DEVELOPMENT OF AN EFFICIENT DEEP LEARNING SYSTEM FOR AUTOMATIC PREDICTION OF POWER DEMAND BASED ON THE FORECASTING OF POWER DISTRIBUTION

T. Aravind[*], *P. Suresh*[†]

Abstract: Electrical load prediction aids electrical producing and allocation firms in planning capacity and management to ensure that all customers get the energy they need on a consistent basis. Despite the fact that numerous prediction models have been created, none of them can be applied to all market trends. As a result, this article provides a practical technique for predicting customer power usage. To address the troubles of power utilization surveying, CRF-based energy utilization choosing strategy conditional random field based powered consumption prediction (CRF-PCP) is proposed. A convolutional brain organization (a technique in view of artificial intelligence) joined with a contingent irregular field is utilized to prepare and foresee the energy consumption (EC) of the districts. The training model's features are extracted using a spatiotemporal texture map (STTM). Supervisory control and data acquisition (SCADA) is utilized to gather and keep up with information on the power utilization of local purchasers. The information given in the cloud is sent to the power circulation framework. Additionally, power utilization expectation utilizing a convolution neural network (CNN) with profound conditional random field (CRF) provides an outcome of 98.9% precision, which is far superior to prior research in the same area. The acquired result demonstrates that the employed machine learning methods are performing at their peak.

Key words: *electricity consumption prediction, convolution neural network, SCADA, conditional random field, spatiotemporal texture map*

Received: October 31, 2023 **DOI:** 10.14311/NNW.2023.33.025
Revised and accepted: December 23, 2023

1. Introduction

Electric power distribution is the last stage in energy delivery. It supplies power from the dissemination framework to existing buyers. The smart grid is a new type

[*]Aravind Thangavel – Corresponding author; Department of Computer Science and Engineering, Muthayammal Engineering College, Rasipuram, Tamilnadu, India, E-mail: taravindcse@gmail.com

[†]Suresh Paramasivam; Department of Mechanical Engineering, Muthayammal Engineering College, Rasipuram, Tamilnadu, India, E-mail: suresh.me2004@gmail.com

of electrical grid technology that is made to control energy use in a way that is sustainable, dependable, and cheap. The gap between supply and demand is reduced by these networks, making it possible to produce and distribute greener energy. Predicting electronic equipment and peak demand is essential for managing and constructing electric power systems [1]. In order to meet short, intermediate, and long-term demand, load forecasting is used. Forecasting helps utilities manage and supervise their customers' supply. Efficient and profitable electricity production and distribution companies rely on load forecasting. It helps them manage their resources and procedures to ensure all consumers receive the energy they need. Predicting energy use has several advantages. It helps utilities plan better and reduces risks by knowing future consumption or load demand. Assessing long-term load helps the company plan and budget for future distribution and load initiatives. It maximizes the utilization of electricity-generating devices while minimizing waste. It helps in the planning of size, location, and type of key generating plants. In areas with high or rising demand, utilities will be more inclined to generate energy near the load. To reduce the size of transmission and distribution infrastructures and associated losses. Predicting energy use helps decision-making and maintenance planning. With this information, the utility may plan repairs to have the least impact on consumers. When most people are at work and demand is low, they might decide to carry out repairs in residential areas round the clock. aides in the identification of the resources required for continuous, cost-effective energy production, including fuel for power plants and other resources. Several factors affect energy forecasts. Climate, temperature, precipitation, and sunshine seasons of the year are temporal or calendar variables. Economic factors include industrial development and growth. Customers are affected by the kind of usage, facility size, electrical equipment, and employee count. In this work, we propose a CRF-based energy utilization expectation strategy CRF-PCP considering these variables to prepare an artificial intelligence (AI) model and foresee power utilization [2]. Considering these elements, the expectation exactness of the proposed framework has been moved along. The contributions of our work are listed as follows.

- By utilizing the CRF-PCP method, the aim is to enhance the accuracy and efficiency of energy consumption (EC) monitoring. This approach takes advantage of the probabilistic modeling capabilities of CRFs, which enable capturing the dependencies and relationships among various factors affecting energy consumption.
- To train and predict the energy consumption of different locations, a combination of a restrictive irregular field and a convolution brain network is employed.
- A spatiotemporal texture map (STTM) is utilized to remove the elements of the preparation model.
- SCADA is used to collect and maintain data about the electricity consumption of customers in the region. The predicted data in the cloud is transmitted to the power distribution system.

The document's structure is broken down as follows: Segment 2 audits a few past undertakings on power dissemination demonstrating and highlight determination.

In Section 3, the proposed CRF-PCP method is shown. Segment 4 examines the viability of the proposed strategy. This study comes to a conclusion in Section 5.

2. Literature survey

A hybrid framework of convolution neural network (CNN) and conditional random field (CRF) to account for spatial and spectral characteristics was proposed for identifying hyperspectral pictures [2]. Deep CRF with CNN built unary and binary potential is used to derive a correlation between image patches. A profound deconvolution network is utilized to work on the presentation of the class map. The South African transmission framework network forecast programming utilizes another artificial intelligence (AI)/ deep learning (DL) cross breed approach [9]. The AI/DL load estimate module is utilized for this review. The impact of temperature on the exhibition of the current mixture of AI and DL models was likewise researched. A one-class support vector machine (OCSVM) was suggested for power disturbance discrimination [11]. This OCSVM is very data-driven. This model can identify disruptions in real time. If any disruptions are detected, classifiers categorize them. A neural network predictor was trained using previous outage data to estimate the duration of unexpected power outages [7]. Based on climatic variables, the initial period estimate is updated based on field reports, which are instantly evaluated through natural language processing.

A method was developed to anticipate the daily use patterns of non-system manager customers and improve distribution network dependability [16]. In order to model models, make predictions, and collect data from smart meters, three machine learning algorithms are utilized to unobserved consumer consumption trends. A study reported the development of the first client-server data categorization protocol using a support vector machine [12]. For both two-class and multi-class problems, the proposed approach safeguards privacy. It utilizes Pailler homomorphic encryption and two-section secure calculation. [3] presented a technique to anticipate traffic flow in the system. The basic model is Least squares support vector regression (LS-SVR) with a Gaussian kernel function. It uses linear LS-SVR to forecast traffic flow. Finally, the harmony search method regulates the linear LS-SVR anticipated restrictions. [6] used energy usage to identify abnormalities in the smart grid. Based on previous purchases, they used long short-term memory (LSTM) to predict customer behavior. This strategy notices customer behavior to identify typical and aberrant patterns.

