Forecasting Energy Demand Using Conditional Random Field and Convolution Neural Network

Aravind Thangavel[1, *], Vijayakumar Govindaraj[2]

[1]Department of Computer Science and Engineering, Muthayammal Engineering College,
Rasipuram, Tamilnadu, India
[2]Department of Electrical Engineering, Muthayammal Engineering College,
Rasipuram, Tamilnadu, India
aravindphd6@gmail.com

Abstract—**Electric load forecasting has been identified as an effective strategy to increase output and revenues in electrical manufacturing and distribution organizations. Several strategies for forecasting power consumption have been suggested; however, they all fail to account for small variations in power demand throughout the prediction. Therefore, the aim of this study was to develop a CRF-based power consumption prediction technique (CRF-PCP) to meet the difficulty of estimating energy consumption (EC). The EC of regions in the area is forecasted using convolution neural networks (CNNs) and conditional random fields (CRFs). Then, using the cloud, the predicted results are delivered to the electricity distribution system. To our knowledge, this is the first attempt to forecast electricity demand using CNN and CRF algorithms. In comparison to state-of-the-art algorithms, this proposed technique achieves 98.9 % accuracy. This recommended work also obtained minimum values of root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and mean bias error (MBE) by using 10-fold cross-validation (CV) and a hold-out (CV) methodology.**

Index Terms—**Conditional random field; Convolutional neural network; Electricity consumption (EC) prediction; Machine learning; SCADA.**

I. INTRODUCTION

The final stage of delivering the electric power is electric power supply, and it is in charge of bringing electricity from the distribution system to individual customers. The "Smart Grid", a contemporary wave of innovative power grid technology, has arisen in recent years to control energy use in a financially sensible, dependable, and secure way. Smart grids are more efficient and productive electric networks that reduce the gap between demand and supply, allowing for eco-friendly power production and distribution. Forecasting electrical equipment and peak demand has become a crucial task and a significant part of the design and expansion of electrical power networks (Abera and Khedkar [1], Fallah, Ganjkhani, Shamshirband, and Chau [2]). Electrical resources are estimated using the load forecasting method that is required to fulfill the delivery in the short, medium, or long term.

Forecasting aids energy businesses in their operations, thereby allowing them to better manage distribution to their customers, and electric load forecasting is found to be a key aspect of the process that might enhance the profitability and the productivity of electrical production and distribution companies. This will help them efficiently manage their resources and operations in terms of efficiently providing energy to all their customers (Haq and Ni [3]; Mamun, Sohel, Mohammad, Haque Sunny, Dipta, and Hossain [4]; Deng Wang, Xu, Xu, Liu, and Zhu [5]; Kwon, Park, and Song [6]; Raju and Laxmi [7]). Predicting electricity consumption has numerous advantages as it encourages the utility provider to plan ahead of time so as to reduce the cost of power system as they are aware of what the anticipated usage or load demand is.

Understanding long-term demand would aid in the development of a business plan and the selection of financially viable choices on generation and transmission investments. It helps maximize the efficiency of power plants and prevents them from underproducing or overproducing. The load forecasting helps in strategic planning for possible capacity, location, and type of the generation facility. The utilities are more likely to produce the electricity close to the load upon locating the regions/areas with increasing or strong demands. It aids in the selection of resources, such as the fuels required to run generating facilities and other resources, to ensure reliable and cost-effective power generation and delivery to customers, and this is complicated for short, medium, and long terms. Machine learning is used to predict future data because of benefits like easy pattern detection from vast amounts of data and control of multi-variety and multi-dimensional data.

Owing to several constraints such as the weather condition and the dynamic behaviour of the occupant, state-of-the-art energy use prediction (ECP) techniques are limited in terms of anticipating the energy use correctly. (Ullah, Ullah, Haq, Rho, and Baik [8], Farsi, Amayri, Bouguila, and Eicker [9]). During recent years, several deep learning models for forecasting load have been introduced, but still only a very few of them have attained the state-of-the-art performance. Therefore, in this paper, a CRF-based power consumption prediction technique (CRF-PCP) has been proposed that accounts for these parameters while predicting the energy usage and training the machine learning models. (Alam, Zhou, Liew, Jia, Chanussot, and

Gao [10]). The accuracy of the suggested system's prediction has improved as a result of these considerations. The contributions of this work are as follows:

– To address the aforementioned challenges with predicting electricity consumption (EC), a CRF-based power consumption prediction approach (CRF-PCP) has been proposed;

– For training an artificial convolution neural network to predict electricity consumption (EC), a machine learning-based technique named "artificial convolution neural networks" has been used;

– The spatiotemporal texture map (STTM) is used to extract features for the training model;

– The accuracy of the proposed CRF-PCP technique is evaluated in terms of MSE, MAPE, MAE, RMSE, and MBE error metrics.

The following is how this paper is structured. Several related works on power distribution modelling and feature extraction are discussed in Section II. The technique for the proposed CRF-PCP is shown in Section III. Section IV contains a description of the efficiency of the suggested method. The conclusions are discussed in Section V.

II. LITERATURE SURVEY

A method for classifying hyperspectral images has been proposed by Alam, Zhou, Liew, Jia, Chanussot, and Gao [10]. It uses both spatial and spectral data in a combined CNN framework, and deep CRN with CNN-based binary and unary potential was calculated to derive a correlation between patches in the image. A deep deconvolution network is also used to improve the performance of the classification map.

Motepe, Hasan, and Stopforth [11] implemented a South African distribution network load forecasting technology using a new hybrid artificial intelligence (AI) and deep learning (DL) system. For predicting the load forecasting, a hybrid AI/DL load prediction technique was utilized. Also, the temperature effect on the recently developed models of the hybrid AI technique and the DL technique was studied and found that this work achieved MAE of 4.78 %, MAPE of 6.35 %, and RMSE of 6.33 %. Several weather parameters such as rain, wind, and humidity are not considered for prediction. To seperate power disturbance from normal values, Parvez, Aghili, Sarwat, Rahman, and Alam [12] proposed a one-class support vector machine (OCSVM), and wavelet filters are applied to detect disturbance. Large datasets are used for training this OCSVM as this model is capable of detecting disturbances in real-time. If disturbances are found, those will be classified into various types with an accuracy of 93 % by the classifiers.

Using historical data on power failures, Jaech, Zhang, Ostendorf, and Kirschen [13] solved the problem of estimating the duration of unplanned power outages to train a sequence of predictors for the neural network. The initial period projection would be based on environmental circumstances, and natural language processing would be usedto modify an automatic interpretation of the text depending on incoming field data. When optimising directly rather than using the gamma distribution, the RMSE results are only marginally improved. Other aspects of evaluation are not covered in this paper. To evaluate day-to-day habits of consumers and to increase the observe ability of delivery networks, Yuan, Dehghanpour, Bu, and Wang [14] proposed a technique with no smart meters (SMs). Three machine learning algorithms are used to extract power consumption patterns from consumers who use smart meters, as well as to forecast the consumption habits of unseen customers. With low error values of 19.47 %, this model was found to achieve great accuracy, and the factors that influence prediction are not taken into account in this study.

