Applied Soft Computing Journal 93 (2020) 106390

Contents lists available at ScienceDirect

Applied Soft Computing Journal

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

Energy load time-series forecast using decomposition and autoencoder integrated memory network

Jatin Bedi [1,*], Durga Toshniwal [2]

Department of Computer Science and Engineering, Indian Institute of Technology Roorkee, Uttarakhand, India

A R T I C L E I N F O

Article history:
Received 21 July 2019
Received in revised form 30 March 2020
Accepted 7 May 2020
Available online 14 May 2020

Keywords:
Load demand forecasting
Seasonal analysis
Long short term memory network
Variational Mode Decomposition

A B S T R A C T

With the increasing population and rising living standard, the demand for energy and materials have increased to a greater extent. The accurate estimation of increasing electricity demand is prerequisite for strategies planning, improving revenue, reducing power wastage and stable operation of the energy demand management system. Recent advancements in the field of electricity load forecasting provide powerful tools to capture non-linear energy demand trends and outperform conventional load prediction models. However, the existing demand prediction models suffer from some significant shortcomings that need to be addressed for improved prediction accuracy. In this context, the current research work proposes a deep learning based hybrid approach which firstly implements Variational Mode Decomposition (VMD) and Autoencoder models to extract meaningful sub-signals/features from the data. Subsequently, a Long Short-term Memory (LSTM) network model is trained for each sub-signal to forecast electricity demand by utilizing historical, seasonal and timestamp data dependencies. The support for incorporating seasonal and timestamp information to LSTM model is provided through the agglomerative clustering algorithm. Furthermore, an error variance modelling strategy is also employed to enhance the prediction accuracy of the proposed approach. The experiments are conducted on electricity consumption dataset of Himachal Pradesh, India. Performance assessment of the proposed approach is made by comparing prediction results with Support Vector Regression (SVR), Recurrent Neural Network (RNN), Deep Belief Network (DBN) and EMD+LSTM models. The experimental results demonstrate that the proposed model outperforms other state-of-the-art demand forecasting models and has the lowest MAPE (3.04%).

1. Introduction

Energy plays a vital role in the social and economic development of society [1,2]. The energy sector has witnessed a rapid growth of the energy demand in the last few decades. The increasing energy demand has been driven by several factors including urbanization, increasing population, increase in the availability of energy consuming devices & appliances, climate changes and globalization. It has been estimated that if this increasing demand trend persists, the energy demand will increase by more than 50% before the year 2030 [3]. Therefore, efficient energy demand management systems are required to provide reliable energy supply, efficient resources usage and energy conservation while keeping operational costs as low as possible.

Smart grids [4] are the system that support various features such as demand management, usage monitoring and demand-supply synchronization. They are also responsible for maintaining a database that records the energy consumption or usage pattern, which in turn are used to forecast the energy demand patterns. The accurate estimation of electricity demand is a crucial step for economic investment management, system reliability, better planning of future strategies, grid construction planning and resource usage management. Depending on the required time horizon of prediction, the energy demand forecasting can be performed at three different levels [5,6]:

- Long-term demand forecasting (*years ahead forecasting, better for capacity planning and generation system installation planning*).
- Medium-term demand forecasting (*one month to 12 months ahead demand forecasting, better for reservoir modelling and scheduling*).
- Short-term demand forecasting (*day to week ahead load demand forecasting, better for daily operations and purchase planning*).

* Corresponding author.
E-mail addresses: jbedi@cs.iitr.ac.in (J. Bedi), durgatoshniwal@gmail.com (D. Toshniwal).
1 S-109, Azad Bhawan, IIT Roorkee, Haridwar, 247667
2 IIT Roorkee, Haridwar, 247667

https://doi.org/10.1016/j.asoc.2020.106390
1568-4946/

Traditionally, various quantitative and qualitative tools have been successfully utilized to forecast electricity demand. These methods implement Delphi curve fitting technique [7] and physics-based dynamic behaviour modelling techniques (white-box methods) to estimate energy demand. In the last two decades, various research studies have motivated the use of data-driven models to capture non-linear trends of energy demand patterns such as Support Vector Regression (SVR) [8], decision-trees, fuzzy-logic, Artificial Neural Networks (ANN) [9], random forest [10] and gradient boosting [11]. Although these methods are very successful at capturing non-linear trends of data, handling historical dependencies in the data require more sophisticated techniques. Recently, several deep learning techniques such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) [12], Deep Belief Network (DBN) [13] have been implemented to effectively forecast the energy demand while taking care of data dependencies. Furthermore, various hybrid approaches that combine signal processing techniques such as wavelet transform, mode decomposition algorithms with deep neural network models [14–17] have also gained a lot of interest for energy demand modelling. The experimental results of various research studies have shown that the hybrid approaches provide more accurate and reliable prediction results than those of other existing state-of-the-art data-driven techniques [18–20]. However, they suffer from several issues such as high computational complexity, high dependency on shifting procedure, handling long-term data dependencies, deciding optimal mode component for noise reduction and so on.

The present research work proposes a multilevel systematic hybrid approach that combines VMD [21] algorithm and Autoencoders model [22] with LSTM multi-input-multi-output network model [23] and error variance modelling strategy [24] for improved prediction accuracy. In addition to this, the proposed approach also provides support for handling the long-run data dependencies by performing timestamp (day-wise) and seasonal analysis. To verify the effectiveness and reliability of the proposed approach, the experimental results are obtained on the ECD data of Himachal Pradesh (HP), India. HP [1] is one of those states in India that has rapidly transformed from the most backward states of the country to one of the advanced states. The state is a rich resource of hydroelectricity due to the presence of perennial rivers. It has an approx. 25% of the national potential in this aspect. The comparative analysis of the experimental results on the HP ECD data shows that the proposed approach surpasses the state-of-the-art demand prediction models.

The remaining of this paper is organized as follows: Section 2 describes the related work done in the field of energy load forecasting. Section 3 details the theoretical background of basic concepts used in the present study. The detailed methodology of the proposed multilevel systematic hybrid approach is given in Section 4. Section 5 demonstrates the application and key results of the proposed approach on the electricity consumption dataset of Himachal Pradesh, India. Finally, Section 6 concludes the paper.

2. Literature review

Energy load data is defined as the time-series with demand observations at equal time intervals [8]. Load time-series prediction models deal with estimating future timestamp demand based on the series of past demand observations. The accurate demand estimation is required for reducing power waste, operative stability of the power grid systems, improving revenue, reducing distribution efficiency and various investment decisions [25]. Since the year 1940s, a wide variety of approaches have been successfully implemented to estimate electricity demand. These approaches can be broadly classified into four major categories: Engineering or traditional models, Statistical or regression models, Artificial Intelligence (AI) based methods and Hybrid models.

2.1. Demand prediction from traditional methods

Traditional demand prediction models are based on solving a set of equations for modelling the behaviour of power grid system [7,26,27] . These methods are also called "white-box methods" as the internal logic is very clear. A number of engineering tools [28–30] have been developed for analysing the energy consumption and conservation measures based on physical behaviour and system-environment interactions. Even though these empirical tools are easy to build and use, the high dependency on input parameters and expert knowledge is a big issue.

2.2. Statistical or regression based prediction models

In contrast to traditional methods, statistical methods avoid any need for prior information. They are based on estimating future demand from the set of influencing variables. Statistical methods including Auto-regressive (AR), Auto-regressive moving average (ARMA), Auto-regressive integrated moving average (ARIMA), linear regression and multi-linear regression models have been widely adopted for time series prediction. In [31], authors used ARIMA model to forecast electricity demand from monthly and quarterly load demand data of Greece. The prediction results proved that the model works well with monthly data by generating a MAPE of 3.78%. Rallapalli [32] performed a comparative analysis between multiplicative seasonal ARIMA model and forecasting technique used by the central electricity authority of India. The experimental results have shown the improved accuracy of multiplicative seasonal ARIMA model.

In [33], authors presented least-square linear regression and reverse error-in variables based methodology to estimate electricity load. Chen et al. [34] developed a multi-linear regression based approach to determine the impact of behaviour variables on energy consumption patterns. Furthermore, Fumo et al. [35] assessed the importance of different regression (linear and quadratic) models and influencing variables in the task of load forecasting. The experimental results of these approaches demonstrated the significance of behaviour and socio-economic variables on the energy consumption variations. Contrary to the physical models, statistical models yield more accurate and promising prediction results. But, they are incapable of dealing with non-linear complexities and fluctuations of the load time series data.

2.3. AI-based demand prediction models

In the past two decades, AI-based methods including ANN, decision trees and SVR have been broadly adopted to forecast electricity demand. These methods have been found very effective in handling non-linearity of the load time series data. This section reviews the applications of AI-based methods in the domain of load forecasting.

Rodrigues et al. [9] developed an ANN-based methodology to estimate electricity demand at the household level. Several input factors such as apartment area, number of occupants and electrical appliances are considered to predict demand. Fu et al. [36] performed a comparative analysis of different demand prediction techniques, including SVM, ANN, ARIMA. Later in the year 2017, Hamzacebi [37] presented an approach to forecast monthly electricity demand of Turkey while considering seasonal trend effects.

In the past few years, various deep learning-based models are being developed and deployed in the field of time series prediction and classification. Dedinec et al. [13] implemented deep belief network with back-prorogation for hourly load forecasting. The approach provided a reduced MAPE of 8.6%. Tong et al. [4]

introduced a novel deep learning-based methodology for day-ahead load demand forecasting. The model combined stacked auto-encoder (for feature extraction) with SVR to forecast load. Chen et al. [38] presented a multi-stage strategy with Monte Carlo dropout to improve the generalization capability of deep neural network based prediction model. Out of various deep learning models, Recurrent Neural Network (RNN) models have been found as the most promising tool to capture the non-linear trends of energy load profiles. A large number of research studies [12,39–41] have successfully implemented recurrent network models to predict electricity load. Bedi et al. [42] proposed a variant of the LSTM model, i.e. LSTM multi-input-multi-output model to forecast electricity demand for the specific time interval of the day. The approach provided support for two major advantages: (a) Firstly, it implemented a stateful LSTM-MIMO model while utilizing memory between the batches and optimal batch_size parameters to support improved accuracy than basic window model. (b) Secondly, the approach provides support for active learning by replacing predicted values with real-time demand observations.