[10] presented a technique for predicting bus arrival using a recurrent neural network (RNN). Multiple passing stations are utilized to "correct" the prediction for a station using RNN with extended short-term memory. A scalable and low-communication distributed support vector machine (SVM) training method was described [5]. A QR decomposition of low-rank approximations is used to compress the kernel matrix, reducing training stage computation and storage needs. [14] used SVM to classify arrhythmic beats into normal and pathological. Delayed errors normalized least mean square (DENLMS) algorithm-based adaptive filter improves filtering efficiency while reducing computation. The pre-processed signal is subjected to discrete wavelet transform, followed by SVM classification.

[4] developed a hybrid prediction method to anticipate 1-day-ahead air conditioner energy usage. It also incorporates linear and nonlinear methods to measure the energy consumption of air conditioners. In a new electricity price forecasting model [15], to reduce feature duplication, a hybrid feature selector based on general combing ability is employed. To reduce dimensionality, the feature extraction method uses kernel principal component analysis (KPCA) and SVM classifiers to forecast power prices. [8] developed a spatiotemporal texture map (STTM) capable of capturing subtle spatial and temporal changes in facial emotions. The dynamic characteristics are extracted using a block-based approach and represented as histograms. The support vector machine classifier then categorizes the characteristics.

3. Methodology

Short-term load allocation and long-term planning for new generation and transmission facilities rely heavily on electricity demand forecasting. A precise prediction also enables us to make more cost-effective and energy-efficient choices. The goal of this study is to forecast consumer power usage. We propose a method called conditional random fields based power consumption prediction (CRF-PCP) for predicting electricity consumption. The CRF-PCP method utilizes CRF to make accurate predictions of energy consumption. By using the CRF-PCP method, we aim to improve the accuracy of electricity consumption prediction. The CRF model takes into account various factors such as historical consumption patterns, weather conditions, time of day, and other relevant variables to make reliable predictions. These predictions are then utilized to distribute the estimated energy consumption values across different regions within the area. There are two moves toward this procedure.

1. Power distribution based on historical data and other factors
2. Power distribution based on predicted data

Here, the whole region is considered and divided into areas. The electricity is consumed in each of the areas in the electricity distribution circle. First, the generated power from the power generator is distributed by the power distributor to the customers based on historical data. Then the CRF based machine learning technique is used to predict electricity consumption. Based on this predicted data, the power is distributed to consumers. Supervisory control and data acquisition (SCADA) are used to maintain and analyze the data about electricity consumption. The diagram depicted in Fig. 1 provides an overview of the CRF based power consumption prediction system, highlighting its key components and workflow. To ensure secure and efficient operations, the system incorporates a secure cloud infrastructure. This secure cloud serves as a centralized platform for storing and processing data related to power distribution. The following steps outline the functioning of the system:

1. Historical data distribution: Initially, the historical data related to power consumption is sent to the power distributor. This data serves as the basis for estimating and distributing power to the consumers in the region.

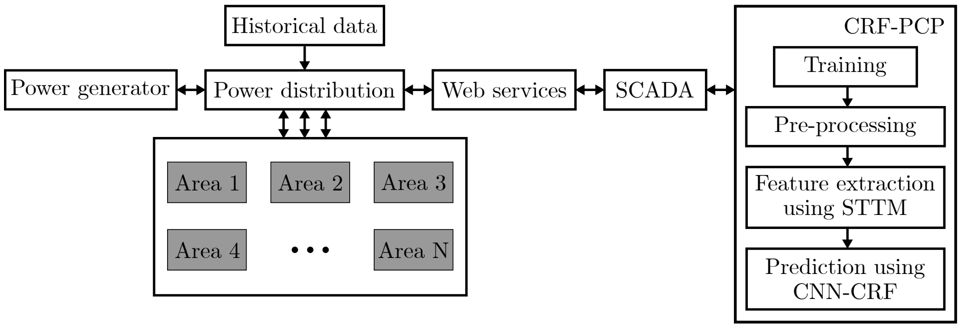

Fig. 1 *System design of CRF based power consumption prediction (CRF-PCP).*

2. Power distribution: In view of the authentic information it gets, the power merchant apportions and conveys capacity to shoppers in the district. This allocation is done considering factors such as past consumption patterns, demand fluctuations, and other relevant parameters.

3. Power consumption data collection: The power distributor collects detailed information about the power consumption from the consumers. This data includes real-time measurements of energy usage by individual consumers or groups of consumers.

4. Data transfer to SCADA: The collected power consumption data is then transferred to the SCADA system. This transfer is typically accomplished through web services, ensuring efficient and secure transmission of the data.

5. Data storage and training: The SCADA system maintains the collected power consumption data in its database. This data is utilized to train the conditional random field-based power consumption prediction (CRF-PCP) approach. The training process involves analyzing the historical data and establishing patterns and correlations between various factors influencing power consumption.

6. Prediction module: The trained CRF-PCP model is utilized for predicting future electricity consumption. The prediction module takes into account various factors such as time of day, weather conditions, consumer behavior, and historical patterns to estimate the expected power demand in the future.

7. Power distribution update: The predicted electricity consumption data is sent back to the power distributor. This updated information helps the power distributor in making informed decisions about power allocation and distribution. It enables the power distributor to distribute the corresponding power to specific areas or consumers in the region, ensuring an efficient and reliable supply.

In summary, the proposed system involves the exchange of historical and real-time power consumption data between the power distributor and consumers. This data is collected, stored, and used for training the CRF-PCP approach, which enables accurate prediction of future electricity consumption. The system ensures effective power distribution to meet the demand and optimize the utilization of available resources.

3.1 Power generator

The most common way of delivering power from essential energy sources is known as the power age. The power generator makes power and conveys it to the power distributor. The SCADA receives these data via transmission.

