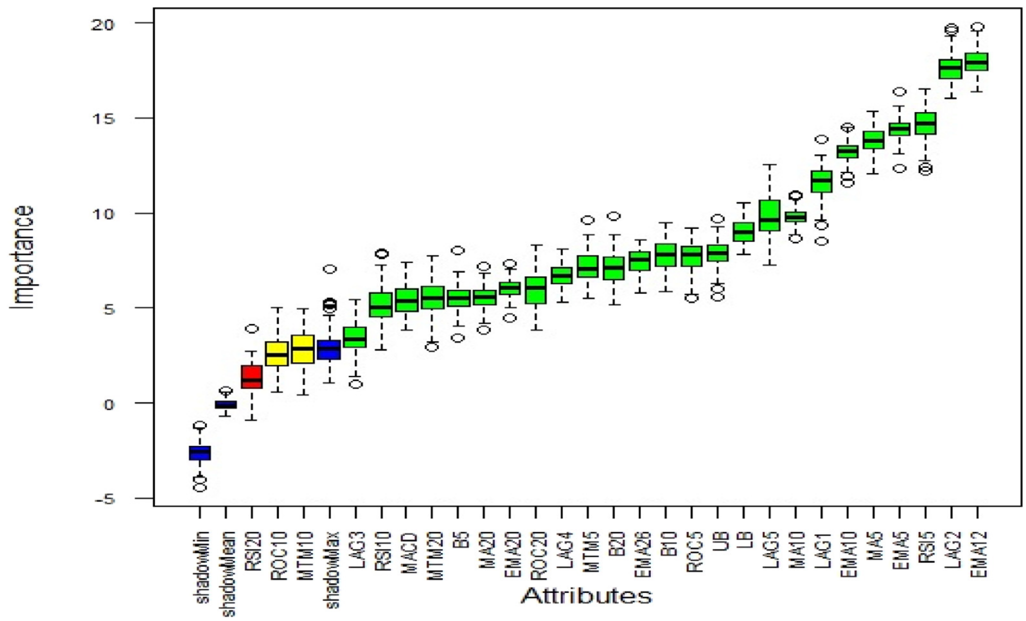

Fig. 8. The outcome of the feature inspection process on industrial energy consumption series.

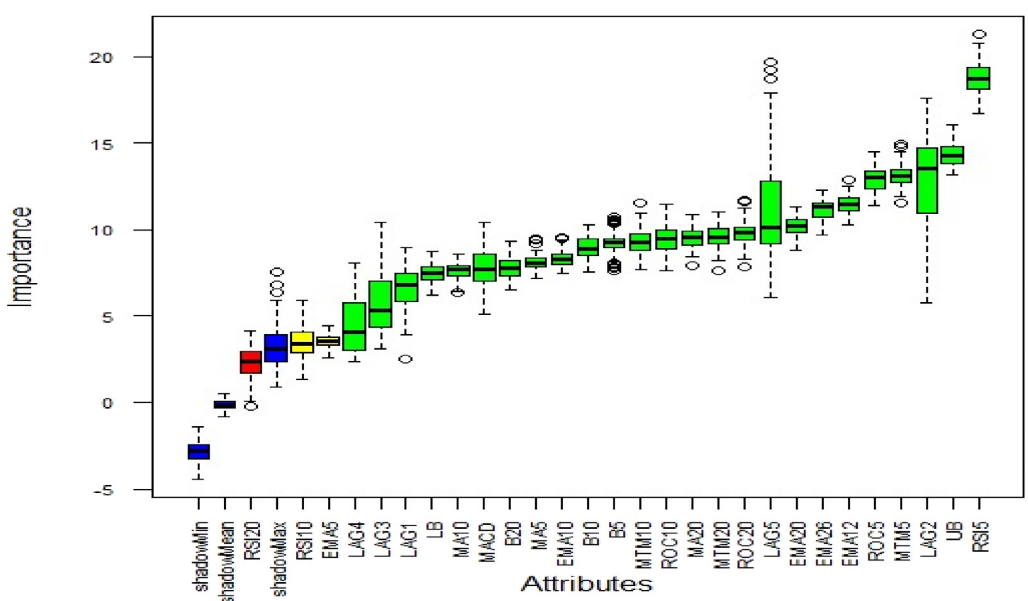

Fig. 9. The outcome of the feature inspection process on transportation energy consumption series.

Table 3
Predictive performance of M-LSTM model.

