Results in Engineering 17 (2023) 100888

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

Forecasting energy consumption demand of customers in smart grid using Temporal Fusion Transformer (TFT)

Amril Nazir [a,*], Abdul Khalique Shaikh [b,*], Abdul Salam Shah [c], Ashraf Khalil [a]

[a] Department of Information Systems and Technology Management, Zayed University, Abu Dhabi, 144534, United Arab Emirates
[b] Department of Information Systems, Sultan Qaboos University, Muscat, 123, Oman
[c] Department of Computer Engineering, University of Kuala Lumpur (UniKl-MIIT), Kuala Lumpur, 50250, Malaysia

A R T I C L E I N F O

Keywords:
Energy consumption prediction
Residential sector
Smart grids
Temporal Fusion Transformer

A B S T R A C T

Energy consumption prediction has always remained a concern for researchers because of the rapid growth of the human population and customers joining smart grids network for smart home facilities. Recently, the spread of COVID-19 has dramatically increased energy consumption in the residential sector. Hence, it is essential to produce energy per the residential customers' requirements, improve economic efficiency, and reduce production costs. The previously published papers in the literature have considered the overall energy consumption prediction, making it difficult for production companies to produce energy per customers' future demand. Using the proposed study, production companies can accurately have energy per their customers' needs by forecasting future energy consumption demands.

Scientists and researchers are trying to minimize energy consumption by applying different optimization and prediction techniques; hence this study proposed a daily, weekly, and monthly energy consumption prediction model using Temporal Fusion Transformer (TFT). This study relies on a TFT model for energy forecasting, which considers both primary and valuable data sources and batch training techniques. The model's performance has been related to the Long Short-Term Memory (LSTM), LSTM interpretable, and Temporal Convolutional Network (TCN) models. The model's performance has remained better than the other algorithms, with mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) of 4.09, 2.02, and 1.50. Further, the overall symmetric mean absolute percentage error (sMAPE) of LSTM, LSTM interpretable, TCN, and proposed TFT remained at 29.78%, 31.10%, 36.42%, and 26.46%, respectively. The sMAPE of the TFT has proved that the model has performed better than the other deep learning models.

1. Introduction

The COVID-19 pandemic has disturbed people's daily life activities in offices and homes. Along with the other problems, energy consumption in the residential sector has increased dramatically due to increased spending time by the residents at home [8]. The other primary reason could be shifting offices to work-from-home mode, which has increased the usage time of electronic equipment and ultimately increased the electric bills. Besides these, the other significant reason for increased energy consumption is the rapid growth in the urban population for work and educational opportunities [55,45]. Energy consumption is increasing day by day, and the wastage of energy by residents is due to negligence and unawareness of peak hours [24,14].

The other significant wastage of energy production resources can be noticed in the companies and smart grids. The wastage in smart grids is due to energy production without considering future demand [13,17]. The smart grid network will add new users in the future; hence it is essential to predict the exact amount of future energy consumption demand of customers to avoid wastage and problems like energy shortfall and optimize the consumption [42,41,33]. Moreover, the shortfall may lead to other crises like a disturbance in the official duties of residents working from home due to COVID-19. Scientists have tackled the energy consumption prediction problem using different prediction algorithms like machine learning and deep learning techniques [15,7,44,5]. Deep learning models have remained successful because they can handle big energy consumption data [46,6,51].

Few models and algorithms have focused on the energy consumption prediction of individual customers in smart grid networks [1,34]. Traditional methods have focused on the energy consumption of the next day, hour, week, month, and year [5,12]. The prediction of exist-

* Corresponding authors.
E-mail addresses: mohd.nazir@zu.ac.ae (A. Nazir), shaikh@squ.edu.om (A.K. Shaikh).

https://doi.org/10.1016/j.rineng.2023.100888
Received 29 September 2022; Received in revised form 27 December 2022; Accepted 9 January 2023
Available online 20 January 2023

ing users' energy consumption is more manageable than future demand due to the unavailability of historical data and the limitations of algorithms. The energy demand of every customer is different, but still, residential customers have similar energy consumption behavior [19]. Hence, it is challenging for researchers and scientists to forecast the energy consumption of individual customers in smart grid networks [26]. Ensemble clustering may play a significant role in clustering customers based on energy consumption [25,43]. The novelty of the proposed model relies on consideration of the customer based data rather than overall consumption considered by the traditional models. This study relies on a TFT model for energy forecasting and carefully considers primary and valuable data sources and batch training. It divides training data-sets into many batches based on their correlations and intrinsic behavior to produce predictability of TFTs. The training samples are reconstructed using auto-encoders for the best possible training; current data is preserved and used accurately. The proposed study uses the customers' historical energy consumption to predict the future energy consumption demand. The model focuses on the customer's daily, weekly, and monthly energy consumption [16,34]. The London household's daily energy consumption data-set contains every customer's energy consumption data from November 2011 to February 2014 [10]. The study considers 169 (28 months for each user) customers with a dynamic time of use (dToU) pricing scheme.

The proposed model has three major components, i.e., pre-processing, prediction, and performance evaluation. The prediction module uses Temporal Fusion Transformer (TFT) as a prediction algorithm. The TFT combines multi-horizon forecasting and interpretable temporal dynamics [31].

This model contributes to the following.

• For demand-side management of consumers (customer based), a model based on TFT is developed for forecasting days, weeks, and months of energy consumption in advance.

• As a way to optimize the performance of a TFT, specialized components are used to choose relevant features and a series of gated layers to suppress the unnecessary features to achieve high accuracy.

• In the proposed model, as part of the training process, the TFT model performs a time-series analysis of the input and maintains time-series behavior.

• This study focuses on developing a high-dimension data processing model that can simulate the behavior pattern of load consumption over a long period to eliminate the problem of over-fitting. For this purpose, a batch-training approach is employed for training the TFT model.

• With TFT, local relationships can be learned using recurrent layers, and long-term relationships can be learned using interpretable layer self-attention.

• With its attention-based architecture, TFT offers a comprehensive representation of the dynamics of time through a combination of high-performance multi-horizon forecasting and interpretable temporal insights.

• The performance evaluation of the proposed TFT procedure is examined with various performance evaluation indicators, including some conventional methods like LSTM, LSTM Interpretable, and TCN.

The rest of the paper is as follows; Section 2 presents the related work, Section 3 discusses the proposed methodology, Section 4 presents results, the detailed discussion in Section 5, and Section 6 presents the study's conclusion.

2. Related work

An analysis of relevant publications has been conducted on energy consumption prediction and efficiency improvements. These studies examine energy efficiency models for smart cars, homes, and grids. The literature published on energy consumption prediction has used deep learning methods due to the deep structure and multiple hidden layers. However, the training time of the algorithms makes deep learning methods costly to be used with massive data-sets. The critical analysis of the studies can be seen in Table 1. The paper by Moradzadeh et al. [36] used Variational Autoencoder Bidirectional Long Short-Term Memory (VAEBiLSTM) for energy consumption forecasting. The model has focused on reducing the noise of data for better accuracy. Further, overfitting has been prevented using batch training. As a result, the VAEBiLSTM model's performance remained better than the LSTM and support vector regression (SVR) for energy consumption forecasting.

The energy consumption activities of residents in houses also help predict residents' energy usage behavior. Lin [32] model has detected and classified residents' home activities using the non-intrusive load monitoring (NILM) technique. By dis-aggregating electricity consumption onto the level of individual appliances, NILM replaces all networked plug-level energy meters. The day ahead scheduling also remained an attractive topic [2] implemented the demand side management (DSM) strategy in a smart grid for the day ahead scheduling to tackle the cost and pollution. Load costs have been reduced by considering the output power of curtailable loads (CLs) and wind turbines (WTs). Yang and Schell [54] used a pre-trained parallel convolutional neural network with a quadruple-branched autoencoder (QCAE) as a forecasting model to develop an optimal design for an autoencoder. In addition to compressing the input, QCAE also compresses the features. Salleh et al. [39] have used data from historical electricity usage for forecasting the identification of optimum historical load using long short-term memory (LSTM) for short-term forecasting. Berriel et al. [5] predicted energy consumption using a deep learning model with normalized data. Le et al. [27] used a convolutional neural network (CNN) and Bi-directional LSTM for the energy consumption prediction. Deep reinforcement learning (DRL) has been used in the model of Xu et al. [53] for building energy management to solve the energy schedule issues of buildings with solar energy storage systems at a minimum electricity price for beneficial people's lifestyle and pocket friendly. However, the DRL agent only makes decisions when measurable information occurs and does not consider the changing environmental conditions.

Deep reinforcement learning has also played a significant role in the model by Li et al. [30] to identify and address the uncertainty in charging power forecasting for electric vehicles. Long short-term memory (LSTM) calculated the electric vehicle charging station (EVCS) charging powerpoint forecasts. The adaptive exploration proximal policy optimization algorithm based on reinforcement learning models changes LSTM cell states using a Markov decision process (MDP).

The multi-objective bi-level optimization model has been used to reduce energy usage with a maximum profit for agents. Jia et al. [22] have examined many stakeholders' allies on agents' pricing methods to gain optimum business profit and progress. The optimization model uses the Stackelberg game.

In addition, the Karush-Kuhn-Tucker (KKT), duality theory, is based on linear relaxation that solves and simplifies the bi-level nesting problem. Lee et al. [28] attempt to describe the life assessment model based on the price of an energy storage system (ESS) and schedule. The depth-of-discharge (DOD) stress model and fatigue analysis techniques determine the life-cycle price function depth in detail. A tool based on reinforcement learning to determine the optimal schedule of the ESS model on its life-cycle cost. Kanellos et al. [23] proposed a system for integrating multiple plug-in electric vehicles (PEVs) in a large cluster and connecting their distribution networks. PEVs are charged using a method that reduces the cost of charging up to the distribution network.

