

Received December 15, 2021, accepted December 31, 2021, date of publication January 6, 2022, date of current version January 14, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3140818

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-Moghaddam¹.

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

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 Stationary wavelet transform (SWT) was combined with the ensemble LSTM to forecast energy consumption [41].

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 patterns) 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

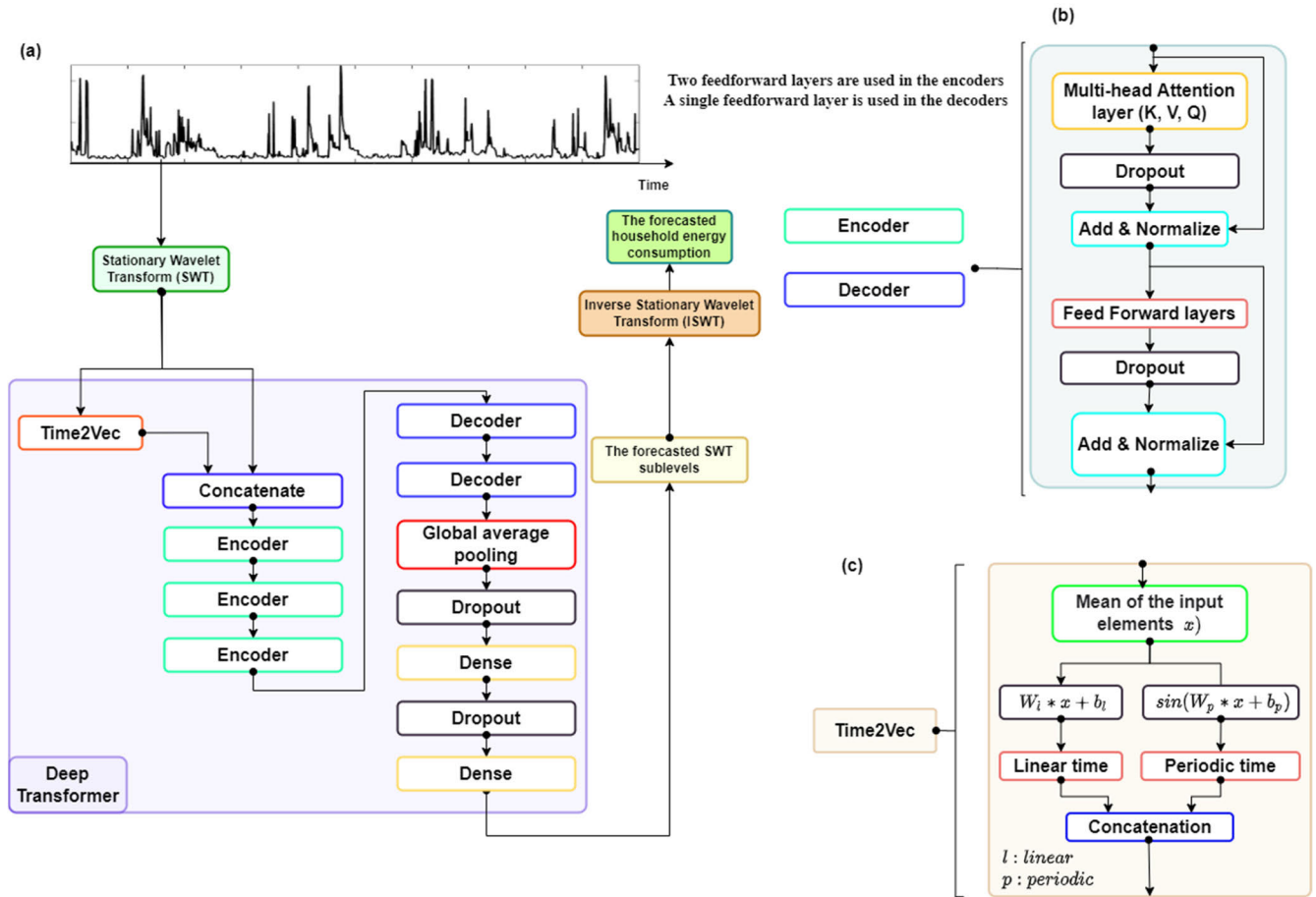


FIGURE 1. (a) The proposed deep transformer SWT model for the household power consumption forecasting, (b) the transformer encoder/decoder, (c) Time2Vec block.

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&Norm 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]:

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

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.:

$$\text{MultiheadAtt.} = \text{Concat}(Att_{.1}, \dots, Att_{.h})W \quad (2)$$

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

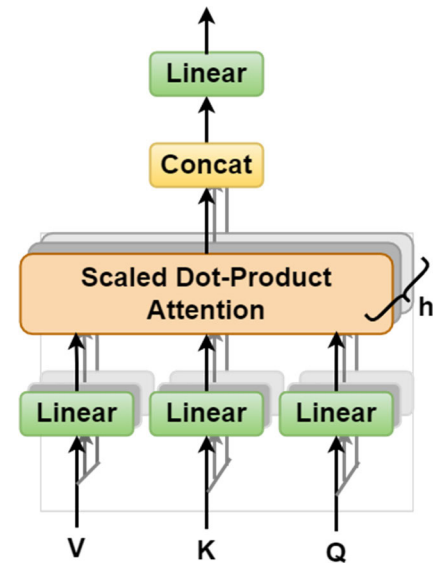


FIGURE 2. Multi-head attention. V, K, and Q are learned linear projections of input data, and h is the number of parallel attention layers [48].