Series	RMSE	NSE	TI	IA	DA
Training dataset					
Residential	0.03007	0.98185	0.03613	0.99997	0.91136
Commercial	0.02754	0.98667	0.02634	0.99998	0.97772
Industrial	0.04789	0.94066	0.04157	0.99992	0.94986
Transportation	0.02924	0.98518	0.02478	0.99998	0.98885
Test dataset					
Residential	0.03157	0.97837	0.03724	0.99994	0.91087
Commercial	0.02893	0.98580	0.02717	0.99996	0.97685
Industrial	0.04904	0.93994	0.04301	0.99989	0.94875
Transportation	0.03136	0.98367	0.02554	0.99996	0.98812

commercial, industrial, and transportation sectors is exceptional. A closer inspection of the performance indicators assists in gaining further insights. The NSE value of the industrial sector is comparatively lesser than the other sectors, while its RMSE is higher than the others. On the other hand, estimated DA value of the residential sector is minimum. In terms of RMSE, NSE, TI, and IA on both training and test datasets, energy consumption patterns in residential sector can be marked less predictable than the commercial and transportation sectors. Hence, the inference can be drawn that the energy consumption dynamics are more systemic in commercial and transportation sectors in the USA than the other two sectors. The findings can help policymakers in framing future policies accordingly.

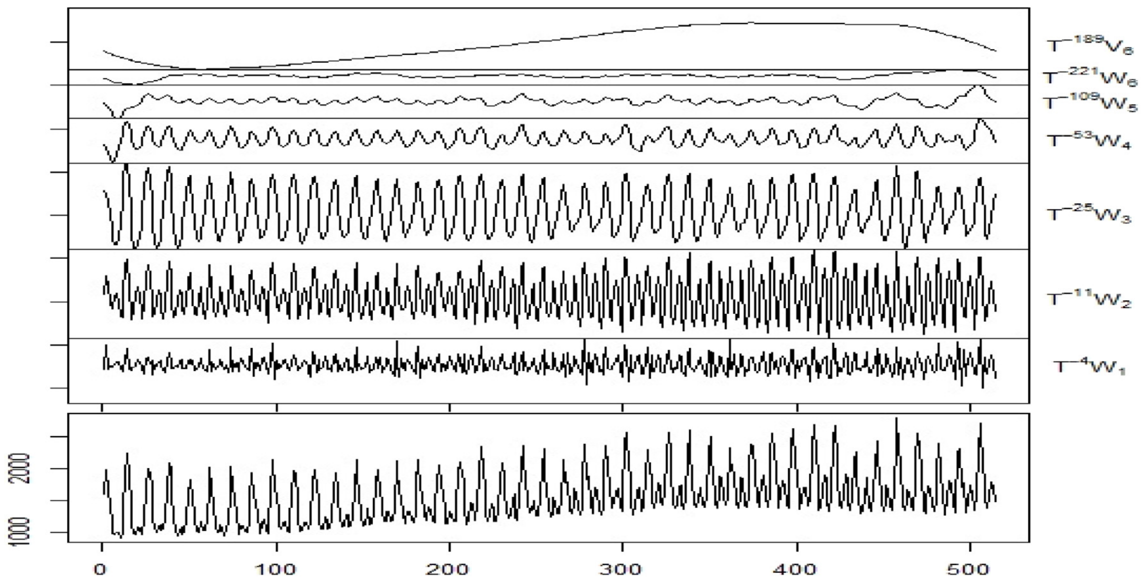

Fig. 10. MODWT decomposition of monthly residential energy intake.

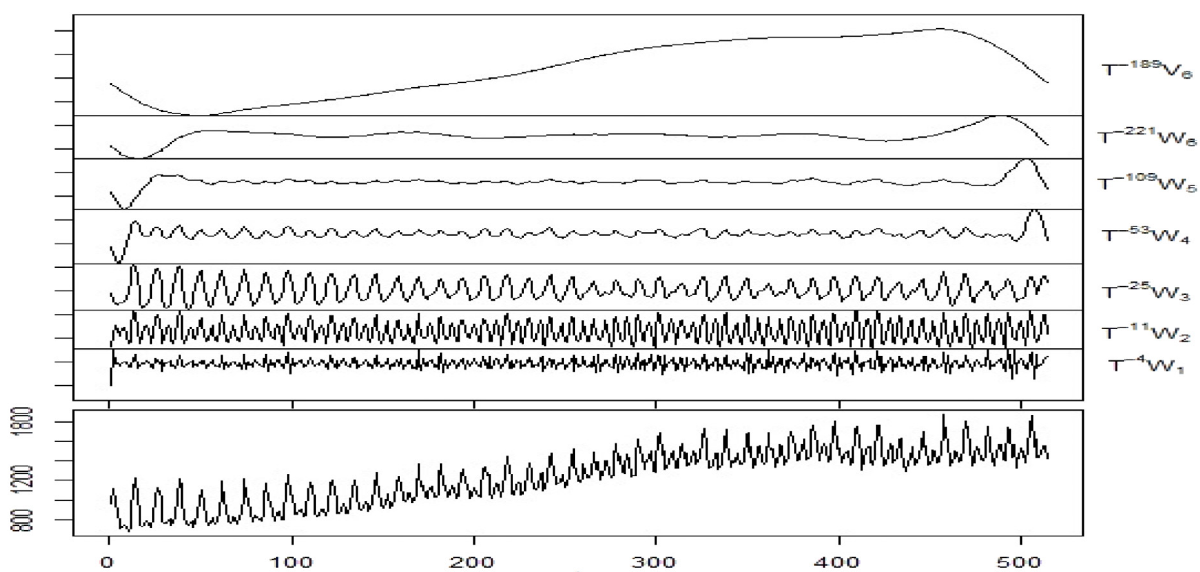

Fig. 11. MODWT decomposition of monthly commercial energy intake.