Utkarsh and Srinivasan [47] are concerned about global warming, and government policies have been developed to encourage the integration of distributed energy resources, or DERs (usually renewable energy sources). The study presents efficient strategies for integrating and coordinating distributed energy resources in medium-to-low voltage distribution networks/micro-grids. Natividade et al. [37] examined the energy efficiency benefits of Energy Performance Contracting

Table 1
Critical analysis of the already proposed methods.

Ref.	Focus/Aim	Method/Algorithm	Schedule	Pre Process	Time Series	Single User Pred.	Multi User Pred.	Big Data	Batch Train.	Total Cons. Pred.	Price Pred.
[36]	**Energy forecast**	**VAEBiLSTM, SVR**	x	✓	✓	x	x	✓	✓	✓	x
[2]	**Demand-side mgmt./ Day ahead schedule**	**PDF, MOWDO, DMM**	✓	x	x	x	x	x	x	✓	x
[54]	**Optimal design for autoen.**	**QCAE**	x	✓	✓	x	x	x	x	✓	x
[39]	**Short-term energy forecast**	**LSTM**	x	✓	✓	x	x	✓	x	✓	x
[5]	**Energy cons. forecast**	**FCN, CNN, LSTM**	x	x	✓	x	x	x	x	✓	x
[27]	**Energy cons. pred.**	**CNN, Bi-LSTM**	x	x	✓	x	x	x	✓	✓	x
[30]	**Charging power forecast of electric vehicles**	**DRL, LSTM-AEPPO, MDP**	x	✓	x	x	x	x	x	✓	x
[28]	**Energy storage system (ESS)**	**MDP, MP**	✓	✓	x	x	x	x	x	✓	x
[23]	**Integrate multiple plug-in electric vehicles (PEVs)**	**PEV, MAS**	x	✓	x	x	x	x	x	x	✓
[37]	**Energy efficiency benefits of energy performance contracting (EnPC)**	**EnPC**	x	✓	x	x	x	x	x	x	✓
Proposed Model	**Energy prediction**	**TCN, LSTM, LSTM Int., and TFT**	x	✓	✓	✓	✓	✓	✓	✓	x

(EnPC). It looks at the potential risks and rewards of contracting with Energy Service Companies (ESCOs) based on traditional and alternative models.

The literature review reveals that the researchers have focused on different dimensions, like energy storage, prediction, optimization, cost reduction, and handling of the pollution problem. Most models have used deep learning and deep reinforcement learning techniques for prediction problems. It is, however, necessary to develop a novel model to predict the future demand for energy consumption by individual customers within a smart grid network.

3. Research design and methods

The proposed model predicts individual customers' demand for energy consumption in a smart grid network. The model contains three modules: pre-processing, prediction, and performance evaluation. The prediction module of the proposed model includes a Temporal Fusion Transformer (TFT) responsible for predicting customers' energy consumption demand.

3.1. Database acquisition and description

The London household's daily energy consumption data-set contains every customer's data from November 2011 to February 2014 [10]. The proposed study considers 169 (28 months for each user) customers with a dynamic time of use (dToU) pricing scheme. The selection of the users is based on the first available 169 customers (some dropped due to high energy consumption) in the data-set. Accordingly, the model considers the daily energy consumption in kWh for the experiment, which means 126 (75%) for training and 43 (25%) for testing. The data of only 169 customers have been considered because of the limitations of the memory and available computation equipment and resources of Google Colab.

Furthermore, with the increased number of customers, the training time of models has increased; hence the training was not completed within the allocated time of GPU and other resources. So the maximum possible number of customers was 169 for the smooth training of models with reasonable time. While for evaluation purposes, this paper mentions the experimental results of one customer.

3.2. Pre-processing

Outliers, or observations below or exceeding Q1 + 1.5 IQR, have been identified using the interquartile range (IQR) with the Python package NumPy. The data has been pre-processed using a moving average to remove the prominent outliers. The normalization and denormalization into 0-1 scale have been carried out using (1) and (2).

$$N(a) = \frac{x(a) - \min(a)}{\max(a) - \min(a)} \tag{1}$$

$$D(a) = N(a) * (\max(a) - \min(a)) + \min(a) \tag{2}$$

Where $N(a)$ and $D(a)$ represent the normalized and denormalized data, and $x(a)$ represents the value that is being normalized. While $\min(a)$ and $\max(a)$ denote the minimum and maximum values of the dataset.

3.3. Prediction module

The most important module of the proposed model is the prediction module. In the initial step, the data has been prepared to be used as input by the prediction module containing the Temporal Fusion Transformer (TFT) as a prediction algorithm [31]. The algorithm uses training data sets and the day series as past covariates. While the input variables are day, month, year and energy consumption in kWh. The lag order of t-2 has been considered for the experimentation for all the scenarios of the experimentation like day, week and month. So for the day the TFT considers energy consumption of the previous 2 days, while for the week it looks into the previous two weeks energy consumption. The same stands for the month scenario. The details can be seen in the Table B.7. Training a machine learning model involves multiple hyper-parameters set before learning begins. The TFT also offers to optimize the parameters and define learning rate using the predefined function of the PyTorch. With PyTorch Lightning learning rate finder, optimize-hyperparameters() determines the learning rate. In the proposed model we have not considered the optimize-parameters() due to various reasons including the complexity. The parameter settings of the algorithm have been carried out based on the trial and error method for finding the best settings. The following parameter settings of the TFT have been used for the experimentation. One layer of the LSTM has 16 neurons in the proposed network. The random state was 42, with 8 as the hidden continuous size. In the experiment, the parameter space for the random state ranged from 1 to 70. There is very little difference in the performance of the algorithm when the random state is changed, so the final model uses 42 random state values. The network uses 30 and 15 input and output chunks, respectively. With a learning rate of le-3, a batch size of 1000 was considered with a 0.1 dropout rate. Model computation costs are heavily influenced by batch size, input chunks, and output chunks. Due to this, the number has been carefully selected. The model considers 30 days of data to predict the next 15 days, increasing accuracy. We have tested batch sizes of 250 to 2000, and it was decided that the batch size of 1000 provides the highest performance at the lowest cost. This is due to larger batches, which reduce computation cost but decrease accuracy. The algorithms training took longer than the other deep learning models, with 150 epochs. The MSE Loss() function uses day series as covariates. The settings of different deep learning algorithms can be seen in Table B.7.

The proposed model uses the TFT implementation in PyTorch forecasting. The model employs past covariates derived from input chunks before prediction time and future covariates derived from output chunks after prediction time [31]. The algorithm uses multi-head attention queries on future inputs from mandatory future covariates. In addition to the past covariates, encoders were employed to generate the day covariates (future covariates) automatically.

3.3.1. Temporal Fusion Transformer (TFT)

The TFT uses covariates known as future inputs in multi-horizon forecasting. The algorithm blends high performance with forecasting in multi-horizon to analyze temporal dynamics [31,21]. The algorithm can do local and long-term processing using recurrent and interpretive self-attention layers. We can control superfluous components to choose relevant attributes from groups of accumulating layers. The adaptive depth for network convolution for the energy consumption data set makes the algorithm suitable to handle big dataset [38]. It provides gating methods to ignore any unnecessary components of the architecture, providing flexibility to networks for variable selection for each time step and specified input variables [52]. Temporal dynamics use encoders to add static features into the network using context vector encoding. The algorithm learns long and short-term correlations of inputs using temporal processing. The sequence-to-sequence layer takes responsibility for the local processing, while the interpretable multi-head attention block deals with long-term dependency [31]. The algorithm uses a deterministic forecast instead of a quantile forecast using MSELoss(). The proposed model can be seen in Fig. 1.

Dynamically generated forecasts are known as deterministic forecasts when they produce forecasts with all the available computing power [20]. The deterministic approach has been adopted as it requires less information with minimum mathematical calculations, reducing the training time and producing detailed forecasts. The deterministic approach remains suitable for short-term energy consumption prediction (hourly, daily) and may not perform better with the one-year ahead forecast [6]. It can be noted that the layers of the model have been shortened for simplicity, and the detailed model readers can refer to the original model proposed by Lim et al. [31].

Lim et al. [31] have already discussed the TFT in extensive detail, including the functionality, detailed mathematical calculations involved in each component, and the complete TFT model flow. These mathematical expressions have been taken from the model explained by Lim et al. [31]. This paper has discussed its highlights to provide a brief overview of how the TFT works.

3.3.2. Gated Residual Network (GRN)

The GRN provides flexibility to the model considering the input a and vector c as in (3).

$$GRN_{\omega}(a,c) = LayerNorm(a + GLU_{\omega}(\eta 1)) \tag{3}$$

$$\eta 1 = W_{1,\omega}\ \eta 2 + b_{1,\omega} \tag{4}$$

$$\eta 2 = ELU(W_{2,\omega} a + W_{3,\omega} c + b_{2,\omega}) \tag{5}$$

The Exponential Linear Unit is represented by ELU [9]. Intermediate layers of the network are denoted by $\eta 1 \epsilon R^{d_{model}}$ and $\eta 2 \epsilon R^{d_{model}}$ as in (4) and (5). The normalization has been represented by the LayerNorm [3]. While the index for weight sharing can be denoted by ω. GLUs (Gated Linear Units) [11] are utilized as component gating layers to control the model's unnecessary components. The input $\gamma \epsilon R^{d_{model}}$ creates following mathematical structure of GLU as seen in (6).

$$GLU_{\omega}(\gamma) = \sigma(W_{4,\omega}\gamma + b_{4,\omega}) \odot (W_{5,\omega}\gamma + b_{5,\omega}) \tag{6}$$

The sigmoid function has been represented by $\sigma(.)$. The weights of the model are represented by $W(.)\epsilon R^{d_{model} * d_{model}}$. The biases of the GLU can be denoted by $b(.)\epsilon R^{d_{model}}$. The d_{model} expresses hidden state size while the Hadamard product is represented $\odot$. As the GLU outputs can all be close to 0, TFT can suppress the nonlinear impact of GRN as much as possible, skipping the layer altogether if necessary.

