Household Energy Consumption Prediction Using the Stationary Wavelet Transform and Transformers

LYES SAAD SAOUD1, HASAN AL-MARZOUQI1, (Senior Member, IEEE), AND RAMY HUSSEIN2

1Electrical and Computer Engineering Department, Khalifa University, Abu Dhabi, United Arab Emirates
2Radiological Sciences Laboratory (RSL), Stanford University, Stanford, CA 94305, USA

Corresponding author: Lyes Saad Saoud (lyes.saoud@ku.ac.ae)

This work was supported by the Information and Communication Technology (ICT) Fund, Telecommunications Regulatory Authority (TRA), Abu Dhabi, United Arab Emirates.

ABSTRACT In this paper, we present a new method for forecasting power consumption. Household power consumption prediction is essential to manage and plan energy utilization. This study proposes a new technique using machine learning models based on the stationary wavelet transform (SWT) and transformers to forecast household power consumption in different resolutions. This approach works by leveraging self-attention mechanisms to learn complex patterns and dynamics from household power consumption data. The SWT and its inverse are used to decompose and reconstruct the actual and the forecasted household power consumption data, respectively, and deep transformers are used to forecast the SWT subbands. Experimental findings show that our hybrid approach achieves superior prediction performance compared to the existing power consumption prediction methods.

INDEX TERMS Household power consumption, transformers, stationary wavelet transform, time series forecasting.

I. INTRODUCTION

Electric energy consumption has recently risen worldwide, driven by economic advancements and increasing population [1]. According to the 2019 World Energy Outlook released by International Energy Agency (IEA), the worldwide electricity demand increases at 2.1% per year to 2040, double the stated policies scenario’s primary energy production rate. Therefore, the total final energy consumption is expected to rise from 19% in 2018 to 24% in 2040 [2]. The housing market accounts for 27% of global electricity demand and significantly affects aggregate electricity usage [3]. Because electricity is used simultaneously during the production at the power plant, it is necessary to forecast energy consumption in advance for a reliable power supply [4]. Over the last few decades, a growing number of models have been developed to predict building energy consumption [4]–[9]. In what follows, we review some of the recently published papers in the literature.

The associate editor coordinating the review of this manuscript and approving it for publication was Amjad Anvari-Moghaddam.

Forecasting energy consumption is a challenging time series prediction problem. Intelligent sensors collect data that may contain redundancy, missing values, outliers, and uncertainties [6]. Moreover, it is hard to predict electrical energy consumption using traditional forecasting techniques since energy usage has erratic trend components, including regular seasonal patterns [4], [10]. Appropriate operating approaches should be implemented in energy control schemes to maximize buildings’ energy efficiency [7]. Therefore, various forecasting techniques have been recently proposed to predict energy consumption. Energy consumption forecasting has been studied using a variety of different techniques that can be divided into conventional and artificial intelligence (AI) models [8]. Wei et al. [8] have reviewed 128 models in 116 published studies used to forecast energy consumption; among them, 62.48% are AI-based models. We have divided energy consumption forecasting systems into three primary categories, statistical models, machine learning models, and hybrid models. Statistical techniques were used mainly in the past to predict energy demand. For example, in [9], the seasonal
The autoregressive integrated moving average (SARIMA) model was compared to the neuro-fuzzy model for forecasting electric load. Both linear regression (with one predictor and multiple predictors) and quadratic regression models were applied to the hourly and daily energy consumption of the research household [11]. Also, in [12], the multiple regression approach together with a genetic engineering technique were proposed to estimate the administration building’s daily energy use. Both models’ significant drawbacks include the unavailability of occupancy data and the reality that none of these models have been studied to estimate comparable buildings’ energy usage. Bootstrap aggregating autoregressive integrated moving average (ARIMA) and the exponential smoothing methods have been used to forecast energy demand for different countries [13]. Generally, the statistical techniques showed their weakness in long-term forecasting and capturing the nonlinear behavior of the energy consumption data.

