Technological Forecasting & Social Change 206 (2024) 123588

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

ForecastExplainer: Explainable household energy demand forecasting by approximating shapley values using DeepLIFT

Md Shajalal [a,b,*], Alexander Boden [a,c], Gunnar Stevens [b,c]

[a] *Fraunhofer Institute for Applied Information Technology FIT, Sankt Augustin, Germany*
[b] *University of Siegen, Germany*
[c] *Bonn-Rhein-Sieg University of Applied Sciences, Bonn, Germany*

ARTICLE INFO

Keywords:
Explainable energy demand forecasting
DeepLIFT
Shapley additive explanation
Deep learning
Human-centered explanation

ABSTRACT

The rapid progress in sensor technology has empowered smart home systems to efficiently monitor and control household appliances. AI-enabled smart home systems can forecast household future energy demand so that the occupants can revise their energy consumption plan and be aware of optimal energy consumption practices. However, deep learning (DL)-based demand forecasting models are complex and decisions from such black-box models are often considered opaque. Recently, eXplainable Artificial Intelligence (XAI) has garnered substantial attention in explaining decisions of complex DL models. The primary objective is to enhance the acceptance, trust, and transparency of AI models by offering explanations about provided decisions. We propose *ForecastExplainer*, an explainable deep energy demand forecasting framework that leverages Deep Learning Important Features (DeepLIFT) to approximate Shapley values to map the contribution of different appliances and features with time. The generated explanations can shed light to explain the prediction highlighting the impact of energy consumption attributes corresponding to time, such as responsible appliances, consumption by household areas and activities, and seasonal effects. Experiments on household datasets demonstrated the effectiveness of our method in accurate forecasting. We designed a new metric to evaluate the effectiveness of the generated explanations and the experiment results indicate the comprehensibility of the explanations. These insights might empower users to optimize energy consumption practices, fostering AI adoption in smart applications.

1. Introduction

Smart home systems offer users the ability to remotely control and access household electrical appliances and monitor the environment through various sensors (Jensen et al., 2018; Shajalal et al., 2022b). Additionally, these systems can autonomously make decisions, such as adjusting the state of connected actuators (e.g., controlling the heating system to adapt room temperature) to enhance the dwellers' comfort (Jensen et al., 2018; Das et al., 2021; Ma et al., 2021). While many smart applications follow simple timetable logic and classic automation paradigms, an increasing number of decisions are made using a combination of machine learning models (Ali et al., 2021; Li et al., 2022b).

Energy demand forecasting using machine learning (ML) models has recently garnered significant attention in the literature. The objective is to make smart home users more aware of their future energy consumption (Kim and Cho, 2020; Zhang et al., 2021; Ma et al., 2021; Kim and Cho, 2019a, 2021). These systems can even forecast the energy consumption for individual appliances (Haq et al., 2020), which enhances household members' awareness and encourages optimal electricity consumption practices. Since increased energy consumption can lead to higher household costs, people are expected to become more cautious and may modify their consumption behavior to decrease energy usage. However, implementing such technology in the real world poses challenges due to its lack of transparency and trust. Users may not fully understand the reasons behind certain predictions and require more trustworthy explanations regarding the facts behind predicted decisions/recommendations. Alongside accurate energy demand predictions, one sensible approach to building trust and increasing transparency and fairness is to explain the predictions by highlighting important factors and time duration.

However, the underlying forecasting models are often black boxes for the end-users (even for AI practitioners), who do not have a clear understanding of the decision-making procedures of these prediction systems. As a consequence, users might want factual explanations for why a particular decision has been taken on their behalf by the

* Corresponding author at: Fraunhofer Institute for Applied Information Technology FIT, Sankt Augustin, Germany.
E-mail addresses: md.shajalal@fit.fraunhofer.de (M. Shajalal), alexander.boden@fit.fraunhofer.de (A. Boden), gunnar.stevens@uni-siegen.de (G. Stevens).

https://doi.org/10.1016/j.techfore.2024.123588
Received 10 April 2023; Received in revised form 27 February 2024; Accepted 23 March 2024
Available online 15 July 2024

system (Ehsan et al., 2021). They might have queries such as "*Why do we need this amount (more/less) of energy in the upcoming week/month*?". The plausible reason behind such an impression is that the system might surprise the user with an unexpected prediction (i.e., forecasting more/less amount of energy for next week/month compared to their expectation). However, providing explanations by highlighting the significant factors corresponding to the time duration might make them understand the reasons behind such predictions. Moreover, to plan the optimal energy consumption in the household, it would be more effective if they know which appliances might be responsible and consume more for the future overall predicted consumption. The relevant question can be "*How can I further optimize my energy consumption?*" (Shajalal et al., 2022b).

Providing comprehensive explanations from the system to address these questions would likely enhance the trust and transparency of AI models for end-users (Shajalal et al., 2022a; Gilpin et al., 2018). Additionally, in accordance with the General Data Protection Regulation (GDPR), citizens of the EU have a civil right to be informed about how AI-based models that pertain to them make decisions (Došilović et al., 2018). The incorporation of explainability to ensure transparency, offering comprehensive explanations employing clear and interpretable facets unveiling DL-based forecasting methods is expected. Explainable forecasting systems have the potential to augment seamlessly the overarching goals centered on technological advancement in AI and comprehending their corresponding societal ramifications.

The research field that focuses on explaining (and/or interpreting) the decision-making process is commonly referred to as eXplainable Artificial Intelligence (XAI). In recent years, there has been a significant interest in interpreting and explaining complex machine learning models (Ribeiro et al., 2016b,a; Lundberg and Lee, 2017; Crabbé and Van Der Schaar, 2021; Crabbé and van der Schaar, 2021; Haque et al., 2023), enabling AI practitioners and developers to enhance models' performance. The application of XAI has also received considerable attention in various fields, including bio-informatics (Karim et al., 2023), healthcare (Adadi and Berrada, 2020), finance (Yang et al., 2023; Ghosh et al., 2023; Efat et al., 2022), inventory management (Shajalal et al., 2022a), natural language processing (Karim et al., 2021; Shajalal et al., 2023) and so on Kim et al. (2022), Wang et al. (2022) and Kim et al. (2023b). However, there remains a need for further research on how to generate *human-centered* explanations that are accessible to end-users with no expert knowledge in AI theory and development (Kim et al., 2023a).

As the human-centered design is highly context-specific, such research would arguably need to take into account the specific user needs of different domains, i.e., studying how explanations can be made meaningful to users in a specific pragmatic context and situated action. In our study, we focus on the domain of smart home technology, where such challenges are prominent but have been hardly studied to our knowledge. In addition, explaining the multivariate time series forecasting model is difficult (Crabbé and Van Der Schaar, 2021), because the explanations might be two-dimensional, including both time and features. Hence, generating human-centered explanations for energy demand forecasting is a challenging task, especially when those are to be understood by end-users.

Let us discuss how different and complex the household energy demand forecasting problems are. Here, we discuss the problem in two different directions. Figs. 1 and 2 indicate the daily total energy consumed by the whole household area and kitchen, respectively. We can see that the energy consumption patterns are quite different. These complex and dissimilar consumption patterns also exist in some other places of the household area including the living room and laundry room. The irregularity in energy consumption for different appliances makes the forecasting problem challenging for any ML models. Moreover, the seasonal consumption patterns for the different household areas are widely varied over time. Therefore, forecasting household energy demand for different areas of the house is challenging.

In addition, time series forecasting models are dependent not only on the values of the features like classical classification or regression tasks but also dependent on time. Since the multivariate forecasting models are dependent on both the features and time, explaining specific predictions by highlighting important factors and corresponding time frames is very challenging. Moreover, unlike classification tasks, the progress of developing XAI tools for understandable explanation is much lower for multivariate forecasting tasks. Therefore, explaining the decision for the time series forecasting model is a more formidable task than other classical classification or regression models. The preliminary idea of having user-centric explanations for household energy demand forecasting has been presented as a poster at the Thirteenth ACM International Conference on Future Energy Systems (Shajalal et al., 2022b).

In this paper, we present the underlying challenges to generate and present explanations for a particular prediction of an energy demand forecasting system. To explain the prediction of the energy demand forecasting system, we propose an explainable framework, *Forecas-Explainer* by approximating Shapley values leveraging DeepLIFT to explain predictions made by the deep learning-based energy demand forecasting model. First, we developed an energy demand forecasting model applying long short-term memory (LSTM) networks, one of the most successful recurrent neural networks (RNN)-based methods in multivariate time series forecasting and then we introduced DeepLIFT-enabled explanation generation technique. We chose LSTM with the objective that most of the audience can understand our explainable framework. However, our explainability method can be applicable to any deep learning-based forecasting model (i.e., CNN, GRU). Due to the complex architecture and working principle, deep time series forecasting models are very opaque (i.e., black-box) and even AI developers struggle to understand the decision.

We approximate shapley values employing DeepLIFT to track the features' contribution in different layers. We apply DeepLIFT which decomposes the complex LSTM energy forecasting model for a specific prediction by back-propagating to compute the contributions of neurons and approximate the Shapley values to generate explanations. Given that multivariate time series forecasting involves both time and features, we address the challenges of mapping feature contributions with corresponding time frames. Finally, we provide explanations by mapping specific time series and feature importance (i.e., highlighting the contribution of different appliances towards the prediction) with different easily understandable visualizations. Compared to the conventional application of XAI tools (i.e., LIME by Ribeiro et al., 2016b) in classification tasks, our method can generate explanations that highlight feature contributions along with their corresponding time frames.

We design a metric to measure the efficiency of the explanations by comparing the highlighted contributions for different appliances towards prediction with the original contributions to the overall consumption. We hypothesize that if the contributions of different features in the explanations for the prediction correlate with the contributions towards the original consumption and have an increasing monotonous relationship, the explanations can be considered effective. The contributions of different appliances towards the overall household consumption can be calculated statistically and represented in vector. Then the contribution vectors for original consumption and prediction are employed to measure effectiveness. If there is a high monotonous correlation between the vectors of contributions for original energy consumption data and Shapley values, we can conclude that the generated contributions using DeepLIFT are analogous. The degree of goodness of the generated explanations can be represented by the correlation coefficient, where the higher the correlation coefficient the better the generated explanations are. However, the major contributions of this research can be summarized as follows:

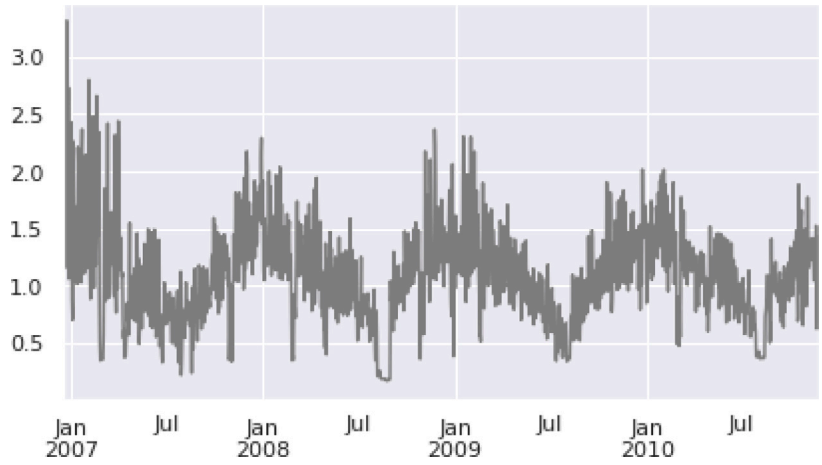

Fig. 1. Daily distribution of global active power.