To discover anomalies in the smart grid, Fenza, Gallo, and Loia [15] focused on electricity usage. They used a long short-term memory network (LSTM) to predict customer behaviour based on previous purchases. This model uses continuous monitoring to differentiate between normal and irregular consumer behaviour, with precision and recall of 78 % and 88 %, respectively. However, there is a lag between the occurrence of the anomaly and its discovery. Chou, Hsu, Ngo, Lin, and Tsui [16] developed a hybrid prediction model to estimate the energy consumed by air conditioners in commercial areas such as offices one day ahead. For recording linear and nonlinear elements from the energy consumed by the air conditioners, both linear and nonlinear methods have been incorporated and the correlation coefficient R of the suggested system was found to be 0.71, while the total error rate found was 4.8 %. Here, in this proposed work, seasonal changes are not considered.

Kamarol, Jaward, Parkkinen, and Parthiban [17] developed a new spatiotemporal texture map (STTM) with minimal computing cost. This map is found to have the ability to capture smaller temporal and spatial variations in face emotions. To derive the dynamic characteristics and show them in histogram format, a block-based approach has been implemented, whereas the support vector machine will classify the features based on the facial emotions.

Zheng, Yang, Yang, Zhang, and Zhang [18] have proposed a two-stage training method both to alleviate the issue of overfitting and to optimize the boundary feature of the deep convolutional neural network (CNN). Although the proposed method achieved better performance results, the anomaly detection was found to be unreliable in high-dimensional space due to the distribution of sparse samples and noise. The two-stage training process will take a long time to complete and increase deep CNN accuracy; Zheng, Yang, Tian, Jiang, and Wang [19] developed a full-stage data augmentation approach that can improve the generalization capacity of deep learning models. According to an experiment finding, these models were found to have a small impact on the convergence ability of the model. Zheng, Tian, Yang, Wu, and Su [20] presented a new pruning method called "drop-path" to minimize the model parameters of a 2D deep CNN. The invariance of the probably approximately correct (PAC)-Bayesian boundary is a key aspect in ensuring the generalisation ability of deep CNN under the constraint of optimizing as much as possible. Eight well-known deep CNNs, including GoogLeNet, AlexNet, VGG-16, and ResNet-34/50/56/110,

all trained on ImageNet and CIFAR-10, produced state-of-the-art results with only a 2 % increase in speed and a 1 % increase in error.

Zheng, Zhao, Li, Wang, and Yang [21] have proposed a two-level data augmentation approach based on interference from the spectrum for deep learning to automatically categorize the modulation. To closely resemble the global input space, the short-time Fourier transform (STFT) and the inverse fast Fourier transform (IFFT) were used to facilitate signal expansion and introduce modifications while maintaining essential characteristics. Even though the data augmentation has increased the generalisation ability of deep learning models for modulation classification, several factors such as weight initialization and model structure have been found.

Zheng, Zhao, Zhang, and Wang [22] have proposed a manifold regularization-based deep convolutional autoencoder (MR-DCAE) model that influences the real performance of neural networks to identify unlicensed broadcast. Original samples are projected onto low-dimensional representations, which are then rebuilt into the input space using manifold invariance as a guarantee. Because MRDCAE did so well in the unlicensed broadcasting identification (UBI) job, it will need to be fine-tuned on a regular basis to stay up with the changing environment.

III. METHODOLOGY

Electricity demand forecasting plays a vital part in allocating the long-term planning and short-term loading process for a new generation and transmission structures, and the accurate prediction can be done exactly when cost and energy efficiencies are considered for better decisions. The goal of this research is to use a CRF-based power consumption prediction algorithm to forecast consumer electricity use (CRF-PCP). Power is allocated to areas within a region based on these expected values. This approach consists of two steps, and they are:

1. Historical data-based power distribution;
2. Predicted data-based power distribution.

A whole region is considered here and is divided into areas where consumers will consume electricity in each area. The power generated will be distributed from the power generator to the consumers according to historical data, and then the level of power consumption of the consumer will be measured using a machine learning-based technique. Power distribution to consumers will be done based on predicted data, and energy consumption data are analysed and maintained using supervisory control and data acquisition (SCADA).

The block diagram of the proposed work is shown in Fig. 1. It consists of a power generator, power distributor, secure cloud, and consumers and its working is described as follows:

1. Historical data are provided to the power distributor for the first time during power distribution;
2. The power distributor distributes power to the region's consumers;
3. The power distributor receives details about the power consumption;
4. The power distributor sends the gathered data to the SCADA via web services;
5. The data collected are stored in SCADA and used to train the CRF-PCP technique;
6. For the next estimate of electricity consumption, the learned data are supplied to the prediction module;
7. The expected data are sent to the power distributor, who then distributes the power to the region's users based on the predicted data. This aids in the distribution of the corresponding power to that specific region.
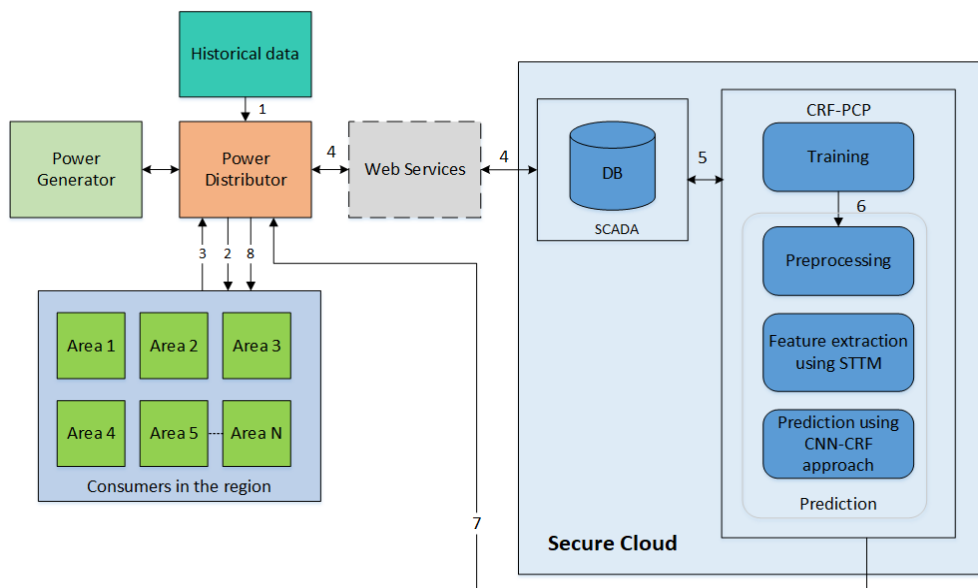

Fig. 1. Block diagram of the proposed work.

A. Dataset

The dataset [23] used in this proposed work is considered from smart meters in London, Kaggle, which comprises data on energy use for 5,567 London households who participated in the UK Power Networks-led Low Carbon London initiative between November 2011 and February 2014. The data from smart meters appears to be mainly related to energy use. During model training, factors that

affect power consumption, such as weather data and bank holidays, are taken into account. The weather factors are wind speed, wind pressure, cloud cover, high/low temperature, wind bearing, dew point, precipitation type, icon, visibility, UV index, humidity, time of sunset and sunrise.