3.3.3. Variable selection networks

Utilizing learning capacity only on the most critical variables can significantly improve model performance. Variable selection networks are used to pick relevant input variables at each step. Additionally, variable selection allows TFT to remove noise inputs that could adversely affect performance and provide insights into what variables are most significant [31]. Feature representations of categorical variables

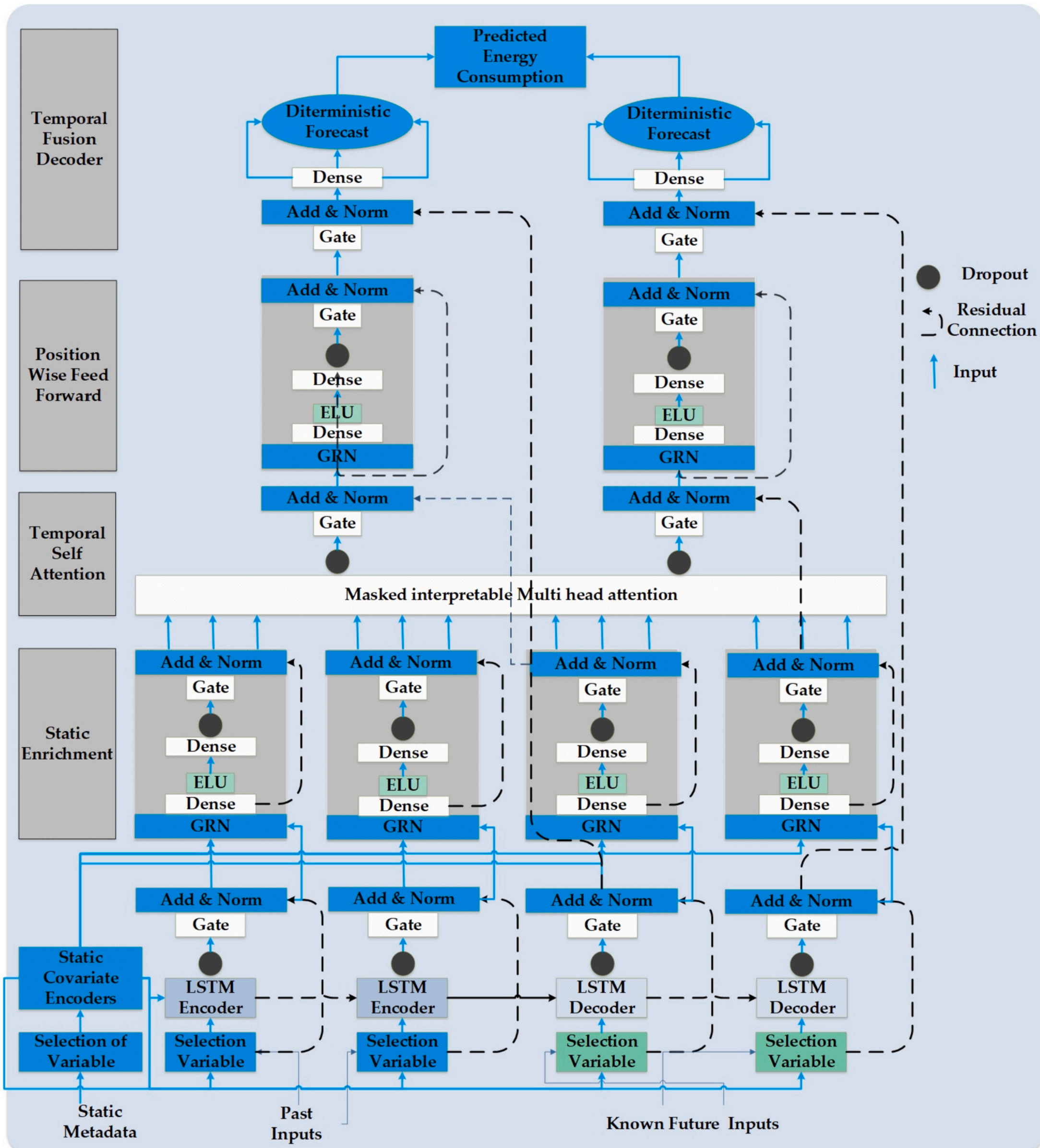

Fig. 1. Proposed TFT model.

are represented through entity embeddings [18], while linear transformations were used to define continuous variables. Variable selection weights are generated by feeding both Ξt and an external context vector c_s through a GRN, followed by a Softmax layer. An external context vector c_s and Ξt are used to generate variable selection weights, using a GRN and Softmax layer (7) where the flattened vector of all past inputs at time t is denoted by Ξt.

$$v_{X_t} = Softmax\, GRN_{vX}(\Xi t, c_s) \tag{7}$$

The vector of variable selection weights is denoted by $v_x t\epsilon\ R\ mx$, wherein c_s is acquired using a static covariate encoder. The nonlinear processing layer at each step feeds $\xi_t^{(j)}$ through GRN as (8).

$$\xi_t^{\sim(j)} = GRN\xi_{(j)}^{\sim}\,(\xi_t^{(j)}) \tag{8}$$

The processed feature vector of j has been represented by $\xi_t^{\sim(j)}$ which are combined as (9).

$$\xi_t^{\sim} = \Sigma_{j=1}^{m_x} v_{xt}^{j}\, \xi_t^{\sim(j)} \tag{9}$$

Where the j element of the vector v_{xt} is represented by v_{xt}^{j}.

3.3.4. Static covariate encoders

The static covariate encoders in the model use context vectors to condition temporal dynamics to integrate static characteristics into the network. The GRN encoders in the TFT network produce temporal variable c_s, temporal features processing c_c, c_h and temporal features enrichment c_e. The static covariate encoders are connected to the model's different layers, as seen in Fig. 1.

3.3.5. Temporal processing

The temporal processing in the model is used to learn both long- and short-term temporal correlations from time-varying inputs (energy consumption data) that are both seen and known. Local processing is handled by a sequence-to-sequence layer, while long-term dependencies are operated by a unique interpretable multi-head attention block [29, 48]. Multi-head attention was first proposed in [48] and modified in [31] to share values in all heads and employ additive aggregation as (10), (11), (12) and (13).

$$IMH(Q,K,V) = H^{\sim} W_H \tag{10}$$

$$H^{\sim} A^{\sim}(Q,K)V\, W_v \tag{11}$$

$$= \{\ 1/H \Sigma_{h=1}^{mH} A(QW_Q^{(h)}, KW_k^{(h)})\ \}\, ssVW_v \tag{12}$$

$$= \{\ 1/H \Sigma_{h=1}^{mH} Attention(QW_Q^{(h)}, KW_k^{(h)})\ \}\, VW_v \tag{13}$$

Where the shared weights across all heads are represented by $W_v \epsilon R^{d\ model\ x\ d_v}$, and $W_H \epsilon R^{d\ attn\ X\ d_{model}}$ denote linear mapping.

3.3.6. Temporal fusion decoder

The temporal fusion decoder uses different layers.

3.3.7. Sequence-to-sequence layer

The sequence-to-sequence layer employs an LSTM encoder and decoders. The model uses cc, ch context vectors for the initialization of cell and hidden state, respectively, of the LSTM. The gated skip connection in the model can be represented as (14).

$$\varnothing^{\sim}(t,n) = LayerNorm(\xi_{t+n}^{\sim} + GLU_{\varnothing^{\sim}}(t,n)) \tag{14}$$

Where $n\epsilon \left[-k, t_{max}\right]$ is a position index.

3.3.8. Static enrichment layer

The static enrichment layer helps to enhance the temporal features. The static enrichment for n index can be represented by (15).

$$\theta(t,n) = GRN_{\theta}(\varnothing^{\sim}(t,n), c_e) \tag{15}$$

The GRN_{θ} weights remain shared with different layers, and c_e denotes the context vector.

3.3.9. Temporal self-attention layer

The multi-head attention can be represented by (16), wherein the $\Theta(t)$ denotes enriched temporal features for the determination of $B(t)$.

$$B(t) = IMH(\Theta(t), \Theta(t), \Theta(t)) \tag{16}$$

TFT uses the self-attention layer to discover long-range dependencies. The decoder masking [29,48] is essential for ensuring that each temporal dimension can only attend to its preceding feature. The additional layer helps to facilitate the training (17).

$$\delta(t,n) = LayerNorm(\theta(t,n) + GLU_{\delta}(\beta(t,n))) \tag{17}$$

3.3.10. Position-wise feed-forward layer

The position-wise feed-forward layer uses GRNs as in (18).

$$\psi(t,n) = GRN\psi(\delta(t,n)) \tag{18}$$

The layers share weights of $GRN\psi$. The gated residual connection provides a direct path to the sequence-to-sequence layer using (19).

$$\psi^{\sim}(t,n) = LayerNorm(\phi^{\sim}(t,n) + GLU_{\psi^{\sim}}(\psi(t,n))) \tag{19}$$

3.3.11. Deterministic forecasts

Calculate the range of likely goal values at each prediction horizon. Only one deterministic forecast has excellent accuracy over the first six days; it is fast, easy to use, and sensitive to new data. The system does not convey confidence, does not produce probability forecasts, and displays little ability to predict the long term. The aim of the proposed model is short-term energy consumption prediction; hence deterministic forecast remains suitable [4]. The sequence-to-sequence layer of the model takes responsibility for the local processing, while the interpretable multi-head attention block helps deal with long-term dependency [31].