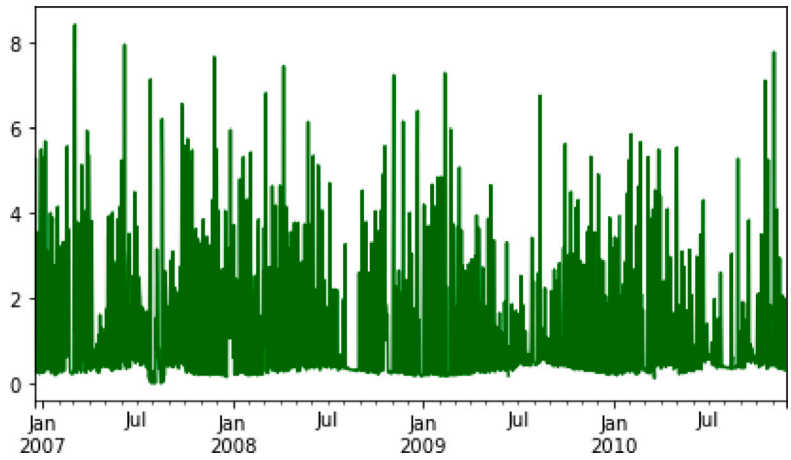

Fig. 2. Daily energy consumption in the kitchen.

- We employed DeepLIFT to enhance the quality of the explanations for the decision provided by the deep multivariate time series forecasting model by mapping the time and the contribution of different features. Note that our method is applicable to explaining the decisions of any other deep learning-based forecasting models, such as CNN, GRU, etc.
- We designed and introduced an evaluation metric to measure the effectiveness of the generated explanation considering the monotonous relationship between the original and predicted impacts on the overall energy consumption. Our framework achieved high efficiency in terms of the designed metric and can capture appliance contributions towards overall household energy consumption.
- We elicit multiple research gaps in providing human-centered explainability by analyzing existing literature on energy demand forecasting and their explainability. These elicited research gaps would provide future directions for HCI and AI practitioners towards making forecasting systems' decisions understandable for users.
- Moreover, the results of multiple experiments on the benchmark datasets with five different households demonstrate that our proposed explainable energy demand forecasting framework achieved effective prediction performance in terms of multiple evaluation metrics.

The rest of the paper is organized as follows: In Section 3, We summarize state-of-the-art methods in energy demand forecasting and advancements in XAI for time series forecasting. We also highlight the research gaps towards achieving explainable energy demand forecasting. We then introduce our proposed explainable deep household energy demand forecasting framework, called *ForecastExplainer*, which approximates SHAP values using DeepLIFT (Section 3). The experimental results with multiple settings on two different datasets are analyzed and discussed in Section 4. We also present the generated explanations and their effectiveness in this section. Conclusions and key findings are presented in Section 5. Finally, Section 6 outlines future research directions, focusing on human-centered evaluation of the explanations and eliciting further requirements through an empirical study from a user-centered perspective in the context of smart home systems.

2. Literature review

This section presents an extensive discussion of the prior works on energy demand forecasting and the progress of explainable artificial intelligence, especially for time series forecasting models. Therefore, we present prior research works reviewing published literature in two different sub-sections. At the end of this section, we highlight the research gaps towards making the energy demand forecasting system explainable.

2.1. Energy demand forecasting

The methods to forecast households' energy demand can range from classical to complex ML and deep neural networks (DNN)-based techniques (Ma et al., 2021; Vanting et al., 2021; Zhang et al., 2021).

In the recent past, there is a huge interest in applying ML and deep learning techniques to forecast household energy demand (Zhang et al., 2021; Li et al., 2022a; Kim and Cho, 2021, 2020, 2019a; Vanting et al., 2021). Kazemzadeh et al. (2020) proposed a hybrid long-term demand forecasting model based on data mining techniques. They applied particle swarm optimization in the hybrid model consisting of support vector regressor, Auto-Regressive Integrated Moving Average (ARIMA), and Artificial Neural Network (ANN). Similar to Kazemzadeh et al. (2020), Yan et al. (2019) proposed an LSTM-based hybrid model for modeling individual household energy consumption. They leveraged the stationary wavelet transform (SWT) technique to increase the dimension of the data and tackle the volatility and then applied the LSTM-based deep learning model. Fu et al. (2018) proposed a data-driven situational awareness framework that monitors energy consumption on the campus. Their framework consists of two different components including energy demand forecasting models and anomaly detection systems to support immediately on the campus. Similar to our research, their energy demand forecasting system is modeled by LSTM-based neural network architecture. However, our goal in this research is to explain the prediction by energy demand forecasting models, not having a new forecasting model. Since LSTM is most widely used and successful in energy demand forecasting tasks, we have applied our explainable framework aiming to generate meaningful explanations for particular prediction highlighting different features corresponding to the time duration.

Kim and Cho (2019b) proposed a CNN-LSTM neural network model combining convolutional neural networks (CNN) and LSTM networks to extract better temporal and spatial features that can predict household energy effectively. A hybrid deep learning framework is proposed by Syed et al. (2021) employing a fully connected neural network followed by a unidirectional LSTM and bi-directional LSTM (Bi-LSTM) model to overcome the temporal dependencies of the energy consumption. Chadoulos et al. (2021) introduced a deep learning model combining recurrent neural networks (RNN) and multi-layer perceptron (MLP) to forecast hourly demand for different households considering consumers' profiles. In addition, the method can capture the past and future impacts of time series and consumer profiles.

Some prior studies (Kim and Cho, 2021, 2020, 2019a; Vanting et al., 2021) modeled the energy demand forecasting task exploiting DNN, CNN, LSTM and auto-encoder, and explained the prediction. Kim and Cho (2021, 2020, 2019a) conducted multiple studies and proposed multiple methods for forecasting household electric demand. In the study (Kim and Cho, 2019a), an auto-encoder-based deep learning model is proposed which can predict the energy demand for each 15, 30, 45, and 60 min for various household scenarios. Similar to the previous study, methods presented in Kim and Cho (2020, 2021) applied also an auto-encoder-based model consisting of four different components. The first component models the past energy consumption, and then a subencoder models consumption information and processes as latent variables. The third component maps the future demand considering the latent variables. Lastly, the final component tried to interpret the important electric information to highlight the model's global interpretability.

Chakraborty et al. (2021) introduced explainable artificial intelligence to predict climate change impact on a scenario-based building cooling energy forecasting. For optimal management of the building's energy, Eseye and Lehtonen (2020) introduced short-term forecasting of heat energy demand with integrated ML models. Their model incorporated a support vector machine (SVM) with an imperialistic competitive algorithm embedding feature selection technique combining binary genetic algorithm and Gaussian process regression. Zhang et al. (2021) proposed an explainable energy forecasting model exploiting AI-based techniques. They trained a surrogate model to mimic the original trained model and interpret the model. Ahmad et al. (2020) proposed a random neural network-based energy prediction model for large buildings. A wide range of experiments was conducted on one-year energy data, and they have achieved better performance than artificial neural networks and support vector machine-based regression techniques.

2.2. Explainable AI in time series forecasting

Though there is some attention to making the model interpretable in time series forecasting, most of those methods attempted to explain only the algorithmic decision-making to increase the model's performance and debugging (Mucha et al., 2020; Kim and Cho, 2020). However, the methods for generating explanations for general users are not so common (Ehsan et al., 2021; Kabir et al., 2021). Assaf and Schumann (2019) proposed a gradient-based technique to explain the prediction from a CNN-based time series model. The explanations are provided via a saliency map considering the time dimension and the features. Their method can identify the specific time duration and highlight the most important factors on the time for the particular prediction.

Similarly, Saadallah et al. (2021, 2022) proposed a CNN-based explainable time series forecasting model using adaptive saliency maps explanations. Prior studies (Rojat et al., 2021; Ekambaram et al., 2020; Barredo Arrieta et al., 2022; Schlegel et al., 2020; Saluja et al., 2021; Rozanec, 2021) surveyed explainable methods on time series data by highlighting the overview, impacts and available methods in the field of explainable models for time series data. Ilic et al. (2021) introduced an explainable boosted regression technique for time series forecasting. Their method provides explanations through regression trees. A heatmap-based explainable technique by Kim and Cho (2020) is presented to explain the auto-encoder-based forecasting model. Zdravković et al. (2022) applied local interpretable model-agnostic explanations (LIME) (Ribeiro et al., 2016b) to explain the heat energy demand forecasting model. LIME and Shapley additive explanation (SHAP) (Lundberg and Lee, 2017) based explainable models are also employed for explaining time series classification not forecasting tasks.

The explanations and their representation interface will surely be different in the case of human-centered explanations for general users (Mucha et al., 2021; Riboni, 2021; Rai, 2020; Kabir et al., 2021). Some explainable methods are also published recently where they focused on human-activity recognition and e-health in smart home environments (Khodabandehloo et al., 2021; Bettini et al., 2021; Das et al., 2021; Arrotta, 2021; Dalvi-Esfahani et al., 2023).

In conclusion, there is still a big gap in human-centered XAI systems for general smart home users, particularly in the energy demand forecasting problem. In this research, we try to explain the complex energy demand forecasting prediction with approximating shapley values incorporating DeepLIFT. Our visualizations towards explaining specific decisions might help general users so that they can build more awareness of consuming energy in their homes. Moreover, these explanations might help towards optimizing their energy consumption considering the factors behind the predictions.

The summary of notable state-of-the-art methods in energy demand forecasting and explaining time series forecasting is depicted in Table 1. We observed that most of the studies in energy demand forecasting are not explainable. The studies focusing on explainability in energy demand forecasting only tried to highlight different features for global interpretability. We also include related works (Assaf and Schumann, 2019; Saadallah et al., 2021, 2022; Ilic et al., 2021) that aimed at explaining time series forecasting models. We observed that few methods tried to explain the forecasting model based on time and features. However, they only focus on global interpretability so that AI practitioners can improve the model's performance. Nevertheless, the explanations are highly technical and not easily understandable by the general users in smart homes. To the best of our knowledge, there is no such model that can provide local explanations for energy demand forecasting model highlighting time and features. In this paper, we try to fill the gap in generating understandable explanations for certain predictions highlighting the time and features in an easily understandable way. The primary goal is to provide such explanations to the user so that they can be more optimal and aware when they utilize a particular appliance.

Table 1
The summary of the state-of-the-art research on household energy demand forecasting with possible research gaps.

Authors and paper (ref.)	Summary of method & contribution	Gaps related to explainability
Kazemzadeh et al. (2020) & Yan et al. (2019)	Both papers proposed hybrid models to predict household energy consumption. Kazemzadeh et al. (2020) applied ARIMA and ANN, whereas Yan et al. (2019) applied a dimensionality reduction approach and LSTM-based DL forecasting technique.	Did not consider explainability
Fu et al. (2018)	Applied a data-driven situational awareness for monitoring energy consumption in a university campus	Did not consider explainability
Kim and Cho (2019b)	Proposed a predictive method combining CNN and LSTM (CNN-LSTM) for extracting better spatial and temporal feature for residential energy demand prediction.	Did not consider explainability
Kim and Cho (2021, 2020, 2019a)	Modeled the energy demand forecasting problem using different neural network-based approaches including CNN, LSTM and auto-encoder.	The proposed models have global interpretability. But the method cannot explain for a particular prediction.
Chakraborty et al. (2021)	Introduced an explainable AI-driven approach to predict climate change impact for building's cooling energy forecasting.	Incorporated Shapley additive explanations for highlighting the feature impacts. But the method cannot explain for a particular prediction.
Eseye and Lehtonen (2020)	Proposed a method for forecasting energy demand for household with several ML models.	Did not consider explainability
Zhang et al. (2021)	Introduced interpretable energy forecasting model by developing a surrogate model that might mimic the original model's performance.	Only provide interpretability about the local mechanism of the model
Assaf and Schumann (2019)	Proposed a CNN-based explainable time series forecasting model via saliency map/heat map.	The model can provide global interpretability with a heatmap highlighting both time and features. But the method cannot explain for particular prediction.
Saadallah et al. (2021, 2022)	Proposed adaptive saliency map-based explanation techniques for time series forecasting models including CNN and ensemble classifiers.	The model can provide global interpretability using a saliency map. But the method cannot explain for particular prediction.
Ilic et al. (2021)	Introduced an explainable boosted regression technique and the explanations can be presented via regression tree.	Explainable only for boosted regression technique.