Table 4
Comparison of M-LSTM and LSTM.

Series	M-LSTM	LSTM
Training dataset		
Residential	0.00090	0.00237
Commercial	0.00076	0.00163
Industrial	0.00229	0.00559
Transportation	0.00086	0.00261
Test dataset		
Residential	0.00162	0.00321
Commercial	0.00105	0.00244
Industrial	0.00340	0.00693
Transportation	0.00157	0.00357

5.4. Comparative performance assessment

The proposed M-LSTM model is compared with the standalone LSTM model for comparing the efficacy. Table 4 reports the figures of MSE of both the models on training and test datasets.

It is amply apparent that the obtained MSE values of M-LSTM model are substantially lower than the standalone LSTM model on both training and test datasets. Hence, the inference can be drawn that in terms of predictive accuracy, M-LSTM outperforms standard LSTM model in forecasting monthly energy consumptions at selected sectors in the USA. Incorporation of Boruta feature selection algorithm and MODWT for granulation in the proposed M-LSTM model is largely responsible for the improvement of predictions. However, rationalizing the efficacy of the M-LSTM model based on the said comparison may not reflect the true nature of the performance.

For ascertaining the effectiveness of M-LSTM model over others, predictive performance must be compared with other well-known models. In this paper, six different models namely, ARIMA, DT, ANN, RNN, MARS, and bagging, are used for accomplishing the task. The same decomposition scheme through MODWT has been incorporated in all the six models. However, Boruta based feature engineering process has not been applied to these models. Mechanisms of ARIMA automatically include the explanatory features,

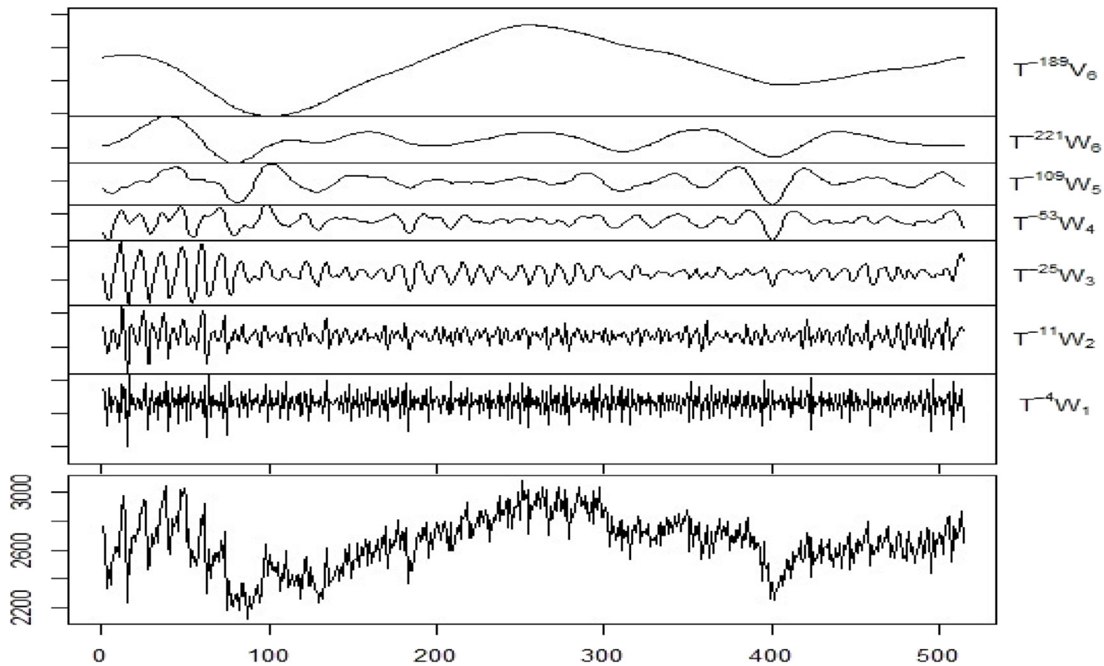

Fig. 12. MODWT decomposition of monthly industrial energy intake.