3.4. Performance evaluation metrics

The model's performance has been measured using mean absolute error (MAE). An MAE also ignores the direction of prediction errors and measures the average magnitude. A specific example is the average absolute difference between a prediction and the actual value across the entire testing set. Each difference is weighed equally in this measurement. The second metric is the RMMSE. A measure of how spread out the residuals are is the root mean square error (RMSE), which calculates the standard deviation of errors. The Symmetric mean absolute percentage error (sMAPE) measures based on percentage. Finally, a metric called mean square error (MSE) describes the average difference between the predicted and actual values [49]. These performance metrics can be mathematically expressed as (20), (21), (22), and (23). Researchers have commonly used these matrices to measure the efficiency of algorithms and models.

$$MAE = \frac{1}{N}\sum_{i=1}^{n} |A_i - P_i| \tag{20}$$

$$sMAPE = \frac{100\%}{N}\sum_{i=1}^{n} \frac{\frac{|P_i - A_i|}{|A_i| + |P_i|}}{2} \tag{21}$$

$$MSE = \frac{1}{N}\sum_{k=0}^{n} (A - P_i)^2 \tag{22}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=0}^{n} (A - P_i)^2} \tag{23}$$

Where N stands for total observations, A for actual values, and P for predicted values.

4. Implementation, results, and performance evaluation

4.1. Implementation

The training of the deep learning algorithms has been carried out using the GPU of Google Colab pro service. The experimentation has been carried out on Tesla P100, a 16 GB graphics card with a RAM of 16 GB. The graphical representation has been carried out in Matlab 2022a.

4.2. Results

The complete testing data set has been used to measure the model's efficiency, while the customer having id 868 was selected for the elaboration and graphical representation. The reason for elaboration using a single customer is to avoid complexity and improve the paper's readability. However, the proposed model lets readers know about the rest of the customer's predicted energy consumption. Customer 868 was chosen randomly; the user was the last user of the testing dataset. The number of days considered is entirely dependent on the data set provided. As a result, the energy usage for customer id 868 is accessible from May to December 2012, January to December 2013, and January to February 2014. Energy consumption prediction has been used for the graphical representation only one day ahead.

Fig. 2 shows LSTM, LSTM interpretable, TCN, and the proposed TFT model's one-day-ahead energy consumption predictions for 2012. The May to June 2012 graph fluctuates with lower energy consumption of around 1 kWh and a maximum of 11 kWh. At the same time, the proposed model has predicted energy consumption in the range of 4 to 7 kWh. The reason is due to frequent fluctuations in the energy consumption behavior of the customer.

Fig. 3 illustrates a similar pattern for energy usage from July to September 2012, with a higher 11 kWh actual energy consumption and a lower 1 kWh. Unlike the preceding graph from May to June 2012, the LSTM, LSTM interpretable, TCN, and proposed models all show the

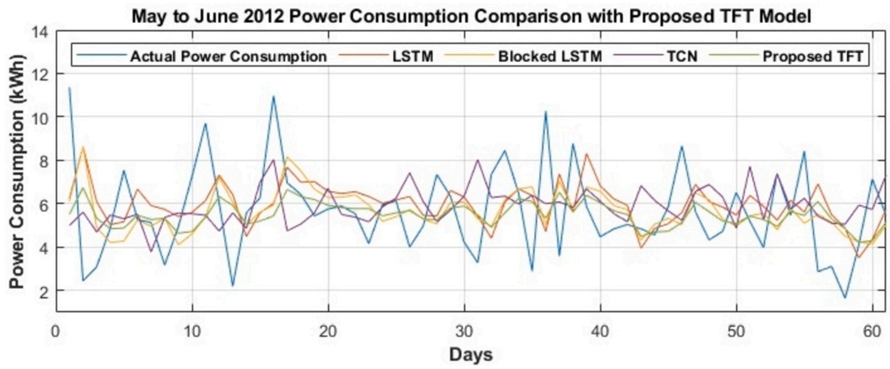

Fig. 2. May to June 2012 power consumption comparison.

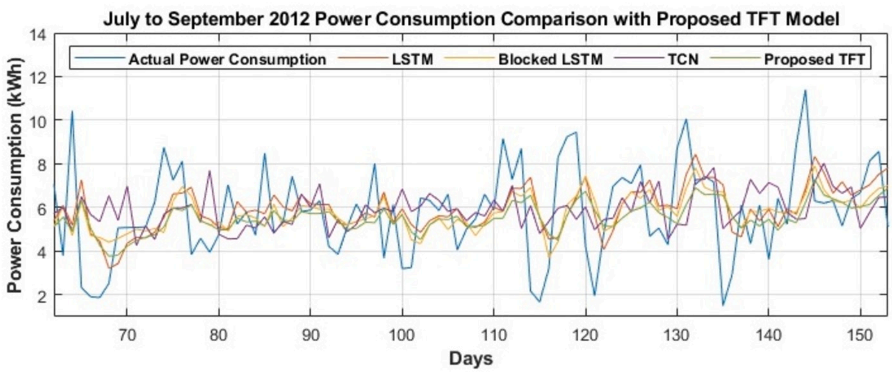

Fig. 3. July to September 2012 power consumption comparison.

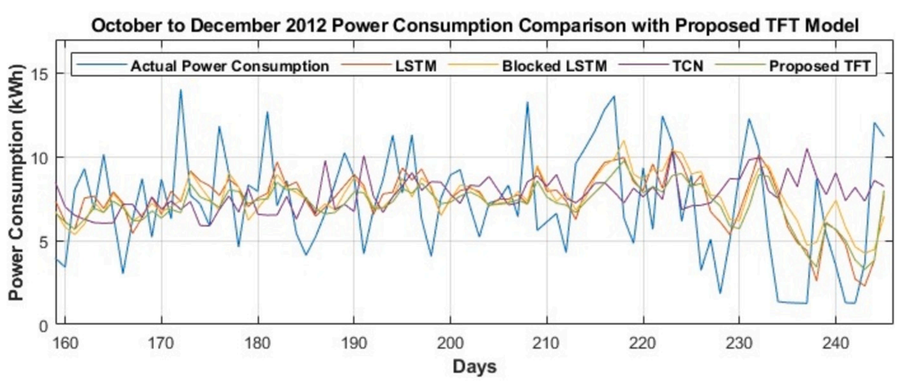

Fig. 4. October to December 2012 power consumption comparison.

same pattern of predicted energy usage for July to September 2012. On the other hand, the proposed model outperformed the May-June 2012 period.

Fig. 4 contains energy consumption prediction for October to December 2012. The higher actual energy consumption can be 14 kWh and lower around 2 kWh. The LSTM, LSTM interpretable, and proposed model have performed better than the TCN model. Overall, the suggested model performed better than the other deep learning interpretable models for one day ahead of energy consumption in 2012.

Fig. 5 displays predicted energy consumption from January to March 2013, with the highest actual consumption of about 15 kWh and the lowest at around 3 kWh. While the LSTM, LSTM interpretable, and proposed model have all successfully predicted energy usage, the TCN graph reveals some irregularities compared to the actual energy consumption.

Fig. 6 depicts the energy consumption prediction from April to June 2013, with the highest actual energy consumption of 10 kWh and the lowest of 1 kWh, indicating variations in the energy consumption. The LSTM, LSTM interpretable, and suggested model show a similar trend; however, the TCN graph differs significantly from the actual energy usage graph.

Fig. 7 shows the estimated energy use from July to September 2013. Around 11 kWh represents the highest actual energy use, while 3 kWh represents the lowest. The LSTM, LSTM interpretable, and suggested model all outperformed TCN, as evidenced by the graph showing that TCN's chart is distinct from the other algorithms. TCN is having difficulty due to data volatility. The proposed model has a deep structure suited for data uncertainties.

Fig. 8 shows the predicted energy consumption for October to December 2013, with the highest actual energy consumption of about 13 kWh and the lowest at around 2 kWh. The graph shows that the proposed model managed better than the LSTM, LSTM interpretable, and TCN models. In this case, the TCN model follows a similar trend, with the straight graph indicating that it has not performed as well as other

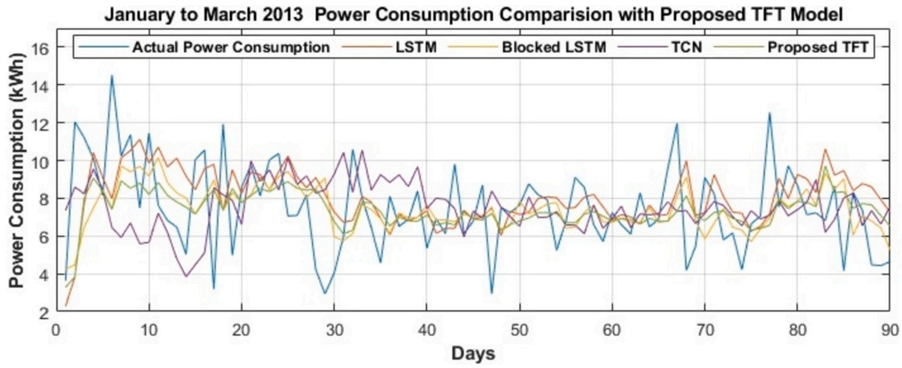

Fig. 5. January to March 2013 power consumption comparison.