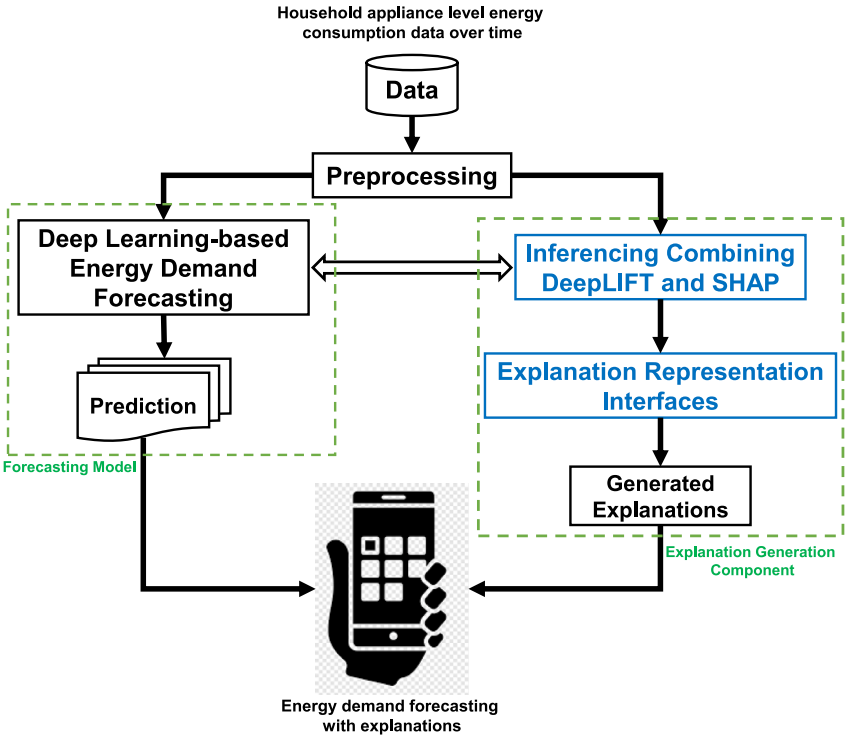

Fig. 3. An overview of explainable energy demand forecasting framework for smart home.

3. Explainable energy demand forecasting framework

This section presents our explainable energy demand forecasting framework. In particular, we have two major components in this framework, (i) a deep LSTM networks-based energy demand forecasting model and (ii) an inference and explanation generation technique by approximating Shapley values applying DeepLIFT. The high-level building blocks of the explainable energy forecasting framework are illustrated in Fig. 3. In summary, we first preprocess the time series data by handling missing values and filtering out the noisy data. The data were collected with 1-min granularity. We re-sampled the data applying the sum on an hourly, daily and weekly basis. Then, we trained an efficient deep LSTM network-based energy demand forecasting model that can predict hourly, daily and weekly energy demand in the household. Finally, we apply DeepLIFT to explain the individual prediction approximating Shapley values and provide comparatively understandable explanations through visualization. However, these two components – the forecasting model and explanation generator – are

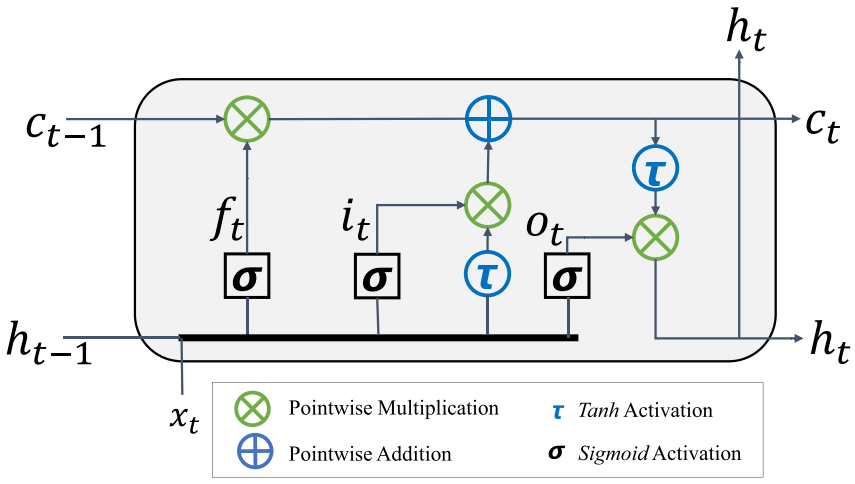

Fig. 4. A LSTM block with forget, input and output gates, f_t, i_t and o_t, respectively.

separate from each other and the prediction performance is not affected both positively and negatively. However, the DeepLIFT-based explanation generation techniques map the impact of features in different layers and can highlight the contribution corresponding to a particular time duration.

With enormous success in predictive modeling, the complex deep neural network (DNN)-based approaches have received huge attention in every sector. Recurrent neural network (RNN) is a successful DNN technique that can model sequential data better. However, traditional RNN faced a problem in memorizing long-term dependency. This is widely called a gradient vanishing problem. To make the presentation of the DeepLIFT-enabled explainable forecasting model simpler for the audience, we chose a long-short term memory (LSTM) network, a widely used variant of RNN that can overcome this long-term dependency problem (Greff et al., 2016). For sequential predictive modeling, LSTM is one of the most successful DNN models, especially for multivariate time-series forecasting. However, LSTM has a very complex architecture and hence the decision-making procedure of these types of predictive model are very opaque, even AI practitioners often fail to understand why a particular decision is being predicted. We applied an efficient explainable LSTM networks-based forecasting model for predicting the household energy demand. Therefore, we first discuss the details of our energy demand forecasting framework applying LSTM networks and then we present our explanation generation technique for specific prediction. Note that our DeepLIFT-enabled explanation techniques can be applicable to other RNN-variants (i.e., GRU)-based DL forecasting models.

3.1. Energy demand forecasting framework with LSTM networks

In contrast to traditional feed-forward DNNs, LSTM networks possess feedback connections that facilitate the processing of sequential data and the retention of crucial information within the sequence. This capability empowers them to effectively handle subsequent data points. Drawing inspiration from the accomplishments of LSTM-based models in addressing text, audio, and time series forecasting challenges, we have developed a sophisticated energy forecasting model employing deep LSTM architecture, featuring multiple LSTM layers. Furthermore, the technique employed to generate explanations for such a successful model holds potential for broader application across diverse domains. It is, however, a reasonable expectation that heightened model performance correlates with increased network depth. Guided by this rationale, we meticulously fine-tuned our deep LSTM model, making meticulous selections of optimal parameters, including the number of hidden layers, the number of hidden units within each LSTM layer, and the size of epochs. In our tiered network, the output of the $(k-1)$th LSTM layer is harnessed as the input for the subsequent kth layer. This intricately layered architecture empowers our model with the capacity to make predictions regarding future energy demand.

To address the issue of gradient vanishing, LSTM cells incorporate three distinct components, known as gates, at a specific time step. The three gates include the forget gate, input gate, and output gate. These gates serve the purpose of regulating the information flow that enters, remains stored, and exits the network, respectively. Each of these gates has its own neural network, functioning as a filter within the LSTM cell. It is important to note that the output of an LSTM cell is dependent on the current input data, the current long-term memory, and the previous hidden state. The diagrammatic representation of an LSTM block, showcasing the functioning of its different gates and states, is depicted in Fig. 4. Let the current input data be x_i at time t and the previous hidden state be h_{t-1}. For each gate, we denote input weight as U, recurrent weight as W and bias as b. First, LSTM processes the *forget gate* at time t, f_t which is the neural network working as follows:

$$f_t = \sigma(X_t \mathbf{U^f} + h_{t-1}\mathbf{W^f} + b_f), \tag{1}$$

where forget get apply a *sigmoid* activation function that returns output in the interval [0,1] and σ represents the *Sigmoid* activation function. When the output of this network is close to 1, the forget gate chooses the input component as relevant. Otherwise, it neglects for output closer to 0 as irrelevant.

After throwing out the irrelevant information, LSTM next decides which information is to store and update the cell state. For this, it employs an input gate as follows:

$$i_t = \sigma(X_t \mathbf{U^i} + h_{t-1}\mathbf{W^i} + b_i). \tag{2}$$

A *tanh* activation function-based layer is then applied to combine the previous hidden state and new input data for generating a new memory update vector $\tilde{C}_t$ as follows

$$\tilde{C}_t = tanh(X_t \mathbf{U^c} + h_{t-1}\mathbf{W^c} + b_c). \tag{3}$$

Applying point-wise multiplication and addition, the old state C_{t-1} is then updated as C_t.

$$C_t = C_{t-1} \otimes f_t \oplus i_t \otimes \tilde{C}_t. \tag{4}$$

With the help of output get, LSTM finally produces the output as the next hidden state by processing cell state and input

$$o_t = \sigma(X_t \mathbf{U^o} + h_{t-1}\mathbf{W^o} + b_o). \tag{5}$$

To produce the final state, a point-wise multiplication is applied between o_t and cell state passed through a *tanh* activation function.

$$h_t = o_t \otimes tanh(C_t). \tag{6}$$

3.2. Explaining predictions with DeepLIFT approximating the Shapley value

To explain the opaque and very complex LSTM-based energy demand forecasting model, we approximate Shapley values employing DeepLIFT (Deep Learning Important FeaTures). DeepLIFT is a method that decomposes the complex deep neural network-based methods for specific prediction by back-propagating to compute the contributions of neurons. For a given prediction, this method provides local explanations summarizing the contributions computing the "*difference in output from some reference output considering the difference in input from some reference input*" (Shrikumar et al., 2017).

Let us assume that we have a neural network prediction model with an input layer with neurons $\{x_1, x_2, x_3, \ldots, x_n\}$, some hidden layers with sets of neurons $\{h_1, h_2, h_3, \ldots, h_n\}$ and a target output neuron t. Consider $f(x)$ to be the activation of a particular neuron and $f(x')$ to be the reference activation (Shrikumar et al., 2017). DeepLIFT calculates the contributions scores as follows:

$$\Delta t = f(x) - f(x') = \sum_{i}^{n} C_{\Delta x_i \Delta t}, \tag{7}$$

where $C_{\Delta x_i \Delta t}$ refers to the contribution score in each neuron x_i. For a given target output t and its reference activation t^0, the difference-from-reference is computed as $\Delta t = t - t^0$. Eq. (7) is also called a *summation-to-delta* property.