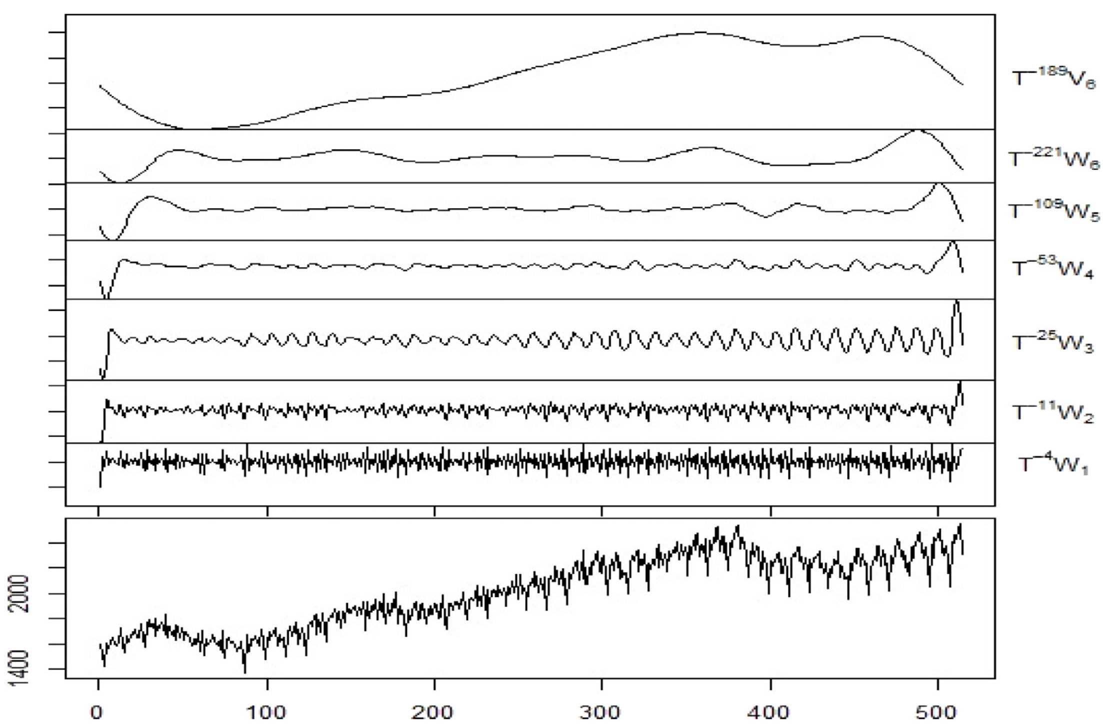

Fig. 13. MODWT decomposition of monthly transportation energy intake.

whereas one-month, two-month, three-month, four-month, and five-month lagged figures of energy consumption of the respective sectors have been chosen as independent features for DT, ANN, RNN, MARS, and bagging. ARIMA model has been implemented using the *'forecast'* package of R software. The standard *'auto.arima*()' function is used for estimating the parameters of the ARIMA model. For DT, the maximum depth of trees has been set to 21, while the minimum number of observations in leaf nodes has been fixed to 7. In ANN model, 1 hidden layer comprising of 50 hidden nodes, back-propagation approach through stochastic gradient descent (SGD) have been used. Sigmoid activation functions are used while the magnitude of learning rate has been set to 0.1. RNN modeling has been carried out through BPTT algorithm with two gated recurrent units of 50 neurons each. The second-order interactions have been considered for executing the MARS model. In implementing the bagging algorithm, 200 decision trees have been deployed for regression as base learners. All programs have been implemented using Intel Core i3-3110M processor (2.40 GHz) with 4.00 GB RAM in 64-bit Windows operating system. Table 5 presents the outcome of the comparative study manifested through MSE measure.

Table 5
Comparative performance assessment.

Series	M-LSTM	ARIMA	DT	ANN	RNN	MARS	Bagging
Training dataset							
Residential	0.00090	0.00813	0.00593	0.00596	0.00581	0.00671	0.00176
Commercial	0.00076	0.00587	0.00478	0.00502	0.00485	0.00513	0.00137
Industrial	0.00229	0.01163	0.00803	0.00879	0.00897	0.00909	0.00622
Transportation	0.00086	0.00694	0.00490	0.00474	0.00484	0.00516	0.00217
Test dataset							
Residential	0.00162	0.00895	0.00661	0.00689	0.00657	0.00792	0.00253
Commercial	0.00105	0.00634	0.00545	0.00580	0.00543	0.00583	0.00191
Industrial	0.00340	0.01301	0.00968	0.01026	0.01083	0.01112	0.00762
Transportation	0.00157	0.00757	0.00573	0.00577	0.00575	0.00625	0.00293

Table 6
The execution time (in second) of respective models.

Series	M-LSTM	ARIMA	DT	ANN	RNN	MARS	Bagging
Training dataset							
Residential	41.68	27.81	44.08	49.73	52.30	44.52	56.76
Commercial	40.92	25.43	43.89	48.06	49.80	44.38	54.30
Industrial	42.35	30.06	46.34	52.18	53.77	46.53	59.52
Transportation	40.51	26.79	42.25	49.44	50.64	45.27	57.84
Test dataset							
Residential	24.79	13.36	28.41	35.48	38.70	31.34	42.32
Commercial	23.84	12.49	29.76	34.19	37.26	30.63	40.19
Industrial	27.73	15.40	32.12	36.76	41.25	33.19	43.46
Transportation	22.66	13.17	27.65	36.81	38.69	32.47	42.39

From Table 5, it is evident that the MSE of the proposed M-LSTM model is much lower than the other models on both training and test datasets. Interestingly, among the contending models, the performance of Bagging is better than the rest of the measures in terms of MSE. The performance of ARIMA has emerged to be the least satisfactory among all the models as its MSE is higher compared to all other measures on both training and test datasets. Hence, the outcomes of the comparative tests justify the superiority of M-LSTM model over the others. Apart from performing comparative assessment in terms of predictive accuracy, comparisons are made on computational time of respective models as well. Table 6 reports the execution time of all the models.