Furthermore, computational approaches have shown limited prediction performance due to the non-stationarity nature and serious trends in the energy demand; therefore, many prediction models have been tested using machine learning methods to improve the forecasting quality [14]–[16]. For instance, Liu et al. [17] have developed a support vector machine (SVM) model to forecast and analyze public buildings’ energy consumption. Driven by the solid nonlinear supporting vector regression capacities, Chen et al. [18] proposed a model that forecasts the electrical load based on the ambient temperature. Energy consumption has been forecasted based on evaluating the usage of aggregated people dynamics in [19]. An artificial neural network-based cuckoo search learning algorithm was proposed to forecast the electricity consumption of the organization of petroleum exporting countries (OPEC) [20]. Pinto et al. [21] proposed an ensemble learning model containing three machine learning algorithms: random forests, gradient boosted regression trees, and AdaBoost to forecast energy consumption. Nevertheless, current machine learning approaches severely suffer from overfitting as the dynamic correlation between variables is challenging, and data characteristics change over time. It is hard to ascertain long-term and reliable usage when overfitting happens.

Likewise, many deep sequential learning neural networks have been established to forecast electricity use. A recurrent neural network model was used to predict medium-to-long term electricity consumption profiles in commercial and residential buildings at one-hour resolution predictions [22]. A pooling-based recurrent neural network (RNN) approach has been proposed to address the over-fitting issue by increasing data diversity and volume [23]. An RNN architecture with long-short term memory (LSTM) cells was used to forecast energy load in [24]. A model based on LSTM networks was also proposed in [25] to forecast regular energy consumption. In addition, an advanced optimization method focused on the bagged echo state network (ESN) and improved by a differential evolution algorithm was proposed in [26] to approximate energy usage. The performance of deep extreme learning machines was investigated for energy consumption prediction in residential buildings [27]. The proposed model outperformed other artificial neural and neuro-fuzzy networks.

To achieve adequate predictability based on the weak knowledge and lack of a multitude of historical evidence in energy consumption, Gao et al. [28] suggested using two deep learning models, a sequence-to-sequence model and a two-dimensional attention-based convolutional neural network model. Deep learning models can extract the crucial and hidden features needed for accurate prediction, even from non-stationary data with dynamic features and/or different biomarkers. However, conventional deep learning models have difficulties identifying the spatiotemporal properties pertinent to energy use [4].

Several variables, such as the market cycle and regional economic policies, have a significant impact on energy usage. Therefore, it is very challenging that a single intelligent algorithm would suffice [29]. Thus, combining efficient pre-processing techniques and feature learning models for forecasting power consumption has a great potential for improving prediction performance. For instance, the stacked autoencoders and extreme learning machines were used to efficiently extract the energy consumption-related features and achieve more robust prediction performance in [5]. AdaBoost ensemble technology was hybridized with a neural network, support vector regression machine, genetic programming, and radial basis function network to better forecast energy consumption [30]. The hybrid SARIMA-metaheuristic firefly algorithm-least squares support vector regression model was used to forecast energy consumption in [8].

Well-known artificial intelligence methods have been used to evaluate energy use in single and ensemble situations. An in-depth study and examination of the hybrid model, integrating forecasting and optimization approaches, were discussed. A thorough analysis revealed that the combination configuration is more reliable than the single and assembly models. A hybrid convolutional neural network - LSTM (CNN-LSTM) model has been established for electricity forecasting [4], [31]. The CNN was used to extract the features, and the LSTM layer was used to deal with the temporal behavior of the time series data. A predictive model of energy consumption using LSTM and sine cosine optimization algorithm was proposed [32]. Hu et al. [33] combined echo state network, bagging, and differential evolution algorithm to forecast energy consumption. Logarithmic Mean Divisia Index, empirical mode decomposition, least-square support vector machine, and particle swarm optimization were hybridized to forecast energy consumption [34]. Kaytez [35] proposed forecasting energy consumption using the least-square SVM (LSSVM) and an autoregressive integrated moving average.

A mixture of three deep reinforcement learning models, including asynchronous advantage Actor-Critic, deep deterministic policy gradient, and recurrent deterministic...
policy gradient, was introduced in [36] to address nonlinear and complex energy consumption forecasting results. An ensemble model was proposed in [37], in which the energy consumption data was divided into stable and stochastic elements. A hybrid model based on ARIMA, artificial neural network, and the combined Particle Swarm Optimization Support Vector Regression, was proposed and used for load and energy forecasting [38]. Complete ensemble empirical mode decomposition with adaptive noise and machine learning model–extreme gradient boosting was proposed to predict building energy consumption [39]. A hybrid model has been proposed by combining CNN with multilayer bi-directional LSTM [40]. The hybrid energy-based sequential learning prediction model that used a coherent structure for the reliable energy usage prediction was brought forward using CNN and Gated Recurrent Units (GRU) [1].