B. Preliminaries

1. Power generator

Electricity generation is the process of generating electricity from the primary energy sources, and this is the stage before transmitting or distributing the power to consumers. For converting the motive power (mechanical power) into electrical energy, an external circuit is used for generating the electricity where the generated power will be distributed by the power generator to the power distributor. The details of power generation will be given to SCADA.

2. Power distributor

The power distributor initially provides power to consumers based on previous data. The power distributor then collects and sends data on consumer electricity usage to SCADA. Then, on the basis of this information, a prediction is made. Finally, a power distributor distributes power to users based on the expected data.

3. Secure cloud

The transport of information to the cloud required extremely safe protection. Cyberattacks are becoming more sophisticated, and cloud computing is no less vulnerable than on-premises technology. When you work with a cloud service, you get best-in-class encryption that has been optimised for this technology. Data on electricity use is predicted in a secure cloud environment.

4. SCADA

For monitoring and controlling data related to power consumption in the cloud, SCADA is used, in which data on power generation and consumption will be stored (Turc [24]). The majority of systems rely on database changes in real time. The SCADA server, which establishes the link between the SCADA applications and the technological system being monitored, is a critical component of the SCADA system. SCADA programmes were deployed as web servers, despite the fact that database processes were represented in structured query language (SQL). SCADA servers power SCADA Web services, which are connected to field equipment via Remote Terminal Units (RTUs) and SCADA clients. For communication between the SCADA server and the technical system, RTUs are used in which the collected data and real-time update present in database will be updated, which will be acting as a central source for the SCADA clients.

For monitoring and controlling data related to power usage, the SCADA system is considered in which the SCADA server acquires data from various components (power generator and power distributor) via RTUs. The real-time data will be present in database. SCADA clients (power generators and distributors) can send requests to web servers for data (pages). When a client views a page that includes service calls, data will flow continuously from the server to the client. Power producers and distributors use web services to access the database.

5. Machine learning-based CRF-PCP technique

To provide predicted data to the power distribution system, a machine learning-based prediction technique named "CRF-PCP" is proposed in this work, which will be used to distribute power to consumers.

6. Consumers

Residential, industrial, and commercial environments all need electricity, and these electricity consumers of electricity are the end customers.

C. Power Distribution Based on Historical Data

The power consumption data of areas are collected by the power distributor and maintained in the SCADA system. The power distribution network, which uses SCADA data, distributes power to all areas of the region prior to prediction. For the next time of power distribution, the output of the CRF-PCP technique (predicted power data) can be used.

D. CRF-based Power Consumption Prediction Technique (CRF-PCP)

In the CRF-PCP technique, the region is classified into areas as shown in Fig. 2. In this work, CNN-based deep CRF is as a machine learning technique used for training and prediction. CRF is a graphical model used for deriving spatiotemporal contextual information of each area.
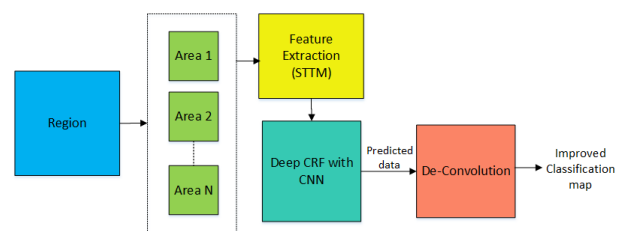

Fig. 2. Flow diagram of the CRF-PCP method.

The proposed technique of CRF-PCP consists of three steps: preprocessing, feature extraction, and CNN with deep CRF-based prediction.

1. Preprocessing

The data collected using SCADA are preprocessed to remove noise and to add extra information if needed. Raw data or real-world data remain imperfect and could not be sent through a system. This may create some errors, and hence, before submitting to a model, preprocessing should be completed.

2. Feature extraction

Features are extracted after preprocessing with low complexity using spatiotemporal texture map (STTM) technique (Kamarol, Jaward, Parkkinen, and Parthiban [17]). This has the ability to extract spatial and temporal variations of areas where spatial variation represents the region of power consumption and temporal variation implies the time of EC for that corresponding region.

The output predicted using machine learning in the last step will be in the form of spatiotemporal data. Hence, spatiotemporal features are extracted using STTM to train the machine learning model for prediction. In the STTM algorithm, the data are represented in linear scale form due to its simplicity and ease of understanding and measuring of the data, and it is shown in (1).

Then, the Gaussian kernel is computed as in (2). Then to extract features, H values are calculated, which μ needs value to be computed. Finally, block-based representation is used to obtain the feature vector. The linear scale-space representation LS was modelled for provided input i. Then, convolution is performed using kernel (matrix) to slide across data and to obtain the desired enhanced output.

The convolution performed for i and the 3D Gaussian kernel can be represented as

$$LS(.; \sigma_t^2, \tau_t^2) = gk(.; \sigma_t^2, \tau_t^2) * i(.). \quad (1)$$

In the above equation, temporal variance and spatial variance of the Gaussian kernel gk are denoted σ_t^2 as τ_t^2. The spatiotemporal Gaussian kernel is given as

$$gk\left(x, y, t; \sigma_t^2, \tau_t^2\right) = \frac{-\exp\left(-\left(x^2 + y^2\right)\right)/2\sigma_t^2 - t^2/2\tau_t^2}{\sqrt{(2\pi)^3 \sigma_t^4 \tau_t^2}}. \quad (2)$$

The spatial domain is represented by the x and y axes of the area in the region. The temporal domain is denoted by the time axis t.

The spatial domain is represented by the x and y axes of the area in the region. The temporal domain is denoted by the time axis t. When there are changes in the consumer's electricity consumption, the electricity data in each area will also be different. Hence, to spot the difference in spatiotemporal domain, convolution is performed between spatiotemporal second-moment matrix and Gaussian kernel function, which is denoted as μ, and from this resulting value of μ, Eigen values will be computed which are high indicating that there is a difference in the way of using the electricity.

As a result, expanded Harris corner functions for the spatiotemporal domain are developed to yield these variations, as illustrated in (3). The extended Harris corner function for the spatiotemporal domain can be defined as

$$H = \det(\mu) - c\,trace^3(\mu) = \lambda_1 \lambda_2 \lambda_3 - c(\lambda_1 + \lambda_2 + \lambda_3)^3, \quad (3)$$

where c denotes the constant value. The H function was standardised to remove variations in the area. The spatiotemporal points of the area were discovered by looking for local positive H maxima. However, employing these points does not produce accurate results, since a considerable difference between the points in the spatiotemporal domain is required. As a result, the H function is used as a texture map to depict the distribution of appearance in space and time. Then, to represent STTM features, each texture map was divided into multiple blocks, with a histogram produced for each block. Finally, all histograms are combined to create a feature vector. The energy consumption of each smart meter, as well as temperature, pressure, humidity, cloud cover, dew point, and wind speed, are among the features collected.