2.4. Hybrid demand prediction models

Hybrid demand prediction models strategically combine conventional methods with AI-based techniques for improved prediction accuracy. This section reviews the applications of hybrid models in the field of load forecasting. Hong et al. [43] used SVR in combination with evolutionary techniques (genetic algorithms) to estimate electricity demand. Chen et al. [44] integrated K-nearest neighbour with improved SVR for better forecasting. Furthermore, in the past few years, hybridization of mode decomposition techniques with AI-based methods has proven to improve the performance of electricity load forecasting models. Several research studies have combined wavelet transform [14,15,45] with AI-based methods including ANN, ANFIS, and SVM to forecast day-ahead electricity load. Empirical Mode Decomposition (EMD) has also been broadly adapted to forecast electricity demand. The integration of EMD with deep learning techniques [6,16,46,47] [17,48] have shown improved performance in comparison of other single and wavelet-based prediction models.

Hybrid demand prediction models overcome the shortfalls of traditional and AI-based prediction models and have shown the potential to enhance the performance of existing load prediction models [2,18–20]. However, there are several associated issues that need to be addressed, such as high computational complexity, determining optimal lag value, deciding optimal mode component for noise reduction and so on. The present work addresses all issues mentioned above while extending our previous work [6,42] done in the field of electricity load forecasting (by utilizing the benefits of both existing approaches). The current research work proposes a multilevel hybrid architecture for estimating electricity load. The approach combines LSTM-MIMO network model with mode decomposition (VMD) and optimal feature subset generation (auto-encoder) algorithms. To further improve the architecture, an error variance modelling strategy is adopted to analyse the interdependence of error components. The performance of the proposed approach is evaluated on HP electricity consumption dataset. The prediction results are compared with existing state-of-the-art prediction models: SVR, RNN, DBN, and EMD+LSTM.

The research gaps and major contributions of the study can be described as follows:

- The recent studies in the field of energy load forecasting have implemented EMD and wavelet transform to decompose load time-series signals into corresponding sub-signals or modes. However, these decomposition algorithms suffer from some significant flaws that need to be addressed, such as high dependence on shifting algorithm, noise sensitivity and finding extremal points. The present research work implements VMD to overcome the limitations of existing mode decomposition algorithms. The proposed approach provides support for handling fluctuations in the original time series by generating an ensemble of nodes band-limited around the specified centre pulsation frequency.
- The existing hybrid time-series forecasting models which implement mode decomposition algorithms perform decomposition of an input signal into several sub-signals or modes. Each extracted sub-signal is then modelled from an independent learning/prediction model. The final prediction output is given by the summation of the results from each prediction model. This process increases the computational complexity of the approach as there is a need to train 'N' prediction models corresponding to 'N' sub-signals. To address this, the present work employs auto-encoder models to generate a reduced representation of the input sub-signals, thereby reducing the computational complexity of the proposed approach.
- The LSTM network model makes use of previous timestamp demand values (lag values) to estimate current timestamp demand. The choice of optimal lag parameter value could significantly affect the LSTM model accuracy. The recent works in the field of load forecasting have not covered this aspect of LSTM network model. The current research work proposes a novel method to determine the optimal lag parameter value effectively.
- The existing demand prediction models have not considered the effect of heteroskedasticity on model accuracy and reliability. The inclusion of an error-variance modelling strategy might positively contribute to the overall model performance. So, the present research work implements an Error Variance Modelling Strategy (EVM-S) to improve the model accuracy.

3. Related concepts

This section briefly discusses the concepts (Variational mode decomposition, Autoencoder and Long short-term memory network) used in the present study.

3.1. Variational mode decomposition (VMD)

VMD [21] is a non-recursive decomposition algorithm that works by decomposing an input time series signal into sub-signals or modes ($u_m : m = 1 : n$ (*number of modes*)) such that each sub-signal is band-limited around the center pulsation ω_m (determined with the decomposition). The method concurrently looks for an ensemble of modes that collectively regenerates the input signal optimally or in a least-square manner.

In the present study, we implement VMD to decompose an input energy demand time series signal (I_t) to their respective modes ($u_m : m = 1 : n$ (*number of modes*)), where u_m denotes the m-th mode and ω_m represents the center frequency of the m-th mode. The process of decomposing an input signal into sub-signals or modes using VMD is as follows [21]:

- Step A: Implement Hilbert Transform to compute the analytic signal of each mode. Formally it is given as:

$$\left\{ \nabla(t) + \frac{j}{\pi . t} \right\} * u_m(t) \tag{1}$$

where ∇ denotes the Dirac distribution.

- Step B: Use the Center frequencies to shift/modulate the spectrum of computed analytic signals:

$$\left[\left\{\nabla(t)+\frac{j}{\pi.t}\right\}*u_m(t)\right].\,e^{-jw_mt} \tag{2}$$

- Step C: Calculate L^2-norm of the demodulated signal for Bandwidth estimation. The resulting constrained optimization problem is given as:

$$min_{\{u_m\},\{w_m\}}\left\{\sum_n\left\|\partial(t)\left[\left\{\nabla(t)+\frac{j}{\pi.t}\right\}*u_m(t)\right].\,e^{-jw_mt}\right\|\right\}$$
$$s.t.\sum_n u_m=I \tag{3}$$

- Step D: The constrained optimization problem can be transformed to unconstrained form by introducing quadratic penalty (to provide reconstruction fidelity) and Lagrangian multiplier ($\mathcal{L}$: to enforce constraints strictly). Formally, it is given as:

$$\mathcal{L}(\{u_k\},\{w_k\},\lambda)=\alpha\left\{\sum_n\left\|\partial(t)\left[\left\{\nabla(t)+\frac{j}{\pi.t}\right\}*u_m(t)\right].\,e^{-jw_mt}\right\|\right\}$$
$$+\ \|I(t)-\sum_k u_k(t)\|_2^2+\left\langle\lambda(t),I(t)-\sum_k u_k(t)\right\rangle \tag{4}$$

- Step E: Implement ADMM (Alternate direction method of multipliers) to update modes, centre frequencies, λ and to provide solution to the optimization problem:

$$\widehat{u}_m^{N+1}(\omega)=\frac{\widehat{I}(\omega)-\sum_{i\neq m}\hat{u}_i(\omega)+\frac{\widehat{\lambda}(\omega)}{2}}{1+2\alpha(\omega-\omega_m)^2} \tag{5}$$

$$\omega_m^{N+1}=\frac{\int_0^\infty\omega|\widehat{u}_m(\omega)|^2d\omega}{\int_0^\infty|\widehat{u}_m(\omega)|^2d\omega} \tag{6}$$

$$\widehat{\lambda}^{N+1}(\omega)=\widehat{\lambda}^N(\omega)+\tau\left(\widehat{I}(\omega)-\sum_{m=1}^n\widehat{u}_m^{N+1}(\omega)\right) \tag{7}$$

where $\widehat{I}(\omega)$, $\widehat{\lambda}(\omega)$ and $\widehat{u}(\omega)$ denotes the Fourier transform of the $I(\omega)$, $\lambda(\omega)$ and $u^{N+1}(\omega)$ respectively.

3.2. Autoencoders

Autoencoders [22] are the special kind of ANN which aims to provide a complexed knowledge representation or approximation of the input i.e. the network tries to generate an output $\widehat{I}$ that is similar to the input I. The overall network architecture of the auto-encoder (as shown in Fig. 1) can be divided into two parts/phases [22,49]:

- Encoder: It aims to generate the generic features of the input vector I and is given by:

$$h=f(I)=\sigma(W.I+b) \tag{8}$$

- Decoder: It aims at reconstructing the input from the abstract representation given by encoder function:

$$\vec{I}=g(h)=\sigma(W'.I+b') \tag{9}$$

where $W\&b$ denotes the weights and bias values for encoder model and $W'\&b'$ represents the weights and bias values for decoder model. The network goes on iterating between the two phases while minimizing the loss function through back-propagation given by:

$$L(I,g(f(I))):\textit{Function that Penalizes for dissimilar Input and Output}. \tag{10}$$

In order to constrain autoencoders model for just copying the input to output without learning any useful information about data distribution or to avoid model over-fitting, a regularization parameter is added to the Loss term and it is given as:

$$L(I,g(f(I)))+\Omega(h,I)\quad\simeq\quad L(I,g(f(I)))+\lambda\sum_j\|\nabla.\,h_j\|^2 \tag{11}$$

3.3. Long short term memory network

RNNs [50] are the special type of ANN that consider this aspect of learning by utilizing previous timestamps demand observations for estimating current load demand. However, due to the vanishing gradient problem they are capable of handling short-term data dependencies only. The issue get resolved with the introduction of a variant of RNN i.e. Long Short-Term Memory Network [51] model. LSTM architecture consists of memory cells and gates: Input, forget and Output gate. The forget gate (f_t) is responsible for removing the information that is no longer required or is irrelevant and is given by [51]:

$$f_t=\sigma(W_f.[x_t,\ h_{t-1}]+b_f) \tag{12}$$

The input gate (i_t) handles addition of new information to the current cell state and is given by [51]:

$$i_t=\sigma(W_i.[h_{t-1},\ x_t]+b_i) \tag{13}$$

The corresponding vector of new candidate values ($\tilde{C}_t$) and updated cell state (C_t) is given by [51]:

$$\tilde{C}_t=tanh(W_C.[h_{t-1},\ x_t]+b_C) \tag{14}$$

$$C_t=f_t\odot C_{t-1}\oplus i_t\odot\tilde{C}_t \tag{15}$$

The output gate (o_t) controls information flow and the corresponding output of memory cell is given as [50,51]:

$$o_t=\sigma(W_o[h_{t-1}\ x_t.]+b_0) \tag{16}$$

$$h_t=o_t\odot tanh(c_t) \tag{17}$$

where x_t denotes current input, h_t represents the output, C_t represents cell state at timestamp t, $\oplus$ represents element-wise summation op, $\odot$ represents element-wise multiplication operator, σ represents sigmoid function, W_i, W_f, W_C represent weights and b_i, b_f, b_C represent bias values.