The k−means clustering based CNN-LSTM (kCNN-LSTM) model was proposed to provide a precise forecast of building energy consumption [42]. The k-CNN-LSTM was found to achieve superior performance when compared to other existing machine learning and deep learning energy demand forecast models. In [43], Liu et al. developed a hybrid model for the short-term predictions of residential electricity consumption based on the Holt-Winters method and Extreme Learning Machine (ELM) network. They also compared their hybrid model with non-hybrid deep learning models such as ELM and LSTM. For a training data set of 50 days, the proposed model reduced the prediction error rate by 53.39-87.98%. Another integrated approach, consisting of feature extraction, optimization, and adaptive deep neural networks (DNNs), was proposed in [44] to forecast week-ahead hourly building energy consumption. The feature extraction procedure was carried out through the k-means clustering technique, while the DNN was the forecasting engine of the proposed model. A genetic algorithm was also deployed to identify the DNN architecture that yields superior prediction performance. The proposed hybrid predictive model was implemented on an actual office building in the UK, and it was found to achieve an 11.9-24.6% decrease in the mean absolute percentage error when compared with other DNNs of fixed architectures. In an attempt to provide more robust forecasting of building energy consumption, the authors of [45] proposed to use an LSTM recurrent neural network together with an improved sine cosine optimization algorithm. They also introduced a novel Haar wavelet-based mutation operator to optimize the hyper-parameters of the LSTM network and improve the divergence of the sine cosine optimization method.

The proposed model showed accurate and reliable predictions for short, mid, and long-term energy consumption forecasting problems. In [46], another integrated machine learning model was proposed to boost the prediction performance of building energy consumption, and it showed lower prediction error rates compared to individual machine learning models. Similarly, a hybrid approach including online search data for household power consumption forecasting was developed to increase forecasting accuracy [47]. To forecast residential electricity usage, an extreme learning machine model optimized by the Jaya algorithm was proposed, along with the selected search keywords. This hybrid model showed the ability to better predict residential electricity consumption.

Recently, transformer networks were introduced to resolve the parallelization issue of the LSTM [48]. With the aid of attention, the intermediate distance between the source and the target sequences is no longer constrained. Rather than producing a single context vector from the last hidden state of the encoder, attention establishes shortcuts between the input sequence and the whole source entry. For each output element, the weights of these shortcuts can be customized.

This approach benefits from eliminating recursion, so those parallel calculations help minimize the training time and tackle the reduction in efficiency related to long-term dependencies and the corresponding vanishing gradient problem. Transformers have been successfully applied to healthcare problems such as influenza-like illness prediction [49]. In general, deep transformers have two limitations: (1) they cannot represent greater than one fixed length of relationships, and (2) the divisions do not generally follow the limitations of the sequence and result in segmentation in the context, which results in ineffective optimization [50].

In contrast to model-aligned sequences, transformer networks don’t handle input in a sequence-ordered way. Instead, it analyses the whole series of information and utilizes mechanisms for self-service to learn dependencies in the sequential data. Transformer-based models do indeed have the ability to describe complicated time-series data dynamics that are difficult for conventional sequence models such as RNNs [49].

For individual homes, energy consumption patterns are usually erratic due to many causes like weather and holidays. Therefore, the use of methodologies based solely on energy consumption data to forecast energy use is unreliable. Univariate time-series data analysis, such as household energy consumption prediction, is challenging even with deep learning models [41]. Thus, integrating other observations (whether the observed point is an anomaly, change point, or part of the pattern) may help improve the prediction performance [51].

The similarities between the different data encoding variables in transformers (e.g., queries and keys) are calculated based on their point-specific values without explicitly considering local contexts [52]. This weakness could be addressed either by introducing new attention algorithms replacing the classical self-attention of the original transformer, e.g., Spring Time Warping Matrix [52], or providing more information about the surroundings of the observed point to the transformer externally [41]. This later approach forms the basis of the proposed method in this paper. In this paper, we propose to use SWT as an efficient pre-processing technique that decomposes a given signal into sub-signals with high and low frequencies, which offers an efficient
representation of the signal’s content and behavior. Next, we use transformer networks to predict SWT sub-bands. Hence, the novelty of this work is in the development of a hybrid approach for household energy consumption forecasting based on SWT and transformers [48].