It can be observed that except for the ARIMA model, the proposed M-LSTM has emerged to be computationally less expensive than the other five models. ARIMA has been found to be most superior in terms of less computational time primarily due to its simpler operational steps. In terms of comparison of MSE based predictive accuracy of M-LSTM and ARIMA on training samples, M-LSTM has emerged to be 9.03, 7.72, 5.08, and 8.07 times superior to ARIMA in residential, commercial, industrial, and transportation sectors, respectively. On the contrary, ARIMA is 1.5, 1.61, 1.41, and 1.51 times faster than M-LSTM for the respective sectors. A similar phenomenon has prevailed in test samples, as well. It should be noted that the present research aims to predictive analysis of macro-level energy consumption, which unlike micro-units, buildings, vehicles, etc. does not belong to real-time modeling of energy intake category. Achieving the higher precision is of paramount importance at the macro level energy modeling for long term strategic policymaking. Lower execution time is of more relevance in the context of macro units where short term decisions assist in energy conservation. Thus, inference can be drawn that the proposed M-LSTM approach has not been too computationally expensive yet has been successfully yielded superior forecasts for all the four sectors.

For drawing the statistical significance of predictive accuracy, the DM test is performed to further solidify the claim. Tables 7–10 summarize the outcome of the DM test.

The DM test works in a pairwise manner. Therefore, the models are stacked with the numbers indicating the order for the sake of better understanding. A significant positive test statistic value implies the performance of the second model is statistically superior to the first model. A significant negative test statistic value suggests the superiority of the first model over the second model. The magnitude of the obtained test statistics clearly indicates that M-LSTM statistically outperforms the other six models. Therefore, the effectiveness of incorporating dedicated feature engineering process through Boruta and deep learning capability of LSTM in modeling the pattern for estimating monthly energy consumption is duly justified. Bagging generates statistically better forecasts than ARIMA, DT, ANN, RNN, and MARS. There exists no significant difference in the performance of DT, ANN, RNN, and MARS. However, they all produce better forecasts than ARIMA. Therefore, the M-LSTM can be ranked the best among all these models followed by Bagging while ARIMA can be placed at the bottom in the list. The outcome of this test is unable to distinguish between the performance of DT, ANN, RNN, and MARS.

6. Conclusions

Predicting energy consumption is of great significance due to its implications on government policies, socio-economic factors, climatic changes, economic growth, etc. This is an extremely arduous task as the energy consumption time series often exhibits a high degree of random movements. This paper proposed M-LSTM, a hybrid deep learning approach for conducting predictive analytics of energy consumption at the macro level. The numerical testing is done on the monthly energy consumption of residential, commercial, industrial and transportation sectors of the USA. The major findings of this research are narrated below:

- The selection of technical indicators in time series theoretical framework has emerged as an excellent approach. They can be effectively used for estimating energy demands in other domains.
- Boruta feature selection algorithm, in conjunction with M-LSTM, produced supreme forecasts on both training and test datasets.
- The proposed M-LSTM model has outperformed models like ARIMA, ANN, DT, RNN, MARS, and bagging. This implies that M-LSTM is a suitable technique for estimating energy demands at the sectoral levels.
- Among the six competitive models, bagging has generated statistically better results than the rest five models. The ensemble nature of Bagging is largely accountable for this.
- In terms of predictions obtained by M-LSTM, an inference can be drawn that commercial and transportation sectors are comparatively more predictable than the other two sectors.
- M-LSTM has been found to be not too expensive in terms of computation time. ARIMA emerged to be computationally least expensive.

In a nutshell, the overall findings of the paper can be effectively exploited by the stakeholders for practical purposes. As residential and industrial sectors are discovered to be less predictable, the energy consumption levels are less stable than the commercial and transportation sectors. Therefore, the policymakers may reform regulations for sustainable developments and escalating developments of alternative renewable energy sources. The obtained results will help the energy investors and decision-makers, as well.

The scope of this paper is restricted to monthly energy demand forecasting. The hybrid granular forecasting model can be tested to estimate daily or hourly energy demand at various units to gain deeper insights into energy dynamics patterns. The paper uses MODWT for granular decomposition. The literature

Table 7
DM test on predictive analysis of residential energy consumption.