4. Proposed methodology

In this section, we explain the methodology of the proposed load time-series forecasting approach. The multi-level hybrid architecture of the proposed approach is shown in Fig. 2. The step by step flow of data in each component of the multi-level architecture (Fig. 2) is explained as follows:

- **Step 1:** Initially, the hierarchical agglomerative clustering algorithm with DTW distance measure is applied to the pre-processed input load time-series data to generate season based segmented data. Subsequently, the load trend characterization is performed to include day-stamp information to the model (detailed explanation in Sections 4.2 and 4.3).
- **Step 2a:** For each of the extracted seasonal day stamp load time series, the VMD algorithm is applied to extract the N-mode components (mod_1, mod_2, mod_3, …, mod_n).
- **Step 2b:** The autoencoder model is implemented to generate latent space representation of the mode components extracted (mod_2, mod_3, …, mod_n) from Step 2a (detailed explanation in Section 4.4).

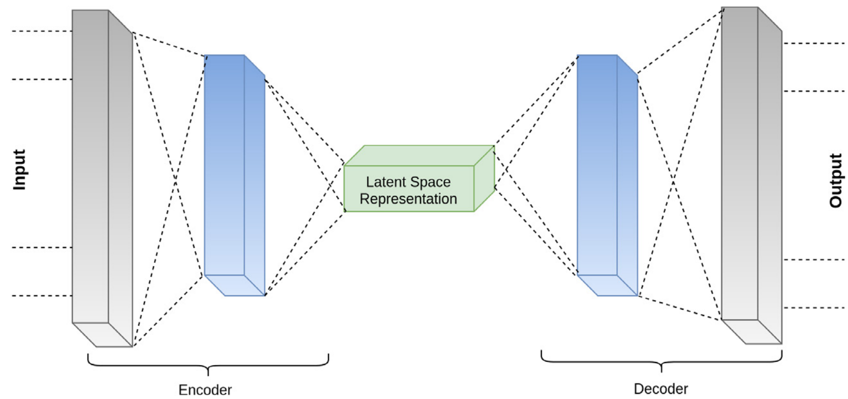


Fig. 1. Architecture of auto-encoders network.

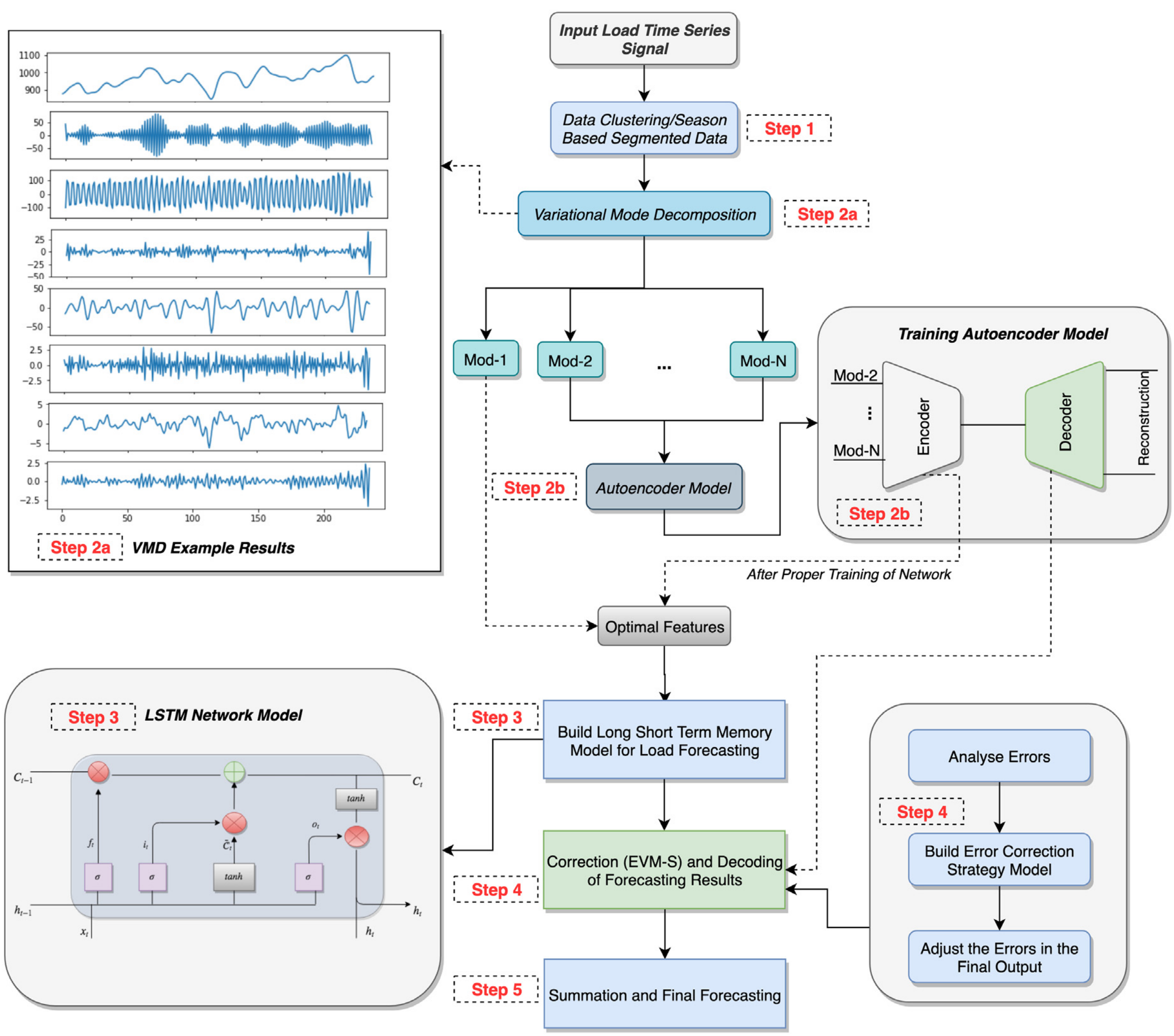


Fig. 2. Flowchart of the proposed model.

- **Step 3:** The optimal features set generated from step 2b along with the mod_1 from step 2a are fed to the LSTM models for load time series forecasting (detailed explanation in Section 4.5).
- **Step 4:** Finally, based on the variance of the forecasting results obtained from step 3, GARCH error variance modelling strategy is adopted to improve the prediction accuracy. (detailed explanation in Section 4.6).

To demonstrate the applicability and reliability of the proposed approach, it is applied to the energy load demand dataset of Himachal Pradesh (HP), India. An in-depth explanation and results of the proposed approach on HP dataset are explained in Section 5.

4.1. Data preprocessing

Data mining [8,52] is defined as the process of extracting valuable information from the large amount of raw data. In the real world scenarios, the raw data collected by means of sensors/metres are susceptible to various errors or inconsistencies such as redundancy, noise, missing values etc. This, in turn, could negatively affect the process of knowledge discovery from the data. Data Preprocessing (DP) plays a substantial role in improving the task of knowledge discovery from the data. It includes several sub-phases which are as follows [6]:

- Data Cleaning: Data cleaning includes several steps such as missing values imputation, noise removal, outlier detection and data smoothing.
- Data Integration and Transformation: Data integration and transformation involves data normalization (scaling features to have a particular range), data aggregation and data conversion (converting data to a useable format).
- Data Reduction: Data Reduction includes several sub-steps such as reducing the number of attributes values or tuples, reducing the number of features (feature selection) and data discretization.

4.2. Data clustering

Clustering [8,52] is an unsupervised learning algorithm which aims to identify natural groups/clusters present in the data. The set of clusters are identified based on some similarity measure among the features of data objects. There exist different type of clustering algorithms which implement different criteria/methodology and similarity measures to determine natural clusters in the data. These clustering algorithms can be broadly classified into the three categories [52] namely, partitioning based clustering (creates independent partitions of data), hierarchical clustering (generate a hierarchy of clusters) and density-based clustering (based on data density space) algorithms. Furthermore, there also exist a number of similarity measures [53] (such as Euclidean distance (ED), Manhattan distance and Kernelized distance) that could be implemented to estimate objects similarity. The choice of optimal clustering algorithm and similarity measure highly depend on the input data requirements and the domain under study.

Algorithm 1 *Hierarchical Clustering* *(Input Time Series Data)*

1: *Initialized each time series signal as a cluster i.e. N clusters if there are N time series.*
2: *Determine the closest pairs of cluster on basis of similarity measure (DTW) and merge them.*
3: *Compute the distance (DTW) between newly formed cluster and each of the old clusters.*
4: *Repeat steps 2 and 3 until all input time series signals get clustered into one.*
5: *Create an output dendrogram to visualize the clustering result*

The present research work implements Hierarchical Agglomerative (HA) clustering [53] algorithm to identify the months of the year with similar energy consumption patterns. As the current study deals with the load time-series signals, ED is not an accurate measure of determining the similarity among time-series due to the distortions in the time domain. Therefore, HA clustering with Dynamic Time Warping (DTW) [54] distance measure is employed to create/identify the hierarchy of objects (time series). The DTW distance measure takes care of the temporal dependencies of the time-series observations. For the two input time series A $\langle A_1, A_2, \ldots A_n \rangle$ and B $\langle B_1, B_2, \ldots B_m \rangle$, the DTW distance between two is given as:

$$DTW(A_i, B_j) = \begin{cases} 0, \\ \quad \text{if } A = \langle\rangle \mathit{and}\ B = \langle\rangle. \\ \infty, \\ \quad \text{if } A = \langle\rangle \mathit{or} B = \langle\rangle. \\ W(A_i, B_j) + min(DTW(A_{i-1}, B_j), DTW(A_i, B_{j-1}), DTW(A_{i-1}, B_{j-1})) \\ \quad \text{Otherwise} \end{cases} \tag{18}$$

where W is a $n \times m$ cost matrix whose $\langle i_{th}, j_{th} \rangle$ entry denotes the distance between A_i and B_j and $\langle\rangle$ represent the empty time series. The step by step procedure of clustering using HA technique is given in Algorithm 1. The output from the hierarchical clustering algorithm is a dendrogram (tree-like structure) that shows the kind of relationship between the clusters. In contrast to the various other clustering algorithms, there are several benefits of implementing hierarchical algorithms for the task of clustering: (a) The algorithm does not require any prior information about the number of clusters (b) The algorithm is easy to implement and provides better visualization of clusters relationship.