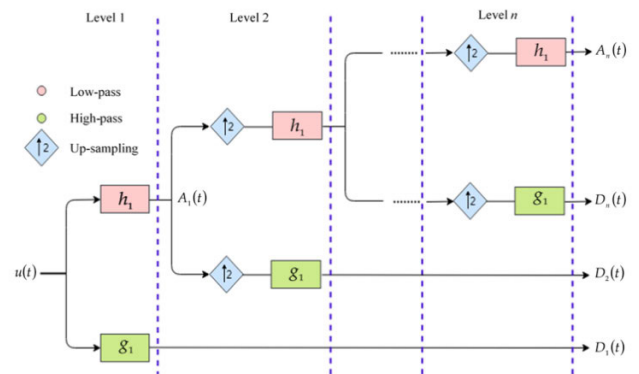


FIGURE 3. n levels decomposition with SWT.

Time2Vec operation implements the following equation [56]:

$$\text{Time2Vec}(\tau)[i] = \begin{cases} \omega_i \tau + \varphi_i, & \text{if } i = 0 \\ \mathfrak{S}(\omega_i \tau + \varphi_i), & \text{if } 1 \leq i \leq k \end{cases} \quad (3)$$

where, $\text{Time2Vec}(\tau)[i]$ is the i^{th} element of $\text{Time2Vec}(\tau)$ that has k elements, \mathfrak{S} is a periodic function, and ω_i and φ_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

TABLE 1. Comparison results of the proposed model with other existing machine learning models for forecasting household energy consumption in five-minute intervals.

Method	Metrics	House 1	House 2	House 3	House 4	House 5	Average
Persistent	RMSE	0.0224	0.0268	0.0329	0.0243	0.0260	0.0265
	MAE	0.0078	0.0082	0.0116	0.0089	0.0080	0.0089
	MAPE	18.340	23.926	18.637	22.266	9.1474	18.463
ARIMA	RMSE	0.0280	0.0199	0.0314	0.0317	0.0334	0.0289
	MAE	0.0114	0.0064	0.0122	0.0145	0.0125	0.0114
	MAPE	27.428	20.880	21.901	37.192	13.793	24.239
MLP	RMSE	0.0202	0.0246	0.0301	0.0228	0.0254	0.0246
	MAE	0.0080	0.0090	0.0122	0.0100	0.0099	0.0098
	MAPE	20.700	30.506	22.284	29.634	13.153	23.255
SVM	RMSE	0.0214	0.0226	0.0329	0.0232	0.0247	0.0250
	MAE	0.0077	0.0077	0.0142	0.0090	0.0086	0.0094
	MAPE	18.221	24.703	26.436	23.313	11.096	20.754
LSTM	RMSE	0.0204	0.0240	0.0301	0.0226	0.0251	0.0244
	MAE	0.0078	0.0087	0.0121	0.0092	0.0086	0.0093
	MAPE	18.985	27.241	20.735	23.914	9.9162	20.158
Deep Transformer	RMSE	0.0125	0.0138	0.0155	0.0141	0.0156	0.0143
	MAE	0.0045	0.0044	0.0078	0.0054	0.0056	0.0055
	MAPE	19.842	21.689	29.647	24.074	10.693	21.189
CNN-LSTM	RMSE	0.0203	0.0248	0.0117	0.0229	0.0254	0.0210
	MAE	0.0077	0.0088	0.0069	0.0094	0.0091	0.0084
	MAPE	18.516	25.830	8.5512	24.807	11.394	17.820
LSTM-SWT	RMSE	0.0064	0.0079	0.0101	0.0070	0.0073	0.0077
	MAE	0.0031	0.0037	0.0032	0.0035	0.0033	0.0034
	MAPE	10.227	15.501	7.3358	12.183	5.1682	10.083
Deep Transformer + SWT	RMSE	0.0027	0.0035	0.0070	0.0035	0.0031	0.0040
	MAE	0.0013	0.0016	0.0034	0.0016	0.0013	0.0018
	MAPE	4.1633	5.2307	7.7096	5.2307	2.0669	4.8802

TABLE 2. Comparison results of the proposed model with other existing machine learning model for forecasting household energy consumption in 10-minute intervals.

Method	Metrics	House 1	House 2	House 3	House 4	House 5	Average
Persistence	RMSE	0.0189	0.0283	0.0429	0.0244	0.0279	0.0285
	MAE	0.0074	0.0111	0.0181	0.0110	0.0103	0.0116
	MAPE	23.735	31.545	16.882	27.709	12.812	22.537
ARIMA	RMSE	0.0252	0.0267	0.0412	0.0204	0.0337	0.0294
	MAE	0.0114	0.0126	0.0192	0.0096	0.0167	0.0139
	MAPE	31.311	28.616	18.933	23.446	20.106	24.482
MLP	RMSE	0.0171	0.0251	0.0436	0.0226	0.0285	0.0274
	MAE	0.0072	0.0103	0.0221	0.0111	0.0126	0.0127
	MAPE	24.558	30.558	19.113	31.228	17.128	24.517
SVM	RMSE	0.0157	0.0224	0.0507	0.0231	0.0263	0.0276
	MAE	0.0064	0.0088	0.0231	0.0105	0.0105	0.0119
	MAPE	20.854	24.551	20.410	25.904	13.692	21.082
LSTM	RMSE	0.0167	0.0246	0.0420	0.0223	0.0245	0.0260
	MAE	0.0070	0.0103	0.0202	0.0106	0.0096	0.0115
	MAPE	23.628	30.366	18.231	27.866	12.287	22.476
Deep Transformer	RMSE	0.0082	0.0138	0.0275	0.0135	0.0146	0.0155
	MAE	0.0033	0.0044	0.0154	0.0059	0.0059	0.0070
	MAPE	22.412	21.689	24.280	28.212	13.422	22.003
LSTM-SWT	RMSE	0.0049	0.0068	0.0313	0.0070	0.0074	0.0115
	MAE	0.0026	0.0035	0.0145	0.0042	0.0040	0.0058
	MAPE	10.495	13.393	12.543	14.432	6.7097	11.515
CNN-LSTM	RMSE	0.0172	0.0244	0.0721	0.0259	0.0321	0.0343
	MAE	0.0073	0.0126	0.0342	0.0137	0.0161	0.0168
	MAPE	24.800	46.085	24.190	38.272	24.065	31.482
Deep Transformer + SWT	RMSE	0.0023	0.0034	0.0111	0.0034	0.0033	0.0047
	MAE	0.0013	0.0015	0.0061	0.0018	0.0016	0.0025
	MAPE	5.2669	6.1220	6.9477	6.0620	2.5819	5.3961