Our contributions are in particular as follows:

- We developed a hybrid SWT-Transformer model for household power consumption time-series forecasting. The developed transformer model forecasts the features produced by the SWT. This combination helps tackle the problem of irregular patterns in the univariate household energy data. To the best of our knowledge, this is the first time SWT and transformers are combined to develop an efficient energy consumption predictive model.
- Experimental results, based on several energy consumption datasets from real-world households, show that our proposed SWT-Transformer approach can accurately forecast household energy usage, achieving superior prediction performance compared to existing methods.

II. THE PROPOSED MODEL

We propose a hybrid approach based on the stationary wavelet transform and deep transformers for forecasting household energy consumption. First, the initial univariate energy input data is decomposed into sub-bands using the SWT to extract the local trends and patterns. Second, the deep transformer is adopted to forecast the next wavelet sub-band. Finally, the inverse SWT is applied to the deep transformer outputs to reconstruct the predicted household energy consumption. The overall proposed method is summarized in Fig. 1a.

A. DATA DESCRIPTION

We use the open-source energy consumption data in five separate family homes in London, UK (UK), under the project name ‘UK-DALE’ [53], to test the validity and strength of the proposed model. To have a fair comparison with other existing models, we used the same strategy and data used in [29], [41], and [54]. They combined multiple entries of the original data collected in 6-seconds intervals and converted it to a dataset with a time interval of 5 minutes.

B. TRANSFORMER-SWT MODEL

Our hybrid energy consumption Transformer-SWT model follows the original Transformer architecture [48], which consists of time to vector, encoder, and decoder layers.
The transformer is a deep learning architecture that exclusively employs attention mechanisms for sequence-based data processing. Therefore, it does not utilize recurrent and convolutional layers that are widely used in sequence modeling. Instead, it maintains an encoder design and employs stacked multi-head self-attention and fully connected layers. Each encoder layer includes a multi-head self-attention layer followed by two feedforward layers. Both multi-head attention and feedforward layers are followed by dropout and Add&Normlize layers.

The encoder consists of sub-encoders that handle the input of each layer sequentially, while the decoder includes layers that do the same with the output of the encoder. Each encoder layer aims to create encodings of critical information on which sections of the inputs are relevant to each other. The encodings are sent to the next encoder layer. Every decoder layer does the reverse, takes all the encodings, and uses them to produce a series of outputs.

To this end, attention is used in each encoder and decoder layer. For each input, attention measures and calls attention to the pertinence of each input. The decoder layer is similar to the encoder layer but uses one feedforward layer rather than two. Encoder and decoder layers have feedforward networks for further output processing and have residual connections and layer normalization processes (Fig. 1b).

Each multi-head attention has three learnable weights, the query weights Q, the key weights K, and the value weights V [48]. Each attention head extracts a layer of ‘relevance’ between input parameters.

In more detail, the multi-head attention module of the transformer performs its calculations in parallel. The attention module performs an attention mechanism several times in parallel (Fig. 2). The independent attention outputs are then concatenated and linearly transformed into the desired dimension. Multi-head attention allows the transformer to encode many associations and subtleties for each input variable. A single attention module output is given by [48]:

$$\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V$$  \hspace{1cm} (1)

where $d_k$ is the dimension of query and key vectors.

The multi-head attention score is the concatenation of the output of $h$ heads given by Eq. (1) multiplied with a learnable projection parameters $W$, i.e.:

$$\text{MultiheadAtt.} = \text{Concat} (\text{Att.}_1, \ldots, \text{Att.}_h) W$$  \hspace{1cm} (2)

The number of parallel attention layers used in the proposed model is $h = 12$.

The Time2Vec [55], [56] is a learnable layer that is an extended version of the original positional encoding of the transformer. It allows learning the input frequencies rather than using a fixed representation. We use this layer because it is invariant to time rescaling and can capture periodic and non-periodic patterns of the input signal (Fig. 1c). The Time2Vec operation implements the following equation [56]:

$$\text{Time2Vec}(\tau)[i] = \begin{cases} \omega_i \tau + \phi_i, & \text{if } i = 0 \\ \Im(\omega_i \tau + \phi_i), & \text{if } 1 \leq i \leq k \end{cases}$$  \hspace{1cm} (3)

where, Time2Vec$(\tau)[i]$ is the $i^{th}$ element of Time2Vec$(\tau)$ that has $k$ elements, $\Im$ is a periodic function, and $\omega_i$ and $\phi_i$ are learnable parameters.