3. CNN with deep CRF-based prediction

The extracted features using STTM from the previous step will be used for predicting the nest power distribution,

and here, the regions are considered as an area groups. Using CNN, the model will be trained, and CNN-based deep CRF was proposed. Unary and pairwise potentials of the CRF can be computed using this CNN-based deep CRF technique to handle the spatiotemporal information along the whole area. Using the mean-field inference algorithm, a classification map was generated and finally the final classification performance was improved by the deconvolution network.

Deep CRF is integrated with spatiotemporal features to use the CNN and CRF properties and to describe spatial and temporal contextual dependencies between the areas through this method. This learning methodology is thought to be particularly appropriate for area research because this integrated model can better leverage the spatiotemporal correlations between area classes to conduct the final classification. Throughout the training phase, the CRF incorporates geographical and temporal contextual information, which is crucial for power distribution applications.

In the feature vector, the individual location denotes the spatial along with the temporal details. Voxels refer to the grid in 3D space, and the voxels in this function vector indicate these spatiotemporal positions. These voxels can be modelled using CNN-CRF, so it is suitable for region data processing. The initial feature vector is used on stacks of CNNs to train deep CRF parameters. In the network, without taking entire region, the inputs for the area groups are considered by CNN-CRF because local spatiotemporal features are represented by initial feature vector. Equations (4)–(8) are derived using (Alam, Zhou, Liew, Jia, Chanussot, and Gao [10]).

A node along a B-area within the feature vector corresponds to each voxel in a CRF network. The voxel labels are shown by the label L. Within nodes, edges are generated. Pairwise connections between neighbouring voxels in CRF are built by linking one node to all of its neighbours. The CRF can be defined as follows

$$P\left(l \mid v_{(d,\lambda)}; v_\lambda\right) = \frac{1}{Z(v_{(d,\lambda)})} \exp\left(-E\left(l, v_{(d,\lambda)}; \theta_\lambda\right)\right). \quad (4)$$

From the above equation, the learning parameters of the network θ of different wavelengths λ can be achieved. The energy function $E(l, v_{(d,\lambda)}; \theta_\lambda)$ is computed to model the compatibility of the input voxel v of (Alam, Zhou, Liew, Jia, Chanussot, and Gao [10]). The spatial coordinates $d = \{x, y\}$ in the time domain λ denote v. The partition function is expressed as $Z(v_{(d,\lambda)}) = \sum \exp\left[-E(l, v_{(d,\lambda)}; \theta_\lambda)\right]$. It is vital to simulate the connectivity between nodes in the CRF network to incorporate the contextual information. Hence, the the energy function for deriving this contextual information can be expressed as a combination of the unary potential and the binary potential function. These functions are discussed in the following subsubsections.

− Unary potential functions

The unary potential function φ is calculated for individual voxels. The unary potential can be determined by adding a stack of CNNs to the node function vectors for each

individual voxel. The final output of the unary potential in every individual voxel along λ can be produced in a fully connected layer. The nodes in the CRF network are represented by each individual voxel. The unary potential function φ can be calculated as follows

$$\phi(l_p, v_p \theta_\lambda) = -\log P(l_p, v_p \theta_\lambda). \qquad (5)$$

In the CNN stack, network parameters, such as θ_λ, can be changed while training deep CRF.

 − *Pairwise potential functions*

Pairwise, for the compatibility between neighbouring voxels, a potential function is constructed. The feature vectors of two adjacent voxels are used for calculating the output ψ. Then pairwise potential output is obtained by applying stack of CNNs on these feature vectors and it is given as

$$\psi(l_p, l_q, v_p, v_q; \theta_\lambda) = \mu(v_p, v_q)\delta_{p,q,l_p,l_q}(f_p, f_q; \theta_\lambda). \qquad (6)$$

The label compatibility function $\mu()$ is used for encoding the possible voxel pair (v_p, v_q), and this pair is labelled as (l_p, l_q) using various combinations of pairs. The CNNs output values δ_{p,q,l_p,l_q} are used on pair of nodes. The initial CNN is used to obtain feature vectors corresponding to each f_p & f_q (pair of nodes). The parameters of the CNN are contained in θ_λ and for the pairwise potential function it should be learned along with the whole area groups λ.

The log-likelihood of a training input-output pair (v, l) can be maximized to estimate the parameters in CRFs. Since marginal distribution calculation of the model is involved, undirected graphical models make it impossible to track actual maximum-likelihood training. It is also difficult during conditional training and prediction for the given observed input voxel v. So, a decision is taken to optimize $p(l \mid v)$ instead of $p(l, v)$. This difficulty can be reduced by training CRF efficiently.

 − *Piecewise CRF training*

The objective of CRF function in proposed CNN-based CRF can be given as

$$\nabla(\theta) = \sum_{\substack{P \in M \\ (p,q) \in N \times N \\ \lambda \in B}} \left(\iota_p, v_p\right)\Psi(\iota_p, \iota_q, v_p, v_q) - Z(v; \theta_\lambda). \qquad (7)$$

By maximizing log-likelihood of (v, ι), the performance will be improved. But the cost is high for training and it is due to log partition function calculation $Z(v)$. This function is dependent on model parameters and the input voxels along with the area groups.

For effective training, the whole model is divided into parts and trained. The weights learned from the parts were then combined and tested. This strategy is known as piecewise training. It is then $p(\iota \mid v)$ calculated on the basis of the number of independent likelihoods for all unary and pairwise potentials

$$p(\iota \mid v; \theta_\lambda) = \prod \varphi \prod_{p \in M} P\varphi(\iota_P \mid v; \theta_\lambda)\prod_\psi \prod_{(p,q) \in N \times N} P_\psi(l_p, l_q \mid v; \theta_\lambda). \qquad (8)$$

Using unary and pairwise potentials, $p_\phi(\iota_p \mid v; \theta_\lambda)$ and $p_\phi(\iota_p, \iota_q \mid v; \theta_\lambda)$ accordingly, it can be calculated. A method called "mean-field inference" is used to train the CRF.

 − *Mean-field inference.*

It is difficult to decrease the energy, as CRF energy function has so many parameters. As a result, the CRF distribution may be constructed using the mean-field approximation approach for maximum posterior marginal inference. The iterative inference technique begins with initialization. In this case, the softmax function is applied to the unary potential across all location labels. The second stage is message forwarding, which is accomplished by convolution using two Gaussian kernels that were previously created for location prediction.

Finally, the normalising step is performed using another softmax operation to generate the final labels in the classification map. The details on electricity consumption in each area will be obtained using the classification map and discussed in detail in the results and discussion section. Deconvolution is applied to the obtained map to obtain a high-resolution classification map.

 − *Deconvolution*

A deconvolution network is employed to generate a high-resolution classification map depending on a weighted inference, and this process is inclusive of unpooling, deconvolution, and rectified linear units (ReLUs).

Classification accuracy is improved by pooling, as it eliminates the noisy activations of the lower layer, thus keeping the activations only in the upper layers. It may abstract the activations with a single value in a receptive region. Evidently, spatial knowledge is lost during pooling within a receptive field. So, it is not easy to localize accurately.