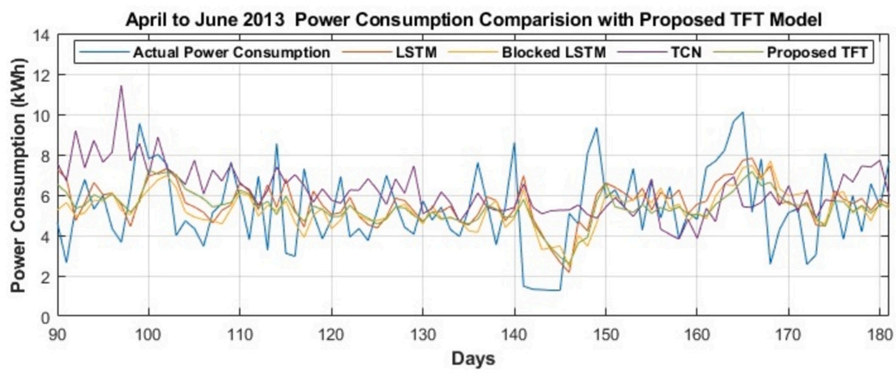

Fig. 6. April to June 2013 power consumption comparison.

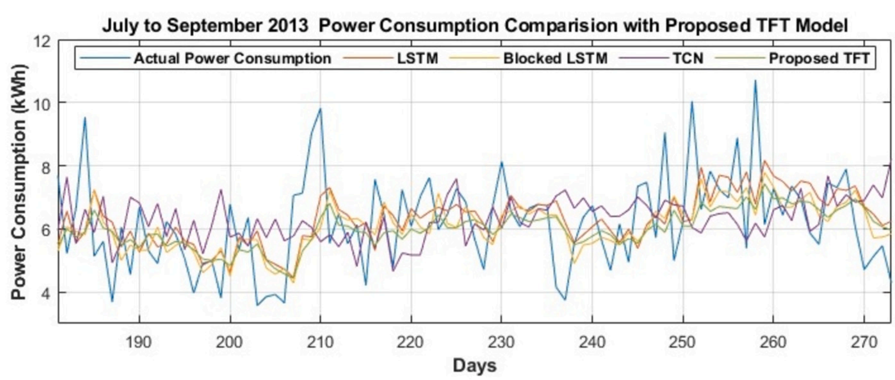

Fig. 7. July to September 2013 power consumption comparison.

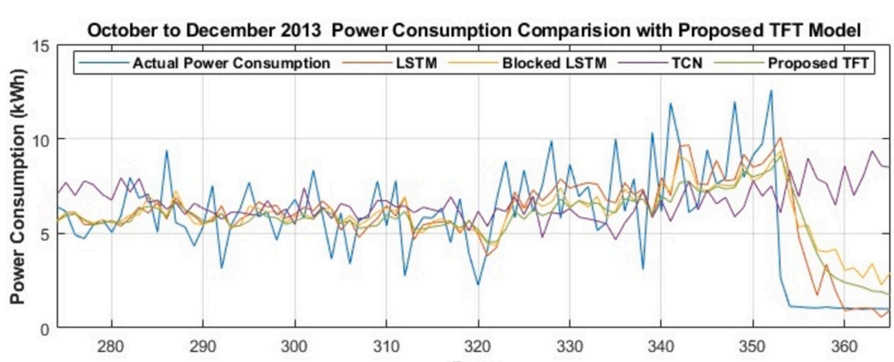

Fig. 8. October to December 2013 power consumption comparison.

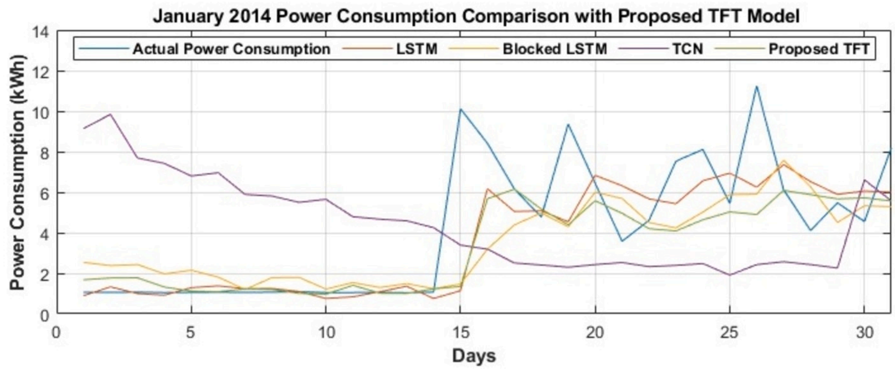

Fig. 9. January 2014 power consumption comparison.

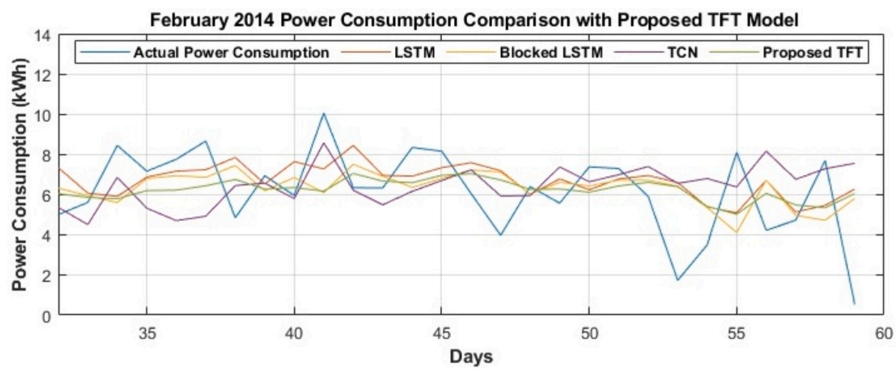

Fig. 10. February 2014 power consumption comparison.

deep learning models. Overall, the proposed model's performance in predicting the customer's energy usage has improved from January to December 2013.

Fig. 9 shows the energy consumption prediction for January 2014, with extremely low energy consumption on the first days of the month, approximately 1 kWh, and several peaks after day 14, with the highest energy consumption of 11 kWh on the 26th day of the month. The TCN graph reveals that it incorrectly projected energy consumption in January 2013. However, the LSTM, LSTM interpretable, and proposed models outperformed the TCN because the actual and predicted energy graphs follow the same pattern.

Fig. 10 depicts the energy consumption forecast for February 2014, with actual consumption remaining higher throughout the month's first days. The highest energy use, estimated at 10 kWh, is noticeable. When we look at the lowest energy, we find around 1 kWh. All algorithms performed better in the February scenario, including the TCN model, which had failed in the previous month's energy consumption estimate.

5. Discussion

Various scenarios have been adopted to predict energy consumption over the year, such as daily, weekly, and monthly. The complexity of the data has created problems for all the algorithms, making prediction challenging. It has been found that the proposed model outperforms other deep learning algorithms. To extensively evaluate the model's performance, the comparison has been divided into daily, weekly, and monthly. Furthermore, the first scenario experiments with the model's performance on the complete testing data set, including data of 40 customers. While for the detailed evaluation and interpretation, one random customer has been evaluated. The precise performance of the model is as follows.

5.1. Comparison of sMAPE

The proposed model performs the best in global performance, especially for testing data based on daily, weekly, and monthly experiments. Energy consumption data in the testing reflect uncertainties; that is, the TFT is best suited for dealing with uncertain data on energy consumption. It is possible to observe and verify this opinion more clearly in Figs. 2 to 10. Table 2 shows the sMAPE performance of the four models for the testing data. Tables 2 and 4 show the TFT's smallest error scale compared to the four comparative methods.

Additionally, this experience indicates that the TFT is more reliable than other comparative approaches when predicting future energy demand. In Figs. 2 to 10, we can also see that TCN has the lowest performance. Tables 2, 3, 4, and 5 show that the TFT approach has the smallest prediction error and highest accuracy for energy consumption forecasting.

There are 35.14%, 33.30%, and 32.17% sMAPE values for the LSTM, LSTM interpretable, and TCN models, respectively. The proposed model outperformed the other algorithms for the May to June 2012 day-ahead energy consumption prediction with an sMAPE of 26.27%.

The LSTM, interpretable LSTM, and TCN have sMAPEs of 30.96%, 29.82%, and 31.50%, respectively, for July to September 2012. The TFT outperformed the other algorithms despite the uncertainty of the data with an sMAPE of 29.44%. Furthermore, it is important to note that the sMAPE errors for October-December 2012 were 36.33%, 37.74%, and 38.79%, respectively, while the TFT errors were 29.71%. The errors of the proposed model for the available data of 2012 remained lower compared to the three deep learning models. Although the data for the entire year of 2012 was unavailable, the model has difficulty learning the consumption pattern.

As the data for 2013 was complete from January to December, the results are expected to be better than those in 2012. From January to March 2013, the sMAPE was 27.51%, 26.12%, and 28.61%. TFT outper-

Table 2
Detailed prediction accuracy of the TFT model in terms of sMAPE.

Model	Forecast Horizon	May-June	July-Sept.	Oct.-Dec.	Jan.-March	April-June	July-Sept.	Oct. to Dec.	Jan.	Feb.	Overall
	1 Day	sMAPE of 2012			sMAPE of 2013				sMAPE of 2014		sMAPE
LSTM		35.14%	30.96%	36.33%	27.51%	32.08%	19.41%	28.55%	28.58%	32.14%	29.78%
Blocked LSTM		33.30%	29.82%	37.74%	26.12%	32.82%	18.96%	34.67%	44.28%	32.84%	31.10%
Temporal Convol. Networks (TCN)		32.17%	31.50%	38.79%	28.61%	35.58%	20.44%	43.35%	107.71%	32.77%	36.42%
Proposed TFT		**26.27%**	**29.44%**	**29.71%**	**23.24%**	**28.21%**	**17.30%**	**27.65%**	**20.46%**	**26.63%**	**26.46%**

formed all other algorithms in terms of error, with the lowest sMAPE of 23.24%. These errors are smaller than those of 2012; thus, it has been proven that algorithms can be trained appropriately only if the data is complete. The TFT achieved an sMAPE of 28.21% from April to June 2013, while errors were 32.08%, 32.82%, and 35.58% for the other three algorithms. From July to September, the TFT achieved an sMAPE of 17.30, lower than 19.41%, 18.96%, and 20.44%. Finally, for October to December 2013, the TFT has shown the lowest value of sMAPE of 27.65%; hence it has outperformed the counterpart algorithms.