Power distributor: The last phase of the power conveyance process is the power merchant. It transports electricity to end users from the power source. It additionally gathers data about customers' energy utilization and sends it to SCADA. The power distributor distributes power to customers after receiving predictive data from SCADA at the subsequent distribution.

Secure cloud: Moving information to the cloud serious area of strength for requires security. Cloud computing is just as vulnerable to security threats as on-premises computing. These threats are constantly evolving and becoming increasingly sophisticated. Working with a cloud service provider that tailors its security to your infrastructure is crucial because of this.

3.2 Supervisory control and data acquisition

Supervisory control and data acquisition (SCADA) monitors and controls data relating to cloud power usage. SCADA has a database to store data. SCADA records data on power production and consumption. SCADA systems typically depend on real-time databases [16]. The primary part, the SCADA server, is the association between the observed equipment framework and the SCADA applications. The SCADA server facilitates data transmission between the technical process and the database. These clients also allow human operators to read and write data to the database. Thus, SCADA programmers are designed to enable database access. While database activities are represented in Structured Query Language (SQL), SCADA applications are implemented as web services. SCADA web services depend on SCADA servers linked to field equipment through remote terminal unit (RTU) and SCADA clients. The RTU associates the SCADA server with the administered specialized framework. The gathered information is put away in a data set that is refreshed progressively to meet the data necessities of SCADA clients. A database stores the real-time data. SCADA clients (generator and wholesaler) can demand information (pages) from web servers. It will open a page with service calls sent to the customer in a continuous stream. Generators and distributors may access the database through web services. The data accessed from the SCADA is preprocessed for prediction.

3.3 Technique for CRF-PCP using machine learning

This paper proposes an AI based gauging procedure called CRF-PCP to give figure information to the power dissemination framework that is utilized to disperse capacity to customers. Domestic, business, and industrial users account for approximately one-third of all energy consumption in the nation. Consumers of electricity include all kinds of end-users. The information about how much power is utilized in every space is accumulated by the power wholesaler and shipped off the SCADA framework. The power dispersion structure starts by apportioning ability to each locale area utilizing data from SCADA. This SCADA also takes the responsibility to maintain the collected data from the region. Along these lines, it is recommended that a simulated intelligence based assumption method known as CRF-PCP be utilized to give data to the system to control dissemination, which will include it in power dispersion.

3.4 CRF based power consumption prediction technique

Fig. 2 shows how the CRF-PCP method segregates the district into locales and uses artificial intelligence to expect power use for each area. In this paper, CNN-CRF method is utilized for arranging and suspicion in mimicked knowledge. For the induction, a graphical model known as CRF is utilized spatiotemporal contextual information of each area. The proposed technique of CRF-PCP consists of three steps:

- Pre-processing,
- Feature extraction,
- CNN with deep CRF based prediction.

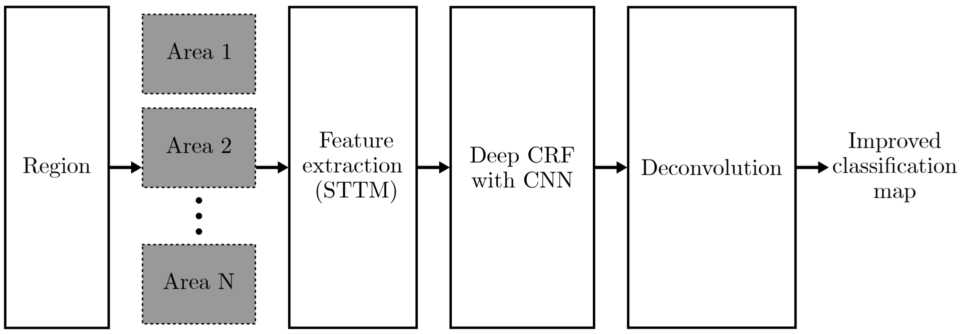

Fig. 2 *Structure of the proposed CRF-PCP technique.*

3.4.1 Pre-processing

During the pre-processing stage, the power distributor collects and manages the power consumption data of different areas through the SCADA system. This data is then utilized by the power distribution network to supply electricity to all regions before making predictions. In the subsequent power distribution cycle, the CRF-PCP technique's output, which includes predicted power data, can be employed. The data undergoes various steps to ensure its quality and prepare it for further analysis using the prediction model.

The primary objectives of pre-processing are to remove noise, handle missing values, and enhance the data's quality and consistency. One common issue with real-world data is the presence of noise, which can be caused by measurement errors, sensor malfunctions, or other external factors. Noise can adversely affect the accuracy of the prediction model, so it is crucial to address it during pre-processing. Locally estimated scatterplot smoothing (LOESS) techniques are applied to reduce or eliminate noise from the data. Another aspect of pre-processing involves handling missing values. In real-world data, certain measurements or attributes may be missing due to various reasons, such as sensor failures or data transmission issues. These missing values can introduce bias or inaccuracies in the analysis. Different strategies, such as mean, median, and mode imputation involve replacing missing values with estimated values and are employed to handle missing data appropriately. Furthermore, pre-processing may involve adding additional information to enhance the predictive capabilities of the model. This can include incorporating features, such as weather data or demographic information that may have a significant impact on electricity consumption patterns. By augmenting the dataset with relevant contextual information, the prediction model can capture more nuanced relationships and improve its accuracy. The LOESS pre-processing steps undertaken will depend on the scatterplot approach and the requirements of the prediction model. It is important to carefully assess the data quality, identify any issues, and apply appropriate techniques to address them effectively. Through robust pre-processing, the data is prepared in a way that minimizes errors and maximizes the performance of the prediction model.