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

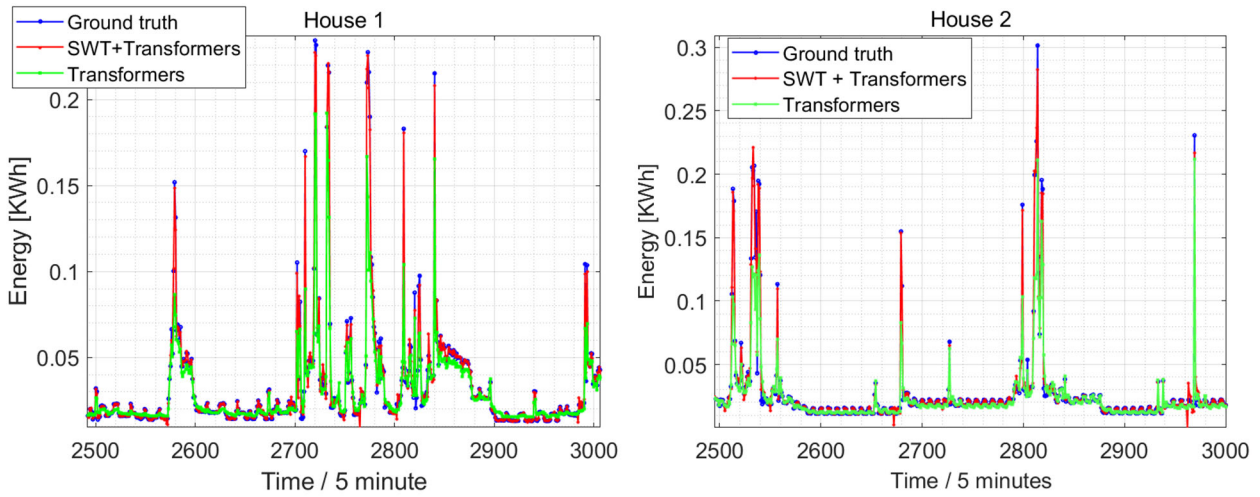


FIGURE 4. A selected portion from the validation part for two houses to show the performance of the proposed method for forecasting 5 minutes step size.

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:

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

$$MAE = \frac{1}{N} \sum_{k=1}^N |y_k - \hat{y}_k| \tag{5}$$

$$MAPE = \frac{100\%}{N} \sum_{k=1}^N \left| \frac{y_k - \hat{y}_k}{y_k} \right| \tag{6}$$

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.

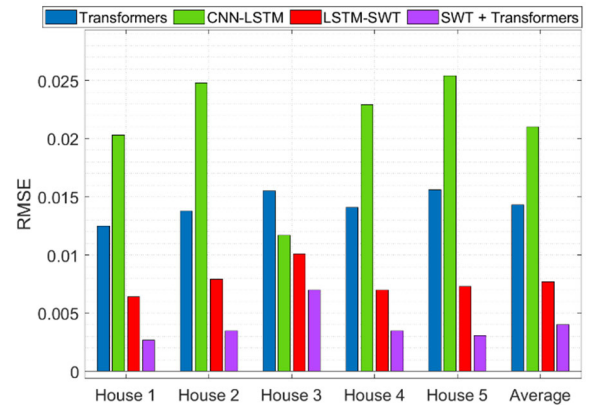


FIGURE 5. Bar graph of the four best results of Table 1.

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 $U = [A_n D_1 D_2 \dots D_n]$. The Transformer input is fed with historical time series of decomposed household energy consumption ($U(t - 11), U(t - 10), \dots U(t)$) to forecast the next SWT levels $U(t + 1)$, which can be described as