As of January 2014, the proposed TFT has outperformed all algorithms with an sMAPE of 20.46%. The other algorithms have sMAPE of 28.58%, 44.28%, and 107.71%. Similar is the story for February, wherein the LSTM, LSTM interpretable, and TCN has sMAPE of 32.14%, 32.84%, and 32.77%. Regarding sMAPE, the proposed TFT outperformed all other algorithms with a value of 26.63%. Energy consumption predictions for LSTM, interpretable LSTM, TCN, and the proposed model follow a similar pattern. Therefore, the suggested model outperformed existing deep learning algorithms. The overall sMAPE of the LSTM was 29.78%; the LSTM interpretable was 31.10%. The TCN sMAPE remained at 36.42%, while the proposed model has outperformed all the algorithms with an sMAPE of 26.46%.

Observing the results, we can conclude that any model's best energy prediction occurred between July and September 2013. At the same time, the worst prediction of any algorithm was for October to December 2012. The worst performance from October to December is due to increased uncertainty in customer energy consumption and complexity. The worst predicted single month by any algorithm was January having an sMAPE error of 107.71% by TCN.

5.2. Comparison of energy consumption

The performance measurement of the model in terms of the energy consumption in kWh provides the overall success of the proposed model. It is more evident from the results that deep learning models can deal with complex tasks. The time series prediction is also one of those complex problems. We can observe the energy consumption of customer 868 in terms of the four seasons of England, i.e. spring (March to May), summer (June to August), Autumn (September to November) and Winter (December to February). For the spring season of 2012, there is no data available, so we can observe energy consumption data from March to May for the spring season of 2013. In February, it is evident that consumption has reduced compared to January, and then during the spring season, consumption remains the same. This means it decreases at night and increases during the day, except for May, where it remains the same day and night. The summer time has higher consumption compared to the spring as seen in June 2013 Fig. 6 and July to August in 7. There is a significant difference in the consumption during Autumn starting from September to November 2013 as seen in Figs. 8 and 9. A similar pattern is observed in the Winter with a drop in consumption between December and the first two weeks of January, followed by a return to normal consumption around 5-8 kWh in February 2014 as seen in 10. The energy consumption prediction with three traditional models has been elaborated in Table 3 compared with the actual energy consumption. The customer having id 868 has 348.08 kWh, 530.16 kWh, and 662.17 kWh actual energy consumption for May to June, July to September, and October to December 2012, while the proposed model has predicted it as 346.25 kWh, 514.72 kWh, and 651.90 kWh, respectively. It can be evident from the comparison that the TCN model has performed worst compared with the other three models.

From January to March 2013, the customer consumed 676.44 kWh of electricity; TFT predicted the same as 682.74 kWh. It can be seen that the best-predicted energy consumption by the proposed model was for April to June 2013, wherein the actual energy consumption was 489.83 kWh, and the proposed model has predicted that as 489.18 kWh. The consumption was reduced to 570.15 kWh from July to September 2013, while the proposed model predicted it as 562.18 kWh. In the fol-

Table 3
Detailed energy prediction of the TFT model.

Model	Forecast Horizon	May to June	July to Sept.	Oct. to Dec.	Jan. to March	April to June	July to Sept.	Oct. to Dec.	Jan.	Feb.	Total
	1 Day	Energy consumption for 2012			Energy consumption for 2013				Energy consumption for 2014		Energy cons. in kWh
Actual		348.08	530.16	662.17	676.44	489.83	570.15	536.36	129.09	172.15	4114.43
LSTM		365.21	537.28	696.34	721.62	509.00	580.83	551.46	113.03	187.19	4261.96
Blocked LSTM		345.74	524.29	696.81	669.25	483.48	560.61	556.51	109.39	177.00	4123.08
Temporal Convol. Networks (TCN)		355.52	543.78	698.66	678.43	564.32	579.95	608.84	138.74	180.11	4348.35
Proposed TFT		**346.25**	**514.72**	**651.90**	**682.74**	**489.18**	**562.18**	**559.82**	**121.51**	**183.05**	**4111.35**

Table 4
Performance comparison of TFT with deep learning models.

Model	Forecast Horizon	Metrics		
	Days Ahead	MSE	RMSE	MAE
LSTM	1	4.94	2.22	1.69
	7	6.69	2.59	1.90
	30	9.68	3.11	2.38
LSTM Interpretable	1	5.06	2.25	1.70
	7	5.94	2.44	1.83
	30	6.74	2.60	1.99
Temporal Convolutional Networks (TCN)	1	7.14	2.67	2.00
	7	6.98	2.64	1.98
	30	6.43	2.54	1.93
Deep Neural Network (DNN) [40]	30	–	–	23.5
Recurrent Neural Network (RNN) [40]	30	–	–	22.4
Gated Recurrent Unit (GRU) [40]	30	–	–	22.5
LSTM [40]	30	–	–	19.2
Proposed TFT	1	4.09	2.02	1.50
	7	5.75	2.40	1.79
	30	6.49	2.55	1.94

lowing three months, from October to December 2013, the customer consumed 536.36 kWh of electricity, and the proposed model predicted it as 559.82 kWh. In January 2014 customer consumed 129.09 kWh, while the proposed model predicted it as 121.51 kWh. During February, consumption increased to 172.15 kWh, and the TFT predicted it as 183.05 kWh. The total actual energy consumption for customer id 868 was 4114.43 kWh, while the LSTM, LSTM interpretable, TCN, and proposed models have predicted it as 4261.96 kWh, 4123.08 kWh, 4348.35 kWh, and 4111.35 kWh, respectively.

So if we observe the month-wise performance, there is a difference in the actual and predicted energy consumption, but in terms of the overall energy consumption, the deep learning models have remained better at predicting the future demand of the customer. The other most prominent algorithm was LSTM interpretable, which performs better than the proposed model. The reason is similar building blocks and usage of LSTM in the layers of the TFT model. Therefore, the proposed model can be implemented in a smart grid for future energy consumption prediction demand for new customers. The model is worth in a smart grid environment because, in everyday situations, the smart grid can predict the future demand, but with the addition of a new customer, it might be challenging to predict the new customer's energy consumption demand. So, in addition to the overall consumption prediction, production companies can predict the future energy consumption of a single customer.

5.3. Prediction errors comparison with traditional methods

Table 4 compares the proposed Temporal Fusion Transformer (TFT) model to Temporal Convolutional Networks (TCN) in terms of performance. Long short-term memory (LSTM) and LSTM interpretable were among the other algorithms. The MSE, the RMSE, the sMAPE, and the MAE matrices are used to statistically analyze the performance of the four prediction models. At the same time, the forecast horizon was considered for the customer's future daily, weekly, and monthly energy consumption. Table A.6 shows the training and validation losses for each algorithm.

5.3.1. Day-ahead energy consumption

Table 4 compares the four deep learning algorithms to predict future energy demand daily. The deep learning models have a strong capability of extracting the patterns of data; hence all models have performed better. However, the LSTM and interpretable LSTM use the same characteristics except for interpretability which increases the complexity of algorithms and training time. Therefore, the general LSTM outperformed the interpretable LSTM in terms of performance. However, the performance is limited to daily energy forecasting; the errors may vary in weekly and monthly predictions. In this scenario, the errors of TCN remained higher compared to the three deep learning algorithms. The proposed TFT model has improved the forecasting and outperformed the other algorithms. However, the improvement is marginal due to the complexity of the data and the unavailability of enough data for each customer to train the algorithms appropriately. The proposed model's interpretable structure is responsible for improved performance. Using the day series as a covariate enhanced the day-ahead energy consumption prediction results for one-day ahead demand of customers in the smart grid network. It helps deep learning algorithms properly adjust the weights and biases according to the changes in time series data.

5.3.2. Week-ahead energy consumption

If we observe the results of weekly energy consumption forecasting, the errors have relatively increased compared to the day ahead demand. When the prediction duration increases in the future, errors are also likely to occur. The close observation of the errors reveals that the TCN algorithm has struggled to predict the accurate future demand in the weekly scenario compared to both versions of LSTM. The reason can be that the data makes it difficult for the algorithm to predict accurately. The TFT has outperformed the other three deep learning algorithms, with decent improvement in the errors. Both versions of LSTM are closer to the TFT algorithm results because the TFT model uses blended LSTM inside the framework; hence, it provides results more comparable to the traditional LSTM. The reason for the improvement in the results of TFT is the GRN units which enhance the model's prediction accuracy and reduce errors. In other words, the static enrichment layer of the model is responsible for the improvements in the model.

5.3.3. Month-ahead energy consumption

The quantitative analysis of the month ahead demand has revealed interesting facts about the interpretable versions of the algorithms. Comparing the results of traditional LSTM models with those of all three algorithms, it is evident that conventional LSTMs provided the worst results. While the interpretable algorithms have shown better prediction accuracy and surprisingly, the TCN algorithm has performed better than the proposed TFT. Although both algorithms have a narrow margin of error, we can conclude that TCN has performed better. Accordingly, the TFT outperformed LSTM and Interpretable LSTM in terms of performance. It is evident from the results that the algorithms have certain limitations; hence the chances of errors have increased once we try to predict the future without having information about the future energy prediction. Still, the interpretable models consider input data in detail with complex calculations, showing improved performance compared to traditional algorithms. Based on these results, it can be concluded that four algorithms can better predict energy for the day ahead than for the week and month ahead. The model has also performed better than the experimental results presented by Schirmer et al. [40] as they used deep learning models for experimental evaluation. We have also compared the results of the TFT and LSTM models using the same data set.