C. THE STATIONARY WAVELET TRANSFORM

The stationary wavelet transform (SWT) is a wavelet transform algorithm proposed by Nason and Silverman [57] to solve the shift-invariance and the non-redundancy issues in the discrete wavelet transform [58]. SWT does not decimate the initial signal. Instead, it changes the filters at each stage by padding zeroes instead of utilizing the down-sampling technique after implementing the low-pass or high-pass filters on the signal [59]. The SWT sub-signals from the decomposition has the same length as the initial signal, which creates an appealing function compared to traditional wavelets. This
feature makes SWT an optimal choice for data used in neural networks and allows for more accurate knowledge of the corresponding approximation and detail coefficients. SWT also demonstrated low-cost computing [60]. We, therefore, adopted SWT to analyze the energy consumption time-series data, produce distinguishable low- and high-frequency components, called approximations and details, and then provide such components as an input to the transformer.

The SWT approximation sub-band reflects the general trend of the time-series, while the detail sub-band indicates...
minor series variations. The SWT breaks down the time series with a hierarchical combination of low-pass and high-pass wavelet filters, enabling the separation of high and low frequencies.

The decomposition is seen as a dyadic tree shape [61]. Fig. 3 gives an illustration of one-dimensional signal decomposition at n-level using SWT. For a given signal $u(t)$ of length $N$, SWT decomposes $u(t)$ into two coefficients: approximation $A_1(t)$ and detail $D_1(t)$. Besides, the approximation coefficients $A_1(t)$ are split down into two pieces using up-sampled low and high-pass filters.

This procedure is repeated until achieving the $n$th decomposition level. We tested the different decomposition levels and wavelet families. We experimented with different wavelet families and different decomposition levels, and our experiments demonstrate that Daubechies (db2) with three levels show the best results, so we use them in all experiments.

D. MODEL EVALUATION CRITERIA

The mean squared error (MSE), the root mean squared error (RMSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE) are chosen as models evaluation metrics. They are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (y_k - \hat{y}_k)^2}$$  \hspace{1cm} (4)

$$\text{MAE} = \frac{1}{N} \sum_{k=1}^{N} |y_k - \hat{y}_k|$$  \hspace{1cm} (5)

$$\text{MAPE} = \frac{100\%}{N} \sum_{k=1}^{N} \left| \frac{y_k - \hat{y}_k}{y_k} \right|$$  \hspace{1cm} (6)

where $y_k$ is the $k$th sample value in $y$, $\hat{y}_k$ is the $k$th forecasted value, and $N$ is the total number of samples.

E. TRAINING

The deep transformer is used to forecast the SWT sublevels signals based on the historical sublevels. Our experiments use twelve lags from the SWT decomposed signal to forecast the next SWT sublevels. Let us take $U$ the SWT n decomposition of the energy consumption time series $U = [A_n D_1 D_2 \ldots D_n]$. The Transformer input is fed with historical time series of decomposed household energy consumption $(U(t - 11), U(t - 10), \ldots, U(t))$ to forecast the next SWT levels $U(t + 1)$, which can be described as

$$\hat{y}(t + 1) = [\hat{A}_n(t + 1) \hat{D}_1(t + 1) \hat{D}_2(t + 1) \cdots \hat{D}_n(t + 1)]$$

In our case, the goal is to forecast one step, and this does not require feeding the decoder with predicted output. Therefore, we ditch the decoder part altogether.

We used the RMSProp optimizer [14], as a learning algorithm to train our model using the parameters: $\eta = 0.001$ and $\beta = 0.999$. The model is regularized using a dropout rate of 0.1 for each sub-layer in the encoder layer, which contains
a multi-attention sub-layer, a feedforward sub-layer, and a normalization sub-layer. The number of query weights Q, key weights K, and value weights V used is 256.

III. RESULTS AND DISCUSSIONS

This study aims to forecast household energy consumption for several time scales. The same problem, i.e., energy consumption prediction, has been addressed in several recent studies [29], [41], allowing comparison with the literature performance. We use the same data as in [54] that consists of five separate houses datasets collected by the UK-DALE project for the whole year of 2015 [55]. From the whole dataset (comprising 36000 samples from each house), we use two-thirds of the samples for training and the remaining one-third of the samples for validation. As described in section II.C, each energy consumption data sample is decomposed into three levels, producing three approximation and three detail sub-signals. The deep transformer is then used to forecast the coefficient representing the next predicted sample from SWT coefficients representing the previous 12 samples, i.e., the approximation and details. Finally, we reconstruct the signal using the inverse SWT to compute the forecasted household energy consumption. The results of the proposed hybrid prediction model are compared with the following state-of-art methods: the persistent method [14], ARIMA [63], the multilayer perceptron (MLP) network [64], SVM [65], LSTM [59], CNN-LSTM [29], the hybrid SWT-LSTM [41], and the deep transformer [49].