4.3. Trend characterization

To perform an effective, reliable and accurate estimation of energy demand patterns, it would be better to characterize the energy consumption patterns at a more abstract/granular level. The load trend characterization aims to provide an understanding of metadata falling to each resultant cluster at different levels of granularity. The characterization can be performed at various levels (weekly analysis, time-span analysis, periodical analysis, seasonal analysis, historical analysis) depending upon the problem under study. In relation to the previous studies in the field of load forecasting [6,42], the trend characterization has been carried out at two different levels:

- Inter-Cluster (Seasonal) Analysis: The resultant clusters generated from the agglomerative clustering algorithm provide an insight of demand patterns over different seasons of a year.
- Intra-Cluster (Timespan) Analysis: The energy demand patterns may vary highly on the day to day basis. So, it would intuitively be better to include additional information (season, day of the week, time-interval of the day) into account to predict current electricity demand. This level of characterization deals with the more deeper level of understanding by providing an overview of the energy demand patterns for different days of a week in each season/cluster.

4.4. Noise filtering and optimal features extraction

EMD [55] is a technique that has been widely adopted to decompose a time series signals into corresponding modes and a residual component in a variety of application domains including climate analysis, utility services, audio engineering and healthcare. However, there are several limitations of the EMD algorithm that needs to be addressed:

- The algorithm is highly dependent on the methods of extremal points finding, interpolation and mode generations.

- The method is highly sensitive to noise and lacks mathematical theory.
- Deciding the optimal mode component for noise reduction is also a major problem.

Hence, in the present work, the VMD algorithm (explained in Section 3.1) is implemented to overcome the shortcomings of the EMD based decomposition algorithm. The method looks for an ensemble of modes (N) centred around the pulsation frequency ω_k. Furthermore, after decomposition by the VMD algorithm, the autoencoders model is trained to generate a latent space representation (consisting M modes) of the input sub-series, where, $M < N$. The auto-encoders model works by extracting and combining the meaningful sub-series components generated by the VMD algorithm while removing the disturbance of irrelevant components. The step by step procedure of noise filtering and features/modes extraction method is as explained in Algorithm 2.

4.5. Long short term memory network model

The latent space features subset (corresponding to each selected day of a season) generated from the output of the previous step is then used as an input for the LSTM network models. The LSTM network requires three-dimensional input of the form $\langle X, \ Y, \ Z \rangle$ where X indicates the number of input samples, Y denotes the sequence length and Z denotes the dimensions of the features space. Generally, the prediction model treats each observation independently, i.e. by treating previous data features values for training the model. However, as the load series observations are temporarily dependent on each other, so this kind of training is not suitable for the current study. According to the recent research studies in the field of time-series forecasting [25, 42] historical load sequence is the most critical input and contain rich information for predicting future load. So, the current study employs LSTM Multi-Input-Multi-Output (MIMO) window strategy to forecast electricity demand for the specific time interval of the day. The strategy utilizes historical timestamps demand to predict demand for the current timestamp. The overall working of the LSTM-MIMO strategy is explained as follows:

Consider an input load series TS (corresponding to a selected day of the particular season/cluster) of the form $TS = \langle ts_1, ts_2, ts_3, \ \ldots \ ts_N \rangle$ where ts_i denotes the demand at ith timestamp. The LSTM-MIMO window model rolls over the input load series while dividing the series into blocks of size $(D+O)$ where D denotes size of the input window and O defines the size of output window. At each particular timestamp, the network uses the current input window load series to estimate demand for the next O (size of output window=1) timestamps. Formally, it can be given as:

$$[ts_{i+1}, \ \ldots \ ts_{i+O}] = F(ts_i, \ \ldots \ ts_{i-d+1}) \tag{19}$$

where *'d'* defines the number of previous observations to be used for estimating the next O timestamps demand. In this way, the network goes on training and iterating over the i/p-o/p windows (while shifting window by one step at each timestamp prediction) until the last window is reached. The last window is used for validating the network model.

There are several things or parameters that need to be defined for providing the accurate prediction result using LSTM-MIMO network model.

- *Determining optimal window size or optimal lag value:* Even though the LSTM-MIMO network window model deals directly with the lagged values and temporal relationships, determining optimal input window size is a major issue that needs to be addressed. The accuracy and reliability of the prediction result highly depend on the window size. However, there does not exist any effective solution to estimate the value of optimal window size. In the present work, we propose an effective method that implements intra-correlation with FFT convolution [56] to determine the optimal window size parameter value. The method computes the Fourier transform of the correlation value by multiplying the Fourier transform of one signal by the complex conjugate Fourier transform of the second signal. i.e. for the given two load time-series signals A and B, the cross-correlation using FFT convolution between them is given by:

$$CC = ifft(fft(A, N) * conju(fft(B, N))) \tag{20}$$

  where $N = size(A) + size(B) - 1$, *'conj'* represents the complex conjugate transform function, *'N'* denotes the length of the input series, *'fft'* and *'ifft'* represent the Fast Fourier transform and Inverse Fourier transform functions respectively. The output of the convolution is a time-series with a peak at the point that reveals valuable information about the optimal lag parameter value. There are two major benefits of implementing correlation for determining the optimal lag value: (a) The method is very fast and has an overall runtime complexity of $O(nlogn)$. (b) The method is very suitable for large time-series sequences.
- Hyper-parameters Selection: Each learning/prediction model has a set of associated hyper-parameters that need to be properly tuned for achieving the best prediction results. The selection of hyper-parameters value is completely data dependent and there does not exist any global method to estimate the value of models hyper-parameters. LSTM model has a number of associated hyper-parameters, namely input window size (S_i), output window size (S_o), number of input neurons (I_i), number of hidden layers (h), number of neurons in the hidden layers (n_h), number of epochs (n_e), memory between batches, batch_size, optimization strategy=*'Adam'* and regularization parameter. Some of these hyper-parameters can be directly determined on the basis of input data features and forecast horizon such as S_i, S_o and I_i. While some other associated hyper-parameters require iterative runs of LSTM training and validation to be accurately determined. Table 2 lists the description of these hyper-parameters. Moreover, the associated value of these hyper-parameters in relation to the prediction model built on the HP load demand dataset are listed in Table 2.
- Performance Measures: In the present work, the following statistical measures are used to analyse the prediction results:

Algorithm 2 : *Feature Extraction* *(Input Time Series Data, Clusters)*

```
1: for each cluster in clusters_formed do
2:     for each day in selected_days do
3:         Apply Variational Mode Decomposition (VMD) algorithm
           to extract modes from the input load time series of the
           selected day.
4:         Train and Build Autoencoder model to provide latent
           representation or optimal features/sub-series of the data
           (Input modes)
5:         Store the latent space representation to be fed as an input
           to the LSTM model
6:     end for
7: end for
```

Table 1
Optimal Value of Lag Parameter for different Clusters.

Cluster_Name	Lag ParameterValue	
	Average demand prediction	*Peak demand prediction*
$Cluster_1$	24	25
$Cluster_2$	32	32
$Cluster_3$	16	16

- *Mean Absolute Percentage Error (% Error)* [52]: It is defined as the percentage (%) deviation of predicted values from the actual values.
- *Correlation Coefficient*: It measures the degree of correlation between the actual and predicted values.

$$\rho = Correl(p_i, \bar{p}_i) = \frac{Covariance(p_i, \bar{p}_i)}{\sigma_{p_i} . \sigma_{\bar{p}_i}} \tag{21}$$

- *RMSE (Root Mean Squared Error)* [52]: It is given as

$$RMSE = \sqrt{\frac{1}{kn}\sum_{t=1}^{k}\sum_{i=1}^{n}(p_{i,t} - \bar{p}_{i,t})^2} \tag{22}$$

where k=1 (forecast horizon) and $p_{i,t}$, $\bar{p}_{i,t}$ denotes the actual and predicted values at timestamp t respectively.

4.6. Error variance modelling strategy (EVM-S)

Data in which the variance of the error term varies highly is said to be suffering from the problem of heteroskedasticity. The proper analysis of such fluctuating variance could be used to improve the reliability and prediction accuracy of the forecasting model. So, the present work implements Generative Auto-regressive Conditional Heteroskedasticity (GARCH) correction strategy [24,57] to support both conditional and time-dependent changes in the variance. The model treats heteroskedasticity as a variance to be modelled and relatively effects prediction by the variance of each term. For the given actual value (p_t) and predicted value ($\bar{p}_t$), the error term ($\mathcal{E}$) is given by:

$$\mathcal{E}_t = p_t - \bar{p}_t \tag{23}$$

The model assumes that the error variance at a timestamp t can be modelled as the weighted average of long-run average variance, variance predicted for the period and the recent squared residuals. Mathematically, it is given as:

$$\delta_t = \alpha_0 + \sum_{k=1}^{p}\alpha_k . \delta_{t-k} + \sum_{l=1}^{q}\beta_l . \mathcal{E}_{l-1}^2 \tag{24}$$

where α, β are the coefficients and p, q represent the orders. In this way, the model allows to support fluctuating volatility in the load time series and improves the prediction results of the LSTM-MIMO model.

5. Case study

5.1. Input dataset

In this study, the Energy Consumption Dataset (ECD) of Himachal Pradesh (HP) is adopted to compare the prediction accuracy of the proposed model with the other state-of-the-art prediction models. Himachal Pradesh is a state that lies in the northern part of India (western Himalayas). The state is bordered by the state Jammu & Kashmir, Punjab, Haryana and Uttarakhand on the north, west, Southwest and Southeast part respectively. It covers an area of approximately 55,673 km^2 and has the highest hydro-power generation capacity due to the presence of perennial rivers [1,58]. The trading, generation, distribution and maintenance of power within the state HP is governed by a board named Himachal Pradesh State Electricity Board Limited (HPSEBL) [58]. An area load dispatch centre of HPSEBL is responsible for recording ECD data of HP at a regular time interval of 15 min (96 entries per day). We have the ECD data (both average and peak demand) of HP for a period of eight years starting from January 2010 to January 2018.