5.3.4. Energy consumption with multiple customers

Table 5 provides statistical metrics to measure the model's performance based on the energy prediction with multiple customers. The experiment was repeated with customers of 10, 20, 30, and 40. Increasing the number of customers has increased error rates caused by the uncertainty of data and the different power consumption profiles of each customer. So the lowest errors can be seen for 10 customers' energy demand predictions. At the same time, the highest was for the 30 customers. Surprisingly the uncertainties have reduced; hence model has performed better with the 40 customers. A similar pattern can be seen for the performance of other models except for TCN, which has performed worst with 30 customers compared to 40. So it is evident from the performance of the models that uncertainties in energy consumption profiles make predictions challenging for the deep learning models [35,50]. Overall, the proposed model's performance remained better than the other deep learning models. The consumption profile clustering will improve performance, but it increases the system's complexity. In further experimentation, the profiles of the customers will be clustered based on similar characteristics and considered for the energy consumption prediction, which will reduce the complexity and challenges caused by the diversity of data. The analysis of the energy consumption profiles provides similar behavior of energy consumption by different customers while the ranges of energy consumption in kWh remain different, so the clustering will help identify the customers with similar energy consumption behavior. Based on the clustering, the customers will be divided into groups based on the lower, high, and very high energy consumption for further experimentation, ultimately improving the system's performance and providing insights into each group of customers while reducing the computation time and complexity of the model design.

Table 5
Performance comparison of TFT with 40 customers.

Model	Customers	Metrics			
		MSE	RMSE	MAE	sMAPE
LSTM	10	4.96	2.23	1.47	0.22
	20	6.96	2.64	1.67	0.23
	30	9.16	3.03	1.97	0.25
	40	8.41	2.90	1.89	0.25
LSTM Interpretable	10	4.84	2.20	1.39	0.20
	20	6.97	2.64	1.61	0.22
	30	9.06	3.00	1.90	0.24
	40	8.36	2.89	1.84	0.24
Temporal Convolutional Networks (TCN)	10	5.81	2.41	1.54	0.22
	20	7.84	2.80	1.75	0.24
	30	10.37	3.22	2.07	0.26
	40	9.68	3.11	1.99	0.26
Proposed TFT	10	4.80	2.19	1.34	0.19
	20	7.29	2.70	1.59	0.21
	30	9.05	3.00	1.83	0.22
	40	8.16	2.86	1.74	0.23

6. Conclusion

The energy consumption prediction problem has always remained a concern for scientists and energy production companies. In recent times, deep learning models have been introduced for the big data of energy consumption. These models need to be improved in terms of accuracy because the dimensionality of the data is a challenge for the algorithms. The intense competition for energy consumption between provinces, cities, urban, rural areas and industries has increased the challenges for energy production companies. There needs to be a reform of the energy consumption structure, a fast-track industrial restructuring, and a renewed commitment to energy conservation. It cannot be overstated how critical it is to invest in clean energy, improve the efficiency of conventional energy sources, and promote energy conservation and emission reductions. The depreciation should be prevented by implementing regional strategies for managing total energy consumption. Establishing a relationship between regional energy consumption and economic performance is necessary to achieve regional energy development goals. The objective of the proposed study is to predict the future energy consumption of different customers. The model has improved the prediction accuracy on a big data set with diverse customer energy consumption profiles. The advantage of the model over traditional methods is the consideration of customer-based energy rather than an overall forecast. Keeping in mind the sensitivity of the problem, the proposed TFT model deals with the future demand of customers' energy consumption in a smart grid environment.

TFT model's performance has remained superior to LSTM, interpretable LSTM, and TCN models. For the one-day ahead prediction of energy consumption, the MSE remained at 4.80; for the seven days, the MSE increased slightly to 5.75; and for the 30 days, the MSE remained at 6.49. There was an RMSE of 2.19 for the next day, 2.40 for the next week, and 2.55 for the next thirty days. One day, one week, and thirty days ahead, the MAE was 1.64, 1.79, and 1.94, respectively. Comparing the proposed TFT model to the LSTM, the interpretable LSTM model, and the TCN model, these errors of the proposed TFT remained the lowest. LSTM has an overall sMAPE of 29.78%, while the Blocked LSTM has an overall sMAPE of 31.10%. There was a slight increase in the sMAPE for the TCN model with a value of 36.42%, whereas the proposed TFT model had an sMAPE value of 29.56%, outperforming all the algorithms.

Sustainability requires prioritizing natural resource energy production. The proposed model will increase the efficiency of existing resources on the supply side. This feature makes the model suitable for use in smart grids and homes. The TFT has a deeper architecture and better learning strategies than other deep learning algorithms. TFT extracted better building energy consumption features from the smaller data set, which improved the prediction accuracy. In contrast, the model does not consider the market prices necessary to ensure a fair pricing system and sufficient energy demand. The total output will continue to rise, but it will be essential to establish a pricing system that is both effective and affordable. The customer-based predictive model development has its advantage as it does not depend on overall energy consumption.

Despite the diversity of data, the model has met its objectives. Model performance may vary as more customers are added, since it has been tested on data from only one customer. For this study, only 169 customers have been considered because of memory constraints, computing equipment, and resource availability. In the future, it would be better to use a more powerful platform or computer to facilitate the running of deep learning algorithms, as the algorithms take a long time to run. Furthermore, the study is limited by a limited number of participants, so more customer data should be input. Future studies can propose hybrid or ensemble methods since they are more accurate than single algorithms.

CRediT authorship contribution statement

The authors hereby confirm that we all have made a substantial contribution to complete this article. **Amril Nazir:** Idea generation, methodology, writing and, reviewing. **Abdul Shaikh:** Writing, reviewing and editing. **Abdul Shah:** Experiment and results. **Ashraf Khalil:** Guidance and supervision. We all read and approved the final manuscript.

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.

Data availability

Data will be made available on request.

Appendix A. Learning loss of deep learning models

Table A.6
Learning loss comparison of TFT with deep learning models.

Parameter	LSTM	Blocked LSTM	TCN	Proposed TFT
Train_Loss	0.00101	0.000823	0.00186	0.00171
Val_Loss	0.000948	0.0013	0.00149	0.00169

Appendix B. Parameters settings of deep learning models

Table B.7
Parameters settings of algorithms.

Parameter	LSTM	Blocked LSTM	TCN	Proposed TFT
Activation Function	ReLU	ReLU	ReLU	
Hidden Size/Widths	20	10		16
Number of Layers	1	1	2	2
Random State	30	42	0	42
Training/Input Length	30	30	30	30
Output Chunk Length	15	15	15	15
Hidden Continuous Size				8
N Epoch Val Period	1	1		
Batch Size	1500	32		1000
Optimizer	Adam	Adam		
Learning Rate	1e-3	1e-3		1e-3
Epochs	200	200	200	150
Droupout Rate	0-0.2	0.1	0.1	0.1
Kernel Size			5	
Number of Filters			3	
Dilation Base			2	
Loss-fn				**torch.nn. MSELoss()**
Likelihood				None
Future/Past Covariates	Day Series	–	–	–

References

[1] S. Ali, D. Kim, Enhanced power control model based on hybrid prediction and preprocessing/post-processing, J. Intell. Fuzzy Syst. 30 (2016) 3399–3410.

[2] S. Ali, K. Ullah, G. Hafeez, I. Khan, F.R. Albogamy, S.I. Haider, Solving day-ahead scheduling problem with multi-objective energy optimization for demand side management in smart grid, Int. J. Eng. Sci. Technol. 36 (2022) 101135, https://doi.org/10.1016/j.jestch.2022.101135.

[3] J.L. Ba, J.R. Kiros, G.E. Hinton, Layer normalization, arXiv preprint arXiv:1607.06450, 2016.

[4] I. Barić, R. Grbić, E.K. Nyarko, Short-term forecasting of electricity consumption using artificial neural networks-an overview, in: 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), IEEE, 2019, pp. 1076–1081.

[5] R.F. Berriel, A.T. Lopes, A. Rodrigues, F.M. Varejao, T. Oliveira-Santos, Monthly energy consumption forecast: a deep learning approach, in: 2017 International Joint Conference on Neural Networks (IJCNN), IEEE, 2017, pp. 4283–4290.

[6] S. Bouktif, A. Fiaz, A. Ouni, M.A. Serhani, Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: comparison with machine learning approaches, Energies 11 (2018) 1636.

[7] C. Chen, Y. Liu, M. Kumar, J. Qin, Y. Ren, Energy consumption modelling using deep learning embedded semi-supervised learning, Comput. Ind. Eng. 135 (2019) 757–765, https://doi.org/10.1016/j.cie.2019.06.052.

[8] C.f. Chen, G.Z. de Rubens, X. Xu, J. Li, Coronavirus comes home? Energy use, home energy management, and the social-psychological factors of Covid-19, Energy Res. Soc. Sci. 68 (2020) 101688.

[9] D.A. Clevert, T. Unterthiner, S. Hochreiter, Fast and accurate deep network learning by exponential linear units (elus), arXiv preprint, arXiv:1511.07289, 2015.

[10] London DataStore, Smartmeter energy consumption data in London households, 2018.

[11] Y.N. Dauphin, A. Fan, M. Auli, D. Grangier, Language modeling with gated convolutional networks, in: Proceedings of the 34th International Conference on Machine Learning, vol. 70, 2017, pp. 933–941.

[12] R.E. Edwards, J. New, L.E. Parker, Predicting future hourly residential electrical consumption: a machine learning case study, Energy Build. 49 (2012) 591–603, https://doi.org/10.1016/j.enbuild.2012.03.010.