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

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

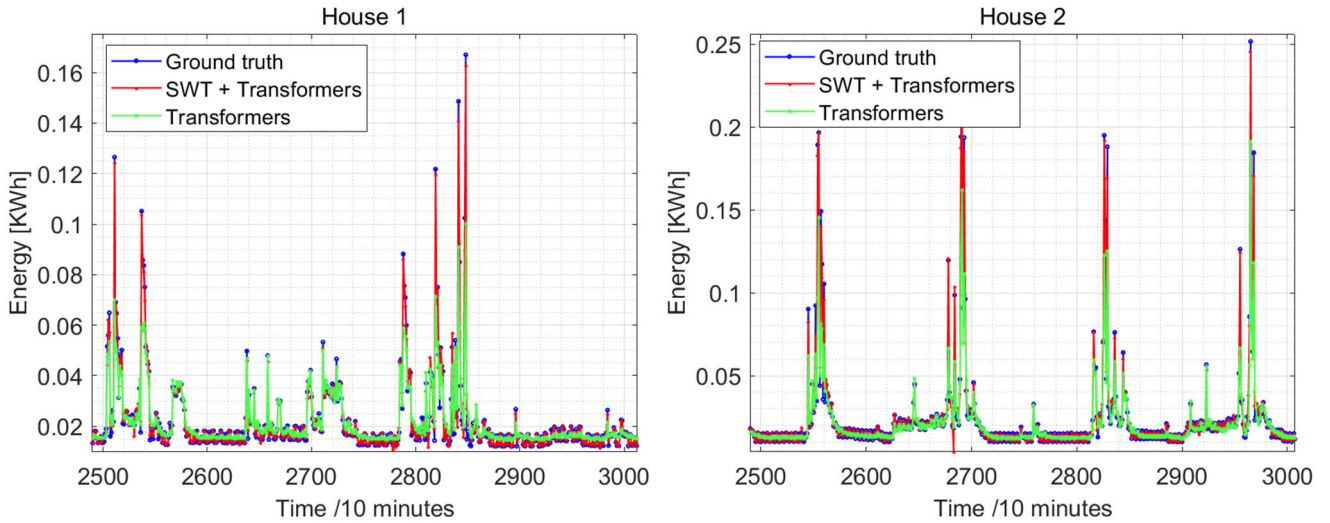


FIGURE 6. Selected portion from the validation part for two houses to show the performance of the proposed method for forecasting 10 minutes step size.

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

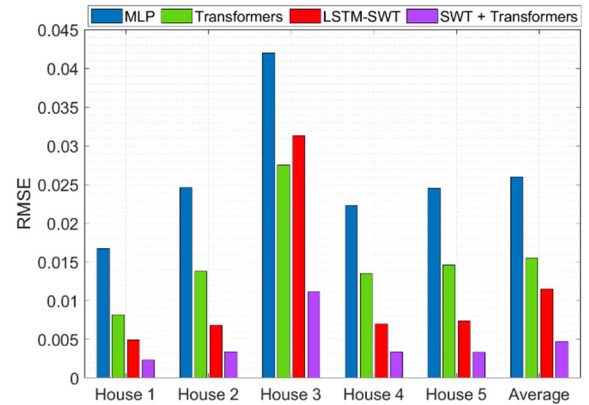


FIGURE 7. Bar graph of the four best results of Table 2.

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.

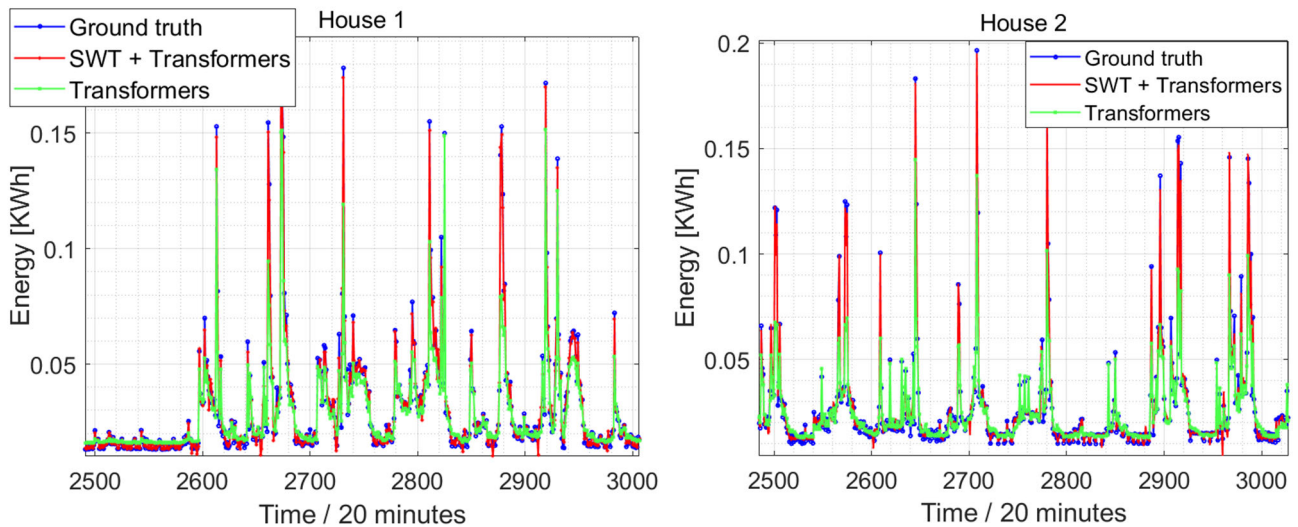


FIGURE 8. Selected portion from the validation part for two houses to show the performance of the proposed method for forecasting 20 minutes step size.

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.