Models	ARIMA (1)	DT (1)	ANN (1)	RNN (1)	MARS (1)	Bagging (1)	M-LSTM (1)
ARIMA (2)	–						
DT (2)	3.86***	–					
ANN (2)	3.95***	0.25#	–				
RNN (2)	3.82***	0.25#	0.22#	–			
MARS (2)	3.86***	0.23#	0.23#	0.26#	–		
Bagging (2)	5.10***	5.12***	5.10***	5.26***	5.27***	–	
M-LSTM (2)	6.18***	6.14***	6.15***	6.24***	6.30***	4.17***	–

*** Significant at 1%, # Not significant.

Table 8
DM test on predictive analysis of commercial energy consumption.

Models	ARIMA (1)	DT (1)	ANN (1)	RNN (1)	MARS (1)	Bagging (1)	M-LSTM (1)
ARIMA (2)	–						
DT (2)	4.18***	–					
ANN (2)	4.20***	0.23#	–				
RNN (2)	3.86***	0.24#	0.19#	–			
MARS (2)	3.83***	0.22#	0.20#	0.23#	–		
Bagging (2)	5.35***	5.12***	5.19***	5.25***	5.27***	–	
M-LSTM (2)	6.24***	6.11***	6.16***	6.27***	6.24***	4.23***	–

*** Significant at 1%, # Not Significant.

Table 9
DM test on predictive analysis of industrial energy consumption.

Models	ARIMA (1)	DT (1)	ANN (1)	RNN (1)	MARS (1)	Bagging (1)	M-LSTM (1)
ARIMA (2)	–						
DT (2)	4.10***	–					
ANN (2)	4.07***	0.23#	–				
RNN (2)	4.11***	0.22#	0.19#	–			
MARS (2)	4.34***	0.22#	0.22#	0.191#	–		
Bagging (2)	4.21***	4.62***	4.64***	4.6356***	4.5779***	–	
M-LSTM (2)	6.39***	6.23***	6.26***	6.6619***	6.4321***	4.6503***	–

*** Significant at 1%, # Not significant.

Table 10
DM test on predictive analysis of transportation energy consumption.

Models	ARIMA (1)	DT (1)	ANN (1)	RNN (1)	MARS (1)	Bagging (1)	M-LSTM (1)
ARIMA (2)	–						
DT (2)	4.12***	–					
ANN (2)	4.21***	0.22#	–				
RNN (2)	4.22***	0.24#	0.21#	–			
MARS (2)	3.98***	0.21#	0.23#	0.21#	–		
Bagging (2)	4.88***	4.78***	4.79***	4.76***	4.67***	–	
M-LSTM (2)	6.61***	6.32***	6.26***	6.26***	6.31***	4.42***	–

*** Significant at 1%, # Not significant.

reports the usage of EEMD for time series prediction at granular level. So, the comparative performance of EEMD and M-LSTM may be carried out. The other deep learning techniques barring LSTM, such as GAN, convolutional neural network, deep belief network, and restricted Boltzmann machine may be used in granular framework by applying different time series decomposition approaches.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.asoc.2020.106091.

Acknowledgments

We are grateful to the editors and anonymous reviewers for offering valuable suggestions that greatly improved the presentation of this article.

References

[1] E. Aydin, D. Brounen, The impact of policy on residential energy consumption, Energy (2019) http://dx.doi.org/10.1016/j.energy.2018.12.030.
[2] Y. Cai, C.Y. Sam, T. Chang, Nexus between clean energy consumption, economic growth and CO2 emissions, J. Cleaner Prod. (2018) http://dx.doi.org/10.1016/j.jclepro.2018.02.035.
[3] M.S. Gorus, M. Aydin, The relationship between energy consumption, economic growth, and CO 2 emission in MENA countries: Causality analysis in the frequency domain, Energy (2019) http://dx.doi.org/10.1016/j.energy.2018.11.139.