The deconvolution network utilizes the unpooling layers for solving this issue in which the reverse process of pooling layers will be done. The unpooling process improves the resolution of an object during pairwise CRF training by reconfiguring the initial scale of the input data, maintaining the complex features connected with the topic of interest. The maximum number of activations chosen during the pooling will be kept typical during the unpooling process to restore the activations in the original pooling positions.

For enhancing the activations, filters from the deconvolution process that are similar to the targeted classes will be utilised as they remove the noisy activations of various classes they compromise. Each layer of the deconvolution network will contribute to the reconstruction of structures at different stages. Lower-layer filters are used to reconstruct the general form of an object, while higher-layer filters are used to reconstruct the specific details of a class. Thus, the usage of a deconvolution network will produce a more enhanced and accurate classification result.

IV. RESULTS AND DISCUSSION

The experiment is performed on a PC with an Intel Core i7 3.5 GHz CPU and 4 GB of RAM with NVIDIA GPU support. Keras and other deep learning frameworks are used to implement the model in Python. Table I indicates the parameters of CNN.

TABLE I. PARAMETERS OF CNN.

Parameters	Values
Learning Rate	0.01
Batch size	64
Epochs	50
Dropout	0.2
Optimizer	Adam
Activation Function	ReLU
Number of filters	64
Filter size	2
Number of Convolution layers	3
Number of Max-pooling layers	3

1. *Sample dataset and prediction over test dataset*

The sample values of the dataset are shown in Table II and contain mean, median, minimum, and maximum values of the electricity consumption in the block 0 of the daily dataset.

Figure 3 shows the prediction performance of the suggested CRF-PCP approach, which delivers the best prediction results over the dataset mentioned above. Only at one point during the prediction of electricity consumption does the anticipated result narrow slightly compared to the actual value of electricity consumption. However, the proposed CRF-PCP approach accurately predicted power usage in all other respects. A global active power is the projected value. The 100-day power consumption data from the dataset were used as a sample.

From the results, it was confirmed that the predicted data, depending on the proposed CRF-PCP technique, yield superior performances, as it nearly predicts the actual data.

At the end of learning using the proposed machine learning model, 100 % accurate data will be achieved. Trends showed a substantial correlation between actual and expected energy consumption, demonstrating the utility of the model in forecasting energy use the following day. As a result, the system is beneficial for users who want to reduce their energy consumption by changing their usage.

2. *Comparison of the accuracy level with other techniques*

The accuracy level of the proposed CRF-PCP technique is compared with other methods such as CNN-LSTM (Kim and Cho [25]), LSTM (Wang, Du, and Wang [26]), bidirectional LSTM (BDLSTM) (He, Chen, Gao, Ma, Xu, and Zhu [27]) and is discussed in this section.

Investigation was carried out at various accuracy levels using the deep learning techniques and is shown in Fig. 4. The proposed CRF-PCP technique forecasts the electricity forecast application accurately with an accuracy of 98.9 % as shown in Fig. 4. By adding the additional deconvolution after classification, the accuracy level was improved. For further improvement in classification performance, deconvolution will be performed.

The role of CNN tends to improve classification accuracy; however, adding more layers produces over fitting and can impair accuracy. The importance of minimising both training and validation losses in a well-trained network is widely acknowledged. When the training loss is low and the validation loss is high, the network will be overfitted and the simulation samples will not generalise well. The network will be overfitted, and the simulation samples will not generalise well.

CNNs are created using the trial-and-error technique, and their rate of learning, number of hidden layers, size of Kernel, and the range of convolutional layers are computed. The model was launched with fewer convolution layers at first and the layers were gradually raised. Training and validation losses were measured using a variety of layer counts. As a result, the categorization accuracy improves.

TABLE II. SAMPLE DATA FROM THE DATASET.

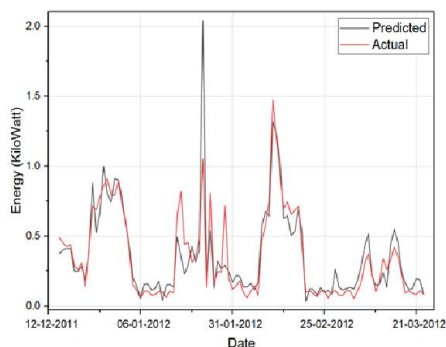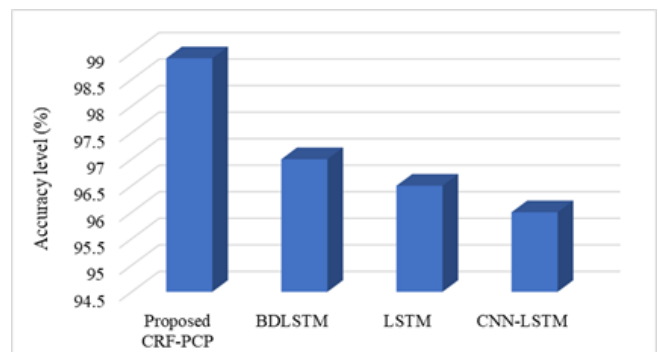| day	energy_median	energy_mean	energy_max	energy_min
12-10-2012	0.1385	0.154304348	0.886	0
13-10-2012	0.18	0.230979167	0.933	0.076
14-10-2012	0.158	0.275479167	1.085	0.07
15-10-2012	0.131	0.2136875	1.164	0.07
16-10-2012	0.145	0.203520833	0.991	0.087

Fig. 3. Visualisation of performance of the proposed CRF-PCP over test data for electricity prediction.

Fig. 4. Comparison of the prediction accuracy of different deep learning techniques.

3. *Performance evaluation-based k-fold cross-validation method*

The performance of the proposed CRF-PCP approach model is evaluated by analysing and verifying its effectiveness utilizing various types of experiments. Most existing approaches test the performance of the system by evaluating its cross value or by evaluating its k-fold CV. As a result, the proposed approach was tested using a 10-fold CV (Ullah, Ullah, Haq, Rho, and Baik [8]). The mean square error (MSE), the root mean square error (RMSE), and the mean absolute error (MAE) are the three basic metrics. These measures are used at rates. Let $\tilde{y}_j$ denotes N power consumption prediction samples and $\tilde{y}_j$ denotes the values that were observed. The MSE, RMSE, and MAE are represented by (9)–(11):

$$MSE = \frac{1}{N} \Sigma_{j}^{N} = 1\left(y_j - \tilde{y}_j \right)^2 , \qquad (9)$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \Sigma_{j}^{N} = 1\left(y_j - \tilde{y}_j \right)^2}, \qquad (10)$$

$$MAE = \frac{1}{N} \Sigma_{j=1}^{N} | y_j - \tilde{y}_j | . \qquad (11)$$

When analysing the forecast performance, the validity of the proposed method can be tested with another statistical metric, mean absolute percentage error (MAPE), which is shown in (12)

$$MAPE = \frac{100\%}{N} \Sigma_{j=1}^{N} \frac{| y_j - \tilde{y}_j |}{y_j}. \qquad (12)$$

The systematic error also called the "average error rate" for overall forecasting will be calculated using additional metrics such as the mean bias error (MBE), and (13) has the MBE metric

$$MBE = \frac{1}{N} \Sigma_{j=1}^{N} \left(y_j - \tilde{y}_j \right). \qquad (13)$$

Underforecasting will be indicated with negative MBE, whereas the overforecasting will be indicated with positive MBE. The forecast may exceed or fall short of the original data depending upon the underforecasting and overforecasting.