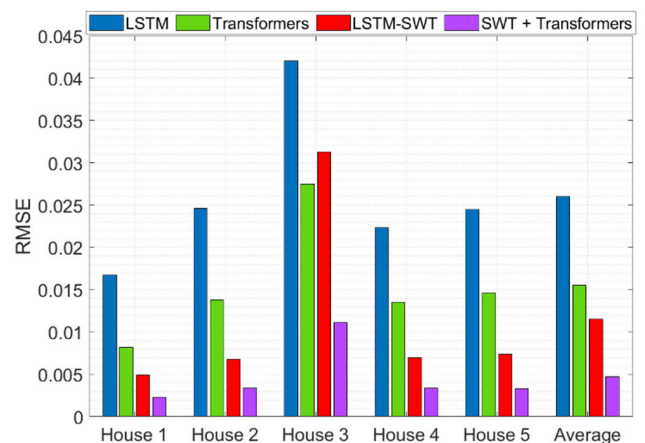


FIGURE 9. Bar graph of the four best results of Table 3.

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

$$S_n = S + \sigma \cdot std(S) \cdot \text{Gaussian}(0, 1) \tag{7}$$

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

TABLE 3. Comparison results of the proposed model with other existing machine learning models for forecasting household energy consumption in 20-minute intervals.

Method	Metrics	House 1	House 2	House 3	House 4	House 5	Average
Persistence	RMSE	0.0230	0.0238	0.0566	0.0241	0.0291	0.0313
	MAE	0.0100	0.0112	0.0293	0.0129	0.0125	0.0152
	MAPE	26.599	35.801	15.128	35.824	17.160	26.102
ARIMA	RMSE	0.0192	0.0233	0.0546	0.0046	0.0389	0.0281
	MAE	0.0100	0.0132	0.0307	0.0036	0.0207	0.0156
	MAPE	31.731	33.827	16.809	21.788	27.732	26.377
MLP	RMSE	0.0204	0.0204	0.0438	0.0221	0.0298	0.0273
	MAE	0.0090	0.0092	0.0257	0.0121	0.0137	0.0139
	MAPE	24.862	29.180	18.873	33.618	18.625	25.032
SVM	RMSE	0.0199	0.0203	0.0757	0.0222	0.0269	0.0330
	MAE	0.0083	0.0092	0.0447	0.0117	0.0122	0.0172
	MAPE	20.518	27.219	23.818	30.814	16.894	23.853
LSTM	RMSE	0.0203	0.0204	0.0538	0.0220	0.0288	0.0291
	MAE	0.0090	0.0102	0.0332	0.0123	0.0130	0.0155
	MAPE	25.054	36.959	20.007	35.257	17.723	27.000
Deep Transformer	RMSE	0.0103	0.0110	0.0328	0.0136	0.0142	0.0164
	MAE	0.0044	0.0043	0.0182	0.0063	0.0067	0.0080
	MAPE	25.105	29.742	19.057	32.435	17.947	24.857
LSTM-SWT	RMSE	0.0057	0.0060	0.0619	0.0069	0.0071	0.0175
	MAE	0.0033	0.0035	0.0304	0.0043	0.0043	0.0092
	MAPE	11.756	13.961	16.054	14.829	7.595	12.839
CNN-LSTM	RMSE	0.0232	0.0243	0.1278	0.0261	0.0307	0.0464
	MAE	0.0117	0.0128	0.0613	0.0147	0.0156	0.0232
	MAPE	34.485	47.182	30.854	41.484	24.013	35.604
Deep Transformer + SWT	RMSE	0.0027	0.0029	0.0325	0.0032	0.0035	0.0090
	MAE	0.0017	0.0018	0.0207	0.0022	0.0020	0.0057
	MAPE	6.0033	8.4809	12.556	8.0486	3.4718	7.7121

TABLE 4. Comparison results of the proposed model with other existing machine learning models for forecasting household energy consumption in 30-minutes intervals.

Method	Metrics	House 1	House 2	House 3	House 4	House 5	Average
Persistence	RMSE	0.0227	0.0222	0.0615	0.0254	0.0286	0.0321
	MAE	0.0113	0.0109	0.0348	0.0139	0.0134	0.0169
	MAPE	30.819	32.349	12.690	38.426	19.178	26.692
ARIMA	RMSE	0.0233	0.0215	0.0610	0.0205	0.0372	0.0327
	MAE	0.0123	0.0131	0.0366	0.0115	0.0225	0.0192
	MAPE	30.862	35.532	13.759	35.089	31.006	29.250
MLP	RMSE	0.0197	0.0196	0.0614	0.0231	0.0263	0.0300
	MAE	0.0097	0.0095	0.0397	0.0126	0.0128	0.0169
	MAPE	26.700	28.511	14.972	34.321	18.270	24.555
SVM	RMSE	0.0193	0.0189	0.0605	0.0227	0.0265	0.0296
	MAE	0.0092	0.0091	0.0354	0.0122	0.0131	0.0158
	MAPE	23.802	27.174	12.714	31.812	18.753	22.851
LSTM	RMSE	0.0197	0.0193	0.0599	0.0225	0.0253	0.0293
	MAE	0.0100	0.0099	0.0391	0.0131	0.0126	0.0169
	MAPE	28.286	31.202	15.504	38.607	18.457	26.411
Deep Transformer	RMSE	0.0100	0.0106	0.0428	0.0131	0.0146	0.0182
	MAE	0.0048	0.0046	0.0302	0.0065	0.0077	0.0108
	MAPE	27.943	31.824	17.766	34.329	22.724	26.917
LSTM-SWT	RMSE	0.0054	0.0053	0.1065	0.0068	0.0071	0.0262
	MAE	0.0035	0.0034	0.0547	0.0046	0.0046	0.0142
	MAPE	12.463	13.823	17.758	15.969	8.5050	13.704
CNN-LSTM	RMSE	0.0230	0.0230	0.2148	0.0269	0.0286	0.0633
	MAE	0.0126	0.0130	0.1074	0.0155	0.0154	0.0328
	MAPE	37.896	45.596	35.454	44.089	23.840	37.375
Deep Transformer + SWT	RMSE	0.0027	0.0029	0.0327	0.0034	0.0033	0.0090
	MAE	0.0018	0.0018	0.0221	0.0025	0.0023	0.0061
	MAPE	6.6681	7.5307	9.6102	9.1719	9.4017	8.4765

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

(inexistence or very high usage of electricity). 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.