3.4.2 Feature extraction

We used a spatiotemporal texture map (STTM) [8] for feature extraction, which is proficient in catching understated spatial and progressive changes in regions while being computationally simple. We first model an input f using its scale-space linear representation L, which is generated by convoluting f with a 3D Gaussian kernel. In a scale-space representation, the data is convolved with kernels of different standard deviations. The kernel serves as a smoothing filter, and by varying the standard deviation, you can control the amount of smoothing applied to the image. By analyzing the data at multiple scales, the representation becomes more robust to changes in scale and can detain information at different levels of detail.

$$L\left(\sigma^2, r^2\right) = \mathbf{Z}\left(\sigma^2, r^2\right) * f\left(\cdot\right), \tag{1}$$

where σ^2 and r^2 are the Gaussian kernel $\mathbf{Z}$ spatial and temporal variances, respectively. The Gaussian kernel is defined by two parameters: the spatial variance and the temporal variance. The spatial variance controls the spread or width of the Gaussian filter in the spatial domain. A larger spatial variance results in a wider filter, leading to more smoothing. Conversely, a smaller spatial variance produces a narrower filter and less smoothing. The general form of the spatio-temporal Gaussian kernel is given by the product of a spatial Gaussian component and a temporal

Gaussian component. Mathematically, it can be expressed as:

$$\mathbf{Z}\left(x, y, t, \sigma, r\right) = \frac{-\exp\left(\frac{-\left(\mathrm{x}^2+\mathrm{y}^2\right)}{2\sigma^2} - \frac{t^2}{2r^2}\right)}{\sqrt{\left(2\pi\right)^t\ \sigma^4 r^2}}. \tag{2}$$

In the spatial domain, x and y represent the x and y axes from the input f, while in the temporal domain, t denotes the time axis. The Gaussian kernel $\mathbf{Z}$ is separable, meaning that it can be expressed as the outer product of a spatial Gaussian kernel and a temporal Gaussian kernel. The spatial part handles the spatial smoothing, and the temporal part handles the temporal smoothing. The fluctuations in the region of the spatiotemporal domain are computed afterwards by merging the determinant and touch [8] to build the modified. The Harris corner measures the local intensity variations in different directions, identifying regions where there are significant changes in intensity. For spatiotemporal data, the Harris corner detection can be extended to capture both spatial and temporal variations. In order to detect disparities in the spatiotemporal domain, a convolution is executed between the spatiotemporal matrix and a Gaussian kernel function, represented as $\mathbf{U}$. Subsequently, the resulting value of $\mathbf{U}$ is used to compute the eigen values $\lambda_1\lambda_2\lambda_3$, which are expected to be high. This signifies that there is a variation in the utilization of electricity. The extension involves considering the spatial and temporal derivatives of the pixel intensities. The Harris point capability $\mathbf{H}$ for the spatiotemporal area is as per the following:

$$\mathbf{H} = \det(\mathbf{U}) - k \cdot trace^3(\mathbf{U}) = \lambda_1\lambda_2\lambda_3 - k\left(\lambda_1+\lambda_2+\lambda_3\right)^3, \tag{3}$$

where k is a constant, $trace(\mathbf{U})$ is the sum of diagonal elements in the matrix $\mathbf{U}$. The $\mathbf{H}$ capability is standardized, which dispenses with region varieties. Changes in time and space are uncovered when the local constructive maxima of $\mathbf{H}$ are found. The surface guide of the area made by STTM presents just the spatiotemporal varieties in the electricity consumption circle. To identify the locations of such changes, a block-based representation is needed. This step guarantees that the texture map's spatial information is preserved. Each texture map is split into various numbers of blocks are taken to reflect the features acquired by STTM. Each block is then given its own histogram, which is then connected to make a component vector addressing the district. As a result, minor changes in the low level may be recorded more effectively. Applying A-law compression to the texture map [8] is one way to do this.

3.5 CNN with deep CRF based prediction

The region is handled as an area group in this phase. It begins by creating a feature map from the area groupings after applying CNN to them. The CNN-based deep CRF is then presented here utilizing the CNN output. The CRF's potentials unary and pairwise are computed by covering a CNN-CRF manner to deal with spatiotemporal data over the whole region. The classification map was then created using a mean-field inference method. To conclude, a deconvolution system-based enhancement to the final classification performance, and equations

4 through 8 were constructed using it [2]. This method combines deep CRF with the spatiotemporal characteristics acquired in the first step to describe the spatial and temporal contextual relationships among the regions by combining the qualities of both CNN and CRF. The integrated models provide an advantage to the spatiotemporal connections between area sets to accomplish the final sorting, making an ideal learning method for area analysis. CRF makes extensive use of geographical and temporal contextual data in the training method, considered significant and helpful in applications of power distribution. The CNN-based deep CRF (CNN-CRF) method will be used to further evaluate the CNN output. It's worth noting that CNN's output comes in the form of feature maps, each of which has a unique location defined by geographical coordinates and temporal information. These spatiotemporal sites are referred to as voxels. Because CNNCRF can model these voxel neighborhoods, it's an excellent choice for analyzing region data. Loads of CNNs functional features to the original map were utilized to train the deep CRF employed constraints in this technique. Meanwhile, the feature original map is previously a strong depiction of local spatiotemporal characteristics and the CNN-CRF uses collective area groups as input to the system, rather than utilizing a whole region. Each voxel in a CRF network is represented by a node in the feature vector. The voxel labels are denoted by the label l_p. Within these nodes, edges are created. To establish pairwise connections between neighboring voxels in the CRF, each node is linked to all of its neighbors. The CRF can be defined in the following manner.

$$\mathrm{P}(l_p \,|\, v_{(d,\lambda)}; v_\lambda) = \frac{1}{z(v_{(d,\lambda)})} \exp\left(-E(l_p, v_{(d,\lambda)}; \theta_\lambda)\right) \tag{4}$$

From the above equation, the energy function E is computed to model the compatibility of the voxel v. $\mathrm{P}(l_p \,|\, v_{(d,\lambda)})$ is the conditional probability of a label l_p given an input sequence $v_{(d,\lambda)}$. The spatial coordinates $d = \{x, y\}$ in the time domain θ_λ denote voxel v.

The spatial coordinates $d = \{x, y\}$ in the time domain θ_λ denote v. The partition function is expressed as $z(v_{(d,\lambda)}) = \sum \exp(-E(l_p, v_{(d,\lambda)}; \theta_\lambda))$. It is crucial to simulate the connectivity among nodes in the CRF network in order to integrate contextual information. Therefore, the energy function that captures this contextual information can be represented as a fusion of the unary potential and the binary potential function.