Let Δx be the difference-from-reference of any input neuron calculated in the same procedure described previously. For target output t and difference from output reference Δt, we can define the multiplayer by averaging the difference as follows:

$$m_{\Delta x \Delta t} = \frac{C_{\Delta x \Delta t}}{\Delta x}. \tag{8}$$

It can be seen as a contribution of Δx to the target difference Δt, computed by dividing with Δx (Shrikumar et al., 2017).

Let $\{x_1, x_2, x_3, \ldots, x_n\}$ be the set of neurons for a complete neural network, $\{h_1, h_2, h_3, \ldots, h_n\}$ be the hidden layers with neurons' set and t is a target output neuron, we can define the contribution multiplier as a chain rule:

$$m\Delta x_i \Delta t = \sum_{j} m_{\Delta x_i \Delta h_j} \cdot m_{\Delta h_j \Delta t}, \tag{9}$$

where the contribution is calculated by applying iterative chain rules in each layer. This can be applicable to any number of layers in the networks. By applying this chain rule, the contribution in terms of multipliers can be computed for a given target employing back-propagation. This is analogous to the chain rule in partial derivatives (Shrikumar et al., 2017).

We employ DeepLIFT to approximate the Shapley values to explain any particular prediction in energy demand forecasting. Here, the multipliers are represented in terms of SHAP values ϕ_i:

$$m_{x_j, f_j} = \frac{\phi_i(f_j, x)}{x_j, \mathbb{E}[x_j]} \tag{10}$$

Similar to the chain rule mentioned in Eq. (9), this can be defined as follows:

$$m_{x_j, f_j} = \sum_{j} m_{x_j, h_j} \cdot m_{h_j, f_j} \tag{11}$$

Here, we approximate the reference value by averaging over the background instances. The approximation is done by summing up the difference between the expected model output of the background model and the output of the current model, $f(x) - \mathbb{E}[x_j]$. The SHAP values are computed as:

$$\phi_i(f, a') = \sum_{z' \subseteq \{a'_1, a'_2, \ldots, a'_n\} \setminus \{a'_i\}} \frac{(|z'|)!(M - |z'| - 1)!}{M!} \cdot [f(z' \cup a'_i) - f(z')], \tag{12}$$

where a is the features vector and z' and a subset of the features employed by the model f. $f(z')$ is the prediction by the model f.

a' is the vector with feature values to be explained and can be defined as $[f(z' \cup x) - f(z')]$ and M is the number of features. The prediction by the model f is denoted by $f(z')$. Moreover, SHAP values are computed by a standard game-theoretical approach and utilized Shapley values to have a unified interpretable model with fast computation. More mathematical and technical details of DeepLIFT and SHAP can be found in the study published by Lundberg and Lee (2017) and Shrikumar et al. (2017), respectively.

4. Experiments and evaluation

This section presents the details about datasets, evaluation metrics, experimental settings, performance in energy demand forecasting and the generated explanations.

4.1. Datasets

We conducted experiments on two public benchmark datasets on household electric energy consumption including EnergyData,[1] and REFIT data (Murray et al., 2017). Here we present the summary of two different datasets.

4.1.1. Household energy consumption dataset (EnergyData)

The data has been collected from a house for 47 months, particularly from December 2006 until November 2010. The dataset consists of different energy consumption measures including global active power, global reactive power, global intensity and consumption in different household areas. A brief description of each feature is summarized in Table 2. *Submetering_1* represents the active power consumed by multiple appliances including a dishwasher, an oven, and a microwave. The active power consumption by the laundry room containing appliances including a washing machine, a tumble-drier, a refrigerator, and a light is represented by *submetering_2*. The power consumed combined by an electric water heater and an air-conditioner is denoted as *submetering_3*. All the above-mentioned measures were collected for every minute.

4.1.2. REFIT smart home dataset

This dataset contains cleaned electrical consumption data for 20 households with different properties (Murray et al., 2017). The dataset includes the aggregate electricity consumption and appliance-level consumption for 9–10 home appliances in watts at an 8-second granularity. It was collected as part of the REFIT project.[2] For our experiments, we selected four diverse households based on different property characteristics as listed in Table 3. The selected households are House 2, House 5, House 8, and House 13. The properties of each household, including the number of occupants, construction year, total owned appliances, type of the house, and size in terms of bedrooms, are summarized in Table 3. Additionally, we provide a list of appliances from which the energy consumption data were collected for each respective house in Table 4.

4.2. Evaluation metrics

We employed different evaluation metrics to validate the performance of our method in forecasting household energy demand. Out of numerous evaluation metrics, we employ four metrics including mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared Error (MSE), and root mean squared error (RMSE) for evaluating the performance. To measure the effectiveness of the generated explanations by DeepLIFT, we introduced a new metric named *contribution monotonicity coefficient, CMC*.

[1] https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption.

[2] Personalized Retrofit Decision Support Tools for UK Homes using Smart Home Technology', Grant Reference EP/K002368/1/1.

Table 2
Description of different variables in EnergyData.

Feature, f_i	Description
Global active power, f_1	Household global minute-averaged active power
Global reactive power, f_2	Household global minute-averaged reactive power
Voltage, f_3	Minute-averaged voltage (in ampere)
Global Intensity, f_4	Household global minute-averaged current intensity
Sub-metering_1, f_5	It corresponds to the kitchen, containing mainly a dishwasher, an oven and a microwave (hot plates are not electric but gas-powered)
Sub-metering_2, f_6	It corresponds to the laundry room, containing a washing machine, a tumble drier, a refrigerator and a light
Sub-metering_3, f_7	It corresponds to an electric water heater and an air-conditioner

Table 3
Properties of the selected households.

House	House information				
	Occupants	Year	Appliance	Type	Size
House 2	4	–	15	Semi-Detached	3 bed
House 5	4	1878	44	Mid-terrace	4 bed
House 8	2	1966	35	Detached	2 bed
House 13	4	Post-2002	28	Detached	4 bed

4.2.1. Evaluation metrics for forecasting

MAE. The average of absolute differences between predicted values and the original values are referred to as mean absolute error (MAE), which can be computed as follows:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}, \tag{13}$$

where y_i denotes the foretasted values and x_i is the original energy consumption. This metric calculates the degree of average error made by the predictive model. The low MAE values close to zero indicate the high accuracy of the predictor. Since this is the arithmetic average, it can be affected by sampling fluctuation.

MAPE. This can be computed by dividing the absolute difference between predicted and original values by original value.

$$MAPE = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{x_t - y_t}{x_t} \right| \tag{14}$$

To address the sampling fluctuation issue, the division is done by x_t for corresponding predicted and original values.

MSE. This is another widely used evaluation metric that calculates the average error by applying the squared difference between the predicted and original values instead of the absolute difference.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2 \tag{15}$$

One of the features of this metric is that it penalizes outliers and/or large errors more than minor differences because of employing a square function. Compared with MAE and MAPE, this evaluation metric is better as it overcomes the extreme and zero value problem.

RMSE. As an extension of MSE, RMSE applies the square-root function over the squared difference between original and predicted values.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{y}_i)^2} \tag{16}$$

This metric makes it easier to understand the performance of any forecasting model than other metrics. However, for all metrics, the lower the value of any metric, the better the performance of the forecasting model is.

4.2.2. Metric to measure explainability

We present the explanations generated by approximating Shapley values with DeepLIFT using different visualizations to comprehend specific predictions from our forecasting methods. The core of these explanations is a list of contributions from various appliances and features towards the predicted future overall consumption. To assess the quality of these explanations, we compare the highlighted contributions for different appliances in households with the original contributions in the test data. We statistically compute the contributions of different appliances to total energy consumption. In other words, how much a particular appliance is responsible for the overall energy consumption for a certain time? The explanations we generate also depict contributions from different appliances towards future consumption, expressed as approximate Shapley values computed by applying DeepLIFT to the deep forecasting model.

Our hypothesis for evaluating the computed contributions is to examine how monotonous and correlated the generated contributions are with the original contributions to overall energy consumption. If there are high monotonous correlations between the vectors of contributions for original energy consumption data and Shapley values, we can conclude that the generated contributions using DeepLIFT are analogous. The degree of goodness of the generated explanations can be represented by the correlation coefficient, where a higher correlation coefficient indicates better-generated explanations.

Following the above-mentioned hypothesis and intuition, we first compute the total energy consumed by a particular appliance A_i, denoted as T_{A_i}, and then calculate the contribution of the appliance C_{A_i} by dividing the total energy consumption of the household T_H by T_{A_i} ($C_{A_i} = \frac{T_{A_i}}{T_H}$). Using the same formula, we compute the contributions for all appliances represented in a vector. On the other hand, we have a contribution vector S_{A_i} in terms of Shapley values by DeepLIFT, representing the predicted contributions for different appliances towards overall predicted consumption. For two given contribution vectors, we can compute the correlation coefficient between them. To do this, we consider the Spearman correlation coefficient to assess the correlation between original and predicted contributions to overall energy consumption. The Spearman Rank-correlation coefficient is chosen because it can identify the monotonous relationship between two vectors. If we find a high correlation and an increasing monotonous relationship, we can conclude that the predicted contributions are analogous to the original contributions.

Contribution monotonicity coefficient (CMC): Given two contribution vectors, $C = \{C_{A_1}, C_{A_2}, C_{A_3}, \ldots, C_{A_n}\}$ and $S = \{S_{A_1}, S_{A_2}, S_{A_3}, \ldots, C_{A_n}\}$ that represent the normalized contributions for different features. The C_{A_i} denotes the real contribution towards the overall consumption and the S_{A_i} represents the predicted contributions by DeepLIFT techniques in terms of Shapley values. We compute the Spearman-ranked correlation coefficient-based measure *contribution monotonicity coefficient*, CMC as follows:

$$\rho_{CMC}(C_A, S_A) = 1 - \frac{6 \sum {d_i}^2}{n(n^2 - 1)}, \tag{17}$$

where d_i is the difference between the ranks of the contribution score C_{A_i} and S_{A_i} and n is the length of the vectors. The higher the value of ρ, the better the explanation is. The positive ρ indicates the increased monotonous relationship between the vectors and the negative indicates decreasing.

Table 4
The list of appliances considered in collecting data for the selected households.

House #	List of appliances
House 2	Fridge-freezer, washing machine, dishwasher, television, microwave, toaster, Hi-Fi, kettle, overhead fan
House 5	Fridge-freezer, tumble dryer, washing machine, dishwasher, desktop computer, television, microwave, kettle, toaster
House 8	Fridge, freezer, washer dryer, washing machine, toaster, computer, television, microwave, kettle
House 13	Television site freezer, washing machine, dishwasher, network site, microwave, microwave, kettle

Table 5
Summary of the parameters of the LSTM-based forecasting model.

Parameter	Hourly	Daily	Weekly
# of features	10	10	10
Sequence length	24	30	7
# of hidden layers	2	2	2
# of hidden units	64	64	64
# of epoch	100	100	50
Learning rate	0.001	0.001	0.001
Batch size	1024	1024	64

Table 6
The performance of our proposed explainable forecasting compared to other methods.

Mode	Method	MAE	MAPE	MSE	RMSE
Hourly	**Our framework**	**0.075**	**48.9**	**0.009**	**0.096**
	Kim and Cho (2019b)	0.077	77.5	0.009	0.098
	Linear regression	0.5022	83.74	0.4247	0.6517
Daily	**Our framework**	**0.052**	**23.3**	**0.005**	**0.069**
	Kim and Cho (2019b)	0.063	28.4	0.006	0.083
	Linear regression	0.3915	52.69	0.2526	0.5026
Weekly	**Our framework**	**0.119**	27.7	**0.019**	**0.138**
	Kim and Cho (2019b)	0.121	**26.4**	0.021	0.146
	Linear regression	0.3199	41.33	0.1480	0.3847

4.3. Experimental setting

The data were collected by measuring the energy consumption in different household areas and appliances for 1 min and 8 s time intervals for EnergyData and REFIT datasets, respectively. We converted the consumption in the datasets in three different forms applying resampling in an hourly, daily and weekly manner. We have summed up the energy consumed hourly, daily and weekly. Then we applied a classical MinMax scaller to transform every feature's values in similar ranging from zero (0) to one (1). Along with this, we also have a descaler function so that we can convert the predicted values (energy forecasting) to original units.