The MAPE and MBE values of the proposed method are compared with those of existing methodologies. Only fundamental measures, such as MSE, RMSE, and MAE, were calculated using the holdout technique. The k-fold CV partitions the dataset into k portions and repeats the hold-out process k times for each segment. Each fold is tested on a subset, with the remaining k-1 subsets being the training set. As each of the data samples is used k times in the validation and training sets, the approach has numerous advantages.

The bias and variance will be greatly reduced when most of the data are used in the fitting process. The machine learning literature suggests a value of k = 10. As a result, the

k value in this procedure is set to 10. The superiority of the suggested method over other deep learning-based approaches like LSTM and BDLSTM was verified by a series of experiments.

The methodologies used are summarized as shown in Table III and the performance results are evaluated using the MAE, MSE, MAPE, MBE, and RMSE error measures, and the loss value will fluctuate when there is an increase in the value of k.

TABLE III. RESULTS OBTAINED FOR THE 10-FOLD CROSS-VALIDATION METHOD WITH VARIOUS DEEP LEARNING METHODS.

Methods	MSE	RMSE	MAE
LSTM	0.4263	0.5130	0.4086
CNN-LSTM	0.4975	0.5816	0.4582
BDLSTM	0.3717	0.4374	0.3530
Proposed CRF-PCP	0.3393	0.4569	0.3174

4. *Performance evaluation based on the hold-out method*

We employed holdout in our trials, which is a CV that divides data into two sets: training and testing. This phase includes fitting the model with the help of a training set and using the model's function to forecast values in the test set for unknown values. This technique was efficient in terms of computing time. A training set of 80 % of the data was used, while a testing set of 20 % was used. The LSTM, BDLSTM, CNN-LSTM, and proposed model are used for experimentation. The prediction models were trained with up to 20 epochs when using the above-mentioned approaches. Table IV presents the results for each deep learning model.

TABLE IV. RESULTS OBTAINED FOR THE HOLD-OUT METHOD WITH VARIOUS DEEP LEARNING METHODS.

Methods	MSE	RMSE	MAE
LSTM	0.4263	0.5130	0.4086
CNN-LSTM	0.4975	0.5816	0.4582
BDLSTM	0.3717	0.4374	0.3530
Proposed CRF-PCP	0.3393	0.4569	0.3174

As the STTM approach used in the previous stage extracted features, the proposed system has a low error rate when compared with other techniques. It is very easy to predict and categorize consumer electricity consumption using this feature extraction technique. Also, the proposed CRF-PCP technique is given piecewise CRF training to obtain fast and accurate results. The proposed technique has a low run time compared to other existing ones.

5. *Comparison of energy consumption with weather factors*

Weather factors such as temperature, pressure, humidity, cloud cover, dew point, and wind speed are compared with electricity consumption.

These factors are considered for better understating of the energy consumption, and the variations of these factors with energy are evaluated as shown in Fig. 5. The x axis shows the recording date of the energy consumption, while the y axis with the red label indicates the weather factor and the y axis with the black label indicates the energy consumption.

It is clearly visible that as the energy consumption increases, the temperature will decrease. Figure 5(a) shows

the energy comparison against temperature. Figure 5(b) shows the comparison of energy with pressure. Figure 5(c) shows the comparison of energy with humidity and shows the decrease in energy consumption with increasing humidity. Figure 5(d) shows the comparison of energy with cloud cover and Fig. 5(e) shows the comparison of energy with the dew point. Figure 5(f) shows the comparison of energy with wind speed.
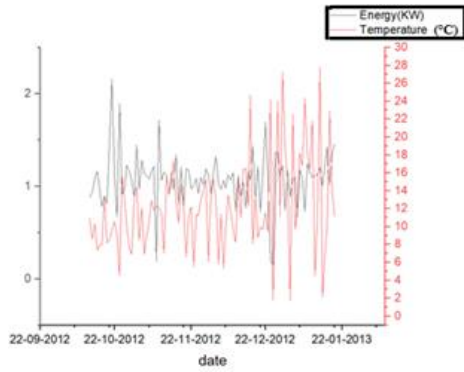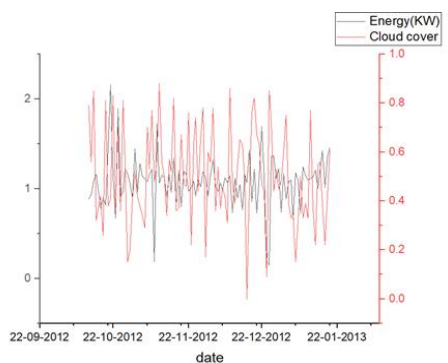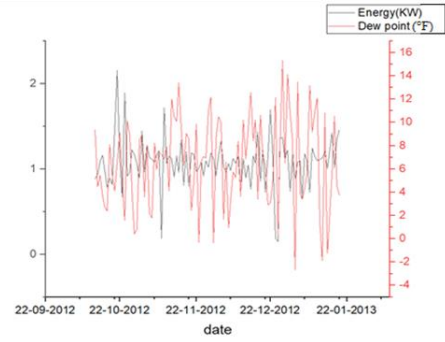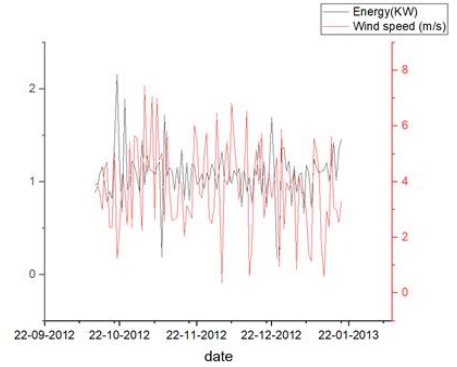

(a)
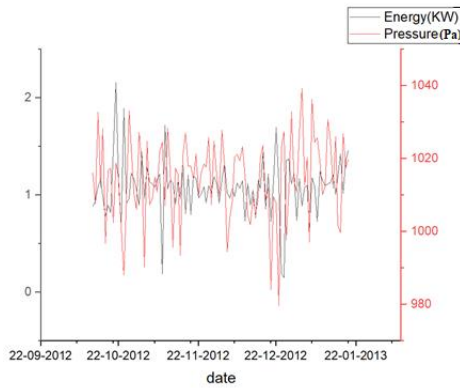
(b)
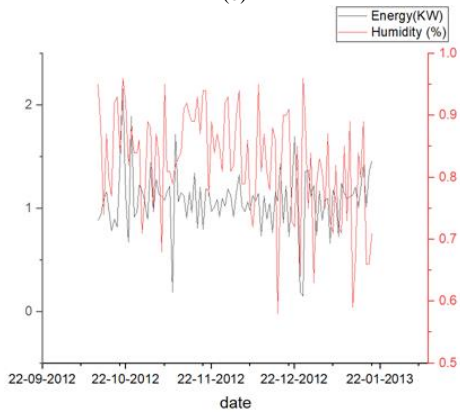
(c)