3.5.1 Unary potential functions

In this proposed CNN-CRF strategy, a CNN stack is utilized to create a completely associated layer of component maps, which is then utilized for the end product, where the potential peculiarity has a place with a solitary individual voxel. To compute the unit capability of each voxel addressing a CRF hub in the chart, a CNN stack was carried out on the hub highlights got from the first element map. The unary potential function φ_i is determined as follows:

$$\varphi_i(l_p, \vartheta_\lambda) = \exp\left(\sum_{j=1}^{M} w_j \cdot f_j(l_p, \vartheta_{\lambda_i})\right). \tag{5}$$

$\varphi_i\,(l_p, \vartheta_\lambda)$ is the unary potential function for the ith variable, which assigns a score to each label l_p given the input sequence ϑ_λ at position i. w_j are the weights associated with different features f_j of the input sequence. l_p is a label assigned to the ith element of the input sequence. ϑ_λ is the input sequence and M is the number of features. These parameters are learned during the training process to minimize the difference between predicted and actual outputs. The statement implies that, during the training of the combined CNN-CRF model, the parameters specific to the CRF part of the network are adjusted.

3.5.2 Pairwise potential functions

For all possible combinations, the set of voxels must be compatible and is taken into account while calculating the pairwise potential functions. Individual voxel feature vectors in the feature map come from the first CNN applied to the whole area. As a consequence, edge features may be generated by concatenating two consecutive voxels feature vectors denoted by δ. The edge highlight vectors are then put through a CNN stack, delivering a potential pairwise yield. The paired potential capability is given as follows:

$$\delta_j, \delta_{j+1}(l_p, l_{p+i}, \vartheta_\lambda) = \exp\left(\sum_{i=1}^{N} v_p \cdot v_q(l_p, l_{p+i}, \vartheta_\lambda)\right). \tag{6}$$

$\delta_j, \delta_{j+1}(l_p, l_{p+i}, \vartheta_\lambda)$ is the pairwise potential function for the jth and $j+1$th variable, which assigns a score to the transition from label l_p to label l_{p+i} given the input sequence ϑ_λ and N is the number of features for the pairwise potential. v_p are the weights associated with different features v_q of the label transition. l_p, l_{p+i} are labels assigned to the jth and $j+1$th variable of the voxels. In CRFs, parameter estimation is done by maximizing a training input-output pair's log-likelihood. For undirected graphical models, exact maximum-likelihood training is used further; it is difficult since the computation requires the model's marginal distribution to be calculated. To minimize the computational complexity, effective CRF training is desirable.

3.5.3 Piecewise CRF training

In the proposed CNN-CRF model for tasks like semantic segmentation, the piecewise CRF function refers to the training and inference processes that are split between the CNN and the CRF. The goal function for the planned CNN-CRF may be defined as follows:

$$\omega(\theta) = \sum_{i=1}^{C} \log \mathrm{P}(l_p^{(i)} | v_{(d,\lambda)}^{(i)}; \theta) \tag{7}$$

where $\omega(\theta)$ are the parameters of the CRF, C is the number of training samples, $v_{(d,\lambda)}^{(i)}$ is the input data, $l_p^{(i)}$ is the ground truth labeling, and $\mathrm{P}(l_p^{(i)} | v_{(d,\lambda)}^{(i)}; \theta)$ is the conditional probability of the labeling input and CRF parameters. The process helps the model to learn features from the CNN while benefiting from the spatial dependencies captured by the CRF. After CRF training, a mean field inference method is used to conduct inference on this model.

3.5.4 Mean-field inference

In reality, precise reduction of CRF energy is virtually difficult owing to the huge parameter numbers including the objective function and limited energy consumption. The distribution of CRF for the maximum subsequent inference margin is calculated using the mean-field approximation method. Softmax is applied to the unary potential at each location label during this iterative inference method's initialization phase. The class map's final labels are the result of another softmax operation finishing the normalization phase. The obtained classification map indicates the details about each area along with electricity consumption by consumers for each day. This is illustrated in a detailed manner in the result and discussion part.

3.5.5 Deconvolution

The deconvolution layers play a crucial role in upsampling the low-resolution classification map to a higher resolution. This process helps generate a more detailed and accurate mean-field inference. By utilizing deconvolution, the network can reconstruct finer spatial details and improve the resolution of the output. In addition to the deconvolution layers, the network incorporates rectified linear unit (ReLU) layers. The ReLU layers help activate and propagate the relevant information, improving the overall performance and accuracy of the network. Connection further develops classification execution by dispensing with uproarious commencements in youngster layers and leaving just the top layers dynamic. It may use a single value to abstract the activations inside a receptive region. During pooling, inside a receptive region, spatial information is clearly vanished. As an outcome, precise localization is difficult. Unpooling layers, which operate in the opposite direction of pooling layers, have been employed in the deconvolution network to address this problem. To abstract the activations of the responsive region, only one value is required. Unfortunately, when pooling, spatial data inside a receptive region is missing. Consequently, pinpoint accuracy is always feasible. By recreating the original input data size and therefore maintaining the intricate constructions of the subject of attention, the unpooling method improves object resolution during CRF paired training. The positions of maximal activations carefully chosen throughout the pooling activity are typically kept track of throughout the unpooling procedure. With this information, activations can be brought back to their original state. Deconvolution channels assist with featuring enactments that are like the objective classes while sifting through commotion from the many classes that make up the districts. Subsequently, different layers of the deconvolution organization will assist with reproducing structures at various levels. While higher layer filters can support more specific object classes, lower layer filters can help restore an item's general shape. As a consequence, using a deconvolution network will result in a better and more accurate classification result. This architecture allows for improved spatial resolution and better feature extraction, leading to more accurate and detailed predictions.