However, we applied 85% of the samples as a training set and the 15% samples left were used for testing the model in all three types of forecasting namely, hourly, daily and weekly. For hourly forecasting, in total, we have household energy consumption data for 34 589 h (i.e., for EnergyData). The sequence lengths for three different forecasting models were 24, 30 and 7, respectively. For all models the number of input features in each sample was the same. Along with regular features including the consumption by different household areas and appliances, we also employ days of the week, month of the year, and quarter of the year as features. Other than that, we subtracted the summation of energy consumed by three different areas (sub-meters) from the total household energy consumption for EnergyData and used it as a new feature. We conducted experiments for both datasets by applying 5-fold cross-validation and applied arithmetic averages to calculate the performance in terms of different evaluation metrics. The details of the parameters of our LSTM models are summarized in Table 5.

After training our LSTM-based forecasting model with adequate training data, we applied our inference component to identify the facts (i.e., the importance of different features corresponding to time). Our inference and explanations interface then visualize the impact of different features with time duration in different forms. This explanation generation component is different from the demand forecasting model. DeepLIFT can identify contributions by approximating shapely values for different features by mapping the changes of the values in the different layers. Finally, we highlight different features' contributions with time duration. To compare the performance of our explainable forecasting framework, we applied the method proposed by Kim and Cho (2019b). They applied a CNN-LSTM-based deep learning model. We designed and conducted the experiments by applying their method with the same feature scaling and normalization techniques.

4.4. Experimental results

Along with the features described in Table 2, we also extracted handcrafted features and introduced four different features including seasonality. Generally, the daily energy consumption is dependent on the type of days. It is expected that the overall energy consumption on the weekend is supposed to be different than on the weekdays. Similarly, the season has a great impact on the overall consumption, i.e., the daily consumption in the winter season will be different from the consumption in summer and the consumption trend will be different in autumn as well. Therefore, we extract three new features namely, *the day of the week*, *the month of the year*, and *the quarter of the year*. Other than these features, we noticed that the total energy measures by three sub-meters are smaller than the total energy consumption. Therefore, we add another new feature named *others* that indicates the energy consumption extracted by subtracting the summation of three sub-meters from the total energy consumption. We conducted experiments to predict hourly, daily and weekly energy demand to validate the performance of our framework.

As we noted earlier, the main objective of our method is to explain the complex forecasting model applying DeepLIFT to approximate the Shapley values that highlight the contribution of different features corresponding to time. Nevertheless, the performance of our framework in forecasting hourly, daily, and weekly energy demand is summarized in Table 6. We conducted experiments by applying 5-fold cross-validation and applied arithmetic average to calculate the metrics. Along with our framework, we also reported the performance of other well-known household energy forecasting models. We can see that the performance of our method is quite consistent and outperformed in predicting the energy demand in the household energy demand.

Table 6 highlighted the performance comparison of our demand forecasting frameworks with some known related works including linear regression and a demand forecasting model by Kim and Cho (2019b) that applied a CNN-LSTM-based deep learning model. We conducted experiments following the proposed model applying the same normalization and scaling techniques. The results show better performance than their approach except in terms of MAPE for weekly prediction. Though the performance difference is not significant (27.7 vs 26.4). However, we can see from the table that the comparison illustrated a consistent performance in forecasting in terms of multiple evaluation metrics.

Along with predicting the total energy consumption, we also carried out experiments to see how our framework performed in predicting consumption in a specific area of the household. Since we have the dataset for three different sub-meters where the energy consumption was measured in the kitchen (i.e., dishwasher, an oven, and a microwave), the laundry room (i.e., containing washing machine, a tumble-drier, a refrigerator and a light), and another room containing

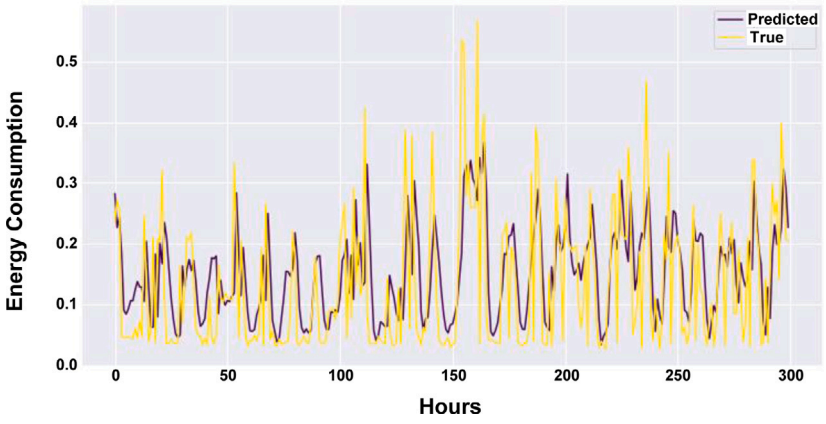

Fig. 5. Hourly prediction of our framework compared to the original consumption. The X-axis represents the hours and the Y-axis represents the original hourly energy consumption and predicted energy demand.

Table 7
Prediction performance in forecasting energy demand for different household areas.

Forecasting	Mode	MAE	MAPE	MSE	RMSE
Submetering_1	Hourly	0.149	0.934	0.028	0.168
	Daily	0.149	0.646	0.031	0.175
	Weekly	0.106	0.244	0.020	0.142
Submetering_2	Hourly	0.150	0.912	0.029	0.170
	Daily	0.164	0.677	0.036	0.189
	Weekly	0.140	0.325	0.031	0.177
Submetering_3	Hourly	0.239	1.479	0.120	0.347
	Daily	0.116	0.495	0.022	0.149
	Weekly	0.098	0.230	0.015	0.121

water heater and air-conditioner. The performance in predicting energy consumption for specific areas on an hourly, daily, and weekly basis is presented in Table 7. In turn, our framework achieved efficient performance since the prediction errors in terms of each evaluation metric are minimal.

The summary of the experimental results compared to two different forecasting methods including linear regression and deep learning models demonstrated the efficiency of our explainable energy demand forecasting model using LSTM. Moreover, the effectiveness of LSTM in time series forecasting is widely known as state-of-the-art in multiple application areas, which concludes the consistency. However, the performance difference is higher than the other baselines in all evaluation metrics. In turn, the prediction performance for the kitchen, an important energy consumption household area for hourly, daily, and weekly consumption is quite consistent and got better performance in all evaluation metrics.

To visualize the prediction performance of our framework more explicitly, we presented the predicted hourly energy consumption of our framework compared to the ground truth, actual energy consumption. We presented the hourly prediction for 300 random hours in Fig. 5. We can see that for the maximum data points, our prediction framework performs with great consistency except for a few sudden fluctuations in actual energy consumption hours.

4.5. Explaining forecasting

The explanations for daily total energy demand forecasting in the household are presented in Fig. 6. We first illustrate the impacts of different household areas to conclude which set of appliances has more responsibility for particular forecasting. Then, we visualize the contributions of seasonality features that reflect the impacts of weather conditions on the final predicted energy consumption. Finally, the contributions of all mentioned features are presented combinedly in different forms of visualizations. We can see that the daily total energy consumption has a strong impact on energy consumption by air-conditioning and water heaters. The next household area with appliances that have a high impact on the household for overall consumption is the laundry room containing appliances including a washing machine, a tumble-drier and a refrigerator. We can also see the impact of time (in the day) that has impact of consumption in different areas. The impacts in previous days are widely different. On weekends, the impacts were comparatively lower than on regular days.

In turn, we try to see that seasonal impact in the forecasting. Fig. 7 illustrates the explanation in terms of seasonal impact. The figure demonstrates that the *quarter of the year* has the highest impact on the final prediction. It makes sense that the quarter of the year, particularly winter, summer and autumn is supposed to have to higher impact on the energy consumption in households. Similarly, particular months and particular days also have an impact on energy consumption. For example, energy consumption on weekend and weekday are supposed to be different. The month and quarter of the year have a high impact on the energy demand forecasting.

Overall, the explanation for daily prediction is visualized in Fig. 8 in terms of all features corresponding to time. To have better visualization and illustration, the same explanation is presented in different formats in Figs. 9 and 10. Presenting this explanation in an easier way to understand would enable users to become more aware of consuming energy in the household. Moreover, with this explanation, users might think of changing their energy use behavior and patterns to save more household energy, hence leading to a decrease in overall carbon footprint.

4.6. Performance robustness

To validate the performance of the forecasting framework, we conducted experiments with another dataset referred to as the "REFIT smart home dataset". Moreover, we also visualized the explanations to illustrate the appliances having impacts corresponding to the times.

4.6.1. Forecasting performance on REFIT data

The performance of our explainable energy demand forecasting system is illustrated in Table 8. We can observe that the forecasting performance for hourly, daily, and weekly aggregate energy consumption across different households in the REFIT dataset remains consistent

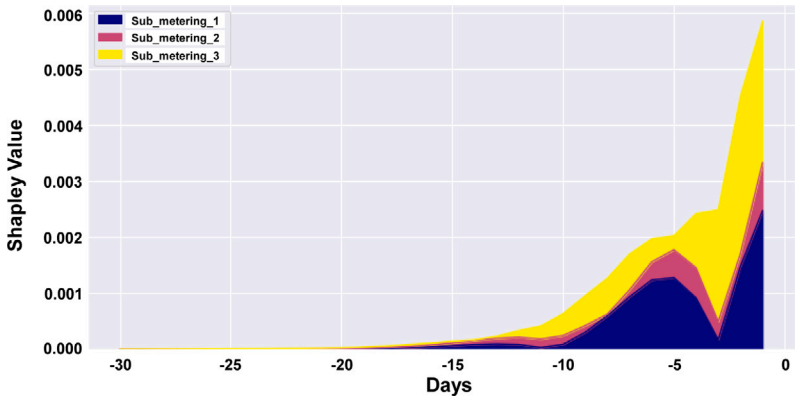

Fig. 6. The impact of the consumption in the different household areas on the total energy consumption prediction corresponding to time. The X-axis represents the time and the Y-axis represents the contributions of household areas in terms of Shapley values.

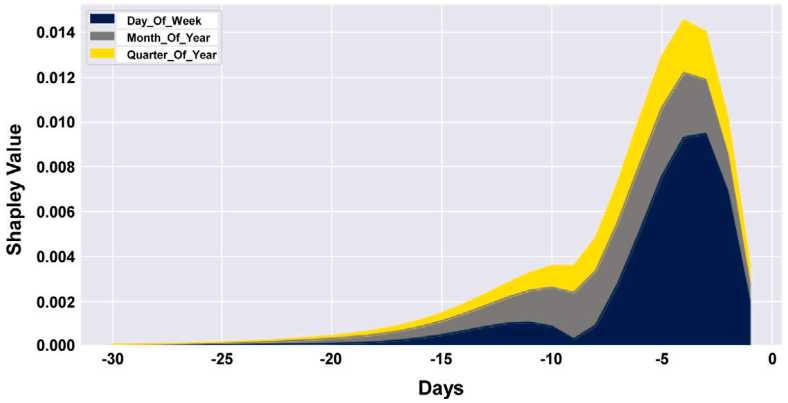

Fig. 7. The seasonal impact on the total energy consumption prediction corresponding to time. The X-axis represents the time and the Y-axis represents the contributions of seasonality features in terms of Shapley values.