[4] P. Agnolucci, V. De Lipsis, T. Arvanitopoulos, Modelling UK sub-sector industrial energy demand, Energy Econ. (2017) http://dx.doi.org/10.1016/j.eneco.2017.08.027.
[5] C. Fan, J. Wang, W. Gang, S. Li, Assessment of deep recurrent neural network-based strategies for short-term building energy predictions, Appl. Energy (2019) http://dx.doi.org/10.1016/j.apenergy.2018.12.004.
[6] B. Mahalingam, W.H. Orman, GDP and energy consumption: A panel analysis of the US, Appl. Energy (2018) http://dx.doi.org/10.1016/j.apenergy.2018.01.036.
[7] S. Sen, H. Vollebergh, The effectiveness of taxing the carbon content of energy consumption, J. Environ. Econ. Manage. (2018) http://dx.doi.org/10.1016/j.jeem.2018.08.017.
[8] T.M.D. Mohr, Fuel poverty in the US: Evidence using the 2009 residential energy consumption survey, Energy Econ. (2018) http://dx.doi.org/10.1016/j.eneco.2018.06.007.
[9] M. Shen, H. Sun, Y. Lu, Household electricity consumption prediction under multiple behavioural intervention strategies using support vector regression, Energy Procedia (2017) http://dx.doi.org/10.1016/j.egypro.2017.12.218.
[10] L.A. Laboissiere, R.A.S. Fernandes, G.G. Lage, Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks, Appl. Soft Comput. J. (2015) http://dx.doi.org/10.1016/j.asoc.2015.06.005.
[11] M.B. Kursa, W.R. Rudnicki, Feature selection with the boruta package, J. Stat. Softw. (2015) http://dx.doi.org/10.18637/jss.v036.i11.
[12] J.L. Ticknor, A Bayesian regularized artificial neural network for stock market forecasting, Expert Syst. Appl. (2013) http://dx.doi.org/10.1016/j.eswa.2013.04.013.
[13] C.H. Su, C.H. Cheng, A hybrid fuzzy time series model based on ANFIS and integrated nonlinear feature selection method for forecasting stock, Neurocomputing (2016) http://dx.doi.org/10.1016/j.neucom.2016.03.068.
[14] Y. Zhao, J. Li, L. Yu, A deep learning ensemble approach for crude oil price forecasting, Energy Econ. 66 (2017) 9–16, http://dx.doi.org/10.1016/j.eneco.2017.05.023.
[15] I. Ghosh, R.K. Jana, M.K. Sanyal, Analysis of temporal pattern, causal interaction and predictive modeling of financial markets using nonlinear dynamics, econometric models and machine learning algorithms, Appl. Soft Comput. J. (2019) http://dx.doi.org/10.1016/j.asoc.2019.105553.
[16] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time Series Analysis: Forecasting and Control: Fourth Edition, 2013, http://dx.doi.org/10.1002/9781118619193.
[17] J.R. Quinlan, Induction of decision trees, Mach. Learn. (1986) http://dx.doi.org/10.1023/A:1022643204877.
[18] G. Montavon, G.B. Orr, K.-R. Mueller, Neural Networks: Tricks of the Trade - Second Edition, 2012, http://dx.doi.org/10.1007/978-3-642-35289-8_34.
[19] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, J. Schmidhuber, A novel connectionist system for unconstrained handwriting recognition, IEEE Trans. Pattern Anal. Mach. Intell. (2009) http://dx.doi.org/10.1109/TPAMI.2008.137.
[20] J.H. Friedman, Multivariate adaptive regression splines, Ann. Statist. 19 (1991) 1–67, http://dx.doi.org/10.1214/aos/1176347963.
[21] N. Simidjievski, L. Todorovski, S. Džeroski, Predicting long-term population dynamics with bagging and boosting of process-based models, Expert Syst. Appl. (2015) http://dx.doi.org/10.1016/j.eswa.2015.07.004.
[22] J. Li, L. Yang, H. Long, Climatic impacts on energy consumption: Intensive and extensive margins, Energy Econ. (2018) http://dx.doi.org/10.1016/j.eneco.2018.03.010.
[23] C. Chen, Y. Liu, M. Kumar, J. Qin, Energy consumption modelling using deep learning technique - a case study of EAF, in: Procedia CIRP, 2018, http://dx.doi.org/10.1016/j.procir.2018.03.095.
[24] T. Papadimitriou, P. Gogas, E. Stathakis, Forecasting energy markets using support vector machines, Energy Econ. (2014) http://dx.doi.org/10.1016/j.eneco.2014.03.017.
[25] A. Mirakyan, M. Meyer-Renschhausen, A. Koch, Composite forecasting approach, application for next-day electricity price forecasting, Energy Econ. (2017) http://dx.doi.org/10.1016/j.eneco.2017.06.020.
[26] S. Al Qadi, B. Sodagar, A. Elnokaly, Estimating the heating energy consumption of the residential buildings in Hebron, Palestine, J. Cleaner Prod. (2018) http://dx.doi.org/10.1016/j.jclepro.2018.06.059.
[27] Z. Farajzadeh, M.A. Nematollahi, Energy intensity and its components in Iran: Determinants and trends, Energy Econ. (2018) http://dx.doi.org/10.1016/j.eneco.2018.05.021.
[28] S. Forin, A. Radebach, J.C. Steckel, H. Ward, The effect of industry delocalization on global energy use: A global sectoral perspective, Energy Econ. (2018) http://dx.doi.org/10.1016/j.eneco.2017.12.031.
[29] S. Gharehbaghi, A.H. Gandomi, S. Achakpour, M.N. Omidvar, A hybrid computational approach for seismic energy demand prediction, Expert Syst. Appl. (2018) http://dx.doi.org/10.1016/j.eswa.2018.06.009.
[30] Y. Hu, J. Yan, D. Yan, Q. Lu, J. Yan, Lightweight energy consumption analysis and prediction for Android applications, Sci. Comput. Program. (2018) http://dx.doi.org/10.1016/j.scico.2017.05.002.