5.2. Data preprocessing

The following preprocessing steps are applied to remove inconsistencies or errors from the ECD dataset of HP.

- Data Cleaning: The set of missing values in the data are interpolated by the average of previous years entries for the same interval.
- Data Normalization: The min–max data normalization is applied to scale the load demand values into a range of zero to one. For the input load series feature l, the new normalized feature l' is given as:

$$l' = \frac{l - min(l)}{max(l) - min(l)} \tag{25}$$

- Data Transformation and Integration: The ECD file format recorded by the HPSEBL area load centre is not suitable for our current analysis. So, there is a need to transform the data in a format suitable for analysis. In the present study, we aim to forecast load at a specific time interval of the day. Therefore, the ECD data corresponding to each day is divided into four time intervals namely Ti_1: 12:00 am to 06:00 am, Ti_2: 06:00 am to 12:00 pm, Ti_3: 12:00 pm to 06:00 pm and Ti_4: 06:00 pm to 12:00 am.

5.3. Data clustering

In the present study, the aim is to forecast load at a specific interval using the historical, seasonal and weekday information. In order to fulfil the requirements, the hierarchical agglomerative clustering with DTW distance measure is applied on the HP-ECD to identify months (clusters) with similar consumption patterns. Fig. 3 shows the resultant dendrogram generated as output from the HA clustering algorithm. From Fig. 3, it is clear that the based on the daily electricity consumption patterns overall ECD can be divided into the three major clusters namely $Cluster_1$ for {*June*, *July*, *August*, *September*}, $Cluster_2$ for {*February*, *March*, *April*, *May*, *October*, *November*} and $Cluster_3$ for {*January*, *December*}.

5.4. Trend characterization

The output of HA clustering algorithm can be used to characterize the demand patterns for the different seasons of a year. In order to further incorporate weekday information to the model for providing more precise and reliable prediction results, it is necessary/required to perform the intra-cluster analysis of the demand patterns. To do so, for each cluster, three days (*week_start_day* : *Monday*, *mid_week_day* : *Wednesday* and *week_end_day* : *Sunday*) are chosen from a week. Furthermore, to provide a demonstration of the variations in the demand patterns for the selected days in each cluster, six random entries (corresponding to each cluster and day) are chosen from the

Table 2
Description of the set of Hyper-parameters associated with proposed approach.

Hyper-parameter	Description	HP Dataset Value/Range
Input window size (S_i)	is equal to the lag value and is calculated by using equation 20.	{$Cluster_1$: 24, $Cluster_2$: 32, $Cluster_3$: 16}
Output window size (S_o)	is equal to the forecast horizon	1
No. of neurons in the Hidden layers (n_h)	determined by performing iterative runs of LSTM model training and validation	[4–64]
No. of neurons in input layer (I_i)	is equal to the input window size	{$Cluster_1$: 24, $Cluster_2$: 32, $Cluster_3$: 16}
No. of Hidden layers (h)	determined by performing iterative runs of LSTM model training and validation	[2-20]
No. of Epochs (n_e)	number of epochs in the prediction model training phase	[100–250]
Optimization Strategy		'*Adam*'

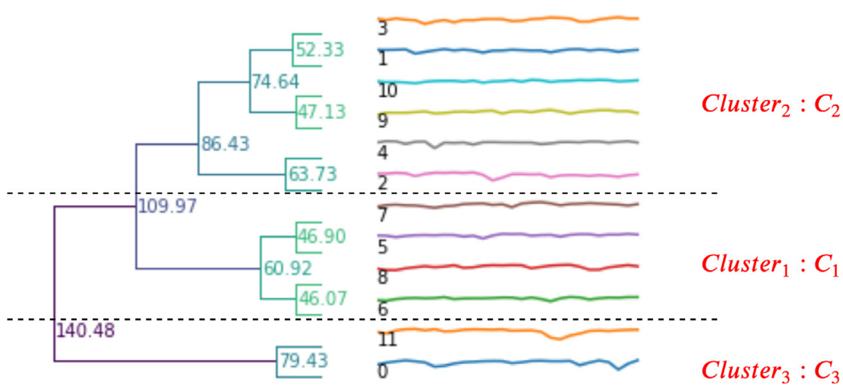


Fig. 3. Hierarchical clustering results (*0*: January, 1: February, 2:March, 3: April, 4:May, 5:June, 6:July, 7:August, 8:September, 9:October, 10:November, 11:December).*

ECD of HP. Fig. 4 provides a graphical representation of the energy demand trend for the selected days in each cluster. From Fig. 4, it is clearly visible that the demand trend varies highly on the day and season basis. For instance, the demand pattern of '$C1 - Monday$' is quite different from the demand patterns of '$C2 - Monday$' and '$C3 - Monday$'. Hence, it is better to include the weekday and cluster information along with historical demand observations into the forecasting model.

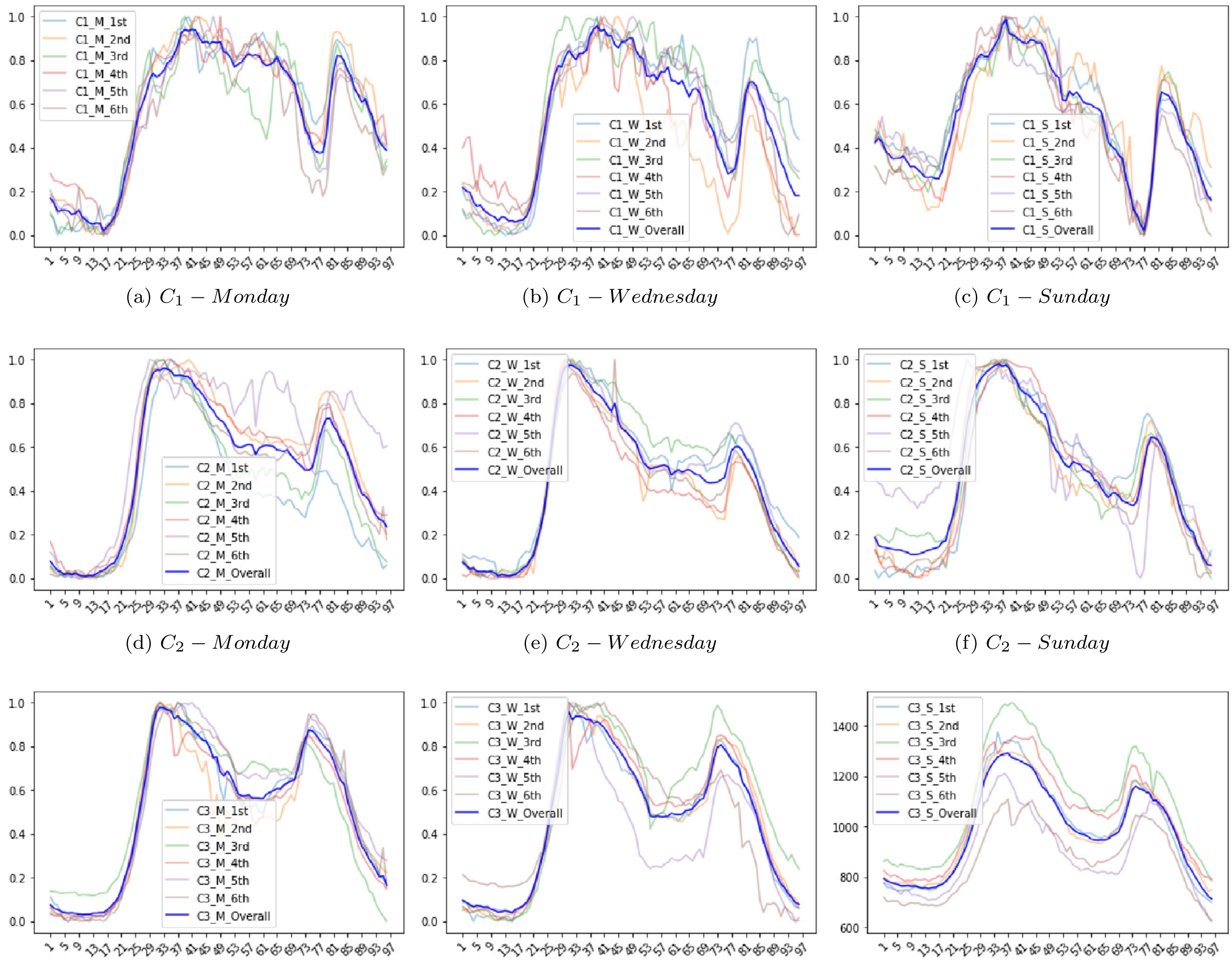
5.5. Noise filtering and feature extraction

The next step after performing load trend characterization is to extract optimal features subset for the input load time-series signals. For the given input load time series signal $C_iD_j = \langle cd_1, cd_2, \dots cd_n \rangle$ of a particular day j and cluster i, the VMD algorithm [21] is applied to decompose the input series into eight sub-series/modes namely $mod_{ij} = < mod_1, mod_2, mod_3, mod_4, mod_5, mod_6, mod_7, mod_8 >$. An example decomposition by the VMD algorithm is illustrated in Fig. 2 (Step 2a). Then, the auto-encoder model [22] is trained and built to provide the latent space representation ($\langle lat_1, lat_2, lat_3, lat_4 \rangle$) of the input sub-series ($\langle mod_2, mod_3, \dots mod_n \rangle$) while regenerating the input load sub-series. The auto-encoder model aims to extract or combine the meaningful sub-series and the resultant extracted optimal features ($\langle lat_1, lat_2, lat_3, lat_4 \rangle$) along with mod_1 are then labelled as $\langle emod_1, emod_2, emod_3, emod_4, emod_5 \rangle$. The above process is repeated for the load time-series signal of each selected day in every cluster.