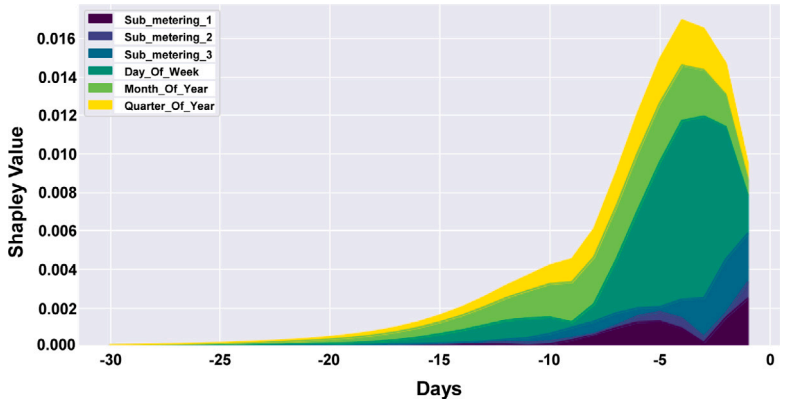

Fig. 8. Explanations in terms of the impact of all features corresponding to time. The X-axis represents the time and the Y-axis represents the contributions of features in terms of Shapley values.

across all evaluation metrics. Moreover, compared to the performance on the previous dataset, we can see that the performance, based on all evaluation metrics, is even better. The bold real numbers in the table indicate the best results achieved across all four households for different forecasting modes (hourly, daily, and weekly).

For hourly forecasting, with the exception of MAPE, we can observe that the forecasting performance is better for House 5 across all evaluation metrics. On the other hand, for daily forecasting, our method achieved the best performance for House 13 in terms of MAE and RMSE, House 8 in terms of MAPE, and House 5 in terms of MSE. The weekly forecasting performance is quite similar to the daily performance, showing better results across all metrics except MAPE for House 13. However, the performance difference across all households is not substantial.

Similar to the previous dataset, we present the hourly predictions for house 8 compared to the original consumption in Fig. 11. The X-axis represents 300 random consecutive hours, and the Y-axis represents the normalized aggregate consumption. The figure demonstrates that our method can accurately forecast future consumption over an extended period, except for a very sudden fluctuation near hour 248.

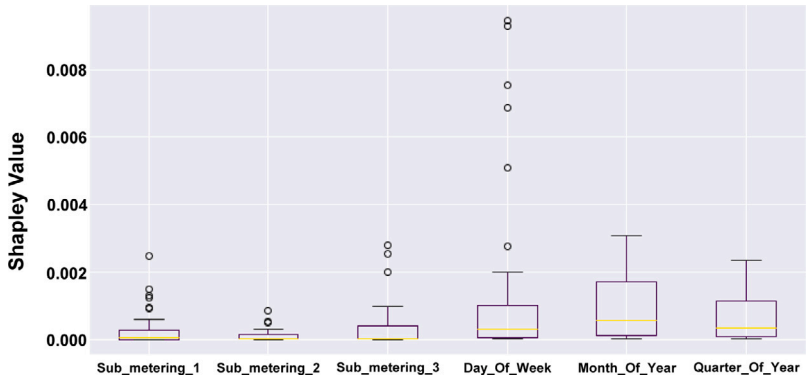

Fig. 9. The global importance of different features presented as explanations in terms of Box Plot.

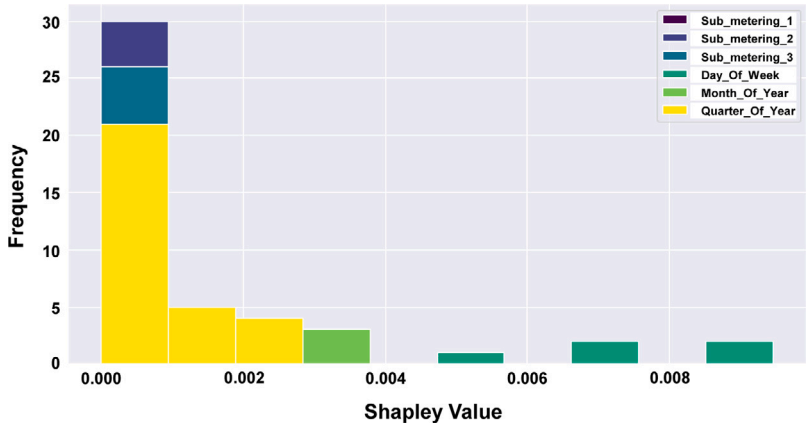

Fig. 10. Explanations in terms of histogram highlighting the impact of all features corresponding to time.

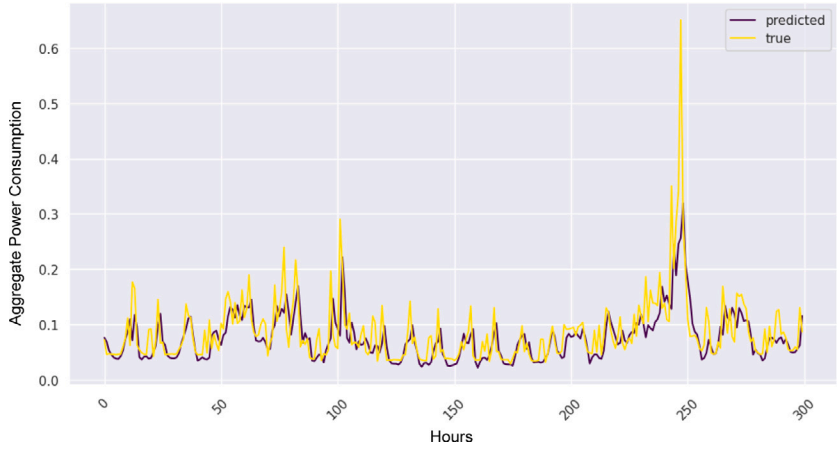

Fig. 11. Hourly prediction of our framework compared to the original consumption on house 8 (REFIT dataset). The X-axis represents the hours and Y-axis represents the original hourly energy consumption and predicted energy demand.

4.6.2. *Explaining forecasting on REFIT data*

The explanations for house 8 are illustrated in Figs. 12 and 13 in terms of area plot and bar chart. The X-axis represents days and the Y-axis represents the contributions/impacts of different appliances with seasonality. We can see that the most influential appliances are the Fridge, Toster, Kettle, Microwave, etc. In terms of seasonality, *Quarter_of_Year* has the highest influence on the weekly prediction.

The contributions or impacts of different feature appliances are illustrated in Fig. 14 in terms of the box plot. We can see that the Fridge and Toaster are the two appliances having the highest contribution towards the model's prediction.

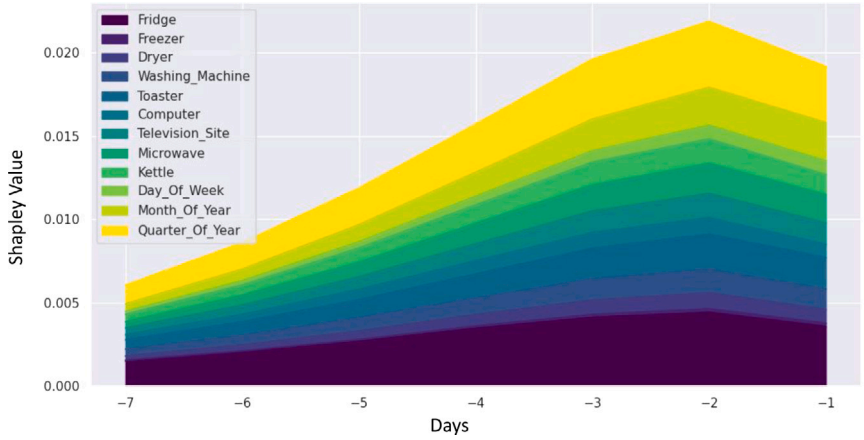

Fig. 12. Contributions of different appliances and features corresponding to times (days) in house 8 towards overall weekly aggregate forecasting.

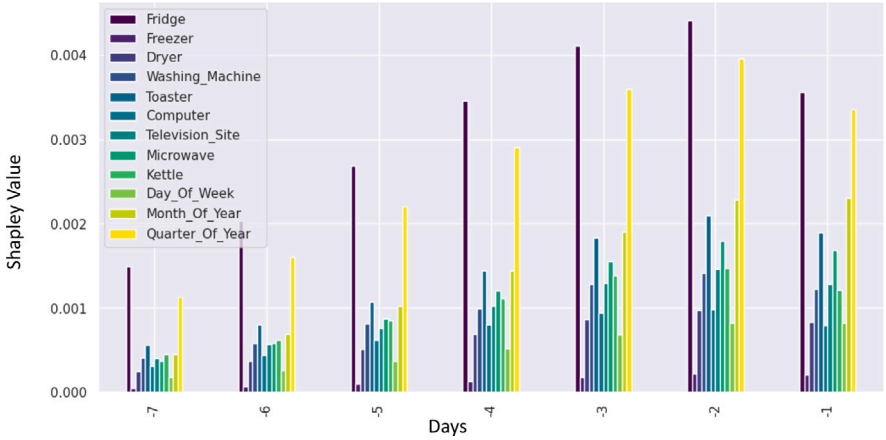

Fig. 13. Contributions of different appliances corresponding to times (days) in house 8 towards overall weekly aggregate forecasting.

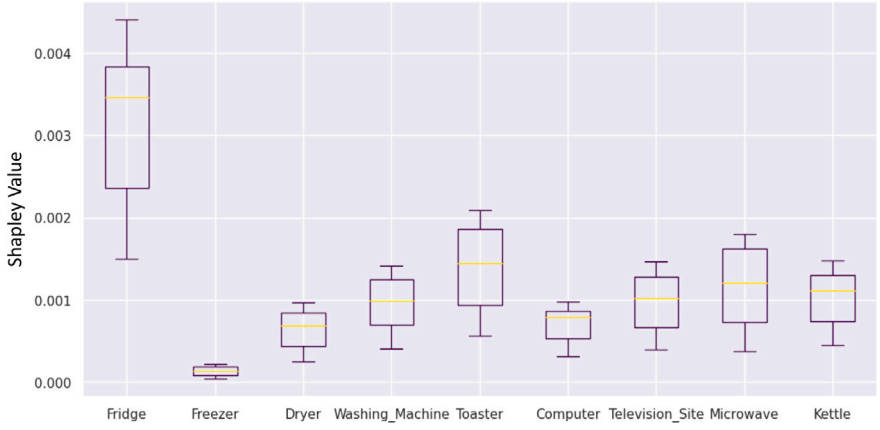

Fig. 14. Contributions of different appliances in house 8 towards overall weekly aggregate forecasting.

In general, kitchen appliances such as the Toaster, Microwave, and Kettle collectively have a more significant impact on the overall energy consumption prediction. However, evaluating the generated explanations is subjective, especially in the context of time series forecasting, where explanations become two-dimensional, making quantitative assessment more challenging.

4.7. Evaluation of generated explanations

We already discussed and reported the forecasting performance on two different datasets which includes data for five different households. We also illustrated the generated explanations in terms of Shapely

Table 8
Prediction performance in forecasting energy demand on 4 households of REFIT dataset (Section 4.1.2).