[13] R.M. Elavarasan, S. Afridhis, R.R. Vijayaraghavan, U. Subramaniam, M. Nurunnabi, Swot analysis: a framework for comprehensive evaluation of drivers and barriers for renewable energy development in significant countries, Energy Rep. 6 (2020) 1838–1864.

[14] A.F. Faisal, A. Rahman, M.T.M. Habib, A.H. Siddique, M. Hasan, M.M. Khan, Neural networks based multivariate time series forecasting of solar radiation using meteorological data of different cities of Bangladesh, Results Eng. 13 (2022) 100365, https://doi.org/10.1016/j.rineng.2022.100365.
[15] C. Fan, F. Xiao, Y. Zhao, A short-term building cooling load prediction method using deep learning algorithms, Appl. Energy 195 (2017) 222–233.
[16] M. Fayaz, I. Ullah, A.S. Shah, D. Kim, An efficient energy consumption and user comfort maximization methodology based on learning to optimization and learning to control algorithms, J. Intell. Fuzzy Syst. 37 (2019) 6683–6706.
[17] A. Fraija, K. Agbossou, N. Henao, S. Kelouwani, M. Fournier, S.S. Hosseini, A discount-based time-of-use electricity pricing strategy for demand response with minimum information using reinforcement learning, IEEE Access 10 (2022) 54018–54028, https://doi.org/10.1109/ACCESS.2022.3175839.
[18] Y. Gal, Z. Ghahramani, A theoretically grounded application of dropout in recurrent neural networks, in: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc., 2016.
[19] S. Gyamfi, S. Krumdieck, T. Urmee, Residential peak electricity demand response—highlights of some behavioural issues, Renew. Sustain. Energy Rev. 25 (2013) 71–77, https://doi.org/10.1016/j.rser.2013.04.006.
[20] Á. Herraiz-Cañete, D. Ribó-Pérez, P. Bastida-Molina, T. Gómez-Navarro, Forecasting energy demand in isolated rural communities: a comparison between deterministic and stochastic approaches, Energy Sustain. Dev. 66 (2022) 101–116.
[21] J. Herzen, F. Lässig, S.G. Piazzetta, T. Neuer, L. Tafti, G. Raille, T. Van Pottelbergh, M. Pasieka, A. Skrodzki, N. Huguenin, et al., Darts: user-friendly modern machine learning for time series, J. Mach. Learn. Res. 23 (2022) 1–6.
[22] S. Jia, K. Peng, X. Zhang, Y. Li, L. Xing, Dynamic pricing strategy and regional energy consumption optimization based on different stakeholders, Int. J. Electr. Power Energy Syst. 141 (2022) 108199, https://doi.org/10.1016/j.ijepes.2022.108199.
[23] F.D. Kanellos, K. Kalaitzakis, I. Psarras, Y. Katsigiannis, Efficient and robust power and energy management for large clusters of plug-in electric vehicles and distribution networks, IET Energy Syst. Integr. (2022).
[24] A. Keshtkar, S. Arzanpour, An adaptive fuzzy logic system for residential energy management in smart grid environments, Appl. Energy 186 (2017) 68–81.
[25] I. Khan, Z. Luo, A.K. Shaikh, R. Hedjam, Ensemble clustering using extended fuzzy k-means for cancer data analysis, Expert Syst. Appl. 172 (2021) 114622, https://doi.org/10.1016/j.eswa.2021.114622.
[26] M. Kuosa, P. Kiviranta, H. Sarvelainen, E. Tuliniemi, T. Korpela, K. Tallinen, H.K. Koponen, Optimisation of district heating production by utilising the storage capacity of a district heating network on the basis of weather forecasts, Results Eng. 13 (2022) 100318, https://doi.org/10.1016/j.rineng.2021.100318.
[27] T. Le, M.T. Vo, B. Vo, E. Hwang, S. Rho, S.W. Baik, Improving electric energy consumption prediction using cnn and bi-lstm, Appl. Sci. 9 (2019) 4237.
[28] W. Lee, M. Chae, D. Won, Optimal scheduling of energy storage system considering life-cycle degradation cost using reinforcement learning, Energies 15 (2022) 2795.
[29] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.X. Wang, X. Yan, Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting, in: Advances in Neural Information Processing Systems, vol. 32, 2019.
[30] Y. Li, S. He, Y. Li, L. Ge, S. Lou, Z. Zeng, Probabilistic charging power forecast of evcs: reinforcement learning assisted deep learning approach, IEEE Trans. Intell. Veh. (2022) 1, https://doi.org/10.1109/TIV.2022.3168577.
[31] B. Lim, S.Ö. Arık, N. Loeff, T. Pfister, Temporal fusion transformers for interpretable multi-horizon time series forecasting, Int. J. Forecast. 37 (2021) 1748–1764.
[32] Y.H. Lin, An advanced smart home energy management system considering identification of adls based on non-intrusive load monitoring, Electr. Eng. (2022) 1–19.
[33] S.X. Lv, L. Wang, Deep learning combined wind speed forecasting with hybrid time series decomposition and multi-objective parameter optimization, Appl. Energy 311 (2022) 118674.
[34] A. Meng, P. Wang, G. Zhai, C. Zeng, S. Chen, X. Yang, H. Yin, Electricity price forecasting with high penetration of renewable energy using attention-based lstm network trained by crisscross optimization, Energy 254 (2022) 124212.
[35] 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.
[36] A. Moradzadeh, H. Moayyed, K. Zare, B. Mohammadi-Ivatloo, Short-term electricity demand forecasting via variational autoencoders and batch training-based bidirectional long short-term memory, Sustain. Energy Technol. Assess. 52 (2022) 102209, https://doi.org/10.1016/j.seta.2022.102209.
[37] J. Natividade, C.O. Cruz, C.M. Silva, Improving the efficiency of energy consumption in buildings: simulation of alternative enpc models, Sustainability 14 (2022) 4228.
[38] R. Phetrittikun, K. Suvirat, T.N. Pattalung, C. Kongkamol, T. Ingviya, S. Chaichulee, Temporal fusion transformer for forecasting vital sign trajectories in intensive care patients, in: 2021 13th Biomedical Engineering International Conference (BMEiCON), 2021, pp. 1–5.
[39] N.S.M. Salleh, A. Suliman, B.N. Jørgensen, Comparison of electricity usage forecasting model evaluation based on historical load dataset duration using long short-term memory architecture, in: R. Alfred, Y. Lim (Eds.), Proceedings of the 8th International Conference on Computational Science and Technology, Springer Singapore, Singapore, 2022, pp. 675–686.
[40] P.A. Schirmer, I. Mporas, I. Potamitis, Evaluation of regression algorithms in residential energy consumption prediction, in: 2019 3rd European Conference on Electrical Engineering and Computer Science (EECS), IEEE, 2019, pp. 22–25.
[41] A.S. Shah, H. Nasir, M. Fayaz, A. Lajis, A. Shah, A review on energy consumption optimization techniques in iot based smart building environments, Information 10 (2019) 108.
[42] A.S. Shah, H. Nasir, M. Fayaz, A. Lajis, I. Ullah, A. Shah, Dynamic user preference parameters selection and energy consumption optimization for smart homes using deep extreme learning machine and bat algorithm, IEEE Access 8 (2020) 204744–204762, https://doi.org/10.1109/ACCESS.2020.3037081.
[43] A.K. Shaikh, M. Al-Shamli, A. Nazir, Designing a relational model to identify relationships between suspicious customers in anti-money laundering (aml) using social network analysis (sna), J. Big Data 8 (2021) 1–22.
[44] H. Shi, M. Xu, R. Li, Deep learning for household load forecasting—a novel pooling deep rnn, IEEE Trans. Smart Grid 9 (2017) 5271–5280.
[45] H. Shrisha, U. Boregowda, An energy efficient and scalable endpoint linked green content caching for named data network based internet of things, Results Eng. 13 (2022) 100345, https://doi.org/10.1016/j.rineng.2022.100345.
[46] J.F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, A. Troncoso, Deep learning for time series forecasting: a survey, Big Data 9 (2021) 3–21.
[47] K. Utkarsh, D. Srinivasan, Distributed machine learning in energy management and control in smart grid, in: Women in Computational Intelligence, Springer, 2022, pp. 219–251.
[48] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, vol. 30, 2017.
[49] F. Wahid, D.H. Kim, Short-term energy consumption prediction in Korean residential buildings using optimized multi-layer perceptron, Kuwait J. Sci. 44 (2017).
[50] H. Wang, Z. Lei, X. Zhang, B. Zhou, J. Peng, A review of deep learning for renewable energy forecasting, Energy Convers. Manag. 198 (2019) 111799.
[51] B. Wu, L. Wang, S. Wang, Y.R. Zeng, Forecasting the US oil markets based on social media information during the COVID-19 pandemic, Energy 226 (2021) 120403.
[52] B. Wu, L. Wang, Y.R. Zeng, Interpretable wind speed prediction with multivariate time series and temporal fusion transformers, Energy 252 (2022) 123990.
[53] C. Xu, W. Li, Y. Rao, B. Qi, B. Yang, Z. Wang, Coordinative energy efficiency improvement of buildings based on deep reinforcement learning, Cyber-Phys. Syst. (2022) 1–13, https://doi.org/10.1080/23335777.2022.2066181.
[54] H. Yang, K.R. Schell, Qcae: a quadruple branch cnn autoencoder for real-time electricity price forecasting, Int. J. Electr. Power Energy Syst. 141 (2022) 108092, https://doi.org/10.1016/j.ijepes.2022.108092.
[55] W. Zheng, P.P. Walsh, Economic growth, urbanization and energy consumption—a provincial level analysis of China, Energy Econ. 80 (2019) 153–162.