4. Result and discussion

The time of electricity consumption of days in a week and the investigation was executed every day of a week. The estimates of each user's electricity use are shown in Tab. I, and the performance of the predicted data was compared to actual energy use. The findings demonstrated that the proposed CRF-PCP method's predicted data performs exceptionally well and almost accurately matches the actual data. 100 percent precise information will be acquired toward the finish of the preparation in view of the proposed AI model. From Monday to Sunday, Fig. 3 shows real and anticipated energy utilization. The models show a serious level of understanding among genuine and anticipated energy utilization, demonstrating the model's precision in foreseeing following day's energy utilization. Let's analyze the results and discuss the percentage-wise increase or decrease in consumption. On Monday, the predicted electricity consumption 1130 kilowatt-hour (kWh) consumption shows a slight increase of 0.71% compared to the actual consumption of 1122 kWh. This

Time (day/week)	Actual data (kWh)	Predicted data (kWh)
Monday	1122	1130
Tuesday	1090	1040
Wednesday	1108	1170
Thursday	1069	980
Friday	1190	1100
Saturday	1147	1190
Sunday	1160	1200

Tab. I *Electricity consumption by consumers in India (2016–2017) vs. time.*

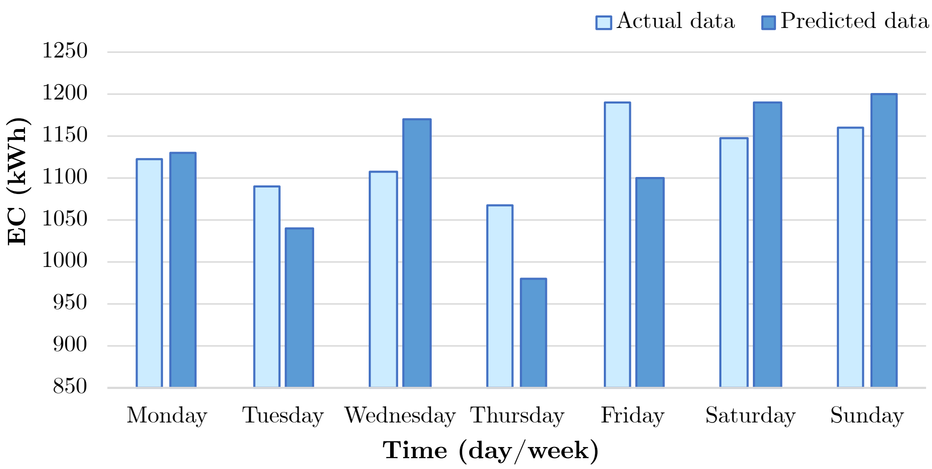

Fig. 3 *Electricity consumption by consumers in India (2016–2017).*

indicates a relatively accurate prediction. Tuesday, however, exhibits a decrease in predicted consumption of 1040 kWh by 4.59% compared to the actual consumption of 1090 kWh. The prediction in this case deviates from the actual values. Moving on to Wednesday, the predicted consumption of 1170 kWh indicates an increase of 5.47% compared to the actual consumption 1108 kWh. This suggests a higher prediction of consumption. Thursday presents a significant decrease in predicted consumption 980 kWh by 8.36% compared to the actual consumption 1069 kWh. The prediction in this case deviates significantly from the actual values. On Friday, the predicted consumption 1100 kWh shows a decrease of 7.56% compared to the actual consumption 1190 kWh. This indicates a lower prediction of consumption. Saturday displays an increase in predicted consumption 1190 kWh by 3.49% compared to the actual consumption 1147 kWh. The prediction in this case slightly overestimates the consumption. Lastly, on Sunday, the predicted consumption 1200 kWh shows an increase of 3.45% compared to the actual consumption 1160 kWh. The prediction in this case slightly overestimates the consumption. In summary, the predicted data generally shows some discrepancies compared to the actual consumption. While Monday and Wednesday have relatively accurate predictions with small increases in consumption, Tuesday, Thursday, and Friday exhibit notable decreases in predicted consumption. Saturday and Sunday show slight overestimations in the predicted consumption. These variations highlight the importance of further refining the prediction models to achieve more accurate results and help in effective electricity management and planning.

The accuracy of CRF-PCP technique is compared with other machine learning algorithms such as SVM Gaussian, complex tree and simple tree.

- SVM Gaussian, a variant of support vector machines, is a popular machine learning algorithm known for its ability to handle both linear and nonlinear data. It can effectively capture complex patterns and relationships in the data, making it suitable for power demand prediction tasks. By applying SVM Gaussian to power demand forecasting, the algorithm can learn from historical data and identify patterns and trends that are indicative of future power consumption. The Gaussian kernel used in SVM allows for the modeling of nonlinear relationships, which is important in capturing the complexities of power demand fluctuations. However, it is important to note that the performance of SVM Gaussian for power demand forecasting depends on various factors, including the quality and representativeness of the training data, the selection of appropriate hyperparameters, and the consideration of other relevant factors that may impact power consumption (such as weather conditions or special events).

- The utilization of a complex tree for power demand prediction in the forecasting of power distribution is an interesting approach. A complex tree, also known as a decision tree algorithm, is a powerful machine learning technique that can handle both classification and regression tasks. In the context of power demand forecasting, complex tree can be trained on historical data to learn patterns and relationships between various factors that influence power consumption. It creates a hierarchical structure of decision nodes based on these factors, enabling it to make predictions for future power demand based on the values of input variables. However, it is important to note that complex tree models can sometimes be prone to overfitting, especially when dealing with complex datasets or datasets with noisy or irrelevant features. Proper care should be taken to prevent overfitting by tuning the model's hyperparameters or applying regularization techniques.

- The utilization of simple tree for power demand prediction in the forecasting of power distribution is a straightforward and intuitive approach. Simple tree, also known as a decision tree algorithm, is a popular machine learning technique that can be applied to various prediction tasks, including power demand forecasting. In the context of power demand prediction, simple tree works by constructing a tree-like model based on historical data. The algorithm analyzes different features or variables that are relevant to power demand, such as time of day, weather conditions, and historical consumption patterns. It then creates a hierarchical structure of decision nodes that recursively split the data based on these features, ultimately leading to predictions of power demand. However, it is important to note that simple tree models can be prone to overfitting, especially when the tree becomes too deep or when the training data is noisy or unrepresentative. Regularization techniques, such as pruning or setting a maximum depth for the tree, can help mitigate this issue.