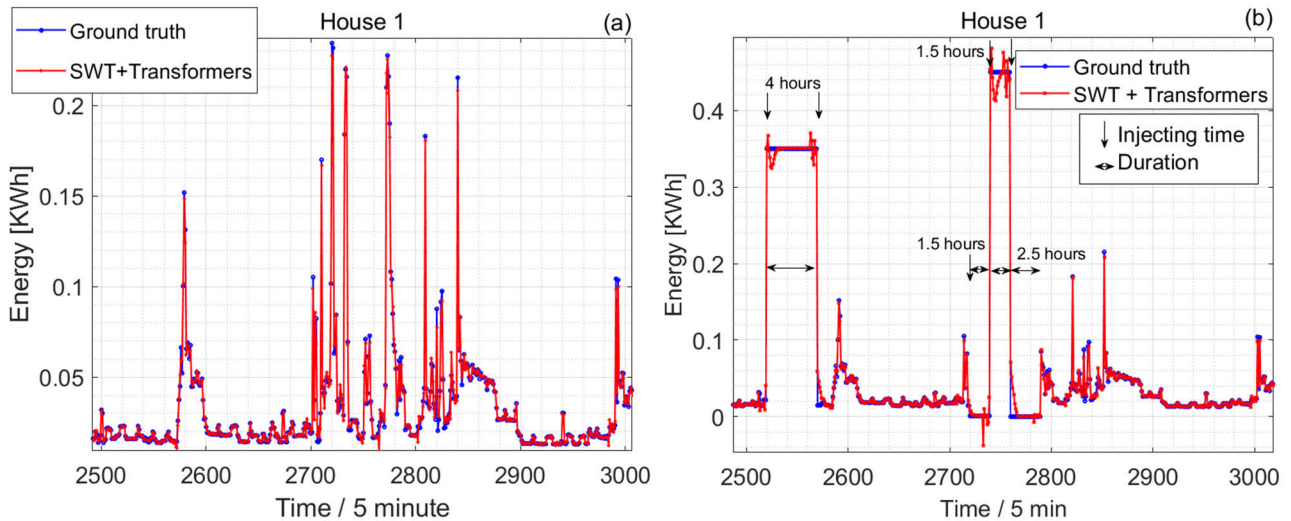


FIGURE 10. (a) Forecasting model without disturbance, (b) Model reaction to constant disturbances with different durations and magnitudes Four hours of high magnitude, 1.5 hours of low magnitude, 1.5 hours of high magnitude, and 2.5 hours of low magnitude.

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 trends 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.