5.6. Prediction results and discussion

The goal of current study is to forecast load demand at a specific time interval using the available historical and seasonal load profiles. So, there is a need to train a separate LSTM-MIMO network model for each selected day in every individual cluster/season, i.e. for the given P clusters, R days (corresponding to each cluster) and k optimal sub-signals (corresponding to each day), there is a need to train P^*R^*k LSTM models. In the present case, the HA clustering algorithm has returned three clusters and for each cluster we have chosen three days, so a total of $9 * k$ LSTM network MIMO models are trained.

In addition to historical and seasonal load profiles, lag parameter and *EVM-S* model also play a critical role in improving the accuracy and reliability of the proposed LSTM based prediction model. The optimal value of the lag parameter could be efficiently utilized to determine the amount of historical information to be used for estimating the current load demand. In the present study, the optimal value of lag parameter is computed from the convolution method described in Section 4.5. Table 1 lists the optimal lag parameter value for the models trained on different clusters. Further, the *EVM-S* is employed to enhance the prediction accuracy of each sub-signal model (within a day). The final demand prediction results are given by the summation of decoded prediction results of all sub-signal models. Initially, the basic correlation functions are utilized to determine the variance of the error value between the actual and predicted values. Subsequently, the generative auto-regressive conditional



(a) $C_1 - Monday$ (b) $C_1 - Wednesday$ (c) $C_1 - Sunday$

(d) $C_2 - Monday$ (e) $C_2 - Wednesday$ (f) $C_2 - Sunday$

(g) $C_3 - Monday$ (h) $C_3 - Wednesday$ (i) $C_3 - Sunday$

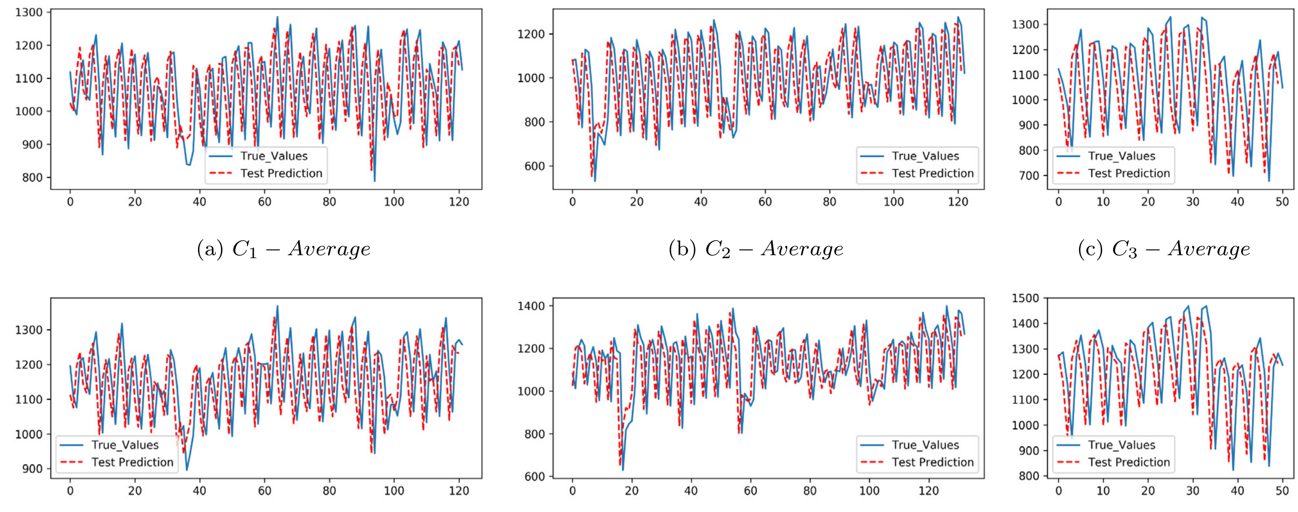
Fig. 4. Trend analysis of demand patterns for the selected days in three clusters (Note*: The overall trend of the each day is shown with the blue line, *x*-axis represents each 15-minutes interval number of a day and *y*-axis denotes the energy load demand value.).

model, as described in Section 4.6, is implemented to improve the prediction results of the proposed model.

To verify the effectiveness and prediction accuracy of the proposed approach, four state-of-the-art prediction models (Support Vector Regression: *SVR*, Recurrent Neural Network: *RNN*, Deep Belief Network: *DBN* and EMD+LSTM) are built to predict the average and peak electricity demand for the state HP. The state-of-the-art prediction models utilize the concept of lag values to estimate the current timestamp demand. The average and peak state demand prediction results of the proposed approach and three other state-of-the-art approaches are listed in Tables 3–8. These tables compare the prediction result of the proposed approach and other existing techniques in terms of various performance measures described in Section 4.5. Fig. 5 shows an example prediction plot of the proposed approach on the selected day (*Monday*) in each cluster. The *x-axis* and *y-axis* in Fig. 5 represents the number of samples or data points and energy load demand values respectively. The blue lines represent the actual load demand values and red lines represent the predicted load output from our proposed approach. Furthermore, Fig. 6 provide a visual representation of the mean absolute prediction error (%) using different models.

From the prediction results listed in Tables 3, 4, 5, 6, 7, 8 and Figs. 6, 7 following inferences can be drawn:

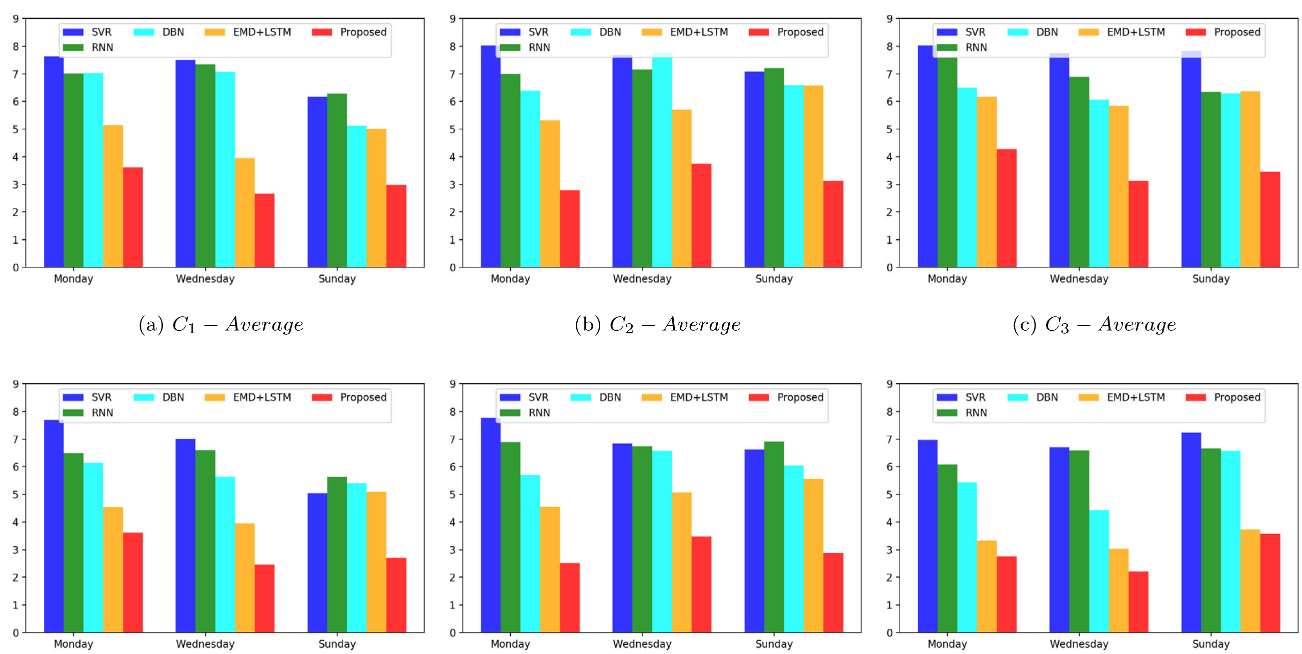
- The inclusion of historical and seasonal load profiles to the prediction model support improved accuracy as the RNN model performs better by providing an overall MAPE of 6.74% as compared to a MAPE of 7.12% by SVM model.
- The integration of mode decomposition algorithm with historical and seasonal load profiles strengthen the prediction accuracy of the model as EMD+LSTM model provides an overall MAPE of 4.94%, while the MAPE of the SVM, RNN and DBN models are 7.12%, 6.74% and 6.15% respectively. This indicates that the decomposition algorithm positively contributes in improving the accuracy of the prediction model.
- The variational mode decomposition algorithm overcomes the shortcomings of the EMD based decomposition. The integration of VMD (decomposition algorithm) with the *LSTM-MIMO* model (historical and seasonal load profiles) and *EVM-S* outperforms other state-of-the-art prediction models (as shown in Fig. 6) by providing an average MAPE of 3.04%. Furthermore, Fig. 7 shows the effect of EVM-S on the overall MAPE (%) of the proposed approach. These results



(a) $C_1 - Average$ (b) $C_2 - Average$ (c) $C_3 - Average$

(d) $C_1 - Peak$ (e) $C_2 - Peak$ (f) $C_3 - Peak$

Fig. 5. Average and peak demand prediction (*test-data*) results for Day_1: Monday in all clusters (Note*: x-axis represents the number of test samples or data points and y-axis represents the load demand value).



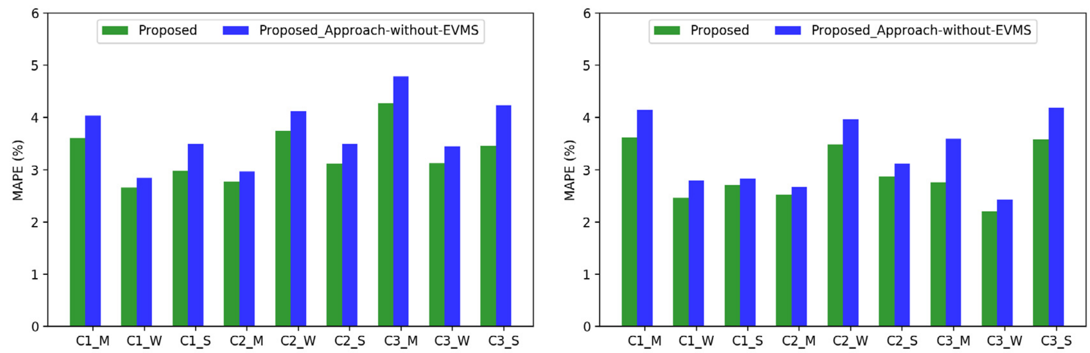
(a) $C_1 - Average$ (b) $C_2 - Average$ (c) $C_3 - Average$

(d) $C_1 - Peak$ (e) $C_2 - Peak$ (f) $C_3 - Peak$

Fig. 6. MAPE (%) comparison of proposed approach and other state-of-the-art techniques.