As indicated in Tab. II, three different machine learning methods were used to predict customer power consumption: SVM, complex tree, and basic tree. The accuracy of various levels of machine learning methods was examined in Fig. 4. For this power determining application, the proposed CRF-PCP strategy predicts accurately with a viable precision of 98.9%, as displayed in Fig. 4. The degree of precision is worked on because of the extra deconvolution step after order. The deconvolution is performed to improve the classification performance further. The impact of depth in CNN helps to increase prediction performance, but introducing too many layers leads to overfitting and may decrease accuracy. In a well-trained network, minimizing training and validation losses is essential. The network is over fitted if the training loss is low, but the validation loss is large. As a result, we used a trial-and-error method to optimize the CNNs, determining the number of hidden layer nodes, rate of learning, size of kernel, and convolution layer numbers. The model is launched with modest convolution layer numbers and progressively increases the number of layers while monitoring the training and validation losses.

Algorithm	Accuracy (%)
Proposed CRF-PCP	98.9
SVM Gaussian	97.0
Complex tree	96.5
Simple tree	96.0

Tab. II *Accuracy comparison.*

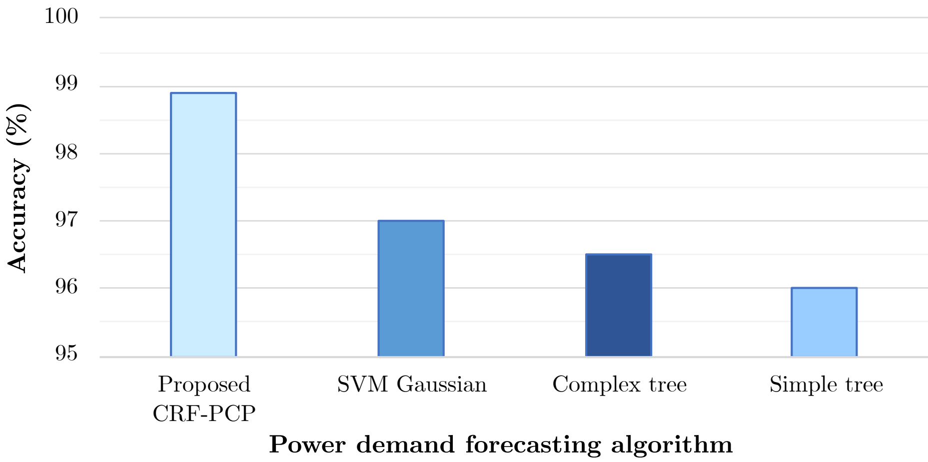

Fig. 4 *Accuracy comparison of proposed and existing algorithms.*

Next, the SVM Gaussian algorithm achieves an accuracy of 97%. Although slightly lower than the proposed CRFPCP algorithm, it still showcases a strong performance in accurately classifying the data. This algorithm proves to be reliable and effective in achieving accurate predictions. The complex tree algorithm attains an accuracy of 96.5%. While it falls slightly behind the top-performing algorithms, it still demonstrates a significant level of accuracy in classification tasks. The complex tree algorithm can be considered as a viable option for achieving reliable results. The simple tree algorithm achieves an accuracy of 96%. Although it exhibits a slightly lower accuracy compared to the other algorithms, it still performs reasonably well in accurately classifying the data. The simple tree algorithm can be a suitable choice for less complex classification tasks. In summary, the proposed CRF-PCP algorithm stands out with the highest accuracy rate of 98.9%. The SVM Gaussian algorithm follows closely with an accuracy of 97%. The complex tree algorithm achieves an accuracy of 96.5%, and the simple tree algorithm attains an accuracy of 96%. These results demonstrate the effectiveness of the algorithms in accurately classifying the data, with each algorithm offering varying levels of accuracy. The choice of algorithm will depend on the specific requirements and complexity of the classification task at hand. The following Tab. III shows the comparison of run time for different datasets. The run time of the proposed CRF-PCP technique is compared with the linear SVM and reduced SVM (RSVM) technique, as shown in Fig. 5.

The working season of the proposed framework is short contrasted with the other two techniques. This is on the grounds that the capability was removed in the past step utilizing the STTM strategy. The prediction and classification of consumers' electricity consumption is made easier by this feature extraction method. Also, the proposed CRF-PCP innovation gives incomplete CRF preparing to quick and precise outcomes. Thusly, the execution season of the proposed CRF-PCP procedure is somewhat low contrasted with other existing techniques. For dataset 1,

Dataset	SVM	RSVM	Proposed CRF-PCP
1	0.600	0.55	0.100
2	0.100	0.30	0.009
3	0.300	0.60	0.010
4	0.100	0.10	0.007
5	0.009	0.50	0.005