(d)

(e)

(f)

Fig. 5. (a) Comparison of energy with temperature; (b) Comparison of energy with pressure; (c) Comparison of energy with humidity; (d) Comparison of energy with cloud cover; (e) Comparison of energy with dew point; (f) Comparison of energy with wind speed.
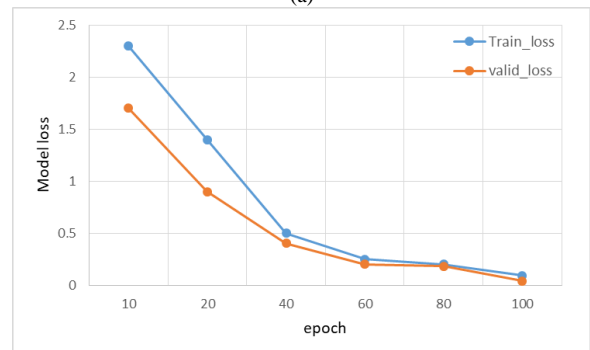
6. *Model loss*

The error presented in the training set of data is called "training loss" and the error after the validation set of data running through the trained network is known as "validation loss". Both loss values should be low.

The comparison of model loss for the proposed CNN-CRF is compared with BDLSTM, LSTM, and CNN-LSTM, and the comparison is shown in Figs. 6(a), 6(b), 6(c), and 6(d), respectively, and
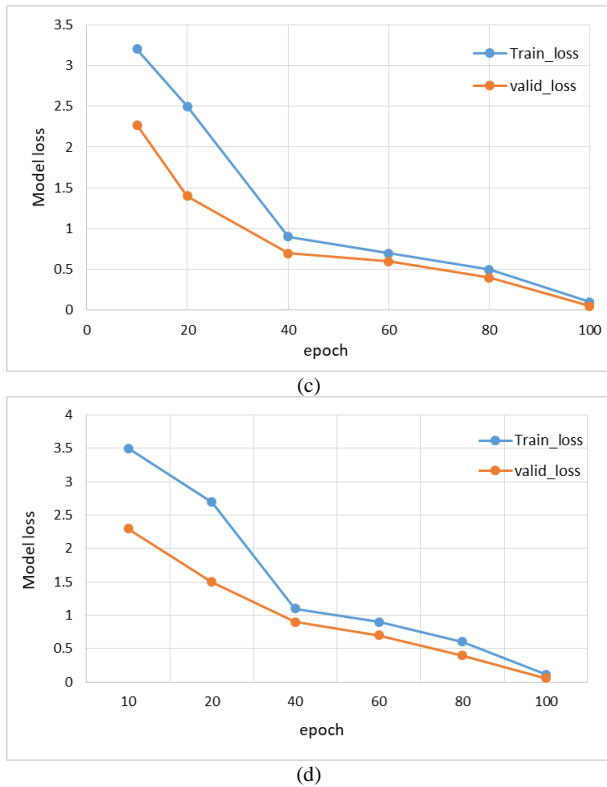

(a)

(b)

(c)

(d)
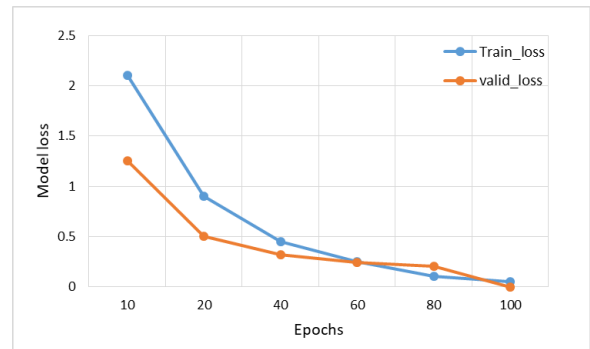
Fig. 6. (a) Model loss for the proposed CNN-CRF method; (b) Model loss for BDLSTM; (c) Model loss for LSTM; (d) Model loss for CNN-LSTM.

It shows the lowest performance due to poor training of the model. When compared with other models, the CNN-CRF model is found to be outperforming well in terms of model loss. During model training, small differences in the data were extracted. The validation loss for the proposed model is low when compared to the training loss. This shows that the model performs effectively when dealing with new data.

7. *Comparison of accuracy with epochs*

Figure 7 shows the training accuracy and validation accuracy of the proposed CNN-CRF with the number of epochs.
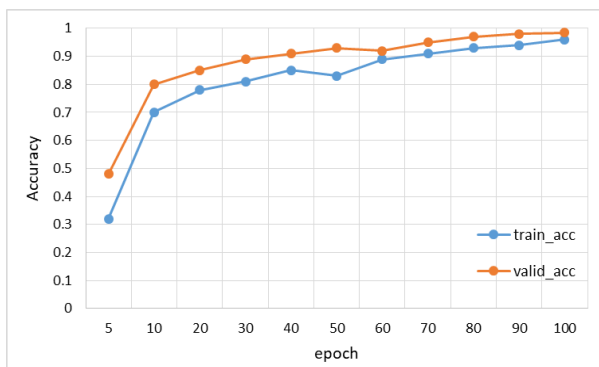

Fig. 7. Accuracy for the proposed CNN-CRF method.

For epoch = 100, CNN-CRF achieved the highest validation accuracy of 98.5 % and the training accuracy of 96 % due to piecewise CRF training. In addition, feature extraction helps reduce the training time, the number of parameters, and the dimensionality problem. Also, the generalization capability of the proposed model was found to increase, whereas the overfitting problem has been significantly avoided.

V. CONCLUSIONS

In this work, a novel CRF-PCP technique is proposed to predict the electricity consumption by consumers in various areas, and utilizing STTM, subtle variations of features of the data were extracted. The suggested model is also robustly trained using CNN. Temperature, pressure, humidity, cloud cover, dew point, and wind speed are all taken into account when predicting electricity use. Using a deconvolution block within the CRF pairwise potential calculations, the first prediction outputs from the CNN-CRF structure were considerably improved. In addition, bigger training group data for CNNs are used to address the issue of overfitting. Total performance was greatly improved as a result of this.

The proposed model achieves accuracy of 98.9 %, which is 2.9 % higher than LSTM, 2.4 % higher than BDLSTM, and 1.9 % higher than CNN-LSTM. The proposed model is also evaluated using 10-fold cross-validation (CV) and a hold-out (CV) method. In the CV method, the proposed CRF-PCP technique achieves low values of MSE, RMSE, MAE, MAPE, and MBE, which are 3.01 %, 8.92 %, 3.7 %, 5 %, and 0.3 % lower than the LSTM model, respectively.

In the hold-out CV approach, the CRF-PCP methodology delivers low MSE, RMSE, and MAE values that are 8.7 % lower than LSTM, and low RMSE and MAE values that are 12 %, 14 % lower than CNN-LSTM models. Weather conditions that affect energy use are also visualised. In general, the suggested system outperformed existing methods by 10 % in accurately predicting electricity use.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] F. Z. Abera and V. Khedkar, "Machine learning approach electric appliance consumption and peak demand forecasting of residential customers using smart meter data", *Wireless Pers. Commun.*, vol. 111, pp. 65–82, 2020. DOI: 10.1007/s11277-019-06845-6.