House #	Mode	MAE	MAPE	MSE	RMSE
House 2	Hourly	0.047	2.125	0.007	0.083
	Daily	0.151	0.701	0.037	0.192
	Weekly	0.134	0.557	0.03	0.174
House 5	Hourly	**0.022**	0.516	**0.001**	**0.034**
	Daily	0.085	0.476	**0.012**	0.109
	Weekly	0.076	**0.396**	0.01	0.1
House 8	Hourly	0.037	**0.436**	0.004	0.065
	Daily	0.173	**0.411**	0.041	0.203
	Weekly	0.19	0.464	0.046	0.214
House 13	Hourly	0.044	0.603	0.006	0.079
	Daily	**0.037**	0.698	0.003	**0.056**
	Weekly	**0.041**	0.812	**0.003**	**0.057**

Table 9
The effectiveness of the generated explanations by DeepLIFT on energy demand forecasting.

Dataset	Household	Mode	ρ_{CMC}	*P*-value
EnergyData (Section 4.1.1)	House 1	Daily	0.8857	0.0188
		Weekly	0.7881	0.0318
REFIT Data (Section 4.1.2)	House 2	Daily	0.8166	0.0072
		Weekly	0.7829	0.0198
	House 5	Daily	0.6985	0.0345
		Weekly	0.7315	0.0280
	House 8	Daily	0.7133	0.0470
		Weekly	0.7315	0.0102
	House 13	Daily	0.7354	0.0297
		Weekly	0.7918	0.0178

values approximated with DeepLIFT. We have seen that the explanations can indicate the factors and appliances associated with future consumption. However, in this section, we computationally report the efficiency of our generated explanations for forecasting in terms of *contribution monotonicity coefficient, CMC*. We designed this evaluation metric that can measure the degree of monotonicity coefficient considering the impacts of different appliances on overall energy consumption on original data and predicted consumption.

We presented the performance of explanations for daily and weekly forecasting in terms of *CMC* for both datasets, encompassing all five households, as shown in Table 9. It can be observed that the explanations for daily energy demand forecasting in the EnergyData dataset achieved the highest Contribution Monotonicity Coefficient (CMC) at 0.8857, with an associated lower *p*-value, numerically 0.0188 (<0.05).

Concerning weekly energy demand forecasting, the effectiveness of the explanations for Household 13 in the REFIT dataset attained the highest CMC score. However, for other households, the correlation between the impacts of different appliances on original consumption is highly consistent with the generated explanations. Except for the explanations for daily prediction for House 5, where the monotonicity coefficient exceeded 70%, indicating a consistently increasing relationship between the impacts of appliances computed by DeepLIFT and the original impacts. The performance on three households data including EnergyData, House 2, and 13 even achieved nearly or more than 80% efficiency in terms of CMC with lower *p*-value. The high correlation between the impact of appliances in explanations and the original overall consumption of the households indicates the efficiency of the generated explanations that might make user sense of any prediction from a deep learning-based forecasting model.

We also compared the impacts of different appliances on future consumption with the findings from the experiments conducted by Stankovic et al. (2016), who investigated the contribution of daily activities to overall energy consumption. They used the same dataset as ours and reported that cooking contributes to 16% of the total energy consumption in House 8. In our generated explanation, we also detected the activity of cooking through the consumption patterns of kitchen appliances like the Toaster, Kettle, and Microwave, and their collective contributions, as shown in the box plot (Fig. 14), are correlated. Combining the contributions of these three cooking appliances shows that cooking has the highest impact on overall consumption.

Stankovic et al. (2016) further reported that the next significant activities impacting energy consumption are laundering (4%) and watching TV (1%). The laundry activity was detected through the consumption of the washing machine and tumble dryer. Our generated explanation aligns with this too, as we observed that the washing machine and dryer collectively have the second-highest contributions, and the television_site also makes considerable contributions.

Based on the careful analysis on the evaluation of the explanations computationally, we can conclude that our method's generated explanations effectively identify the impact of different appliances on energy consumption. With further empirical studies involving smart home users, we aim to enhance and validate the explanation quality, enabling users to optimize their energy consumption effectively. The predicted future energy demand and the explanation can help the users with energy consumption literacy (Schwartz et al., 2013a,b) and the policymaker can think of adopting our method in dynamic pricing and energy policy optimization.

5. Conclusion

This paper presents an explainable energy demand forecasting system where we attempt to generate easy-to-understand explanations for forecasting decisions for smart home users. For doing so, we approximate the SHAP values by applying DeepLIFT to identify the feature's contributions in each neuron of an LSTM-based model. Our LSTM-based energy demand forecasting model was used to predict hourly, daily and weekly energy demand effectively on two different datasets for five different households in terms of all evaluation metrics. The major goal of this study was to explain the predictions in such a way that users can have a clear understanding of why a particular decision has been predicted. Our framework applied DeepLIFT to approximate the SHAP values to generate easy-to-understand explanations. These explanations generation technique combining DeepLIFT and SHAP can be applied to interpret the predictions for any deep learning-based forecasting models. The explanations can highlight both the time or season and the impact of different attributes (features) for a particular prediction at the same time. Based on our introduced evaluation metric named contribution monotonicity coefficient, the generated explanations achieved high efficiency and the relationship with original contributions of different appliances towards the total consumption is monotonous. We also observed that the explanations for household energy forecasting can identify the impacts of appliances for corresponding energy consumption activities that are aligned and correlated with the findings of the previous study (Stankovic et al., 2016). With these explanations, users might be more aware of and think of optimizing their energy consumption practice by considering the most responsible factors for their upcoming energy consumption demand. The predicted future energy demand and the explanation can help the users with energy consumption literacy (Schwartz et al., 2013a,b) and the policymaker can think of adopting our method in dynamic pricing and energy policy optimization.

6. Future direction

In the pursuit of creating functional user interfaces tailored to smart home users, we will adopt a user-centric design methodology, akin to the approach advocated by Jensen et al. (2018). Our current trajectory involves the development of a prototype for our proposed system, which aims to offer transparent insights into energy demand prediction and forecasting. Our underlying assumption is that by allowing smart

home users to interact with our prototype in their daily lives, we can glean insights into both the domain and the technology. This interaction will provide them the opportunity to articulate the types of explanations they consider vital and valuable. This approach is especially significant in light of existing systems, often geared towards developers and AI experts, which may exhibit certain limitations. In this regard, we have outlined a set of inquiries enumerated below, which we intend to pose as we construct our explainable prediction system with a strong emphasis on human-centered design.

1. Are these explanations helpful for you in understanding the decision-making process?
2. What open or further questions would you like to have answered, if any?
3. Do you find the presented user interface useful for engaging with the presented explanations?
4. What problems or areas for improvement would you see in this respect, if any?
5. Thinking aloud, would you please walk us through the explanation interface, reflecting on a particular prediction that is presented there?

We believe that through a user-centered prototyping approach with different kinds of explanation visualizations, we can learn more about the specific user needs in the energy domain, and elicit requirements and insights towards building a collaborative, human-centered explainable energy demand forecasting system. Hence, the system will increase transparency, fairness, and accountability to end-users.

Funding information

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955422

CRediT authorship contribution statement

Md Shajalal: Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Alexander Boden:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis. **Gunnar Stevens:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis.

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.

References

Adadi, A., Berrada, M., 2020. Explainable AI for healthcare: from black box to interpretable models. In: Embedded Systems and Artificial Intelligence: Proceedings of ESAI 2019, Fez, Morocco. Springer, pp. 327–337.

Ahmad, J., Tahir, A., Larijani, H., Ahmed, F., Aziz Shah, S., Hall, A.J., Buchanan, W.J., 2020. Energy demand forecasting of buildings using random neural networks. J. Intell. Fuzzy Systems 38 (4), 4753–4765.

Ali, U., Shamsi, M.H., Hoare, C., Mangina, E., O'Donnell, J., 2021. Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis. Energy Build. 246, 111073.

Arrotta, L., 2021. Multi-inhabitant and explainable activity recognition in smart homes. In: 2021 22nd IEEE International Conference on Mobile Data Management. MDM, IEEE, pp. 264–266.

Assaf, R., Schumann, A., 2019. Explainable deep neural networks for multivariate time series predictions. In: IJCAI. pp. 6488–6490.

Barredo Arrieta, A., Gil-Lopez, S., Laña, I., Bilbao, M.N., Del Ser, J., 2022. On the post-hoc explainability of deep echo state networks for time series forecasting, image and video classification. Neural Comput. Appl. 34 (13), 10257–10277.

Bettini, C., Civitarese, G., Fiori, M., 2021. Explainable activity recognition over interpretable models. In: 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom Workshops). IEEE, pp. 32–37.

Chadoulos, S., Koutsopoulos, I., Polyzos, G.C., 2021. One model fits all: Individualized household energy demand forecasting with a single deep learning model. In: Proceedings of the Twelfth ACM International Conference on Future Energy Systems. pp. 466–474.

Chakraborty, D., Alam, A., Chaudhuri, S., Başağaoğlu, H., Sulbaran, T., Langar, S., 2021. Scenario-based prediction of climate change impacts on building cooling energy consumption with explainable artificial intelligence. Appl. Energy 291, 116807.

Crabbé, J., van der Schaar, M., Supplementary materials for explaining time series predictions with dynamic masks.

Crabbé, J., Van Der Schaar, M., 2021. Explaining time series predictions with dynamic masks. In: International Conference on Machine Learning. PMLR, pp. 2166–2177.

Dalvi-Esfahani, M., Mosharaf-Dehkordi, M., Leong, L.W., Ramayah, T., Kanaan-Jebna, A.M.J., 2023. Exploring the drivers of XAI-enhanced clinical decision support systems adoption: Insights from a stimulus-organism-response perspective. Technol. Forecast. Soc. Change 195, 122768.

Das, D., Nishimura, Y., Vivek, R.P., Takeda, N., Fish, S.T., Ploetz, T., Chernova, S., 2021. Explainable activity recognition for smart home systems. arXiv preprint arXiv:2105.09787.

Došilović, F.K., Brčić, M., Hlupić, N., 2018. Explainable artificial intelligence: A survey. In: 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics. MIPRO, IEEE, pp. 0210–0215.

Efat, M.I.A., Hajek, P., Abedin, M.Z., Azad, R.U., Jaber, M.A., Aditya, S., Hassan, M.K., 2022. Deep-learning model using hybrid adaptive trend estimated series for modelling and forecasting sales. Ann. Oper. Res. 1–32.

Ehsan, U., Wintersberger, P., Liao, Q.V., Mara, M., Streit, M., Wachter, S., Riener, A., Riedl, M.O., 2021. Operationalizing human-centered perspectives in explainable AI. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. pp. 1–6.

Ekambaram, V., Manglik, K., Mukherjee, S., Sajja, S.S.K., Dwivedi, S., Raykar, V., 2020. Attention based multi-modal new product sales time-series forecasting. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 3110–3118.

Eseye, A.T., Lehtonen, M., 2020. Short-term forecasting of heat demand of buildings for efficient and optimal energy management based on integrated machine learning models. IEEE Trans. Ind. Inform. 16 (12), 7743–7755.

Fu, H., Kampezidou, S., Sung, W., Duncan, S., Mavris, D.N., 2018. A data-driven situational awareness approach to monitoring campus-wide power consumption. In: 2018 International Energy Conversion Engineering Conference. p. 4414.