- [6] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, "A review on time series forecasting techniques for building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 74, pp. 902–924, Jul. 2017.
- [7] J.-S. Chou and D.-S. Tran, "Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders," *Energy*, vol. 165, pp. 709–726, Dec. 2018.
- [8] N. Wei, C. Li, X. Peng, F. Zeng, and X. Lu, "Conventional models and artificial intelligence-based models for energy consumption forecasting: A review," *J. Petroleum Sci. Eng.*, vol. 181, Oct. 2019, Art. no. 106187.
- [9] V. G. Tran, V. Debusschere, and S. Bacha, "One week hourly electricity load forecasting using neuro-fuzzy and seasonal ARIMA models," *IFAC Proc. Volume*, vol. 45, no. 21, pp. 97–102, 2012.
- [10] M. I. Ahmad, "Seasonal decomposition of electricity consumption data," *Rev. Integr. Bus. Econ. Res.*, vol. 6, no. 4, pp. 271–275, 2017.
- [11] N. Fumo and M. A. R. Biswas, "Regression analysis for prediction of residential energy consumption," *Renew. Sustain. Energy Rev.*, vol. 47, pp. 332–343, Jul. 2015.
- [12] K. P. Amber, M. W. Aslam, and S. K. Hussain, "Electricity consumption forecasting models for administration buildings of the UK higher education sector," *Energy Buildings*, vol. 90, pp. 127–136, Mar. 2015.
- [13] E. M. de Oliveira and F. L. C. Oliveira, "Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods," *Energy*, vol. 144, pp. 776–788, Feb. 2018.
- [14] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor, "Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy," *Appl. Energy*, vol. 123, pp. 168–178, Jun. 2014.
- [15] Y. Yaslan and B. Bican, "Empirical mode decomposition based denoising method with support vector regression for time series prediction: A case study for electricity load forecasting," *Measurement*, vol. 103, pp. 52–61, Jun. 2017.
- [16] R. Rueda, M. P. Cuéllar, M. C. Pegalajar, and M. Delgado, "Straight line programs for energy consumption modelling," *Appl. Soft Comput.*, vol. 80, pp. 310–328, Jul. 2019.
- [17] Y. Liu, H. Chen, L. Zhang, X. Wu, and X.-J. Wang, "Energy consumption prediction and diagnosis of public buildings based on support vector machine learning: A case study in China," *J. Cleaner Prod.*, vol. 272, Nov. 2020, Art. no. 122542.
- [18] Y. Chen, P. Xu, Y. Chu, W. Li, Y. Wu, L. Ni, Y. Bao, and K. Wang, "Short-term electrical load forecasting using the support vector regression (SVR) model to calculate the demand response baseline for office buildings," *Appl. Energy*, vol. 195, pp. 659–670, Jun. 2017.
- [19] A. Bogomolov, B. Lepri, R. Larcher, F. Antonelli, F. Pianesi, and A. Pentland, "Energy consumption prediction using people dynamics derived from cellular network data," *EPJ Data Sci.*, vol. 5, no. 1, pp. 1–13, Dec. 2016.
- [20] A. Khan, H. Chiroma, M. Imran, A. Khan, J. I. Bangash, M. Asim, M. F. Hamza, and H. Aljuaid, "Forecasting electricity consumption based on machine learning to improve performance: A case study for the organization of petroleum exporting countries (OPEC)," *Comput. Electr. Eng.*, vol. 86, Sep. 2020, Art. no. 106737.
- [21] T. Pinto, I. Praça, Z. Vale, and J. Silva, "Ensemble learning for electricity consumption forecasting in office buildings," *Neurocomputing*, vol. 423, pp. 747–755, Jan. 2021.
- [22] A. Rahman, V. Srikanth, and A. D. Smith, "Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks," *Appl. Energy*, vol. 212, pp. 372–385, Feb. 2018.
- [23] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—A novel pooling deep RNN," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Sep. 2018.
- [24] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.
- [25] J. Q. Wang, Y. Du, and J. Wang, "LSTM based long-term energy consumption prediction with periodicity," *Energy*, vol. 197, Apr. 2020, Art. no. 117197.
- [26] H. Hu, L. Wang, L. Peng, and Y.-R. Zeng, "Effective energy consumption forecasting using enhanced bagged echo state network," *Energy*, vol. 193, Feb. 2020, Art. no. 116778.
- [27] M. Fayaz and D. Kim, "A prediction methodology of energy consumption based on deep extreme learning machine and comparative analysis in residential buildings," *Electronics*, vol. 7, no. 10, p. 222, Sep. 2018.
- [28] Y. Gao, Y. Ruan, C. Fang, and S. Yin, "Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data," *Energy Buildings*, vol. 223, Sep. 2020, Art. no. 110156.
- [29] G.-F. Fan, X. Wei, Y.-T. Li, and W.-C. Hong, "Forecasting electricity consumption using a novel hybrid model," *Sustain. Cities Soc.*, vol. 61, Oct. 2020, Art. no. 102320.
- [30] J. Xiao, Y. Li, L. Xie, D. Liu, and J. Huang, "A hybrid model based on selective ensemble for energy consumption forecasting in China," *Energy*, vol. 159, pp. 534–546, Sep. 2018.
- [31] K. Yan, X. Wang, Y. Du, N. Jin, H. Huang, and H. Zhou, "Multi-step short-term power consumption forecasting with a hybrid deep learning strategy," *Energies*, vol. 11, no. 11, p. 3089, Nov. 2018.
- [32] N. Somu, M. R. G. Raman, and K. Ramamritham, "A hybrid model for building energy consumption forecasting using long short term memory networks," *Appl. Energy*, vol. 261, Mar. 2020, Art. no. 114131.
- [33] H. Hu, L. Wang, and S.-X. Lv, "Forecasting energy consumption and wind power generation using deep echo state network," *Renew. Energy*, vol. 154, pp. 598–613, Jul. 2020.
- [34] C. Xia and Z. Wang, "Drivers analysis and empirical mode decomposition based forecasting of energy consumption structure," *J. Cleaner Prod.*, vol. 254, May 2020, Art. no. 120107.
- [35] F. Kaytez, "A hybrid approach based on autoregressive integrated moving average and least-square support vector machine for long-term forecasting of net electricity consumption," *Energy*, vol. 197, Apr. 2020, Art. no. 117200.
- [36] T. Liu, Z. Tan, C. Xu, H. Chen, and Z. Li, "Study on deep reinforcement learning techniques for building energy consumption forecasting," *Energy Buildings*, vol. 208, Feb. 2020, Art. no. 109675.
- [37] G. Zhang, C. Tian, C. Li, J. J. Zhang, and W. Zuo, "Accurate forecasting of building energy consumption via a novel ensemble deep learning method considering the cyclic feature," *Energy*, vol. 201, Jun. 2020, Art. no. 117531.
- [38] M.-R. Kazemzadeh, A. Amjadian, and T. Amraee, "A hybrid data mining driven algorithm for long term electric peak load and energy demand forecasting," *Energy*, vol. 204, Aug. 2020, Art. no. 117948.
- [39] H. Lu, F. Cheng, X. Ma, and G. Hu, "Short-term prediction of building energy consumption employing an improved extreme gradient boosting model: A case study of an intake tower," *Energy*, vol. 203, Jul. 2020, Art. no. 117756.
- [40] F. U. M. Ullah, A. Ullah, I. U. Haq, S. Rho, and S. W. Baik, "Short-term prediction of residential power energy consumption via CNN and multi-layer bi-directional LSTM networks," *IEEE Access*, vol. 8, pp. 123369–123380, 2020.
- [41] K. Yan, W. Li, Z. Ji, M. Qi, and Y. Du, "A hybrid LSTM neural network for energy consumption forecasting of individual households," *IEEE Access*, vol. 7, pp. 157633–157642, 2019, doi: [10.1109/ACCESS.2019.2949065](https://doi.org/10.1109/ACCESS.2019.2949065).
- [42] N. Somu, M. R. G. Raman, and K. Ramamritham, "A deep learning framework for building energy consumption forecast," *Renew. Sustain. Energy Rev.*, vol. 137, Mar. 2021, Art. no. 110591.
- [43] C. Liu, B. Sun, C. Zhang, and F. Li, "A hybrid prediction model for residential electricity consumption using holt-winters and extreme learning machine," *Appl. Energy*, vol. 275, Oct. 2020, Art. no. 115383.
- [44] X. J. Luo, L. O. Oyedele, A. O. Ajayi, O. O. Akinade, H. A. Owolabi, and A. Ahmed, "Feature extraction and genetic algorithm enhanced adaptive deep neural network for energy consumption prediction in buildings," *Renew. Sustain. Energy Rev.*, vol. 131, Oct. 2020, Art. no. 109980.
- [45] N. Somu, M. R. G. Raman, and K. Ramamritham, "A hybrid model for building energy consumption forecasting using long short term memory networks," *Appl. Energy*, vol. 261, Mar. 2020, Art. no. 114131.
- [46] R. Wang, S. Lu, and W. Feng, "A novel improved model for building energy consumption prediction based on model integration," *Appl. Energy*, vol. 262, Mar. 2020, Art. no. 114561.
- [47] F. Gao, H. Chi, and X. Shao, "Forecasting residential electricity consumption using a hybrid machine learning model with online search data," *Appl. Energy*, vol. 300, Oct. 2021, Art. no. 117393.
- [48] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. U. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2017, pp. 5998–6008.
- [49] N. Wu, B. Green, X. Ben, and S. O'Banion, "Deep transformer models for time series forecasting: The influenza prevalence case," 2020, *arXiv:2001.08317*.