[2] S. N. Fallah, M. Ganjkhani, S. Shamshirband, and K.-w. Chau, "Computational intelligence on short-term load forecasting: A methodological overview", *Energies*, vol. 12, no. 3, p. 393, 2019. DOI: 10.3390/en12030393.

[3] Md. R. Haq and Z. Ni, "A new hybrid model for short-term electricity load forecasting", *IEEE Access*, vol. 7, pp. 125413–125423, 2019. DOI: 10.1109/ACCESS.2019.2937222.

[4] A. A. Mamun, Md. Sohel, N. Mohammad, Md. S. Haque Sunny, D. R. Dipta, and E. Hossain, "A comprehensive review of the load forecasting techniques using single and hybrid predictive models", *IEEE Access*, vol. 8, pp. 134911–134939, 2020. DOI: 10.1109/ACCESS.2020.3010702.

[5] Z. Deng, B. Wang, Y. Xu, T. Xu, C. Liu, and Z. Zhu, "Multi-scale convolutional neural network with time-cognition for multi-step short-term load forecasting", *IEEE Access*, vol. 7, pp. 88058–88071, 2019. DOI: 10.1109/ACCESS.2019.2926137.

[6] B.-S. Kwon, R.-J. Park, and K.-B. Song, "Short-term load forecasting based on deep neural networks using LSTM layer", *J. Electr. Eng. Technol.*, vol. 15, no. 4, pp. 1501–1509, 2020. DOI: 10.1007/s42835-020-00424-7.

[7] M. P. Raju and A. J. Laxmi, "IOT based online load forecasting using machine learning algorithms", *Procedia Comput. Sci.*, vol. 171, pp. 551–560, 2020. DOI: 10.1016/j.procs.2020.04.059.

[8] 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. DOI: 10.1109/ACCESS.2019.2963045.

[9] B. Farsi, M. Amayri, N. Bouguila, and U. Eicker, "On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach", *IEEE Access*, vol. 9, pp. 31191–31212, 2021. DOI: 10.1109/ACCESS.2021.3060290.

[10] F. I. Alam, J. Zhou, A. W.-C. Liew, X. Jia, J. Chanussot, and Y. Gao, "Conditional random field and deep feature learning for hyperspectral image classification", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1612–1628, 2019. DOI: 10.1109/TGRS.2018.2867679.

[11] S. Motepe, A. N. Hasan, and R. Stopforth, "Improving load forecasting process for a power distribution network using hybrid AI and deep learning algorithms", *IEEE Access*, vol. 7, pp. 82584–82598, 2019. DOI: 10.1109/ACCESS.2019.2923796.

[12] I. Parvez, M. Aghili, A. I. Sarwat, S. Rahman, and F. Alam, "Online power quality disturbance detection by support vector machine in smart meter", *J. Mod. Power Syst. Clean Energy*, vol. 7, pp. 1328–1339, 2019. DOI: 10.1007/s40565-018-0488-z.

[13] A. Jaech, B. Zhang, M. Ostendorf, and D. S. Kirschen, "Real-time prediction of the duration of distribution system outages", *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 773–781, 2019. DOI: 10.1109/TPWRS.2018.2860904.

[14] Y. Yuan, K. Dehghanpour, F. Bu, and Z. Wang, "A multi-timescale data-driven approach to enhance distribution system observability", *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3168–3177, 2019. DOI: 10.1109/TPWRS.2019.2893821.

[15] G. Fenza, M. Gallo, and V. Loia, "Drift-aware methodology for anomaly detection in smart grid", *IEEE Access*, vol. 7, pp. 9645–9657, 2019. DOI: 10.1109/ACCESS.2019.2891315.

[16] J.-S. Chou, S.-C. Hsu, N.-T. Ngo, C.-W. Lin, and C.-C. Tsui, "Hybrid machine learning system to forecast electricity consumption of smart grid-based air conditioners", *IEEE Systems Journal*, vol. 13, no. 3, pp. 3120–3128, 2019. DOI: 10.1109/JSYST.2018.2890524.

[17] S. K. A. Kamarol, M. H. Jaward, J. Parkkinen, and R. Parthiban, "Spatiotemporal feature extraction for facial expression recognition", *IET Image Processing*, vol. 10, no. 7, pp. 534–541, 2016. DOI: 10.1049/iet-ipr.2015.0519.

[18] Q. Zheng, M. Yang, J. Yang, Q. Zhang, and X. Zhang, "Improvement of generalization ability of deep CNN via implicit regularization in two-stage training process", *IEEE Access*, vol. 6, pp. 15844–15869, 2018. DOI: 10.1109/ACCESS.2018.2810849.

[19] Q. Zheng, M. Yang, X. Tian, N. Jiang, and D. Wang, "A full stage data augmentation method in deep convolutional neural network for natural image classification", *Discrete Dynamics in Nature and Society*, vol. 2020, art. ID 4706576, 2020. DOI: 10.1155/2020/4706576.

[20] Q. Zheng, X. Tian, M. Yang, Y. Wu, and H. Su, "PAC-Bayesian framework based drop-path method for 2D discriminative convolutional network pruning", *Multidim. Syst. Sign. Process.*, vol. 31, pp. 793–827, 2020. DOI: 10.1007/s11045-019-00686-z.

[21] Q. Zheng, P. Zhao, Y. Li, H. Wang, and Y. Yang, "Spectrum interference-based two-level data augmentation method in deep learning for automatic modulation classification", *Neural Comput. & Applic.*, vol. 33, pp. 7723–7745, 2021. DOI: 10.1007/s00521-020-05514-1.

[22] Q. Zheng, P. Zhao, D. Zhang, and H. Wang, "MR-DCAE: Manifold regularization-based deep convolutional autoencoder for unauthorized broadcasting identification", *International Journal of Intelligent Systems*, vol. 36, no. 12, pp. 7204–7238, 2021. DOI: 10.1002/int.22586.

[23] Smart meters in London, (n.d.), Kaggle: Your Machine Learning and Data Science Community. [Online]. Available: https://www.kaggle.com/jeanmidev/smart-meters-in-london?select=acorn_details.csv

[24] T. Turc, "Using WEB services in SCADA applications", *Procedia Technology*, vol. 19, pp. 584–590, 2015. DOI: 10.1016/j.protcy.2015.02.083.

[25] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks", *Energy*, vol. 182, pp. 72–81, 2019. DOI: 10.1016/j.energy.2019.05.230.

[26] J. Q. Wang, Y. Du, and J. Wang, "LSTM based long-term energy consumption prediction with periodicity", *Energy*, vol. 197, art. 117197, 2020. DOI: 10.1016/j.energy.2020.117197.

[27] Y.-L. He, L. Chen, Y. Gao, J.-H. Ma, Y. Xu, and Q.-X. Zhu, "Novel double-layer bidirectional LSTM network with improved attention mechanism for predicting energy consumption", *ISA Transactions*, vol. 27, pp. 350–360, 2022. DOI: 10.1016/j.isatra.2021.08.030.