The proposed architecture has been implemented in Tensorflow with Keras backend [67], [68]. The coefficients of SWT approximation and detail sub-bands were standardized to have zero arithmetic mean and standard deviation of 1. This pre-processing step helps speed up the training of deep neural networks.

The prediction results for the time steps of 5 minutes, 10 minutes, 20 minutes, and 30 minutes are presented in tables 1, 2, 3, and 4, respectively. Table 1 shows the obtained results for the case of 5 minutes step. We can see that all other strategies underperform our model and for all metrics. According to the data provider, houses 1 and 3 are relatively more active, whereas houses 2, 4, and 5 are less volatile [41]. We show a selected timestamp from the validation part houses 1 and 2, representing both categories, using our model and deep transformer in Fig. 4.

One can see that our model forecasts better energy consumption than the transformer-based model solely without SWT, which predicts the global and the local features of energy consumption. Fig. 3 demonstrates that the proposed approach well forecasts the irregular energy consumption pattern in the case of 5 minutes forecasting. Fig. 4 presents the bar graph of the four best models of Table 1, which are in decreasing order, CNN-LSTM, deep transformer, LSTM-SWT, and transformer-SWT. Our model improves the average RMSE, MAE, and MAPE values by 48%, 47% and, 51%, respectively, compared to the LSTM-SWT model [41] that produces the nearest results.

FIGURE 6. Selected portion from the validation part for two houses to show the performance of the proposed method for forecasting 10 minutes step size.

FIGURE 7. Bar graph of the four best results of Table 2.
Results for forecasting 10 minutes timestep are given in Table 2, in which we can see that our model outperforms all the other models by at least 59%, 56%, and 53% in RMSE, MAE, and MAPE, respectively. The performance of the proposed model is shown in Fig. 6 and Fig. 7. In Fig. 6, we plot the actual energy consumption and the two forecasted outputs based on our model and the deep transformer that present the nearest result we have run, in this case. Fig. 7 shows the bar graph of the four best results in Table 2: MLP, deep transformer, LSTM-SWT, and our model.

The results for the cases of forecasting 20 and 30 minutes times steps are presented in tables 3 and 4. Again, we can see that our model achieves superior prediction performance of total energy consumption.

Fig. 8 shows comparison results between the proposed model and the transformer without SWT for the prediction of 20 minutes step size, in which we can see the improvements provided by using the SWT. We have to note that, in the cases of 20 and 30 minutes steps, the transformer alone performs better than the hybrid models of CNN-LSTM [30], as shown in the bar graph of Fig. 9.

Our model improves forecasting quality by 48%, 38%, and 40% in RMSE, MAE, and MAPE values compared to the LSTM-SWT model, which is the hybrid model that presents the nearest results for the case of 20 minutes time step.

For the case of the forecasting 30 minutes step, our model improves the RMSE, MAE, and MAPE values by 65 %, 57%, and 38%, respectively.

IV. ANALYSIS
In this section, we study the robustness and performance of the proposed model in different situations that could happen in real-life situations, like noise, magnitude, and dips disturbances. House 1 is the sample used in this analysis.

We first injected noisy Gaussian signal with different values of standard deviation $\sigma$ to a signal from the testing dataset $S$ according to the following equation

$$S_n = S + \sigma \cdot \text{std}(S) \cdot \text{Gaussian}(0, 1)$$

where $\text{std}(S)$ is the standard of deviation of the testing signal, $S_n$ is the noisy signal, and Gaussian $(0,1)$ is a Gaussian signal with zero mean and unit standard deviation.

Forecasting results reported in Table 5 shows that the performance of the model understandably decreases with higher level of noise, but maintains a robust energy consumption prediction performance under very high levels of noise (e.g. $\sigma = 2$ and $\sigma = 3$).