Table 3
Average Energy Demand Prediction Results for $Cluster_1$ (N: Normalized Values, UN: Un-normalized Values).

Models	Monday					Wednesday					Sunday				
	RMSE (Train)		RMSE (Test)		ρ	RMSE (Train)		RMSE (Test)		ρ	RMSE (Train)		RMSE (Test)		ρ
	N	UN	N	UN		N	UN	N	UN		N	UN	N	UN	
SVR	0.07	73.15	0.08	76.15	0.71	0.08	65.96	0.10	76.28	0.72	0.07	55.71	0.08	59.81	0.76
RNN	0.07	67.38	0.07	70.07	0.76	0.08	65.43	0.10	74.88	0.76	0.07	54.94	0.08	60.75	0.74
DBN	0.06	62.06	0.07	70.22	0.82	0.07	54.48	0.10	72.15	0.84	0.06	46.96	0.07	49.50	0.81
EMD+LSTM	0.06	43.09	0.06	51.38	0.91	0.04	33.67	0.05	40.27	0.86	0.05	41.33	0.06	48.45	0.87
Proposed Approach	**0.04**	**29.95**	**0.05**	**36.04**	**0.94**	**0.03**	**19.46**	**0.04**	**27.10**	**0.92**	**0.03**	**21.46**	**0.04**	**28.85**	**0.89**



(a) *Average Demand Prediction Results* (b) *Peak Demand Prediction Results*

Fig. 7. Effects of *EVM-S* strategy on the overall MAPE (%) of the proposed approach (Note*- M:Monday, W:Wednesday, S:Sunday).

Table 4
Average Energy Demand Prediction Results for $Cluster_2$ (N: Normalized Values, UN: Un-normalized Values).

Models	Monday					Wednesday					Sunday				
	RMSE (Train)		RMSE (Test)		ρ	RMSE (Train)		RMSE (Test)		ρ	RMSE (Train)		RMSE (Test)		ρ
	N	UN	N	UN		N	UN	N	UN		N	UN	N	UN	
SVR	0.07	63.40	0.09	77.49	0.84	0.07	68.32	0.09	75.19	0.77	0.07	53.45	0.09	65.95	0.80
RNN	0.06	58.40	0.07	67.53	0.85	0.06	64.03	0.09	70.10	0.81	0.07	54.33	0.09	67.06	0.81
DBN	0.06	52.43	0.07	61.67	0.90	0.06	63.82	0.09	76.10	0.84	0.06	46.83	0.08	62.67	0.85
EMD+LSTM	0.04	42.97	0.06	53.13	0.87	0.05	51.34	0.06	55.86	0.89	0.05	44.13	0.08	61.10	0.87
Proposed Approach	**0.03**	**22.93**	**0.03**	**26.82**	**0.96**	**0.03**	**25.10**	**0.04**	**36.77**	**0.96**	**0.03**	**21.24**	**0.04**	**29.03**	**0.98**

Table 5
Average Energy Demand Prediction Results for $Cluster_3$ (N: Normalized Values, UN: Un-normalized Values).

Models	Monday					Wednesday					Sunday				
	RMSE (Train)		RMSE (Test)		ρ	RMSE (Train)		RMSE (Test)		ρ	RMSE (Train)		RMSE (Test)		ρ
	N	UN	N	UN		N	UN	N	UN		N	UN	N	UN	
SVR	0.07	62.11	0.11	81.11	0.87	0.08	62.31	0.11	80.21	0.86	0.07	54.33	0.10	75.66	0.86
RNN	0.07	60.50	0.10	76.98	0.85	0.08	61.35	0.09	71.37	0.84	0.07	54.10	0.08	62.43	0.86
DBN	0.05	50.66	0.08	65.67	0.96	0.07	52.13	0.08	62.74	0.86	0.07	50.18	0.08	61.93	0.92
EMD+LSTM	0.05	49.79	0.07	62.28	0.94	0.07	45.13	0.08	60.57	0.90	0.07	49.73	0.08	61.78	0.95
Proposed Approach	**0.04**	**32.32**	**0.05**	**43.28**	**0.96**	**0.04**	**26.19**	**0.06**	**32.43**	**0.92**	**0.03**	**23.98**	**0.06**	**33.51**	**0.97**

Table 6
Peak Energy Demand Prediction Results for $Cluster_1$ (N: Normalized Values, UN: Un-normalized Values).

Models	Monday					Wednesday					Sunday				
	RMSE (Train)		RMSE (Test)		ρ	RMSE (Train)		RMSE (Test)		ρ	RMSE (Train)		RMSE (Test)		ρ
	N	UN	N	UN		N	UN	N	UN		N	UN	N	UN	
SVR	0.11	78.71	0.11	83.49	0.77	0.08	64.92	0.10	77.35	0.76	0.07	48.80	0.07	52.50	0.71
RNN	0.09	65.40	0.10	70.39	0.81	0.08	63.40	0.10	72.86	0.77	0.07	47.93	0.08	58.58	0.73
DBN	0.08	56.57	0.09	66.54	0.81	0.07	54.48	0.08	62.15	0.73	0.07	44.71	0.08	56.34	0.76
EMD+LSTM	0.05	36.51	0.07	49.50	0.87	0.06	35.51	0.07	43.59	0.84	0.06	37.04	0.07	53.10	0.85
Proposed Approach	**0.03**	**25.54**	**0.05**	**39.60**	**0.95**	**0.04**	**21.95**	**0.06**	**27.34**	**0.87**	**0.05**	**22.71**	**0.05**	**28.18**	**0.94**

reflect that the inclusion of EVM-S model provide a positive impact on the overall prediction accuracy.

6. Conclusion

The current research work has proposed a novel deep learning based multilevel hybrid approach to estimate energy demand. Firstly, the approach performs cluster and load trend analysis to generate season based timestamp data for modelling purpose. Secondly, the Variational Mode Decomposition algorithm and Autoencoder model are trained to generate optimal features subset for season based timestamp data. Thirdly, the LSTM network models are built for the extracted optimal sub-signals. Finally, an EVM-S is incorporated to capture the variations of error-term for each sub-signal and the final output prediction results are given by summation. The proposed approach is applied to energy consumption data of HP, India and the prediction results are evaluated using popular performance measures (RMSE, ρ and

Table 7
Peak Energy Demand Prediction Results for $Cluster_2$ (N: Normalized Values, UN: Un-normalized Values).

Models	Monday					Wednesday					Sunday				
	RMSE (Train)		RMSE (Test)		ρ	RMSE (Train)		RMSE (Test)		ρ	RMSE (Train)		RMSE (Test)		ρ
	N	UN	N	UN		N	UN	N	UN		N	UN	N	UN	
SVR	0.07	67.13	0.10	84.13	0.84	0.07	66.19	0.09	75.20	0.72	0.07	60.80	0.08	68.52	0.74
RNN	0.07	65.42	0.09	74.62	0.85	0.07	64.23	0.09	74.37	0.73	0.08	64.33	0.09	71.42	0.79
DBN	0.06	53.80	0.07	61.67	0.90	0.07	63.82	0.08	72.43	0.78	0.06	47.62	0.08	62.52	0.86
EMD+LSTM	0.05	32.12	0.06	49.24	0.87	0.06	51.34	0.07	55.86	0.88	0.06	35.01	0.07	57.61	0.90
Proposed Approach	**0.04**	**22.95**	**0.05**	**27.34**	**0.96**	**0.05**	**26.11**	**0.06**	**38.40**	**0.95**	**0.05**	**21.61**	**0.06**	**29.75**	**0.94**

Table 8
Peak Energy Demand Prediction Results for $Cluster_3$ (N: Normalized Values, UN: Un-normalized Values).

Models	Monday					Wednesday					Sunday				
	RMSE (Train)		RMSE (Test)		ρ	RMSE (Train)		RMSE (Test)		ρ	RMSE (Train)		RMSE (Test)		ρ
	N	UN	N	UN		N	UN	N	UN		N	UN	N	UN	
SVR	0.08	69.09	0.11	81.17	0.87	0.08	69.29	0.09	79.87	0.85	0.08	72.11	0.09	79.94	0.87
RNN	0.08	67.35	0.08	70.84	0.91	0.08	65.06	0.09	78.47	0.84	0.07	66.56	0.08	73.72	0.84
DBN	0.06	53.01	0.08	63.29	0.90	0.07	52.13	0.07	52.74	0.92	0.06	64.18	0.07	72.68	0.91
EMD+LSTM	0.05	28.01	0.06	38.58	0.91	0.05	32.47	0.06	36.14	0.94	0.05	36.59	0.05	41.30	0.92
Proposed Approach	**0.05**	**25.46**	**0.05**	**32.19**	**0.93**	**0.04**	**26.29**	**0.05**	**31.26**	**0.98**	**0.04**	**29.64**	**0.06**	**39.61**	**0.90**

MAPE). Four benchmark demand prediction models have been compared to signify the effectiveness of the proposed approach: SVR, RNN, DBN and EMD+LSTM. From the comparative analysis of prediction results, the following observations can be drawn:

- The decomposition-based hybrid approaches work better than other load demand prediction models, but they are slower to train. The approach proposed in the current work provides support for both improved accuracy and reduced complexity.
- The proposed approach shows better performance in handling compound relationship across time intervals.
- The proposed approach outperforms other benchmark demand prediction models and supports improved prediction accuracy (*MAPE: 3.04%*). So, it can be effectively implemented for regulatory and market planning activities.
- The inclusion of decomposition and error variance modelling strategy strengthens the prediction accuracy of the proposed approach.

CRediT authorship contribution statement

Jatin Bedi: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. **Durga Toshniwal:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. Publication, Electrical India, 1961, URL https://www.electricalindia.in/hydro-power-scenario-in-himachal-pradesh/.