- [50] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
- [51] 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.
- [52] 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.
- [53] 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.
- [54] K. Yan. *CNN-LSTM Framework*. Accessed: Aug. 3, 2021. [Online]. Available: <http://keddiyan.com/files/PowerForecast.html>
- [55] S. M. Kazemi, R. Goel, S. Eghbali, J. Ramanan, J. Sahota, S. Thakur, S. Wu, C. Smyth, P. Poupart, and M. Brubaker, "Time2Vec: Learning a vector representation of time," 2019, *arXiv:1907.05321*.
- [56] 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/ICKG50248.2020.00042](https://doi.org/10.1109/ICKG50248.2020.00042).
- [57] 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.
- [58] 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.
- [59] 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.
- [60] 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.
- [61] S. Supratid, T. Aribarg, and S. Supharatid, "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.
- [62] 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
- [63] B. Nepal, M. Yamaha, 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.
- [64] 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. 155014771987761.
- [65] Z. Ma, C. Ye, H. Li, and W. Ma, "Applying support vector machines to predict building energy consumption in China," *Energy Proc.*, vol. 152, pp. 780–786, Oct. 2018.
- [66] 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.
- [67] *Keras: The Python Deep Learning Library*. Accessed: Aug. 3, 2021. [Online]. Available: <http://keras.io/>
- [68] 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.



LYES SAAD SAOUD received the D.Sc. and M.C.A. degrees (empowerment to direct research) in physics, specialty in renewables energies, from the University of Boumerdès, Algeria, in January 2015 and February 2017, respectively. He was an Associate Professor with the University of Boumerdès. From February 2019 to October 2019, he was the Head and a Scientific Member of the Department of Electrical Systems Engineering and a Scientific Member of the Faculty of Technology, Boumerdès. From October 2019, he joined Khalifa University as a postdoctoral fellow in artificial intelligence and computer vision. He is currently an Engineer and a Magister in electronics. His recent research interests include advanced neural networks, fuzzy logic, hybrid systems, renewable energy, sustainability and the environment, energy engineering, and cybersecurity. He has served as a member of the Technical Committee for several international specialized conferences. He is a Reviewer of IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE ACCESS, *Transactions of the Institute of Measurement and Control* (SAGE Publications), *Journal of Renewable and Sustainable Energy* (American Institute of Physics), *Neurocomputing* (Elsevier), and *Neural Processing Letters* (Springer).



HASAN AL-MARZOUQI (Senior Member, IEEE) received the bachelor's (Hons.) and M.S. degrees in electrical and computer engineering from Vanderbilt University, Nashville, TN, USA, in 2004 and 2006, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, in 2014. He is currently an Assistant Professor with the Electrical and Computer Engineering Department, Khalifa University, Abu Dhabi, United Arab Emirates. His

research interests include machine learning, image and video processing, and digital rock physics. He is an Organizing Committee Member and the Challenge Sessions Co-Chair of ICIP 2020 in Abu Dhabi and a Technical Program Committee Member of EUVIP 2020 in Paris. He is an active reviewer of many international conferences and journals.



RAMY HUSSEIN received the Ph.D. degree in electrical and computer engineering from The University of British Columbia, Canada. He is currently a Research Scientist with the Center for Artificial Intelligence in Medicine & Imaging (AIMI), Stanford University. He is interested in developing and optimizing artificial intelligence solutions for the early diagnosis and prediction of cerebrovascular and neurodegenerative diseases, with more focus on ischemic stroke and alzheimer's disease. He combines computer vision with signal/image processing for better acquisition, detection, and also prediction of biomedical time-series data, CT scans, MRI, and also PET images.

• • •