[31] J. Li, R.E. Just, Modeling household energy consumption and adoption of energy efficient technology, Energy Econ. (2018) http://dx.doi.org/10.1016/j.eneco.2018.04.019.
[32] Z. Wang, Y. Wang, R. Zeng, R.S. Srinivasan, S. Ahrentzen, Random forest based hourly building energy prediction, Energy Build. (2018) http://dx.doi.org/10.1016/j.enbuild.2018.04.008.
[33] J. Xiao, Y. Li, L. Xie, D. Liu, J. Huang, A hybrid model based on selective ensemble for energy consumption forecasting in China, Energy (2018) http://dx.doi.org/10.1016/j.energy.2018.06.161.
[34] F. Belaïd, D. Roubaud, E. Galariotis, Features of residential energy consumption: Evidence from France using an innovative multilevel modelling approach, Energy Policy (2019) http://dx.doi.org/10.1016/j.enpol.2018.11.007.
[35] I. Ghosh, M.K. Sanyal, R.K. Jana, An integrated approach incorporating nonlinear dynamics and machine learning for predictive analytics and delving causal interaction, in: V. Bhateja, C. Coello Coello, S. Satapathy, P. Pattnaik (Eds.), Intelligent Engineering Informatics, in: Advances in Intelligent Systems and Computing, vol. 695, Springer, Singapore, 2018, http://dx.doi.org/10.1007/978-981-10-7566-7_29.
[36] I. Ghosh, M.K. Sanyal, R.K. Jana, Fractal inspection and machine learning-based predictive modelling framework for financial markets, Arab. J. Sci. Eng. (2018) http://dx.doi.org/10.1007/s13369-017-2922-3.
[37] W.R. Rudnicki, M. Kierczak, J. Koronacki, J. Komorowski, A statistical method for determining importance of variables in an information system, 2006, http://dx.doi.org/10.1007/11908029_58.
[38] R. Sircar, An introduction to wavelets and other filtering methods in finance and economics, Waves Random Media (2002) http://dx.doi.org/10.1088/0959-7174/12/3/701.
[39] R. Gençay, F. Selçuk, B. Whitcher, Multiscale systematic risk, J. Int. Money Financ. (2005) http://dx.doi.org/10.1016/j.jimonfin.2004.10.003.
[40] D. Das, P. Bhowmik, R.K. Jana, A multiscale analysis of stock return co-movements and spillovers: Evidence from Pacific developed markets, Physica A 502 (2018) 379–393, http://dx.doi.org/10.1016/j.physa.2018.02.143.
[41] C.R. Cornish, C.S. Bretherton, D.B. Percival, Maximal overlap wavelet statistical analysis with application to atmospheric turbulence, Bound.-Layer Meteorol. (2006) http://dx.doi.org/10.1007/s10546-005-9011-y.
[42] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. (1997) http://dx.doi.org/10.1162/neco.1997.9.8.1735.
[43] T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, European J. Oper. Res. (2018) http://dx.doi.org/10.1016/j.ejor.2017.11.054.
[44] D.P. Kingma, J.L. Ba, Adam: A method for stochastic gradient descent, ICLR Int. Conf. Learn. Represent. (2015).
[45] J.E. Nash, J.V. Sutcliffe, River flow forecasting through conceptual models part I - A discussion of principles, J. Hydrol. (1970) http://dx.doi.org/10.1016/0022-1694(70)90255-6.
[46] C.J. Willmott, D.E. Wicks, An empirical method for the spatial interpolation of monthly precipitation within california, Phys. Geogr. (1980) http://dx.doi.org/10.1080/02723646.1980.10642189.
[47] E. Ruist, H. Theil, Applied economic forecasting, Swedish J. Econ. (1967) http://dx.doi.org/10.2307/3439127.
[48] L. Chen, Z. Qiao, M. Wang, C. Wang, R. Du, H.E. Stanley, Which artificial intelligence algorithm better predicts the chinese stock market? IEEE Access (2018) http://dx.doi.org/10.1109/ACCESS.2018.2859809.
[49] F.X. Diebold, R.S. Mariano, Comparing predictive accuracy, J. Bus. Econom. Statist. (1995) http://dx.doi.org/10.1080/07350015.1995.10524599.

Rabin K. Jana is an Assistant Professor of Operations & Quantitative Methods at Indian Institute of Management Raipur, India. He has earned his Ph.D. from IIT Kharagpur, India. He was a postdoctoral researcher at George Mason University, USA, and National University of Singapore. Dr. Jana is a senior member of Operational Research Society of India, a member of Decision Sciences Institute, USA, and Indian Statistical Institute, Kolkata. His research interest includes optimization, artificial intelligence, machine learning, and time series forecasting.

Indranil Ghosh is an Assistant Professor of Operations Management & IT at Calcutta Business School, India. He has earned his B.E. degree in Information Technology and M. Tech degree in Industrial Engineering & Management from WBUT, India. His areas of interest are data mining & pattern recognition, AI techniques, statistical data analysis, design and analysis of algorithm, supply chain management & ERP, image processing & computer vision, and social network analysis.

Manas K. Sanyal is a Professor in the Department of Business Administration, University of Kalyani, India. He has earned his M. Tech degree and Ph.D. in Information Technology. Professor Sanyal has published several research papers in international journals of repute and co-authored a number of books.