[2] H. Son, C. Kim, Forecasting short-term electricity demand in residential sector based on support vector regression and fuzzy-rough feature selection with particle swarm optimization, Procedia Eng. 118 (2015) 1162–1168.

[3] L. Suganthi, A.A. Samuel, Energy models for demand forecasting—A review, Renew. Sustain. Energy Rev. 16 (2) (2012) 1223–1240.

[4] C. Tong, J. Li, C. Lang, F. Kong, J. Niu, J.J. Rodrigues, An efficient deep model for day-ahead electricity load forecasting with stacked denoising auto-encoders, J. Parallel Distrib. Comput. 117 (2018) 267–273, http://dx.doi.org/10.1016/j.jpdc.2017.06.007.

[5] E. Mocanu, P.H. Nguyen, M. Gibescu, W.L. Kling, Deep learning for estimating building energy consumption, Sustain. Energy Grids Netw. 6 (2016) 91–99.

[6] J. Bedi, D. Toshniwal, Empirical mode decomposition based deep learning for electricity demand forecasting, IEEE Access 6 (2018) 49144–49156.

[7] L. Suganthi, T. Jagadeesan, Energy substitution methodology for optimum demand variation using Delphi technique, Int. J. Energy Res. 16 (9) (1992) 917–928.

[8] J. Han, J. Pei, M. Kamber, Data Mining: Concepts and Techniques, Elsevier, 2011.

[9] F. Rodrigues, C. Cardeira, J.M.F. Calado, The daily and hourly energy consumption and load forecasting using artificial neural network method: a case study using a set of 93 households in Portugal, Energy Procedia 62 (2014) 220–229.

[10] A. Lahouar, J.B.H. Slama, Day-ahead load forecast using random forest and expert input selection, Energy Convers. Manage. 103 (2015) 1040–1051.

[11] S.B. Taieb, R.J. Hyndman, A gradient boosting approach to the Kaggle load forecasting competition, Int. J. Forecast. 30 (2) (2014) 382–394.

[12] H. Shi, M. Xu, R. Li, Deep learning for household load forecasting—a novel pooling deep RNN, IEEE Trans. Smart Grid 9 (5) (2018) 5271–5280, http://dx.doi.org/10.1109/tsg.2017.2686012.

[13] A. Dedinec, S. Filiposka, A. Dedinec, L. Kocarev, Deep belief network based electricity load forecasting: An analysis of Macedonian case, Energy 115 (2016) 1688–1700, http://dx.doi.org/10.1016/j.energy.2016.07.090.

[14] R.-A. Hooshmand, H. Amooshahi, M. Parastegari, A hybrid intelligent algorithm based short-term load forecasting approach, Int. J. Electr. Power Energy Syst. 45 (1) (2013) 313–324.

[15] D. Chaturvedi, A. Sinha, O. Malik, Short term load forecast using fuzzy logic and wavelet transform integrated generalized neural network, Int. J. Electr. Power Energy Syst. 67 (2015) 230–237.

[16] Y. Ren, P. Suganthan, N. Srikanth, A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods, IEEE Trans. Sustain. Energy 6 (1) (2015) 236–244.

[17] N.A. Agana, E. Oleka, G. Awogbami, A. Homaifar, Short-term load forecasting based on a hybrid deep learning model, in: SoutheastCon 2018, IEEE, 2018, pp. 1–6.

[18] C. Deb, F. Zhang, J. Yang, S.E. Lee, K.W. Shah, A review on time series forecasting techniques for building energy consumption, Renew. Sustain. Energy Rev. 74 (2017) 902–924.

[19] H.-F. Yang, Y.-P.P. Chen, Hybrid deep learning and empirical mode decomposition model for time series applications, Expert Syst. Appl. 120 (2019) 128–138.
[20] J. Wu, Z. Cui, Y. Chen, D. Kong, Y.-G. Wang, A new hybrid model to predict the electrical load in five states of Australia, Energy 166 (2019) 598–609.
[21] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Trans. Signal Process. 62 (3) (2014) 531–544.
[22] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, Deep Learning, Vol. 1, MIT Press Cambridge, 2016.
[23] JAKOB AUNGIERS, 2018, http://www.jakob-aungiers.com/articles/a/LSTM-Neural-Network-for-Time-Series-Prediction. (Accessed 15 January 2018).
[24] R.C. Garcia, J. Contreras, M. Van Akkeren, J.B.C. Garcia, A GARCH forecasting model to predict day-ahead electricity prices, IEEE Trans. Power Syst. 20 (2) (2005) 867–874.
[25] W. He, Load forecasting via deep neural networks, Procedia Comput. Sci. 122 (2017) 308–314.
[26] N. Dalkey, O. Helmer, An experimental application of the Delphi method to the use of experts, Manage. Sci. 9 (3) (1963) 458–467.
[27] H.-x. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (6) (2012) 3586–3592.
[28] J.A. White, R. Reichmuth, Simplified method for predicting building energy consumption using average monthly temperatures, in: Energy Conversion Engineering Conference, 1996. IECEC 96., Proceedings of the 31st Intersociety, Vol. 3, IEEE, 1996, pp. 1834–1839.
[29] C.S. Barnaby, J.D. Spitler, Development of the residential load factor method for heating and cooling load calculations, ASHRAE Trans. 111 (1) (2005).
[30] A. Rice, S. Hay, D. Ryder-Cook, A limited-data model of building energy consumption, in: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, ACM, 2010, pp. 67–72.
[31] E. Tserkezos, Forecasting residential electricity consumption in Greece using monthly and quarterly data, Energy Econ. 14 (3) (1992) 226–232.
[32] S.R. Rallapalli, S. Ghosh, Forecasting monthly peak demand of electricity in India—A critique, Energy Policy 45 (2012) 516–520.
[33] A. Papalexopoulos, T. Hesterberg, A regression-based approach to short-term system load forecasting, IEEE Trans. Power Syst. 5 (4) (1990) 1535–1547.
[34] J. Chen, X. Wang, K. Steemers, A statistical analysis of a residential energy consumption survey study in Hangzhou, China, Energy Build. 66 (2013) 193–202.
[35] N. Fumo, M.R. Biswas, Regression analysis for prediction of residential energy consumption, Renew. Sustain. Energy Rev. 47 (2015) 332–343.
[36] Y. Fu, Z. Li, H. Zhang, P. Xu, Using support vector machine to predict next day electricity load of public buildings with sub-metering devices, Procedia Eng. 121 (2015) 1016–1022.
[37] C. Hamzaçebi, H.A. Es, R. Çakmak, Forecasting of Turkey's monthly electricity demand by seasonal artificial neural network, Neural Comput. Appl. (2017).
[38] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, J. He, Short-term load forecasting with deep residual networks, IEEE Trans. Smart Grid (2018).
[39] W. Kong, Z.Y. Dong, Y. Jia, D.J. Hill, Y. Xu, Y. Zhang, Short-term residential load forecasting based on LSTM recurrent neural network, IEEE Trans. Smart Grid (2017).
[40] D.L. Marino, K. Amarasinghe, M. Manic, Building energy load forecasting using deep neural networks, in: Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE, IEEE, 2016, pp. 7046–7051.
[41] A. Rahman, V. Srikumar, A.D. Smith, Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks, Appl. Energy 212 (2018) 372–385.
[42] J. Bedi, D. Toshniwal, Deep learning framework to forecast electricity demand, Appl. Energy 238 (2019) 1312–1326, http://dx.doi.org/10.1016/j.apenergy.2019.01.113, URL http://www.sciencedirect.com/science/article/pii/S0306261919301217.
[43] W.-C. Hong, Hybrid evolutionary algorithms in a SVR-based electric load forecasting model, Int. J. Electr. Power Energy Syst. 31 (7–8) (2009) 409–417.
[44] T.-T. Chen, S.-J. Lee, A weighted LS-SVM based learning system for time series forecasting, Inform. Sci. 299 (2015) 99–116.
[45] S. Li, L. Goel, P. Wang, An ensemble approach for short-term load forecasting by extreme learning machine, Appl. Energy 170 (2016) 22–29.
[46] X. Qiu, P. Suganthan, G.A. Amaratunga, Short-term electricity price forecasting with empirical mode decomposition based ensemble kernel machines, Procedia Comput. Sci. 108 (2017) 1308–1317.
[47] Y. luo YU, W. LI, A hybrid short-term load forecasting method based on improved ensemble empirical mode decomposition and back propagation neural network, J. Zhejiang Univ. Sci. A 17 (2016) 101–114.
[48] 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 103 (2017) 52–61.
[49] U. Tutorial, Autoencoder, 2018, URL http://ufldl.stanford.edu/tutorial/unsupervised/Autoencoders/. (Accessed 25 December 2018).
[50] K. Bandara, C. Bergmeir, S. Smyl, Forecasting across time series databases using long short-term memory networks on groups of similar series, 2017, arXiv preprint arXiv:1710.03222.
[51] F. Gers, Long Short-Term Memory in Recurrent Neural Networks, Unpublished Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2001.
[52] P.-N. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2016.
[53] T.W. Liao, Clustering of time series data—a survey, Pattern Recognit. 38 (11) (2005) 1857–1874.
[54] E. Keogh, Efficiently finding arbitrarily scaled patterns in massive time series databases, in: European Conference on Principles of Data Mining and Knowledge Discovery, Springer, 2003, pp. 253–265.
[55] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, in: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, Vol. 454, No. 1971, The Royal Society, 1998, pp. 903–995.
[56] D. Related, Correaltion analysis, 2018, URL https://www.dsprelated.com/freebooks/mdft/Correlation_Analysis.html. (Accessed 9 August 2018).
[57] C. Bikcora, L. Verheijen, S. Weiland, Density forecasting of daily electricity demand with ARMA-GARCH, CAViaR, and CARE econometric models, Sustain. Energy Grids Netw. 13 (2018) 148–156.
[58] M. of Statistics, P.I. GOI, Energy statistics, 2018, URL http://mospi.nic.in/sites/default/files/publication_reports/Energy_Statistics_2018.pdf. (Accessed 19 October 2018).