Next, we studied the performance of the developed model after injecting constant magnitude disturbances with different durations, as depicted in Fig. 10. The injection was done in both parts of the day (night and day). First, we applied...
a disturbance with a large magnitude during the night for more than 4 hours of duration. Second, we injected a low-
high-low magnitude sequence of disturbances during the day to see the model’s response to sudden electricity changes.

One can observe that the model reacts to these disturbances and can adequately forecast them even if the injected signals are significantly different from usual energy consumption data.
Figure 10. (a) Forecasting model without disturbance, (b) Model reaction to constant disturbances with different durations and magnitudes Four hours of high magnitude, 1.5 hours of low magnitude, 1.5 hours of high magnitude, and 2.5 hours of low magnitude.

Table 5. Robustness of the proposed model checking for house 1.

| Time Resolution | Metrics | $\sigma = 0.1$ | $\sigma = 0.25$ | $\sigma = 0.5$ | $\sigma = 1$ | $\sigma = 2$ | $\sigma = 3$ |
|-----------------|---------|----------------|----------------|----------------|----------------|----------------|----------------|
| 5 min           | RMSE    | 0.0030         | 0.0039         | 0.0057         | 0.0086         | 0.0143         | 0.0258         |
|                 | MAE     | 0.0018         | 0.0028         | 0.0043         | 0.0068         | 0.0110         | 0.0163         |
| 10 min          | RMSE    | 0.0024         | 0.0032         | 0.0047         | 0.0070         | 0.0163         | 0.0183         |
|                 | MAE     | 0.0015         | 0.0023         | 0.0035         | 0.0056         | 0.0096         | 0.0128         |

V. LIMITATIONS

Despite the proposed model’s effectiveness, we still have two limitations. First, it is a learning-based system and can fail when faced with unknown circumstances. One possible way to alleviate this issue is to dynamically update the model with new training data to increase the size and variability of input data. The second limitation is that the decomposition method expects a regularly spaced signal, and this makes signal reconstruction difficult in multistep prediction problems. The use of the recursive predicted output in the SWT reconstruction may resolve this issue. Our future work will focus on resolving these two issues.

VI. CONCLUSION

In this study, we have proposed a hybrid predictive model based on SWT and deep transformers for reliable forecasting of residential energy consumption. Our model forecasts the local feature of the electrical energy consumption by using SWT and modeling the local tends through the deep transformer. Comparison with other existing machine learning models has shown the utility and superiority of the proposed model. For three significant factors, the benefit of using our model over other current models can be eligible.

- SWT can efficiently analyze the energy usage time-series data. Thus, each aspect, trend, or biomarker in the data can be captured more quickly and precisely.
- The deep transformer can predict different frequency levels by SWT of energy consumption rather than predicting the entire signal that includes all sub-frequency signals.
- Taking the benefit of our approach, the hybrid model can well catch the sophisticated features of energy usage and produce more precise forecasting performance for the four-time scales with an average improvement of more than 45 % in RMSE.

As a general conclusion, electric energy forecasting has important consequences for reliable power supply, effective operation, and electricity generation systems sustainability. The proposed strategy would incorporate a forecasting approach to reduce costs and control rising energy consumption. The proposed approach might also mitigate economic loss from unplanned activities of power plants.

REFERENCES

[1] M. Sajjad, Z. A. Khan, A. Ullah, T. Hussain, W. Ullah, M. Y. Lee, and S. W. Baik, “A novel CNN-GRU-based hybrid approach for short-term residential load forecasting,” IEEE Access, vol. 8, pp. 143759–143768, 2020, doi: 10.1109/ACCESS.2020.3009537.
[2] World Energy Outlook 2019, IEA, Paris, France, 2019. [Online]. Available: https://www.iea.org/reports/world-energy-outlook-2019
[3] P. Nejat, F. Jomehzadeh, M. M. Taheri, M. Gohari, and M. Z. A. Majid, “A global review of energy consumption, CO₂ emissions and policy in the residential sector (with an overview of the top ten CO₂ emitting countries),” Renew. Sustain. Energy Rev., vol. 43, pp. 843–862, Mar. 2015.
[4] T.-Y. Kim and S.-B. Cho, “Predicting residential energy consumption using CNN-LSTM neural networks,” Energy, vol. 182, pp. 72–81, Sep. 2019.
[5] C. Li, Z. Ding, D. Zhao, J. Yi, and G. Zhang, “Building energy consumption prediction: An extreme deep learning approach,” Energies, vol. 10, no. 10, p. 1525, Oct. 2017.
J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” 2018, arXiv:1810.04805.