Ghosh, I., Jana, R.K., Abedin, M.Z., 2023. An ensemble machine learning framework for Airbnb rental price modeling without using amenity-driven features. Int. J. Contemp. Hosp. Manag..

Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M., Kagal, L., 2018. Explaining explanations: An overview of interpretability of machine learning. In: 2018 IEEE 5th International Conference on Data Science and Advanced Analytics. DSAA, IEEE, pp. 80–89.

Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., Schmidhuber, J., 2016. Lstm: a search space odyssey. IEEE Trans. Neural Netw. Learn. Syst. 28 (10), 2222–2232.

Haq, E.U., Lyu, X., Jia, Y., Hua, M., Ahmad, F., 2020. Forecasting household electric appliances consumption and peak demand based on hybrid machine learning approach. Energy Rep. 6, 1099–1105.

Haque, A.B., Islam, A.N., Mikalef, P., 2023. Explainable artificial intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. Technol. Forecast. Soc. Change 186, 122120.

Ilic, I., Görgülü, B., Cevik, M., Baydoğan, M.G., 2021. Explainable boosted linear regression for time series forecasting. Pattern Recognit. 120, 108144.

Jensen, R.H., Strengers, Y., Kjeldskov, J., Nicholls, L., Skov, M.B., 2018. Designing the desirable smart home: A study of household experiences and energy consumption impacts. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 1–14.

Kabir, M.H., Hasan, K.F., Hasan, M.K., Ansari, K., 2021. Explainable artificial intelligence for smart city application: A secure and trusted platform. arXiv preprint arXiv:2111.00601.

Karim, M.R., Dey, S.K., Islam, T., Sarker, S., Menon, M.H., Hossain, K., Hossain, M.A., Decker, S., 2021. Deephateexplainer: Explainable hate speech detection in under-resourced bengali language. In: 2021 IEEE 8th International Conference on Data Science and Advanced Analytics. DSAA, IEEE, pp. 1–10.

Karim, M.R., Islam, T., Shajalal, M., Beyan, O., Lange, C., Cochez, M., Rebholz-Schuhmann, D., Decker, S., 2023. Explainable ai for bioinformatics: Methods, tools and applications. Brief. Bioinform. 24 (5), bbad236.

Kazemzadeh, M.-R., Amjadian, A., Amraee, T., 2020. A hybrid data mining driven algorithm for long term electric peak load and energy demand forecasting. Energy 204, 117948.

Khodabandehloo, E., Riboni, D., Alimohammadi, A., 2021. HealthXAI: Collaborative and explainable AI for supporting early diagnosis of cognitive decline. Future Gener. Comput. Syst. 116, 168–189.
Kim, J.-Y., Cho, S.-B., 2019a. Electric energy consumption prediction by deep learning with state explainable autoencoder. Energies 12 (4), 739.
Kim, T.-Y., Cho, S.-B., 2019b. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 182, 72–81.
Kim, J.-Y., Cho, S.-B., 2020. Electric energy demand forecasting with explainable time-series modeling. In: 2020 International Conference on Data Mining Workshops. ICDMW, IEEE, pp. 711–716.
Kim, J.-Y., Cho, S.-B., 2021. Explainable prediction of electric energy demand using a deep autoencoder with interpretable latent space. Expert Syst. Appl. 186, 115842.
Kim, J., Lee, G., Lee, S., Lee, C., 2022. Towards expert–machine collaborations for technology valuation: An interpretable machine learning approach. Technol. Forecast. Soc. Change 183, 121940.
Kim, D., Song, Y., Kim, S., Lee, S., Wu, Y., Shin, J., Lee, D., 2023a. How should the results of artificial intelligence be explained to users?-Research on consumer preferences in user-centered explainable artificial intelligence. Technol. Forecast. Soc. Change 188, 122343.
Kim, B.R., Srinivasan, K., Kong, S.H., Kim, J.H., Shin, C.S., Ram, S., 2023b. ROLEX: A novel method for interpretable machine learning using robust local explanations. MIS Q. 47 (3).
Li, T., Wang, Z., Zhao, W., 2022a. Comparison and application potential analysis of autoencoder-based electricity pattern mining algorithms for large-scale demand response. Technol. Forecast. Soc. Change 177, 121523.
Li, G., Wu, Y., Liu, J., Fang, X., Wang, Z., 2022b. Performance evaluation of short-term cross-building energy predictions using deep transfer learning strategies. Energy Build. 275, 112461.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.
Ma, Y., Chen, X., Wang, L., Yang, J., 2021. Study on smart home energy management system based on artificial intelligence. J. Sens. 2021.
Mucha, H., Robert, S., Breitschwerdt, R., Fellmann, M., 2020. Towards participatory design spaces for explainable ai interfaces in expert domains. In: 43rd German Conference on Artificial Intelligence, Bamberg, Germany.
Mucha, H., Robert, S., Breitschwerdt, R., Fellmann, M., 2021. Interfaces for explanations in human-AI interaction: Proposing a design evaluation approach. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. pp. 1–6.
Murray, D., Stankovic, L., Stankovic, V., 2017. An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. Sci. Data 4 (1), 1–12.
Rai, A., 2020. Explainable AI: From black box to glass box. J. Acad. Mark. Sci. 48 (1), 137–141.
Ribeiro, M.T., Singh, S., Guestrin, C., 2016a. Model-agnostic interpretability of machine learning. arXiv preprint arXiv:1606.05386.
Ribeiro, M.T., Singh, S., Guestrin, C., 2016b. "Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1135–1144.
Riboni, D., 2021. Keynote: Explainable AI in pervasive healthcare: Open challenges and research directions. In: 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom Workshops). IEEE, p. 1.
Rojat, T., Puget, R., Filliat, D., Del Ser, J., Gelin, R., Díaz-Rodríguez, N., 2021. Explainable artificial intelligence (xai) on timeseries data: A survey. arXiv preprint arXiv:2104.00950.
Rozanec, J.M., 2021. Explainable demand forecasting: A data mining goldmine. In: Companion Proceedings of the Web Conference 2021. pp. 723–724.
Saadallah, A., Jakobs, M., Morik, K., 2021. Explainable online deep neural network selection using adaptive saliency maps for time series forecasting. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, pp. 404–420.
Saadallah, A., Jakobs, M., Morik, K., 2022. Explainable online ensemble of deep neural network pruning for time series forecasting. Mach. Learn. 1–29.
Saluja, R., Malhi, A., Knapič, S., Främling, K., Cavdar, C., 2021. Towards a rigorous evaluation of explainability for multivariate time series. arXiv preprint arXiv:2104.04075.
Schlegel, U., Oelke, D., Keim, D.A., El-Assady, M., 2020. An empirical study of explainable AI techniques on deep learning models for time series tasks. arXiv preprint arXiv:2012.04344.
Schwartz, T., Denef, S., Stevens, G., Ramirez, L., Wulf, V., 2013a. Cultivating energy literacy: results from a longitudinal living lab study of a home energy management system. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1193–1202.
Schwartz, T., Stevens, G., Ramirez, L., Wulf, V., 2013b. Uncovering practices of making energy consumption accountable: A phenomenological inquiry. ACM Trans. Comput.-Hum. Interact. 20 (2), 1–30.
Shajalal, M., Boden, A., Stevens, G., 2022a. Explainable product backorder prediction exploiting CNN: Introducing explainable models in businesses. Electron. Mark. 32 (4), 2107–2122.
Shajalal, M., Boden, A., Stevens, G., 2022b. Towards user-centered explainable energy demand forecasting systems. In: Proceedings of the Thirteenth ACM International Conference on Future Energy Systems. pp. 446–447.
Shajalal, M., Denef, S., Karim, M.R., Boden, A., Gunnar, S., 2023. Unveiling Black-boxes: Explainable deep learning models for patent classification. In: The World Conference on EXplainable Artificial Intelligence 2023. Springer, pp. 1–18.
Shrikumar, A., Greenside, P., Kundaje, A., 2017. Learning important features through propagating activation differences. In: International Conference on Machine Learning. PMLR, pp. 3145–3153.
Stankovic, L., Stankovic, V., Liao, J., Wilson, C., 2016. Measuring the energy intensity of domestic activities from smart meter data. Appl. Energy 183, 1565–1580.
Syed, D., Abu-Rub, H., Ghrayeb, A., Refaat, S.S., 2021. Household-level energy forecasting in smart buildings using a novel hybrid deep learning model. IEEE Access 9, 33498–33511.
Vanting, N.B., Ma, Z., Jørgensen, B.N., 2021. A scoping review of deep neural networks for electric load forecasting. Energy Inform. 4 (2), 1–13.
Wang, Z., Jiang, C., Zhao, H., 2022. Know where to invest: Platform risk evaluation in online lending. Inf. Syst. Res. 33 (3), 765–783.
Yan, K., Li, W., Ji, Z., Qi, M., Du, Y., 2019. A hybrid LSTM neural network for energy consumption forecasting of individual households. Ieee Access 7, 157633–157642.
Yang, C., Abedin, M.Z., Zhang, H., Weng, F., Hajek, P., 2023. An interpretable system for predicting the impact of COVID-19 government interventions on stock market sectors. Ann. Oper. Res. 1–28.
Zdravković, M., Ćirić, I., Ignjatović, M., 2022. Explainable heat demand forecasting for the novel control strategies of district heating systems. Annu. Rev. Control.
Zhang, W., Liu, F., Wen, Y., Nee, B., 2021. Toward explainable and interpretable building energy modelling: an explainable artificial intelligence approach. In: Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. pp. 255–258.

Md Shajalal is a Marie Skłodowska-Curie Research Fellow at Fraunhofer FIT, Germany, and a Ph.D. candidate in Explainable AI at the University of Siegen, Germany. He has served as a visiting research fellow at the University of Strathclyde, Glasgow, UK, and completed an industrial secondment at AGENTS.Inc, Berlin, Germany. Shajalal is also academic faculty member of Computer Science (currently on leave) at Hajee Mohammad Danesh Science and Technology University, Bangladesh. He holds a B.Sc. in Computer Science from the University of Chittagong (CU), Bangladesh, and an M.Eng. degree in Computer Science from Toyohashi University of Technology (TUT), Japan. Shajalal was awarded prestigious Marie Skłodowska-Curie Fellowship and MEXT scholarship for his doctoral and master's studies, funded by the European Union (EU) and the Ministry of Education, Culture, Sports, Science, and Technology, Japan, respectively. His research interests include human-centered explainable AI (XAI), natural language processing (NLP), information retrieval (IR), and applied machine/deep learning.

Alexander Boden is a professor of Software Engineering at Bonn-Rhein-Sieg University of Applied Sciences (HBRS), Bonn, Germany. He received his Ph.D. in business informatics from the University of Siegen, Germany in an interdisciplinary topic between software engineering and social sciences. In 2019, he received his habilitation in business informatics at the University of Siegen. He is the director of the Institute for Consumer Informatics, HBRS. Professor Alex and his research team broadly focus on Humancentered AI, Human-computer interaction, and Business informatics. He is also affiliated with the Department of User-centered Computing at Fraunhofer Institute for Applied Information Technology FIT, Germany where he leads a research team in the field of Human-Centered Computing.

Gunnar Stevens is a professor of Information Systems and New Media at the University of Siegen, Germany. He is the divisional director of IT Security and Consumer Informatics. Professor Gunnar is leading a multidisciplinary research team that is broadly concerned with user-centered computing and IT security, human–computer interaction, and applied machine/deep learning. He is also serving as a Head of the Consumer Informatics/Competence Center SME 4.0 Usability at Bonn-Rhein-Sieg University of Applied Sciences, Bonn, Germany.