S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang, and X. Yan, “Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting,” in Proc. Conf. Neural Inf. Process. Syst. (NeurIPS), 2019, pp. 5243–5253.

Y. Lin, I. Koprinska, and M. Rana, “SpringNet: Transformer and spring DTW for time series forecasting,” in Neural Information Processing (Lecture Notes in Computer Science), vol. 12534, H. Yang, K. Pasupa, A. C. S. Leung, J. T. Kwok, J. H. Chan, and I. King, Eds, Cham, Switzerland: Springer, 2020.

J. Kelly and W. Knottenbelt, “The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes,” Sci. Data, vol. 2, no. 1, Mar. 2015, Art. no. 150007.

K. Yan. CNN-LSTM Framework. Accessed: Aug. 3, 2021. [Online]. Available: http://keddiyan.com/files/PowerForecast.html

S. M. Kazemi, R. Goel, S. Eghbali, J. Ramanan, J. Sahota, S. Thakur, S. Wu, C. Smyth, P. Poupard, and M. Brubaker, “Time2Vec: Learning a vector representation of time,” 2019, arXiv:1907.05321.

Y. Shen, X. Jiang, Y. Wang, X. Jin, and X. Cheng, “Dynamic relation extraction with a learnable temporal encoding method,” in Proc. IEEE Int. Conf. Knowl. Graph (ICKG), Aug. 2020, pp. 235–242, doi: 10.1109/CKB50248.2020.00042.

G. P. Nason and B. W. Silverman, “The stationary wavelet transform and some statistical applications,” in Wavelets and Statistics (Lecture Notes in Statistics), vol. 103, A. Antoniadis and G. Oppenheim, Eds. New York, NY, USA: Springer, 1995, pp. 281–299.

U. B. Tayab, A. Zia, F. Yang, J. Lu, and M. Kashif, “Short-term load forecasting for microgrid energy management system using hybrid HHO-FNN model with best-basis stationary wavelet packet transform,” Energy, vol. 203, Jul. 2020, Art. no. 117857.

C. Yang, P. Liu, G. Yin, H. Jiang, and X. Li, “Defect detection in magnetic tile images based on stationary wavelet transform,” NDT & E Int., vol. 83, pp. 78–87, Oct. 2016.

R. R. Coifman and D. L. Donoho, “Translation-invariant denoising,” in Wavelets and Statistics (Lecture Notes in Statistics), vol. 103, New York, NY, USA: Springer-Verlag, 1995, pp. 125–150.

S. Supratid, T. Arisbarg, and S. Sapharadit, “An integration of stationary wavelet transform and nonlinear autoregressive neural network with exogenous input for baseline and future forecasting of reservoir inflow,” Water Resour. Manage., vol. 31, no. 12, pp. 4023–4043, Sep. 2017.

G. Hinton, “Neural networks for machine learning—Lecture 6a—Overview of mini-batch gradient descent,” Univ. Toronto, Toronto, ON, Canada, 2012. Accessed: Dec. 20, 2020. [Online]. Available: https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf

B. Nepal, M. Yahama, A. Yokoe, and T. Yamaji, “Electricity load forecasting using clustering and ARIMA model for energy management in buildings,” Jpn. Archit. Rev., vol. 3, no. 1, pp. 62–76, Jan. 2020.

J. Moon, S. Park, S. Rho, and E. Hwang, “A comparative analysis of artificial neural network architectures for building energy consumption forecasting,” Int. J. Distrib. Sensor Netw., vol. 15, no. 9, Sep. 2019, Art. no. 150147719887761.

Z. Ma, C. Ye, H. Li, and W. Ma, “Applying support vector machines to predict building energy consumption in China,” Energy Prog., vol. 152, pp. 780–786, Oct. 2018.

D. L. Marino, K. Amarasinghe, and M. Manic, “Building energy load forecasting using deep neural networks,” in Proc. 42nd Annu. Conf. IEEE Ind. Electron. Soc., Oct. 2016, pp. 7046–7051.

Keras: The Python Deep Learning Library, Accessed: Aug. 3, 2021, [Online]. Available: http://keras.io/

M. Abadi et al., “TensorFlow: A system for large-scale machine learning,” in Proc. 12th USENIX Conf. Oper. Syst. Design Implement., Savannah, GA, USA, Nov. 2016, pp. 265–283.