Tab. III *Comparison of run time in seconds for different datasets.*

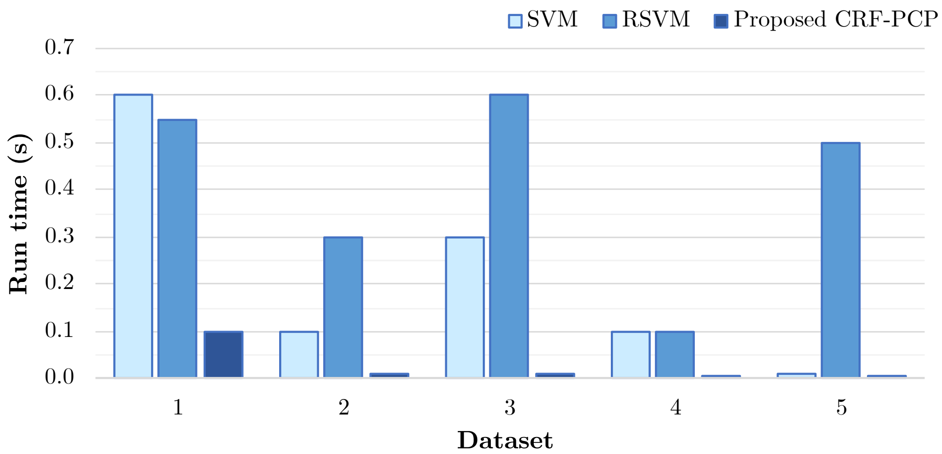

Fig. 5 *Comparison of run time in seconds for different datasets.*

the SVM algorithm takes 0.6 seconds, the RSVM algorithm takes 0.55 seconds, and the proposed CRF-PCP algorithm takes only 0.1 seconds. This indicates a significant improvement in run time with the proposed CRF-PCP algorithm compared to SVM and RSVM, resulting in a decrease of 83.3% and 81.8% respectively. Moving to dataset 2, the SVM algorithm takes 0.1 seconds, the RSVM algorithm takes 0.3 seconds, and the proposed CRF-PCP algorithm takes 0.009 seconds. Here, we observe a decrease in run time by 90% with the proposed CRF-PCP algorithm compared to SVM, and a decrease of 97% compared to RSVM. Dataset 3 shows that the SVM algorithm takes 0.3 seconds, the RSVM algorithm takes 0.6 seconds, and the proposed CRF-PCP algorithm takes 0.01 seconds. The proposed CRF-PCP algorithm demonstrates a decrease of 96.7% in run time compared to SVM, and a decrease of 98.3% compared to RSVM. For dataset 4, the SVM algorithm takes 0.1 seconds, the RSVM algorithm takes 0.1 seconds, and the proposed CRF-PCP algorithm takes 0.007 seconds. Here, the proposed CRF-PCP algorithm showcases a decrease of 93% in run time compared to both SVM and RSVM. Lastly, dataset 5 reveals that the SVM algorithm takes 0.009 seconds, the RSVM algorithm takes 0.5 seconds, and the proposed CRF-PCP algorithm takes 0.005 seconds. The proposed CRF-PCP algorithm displays a decrease of 44.4% in run time compared to SVM, and a decrease of 99% compared to RSVM. In summary, the proposed CRF-PCP algorithm consistently outperforms both SVM and RSVM algorithms in terms of run time across all datasets. It demonstrates significant improvements, with percentage-wise decreases ranging from 81.8% to 99% compared to SVM and RSVM algorithms. These results highlight the efficiency and effectiveness of the proposed CRF-PCP algorithm in terms of run time, making it a favorable choice for applications where fast processing is crucial.

5. Conclusion

This research paper introduces a novel technique called CRF-PCP for forecasting electricity consumption by consumers in different areas. The proposed system demonstrates significant improvements in predicting electricity consumption, achieving 10% higher efficiency. Additionally, the proposed CRF-PCP technique exhibits shorter run time, as validated through comparison with linear SVM and RSVM techniques. This efficiency is achieved through feature extraction using the STTM technique, which facilitates easier prediction and classification of electricity consumption by consumers themselves. In addition, piecewise CRF training is included in the proposed CRF-PCP method to guarantee quick and accurate results. In a resulting work, the review means to explore and dissect the variables impacting power utilization in more detail, in this manner working on the estimating of power utilization.

References

[1] ABERA F.Z., KHEDKAR V. Machine Learning Approach Electric Appliance Consumption and Peak Demand Forecasting of Residential Customers Using Smart Meter Data. Wireless PersCommun, 2020, 111, pp. 65–82, doi: 10.1007/s11277-019-06845-6.

[2] ALAM F.I., ZHOU J., LIEW A.W., JIA X., CHANUSSOT J., GAO Y. Conditional Random Field and Deep Feature Learning for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(3), pp. 1612–1628, doi: 10.1109/TGRS.2018.2867679.

[3] CHEN X., CAI X., LIANG J., LIU Q. Ensemble Learning Multiple LSSVR with Improved Harmony Search Algorithm for Short-Term Traffic Flow Forecasting. IEEE Access, 2018, 6, pp. 9347–9357, doi: 10.1109/ACCESS.2018.2805299.

[4] CHOU J., HSU S., NGO N., LIN C., TSUI C. Hybrid Machine Learning System to Forecast Electricity Consumption of Smart Grid-Based Air Conditioners. IEEE Systems Journal, 2019, 13(3), pp. 3120–3128, doi: 10.1109/JSYST.2018.2890524.

[5] DASS J., SARIN V., MAHAPATRA R.N. Fast and Communication-Efficient Algorithm for Distributed Support Vector Machine Training. IEEE Transactions on Parallel and Distributed Systems, 2019, 30(5), pp. 1065–1076, doi: 10.1109/TPDS.2018.2879950.

[6] FENZA G., GALLO M., LOIA V. Drift-Aware Methodology for Anomaly Detection in Smart Grid. IEEE Access, 2019, 7, pp. 9645–9657, doi: 10.1109/ACCESS.2019.2891315.

[7] ZHANG J.B., OSTENDORF M., KIRSCHEN D.S. Real-Time Prediction of the Duration of Distribution System Outages. IEEE Transactions on Power Systems, 2019, 34(1), pp. 773–781, doi: 10.1109/TPWRS.2018.2860904.

[8] KAMAROL S.K.A., JAWARD M.H., PARKKINEN J., PARTHIBAN R. Spatiotemporal feature extraction for facial expression recognition. IET Image Processing, 2016, 10(7), pp. 534–541, doi: 10.1049/iet-ipr.2015.0519.

[9] MOTEPE S., HASAN A.N., STOPFORTH R. Improving Load Forecasting Process for a Power Distribution Network Using Hybrid AI and Deep Learning Algorithms. IEEE Access, 2019, 7, pp. 82584–82598, doi: 10.1109/ACCESS.2019.2923796.

[10] PANG J., HUANG J., DU Y., YU H., HUAN Q., YIN B. Learning to Predict Bus Arrival Time from Heterogeneous Measurements via Recurrent Neural Network. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(9), pp. 3283–3293, doi: 10.1109/TITS.2018.2873747.

[11] PARVEZ I., AGHILI M., SARWAT A.I., RAHMAN S., ALAM F. Online power quality disturbance detection by support vector machine in smart meter. J. Mod. Power Syst. Clean Energy, 2019, 7, pp. 1328–1339, doi: 10.1007/s40565-018-0488-z.

[12] RAHULAMATHAVAN Y., PHAN R.C., VELURU S., CUMANAN K., RAJARAJAN M. Privacy-Preserving Multi-Class Support Vector Machine for Outsourcing the Data Classification in Cloud. IEEE Transactions on Dependable and Secure Computing, 2014, 11(5), pp. 467–479, doi: 10.1109/TDSC.2013.51.

[13] TURC T. Using WEB Services in SCADA Applications. Procedia Technology, Elsevier, 2015, 19, pp. 584–590, doi: 10.1016/j.protcy.2015.02.083.

[14] VENKATESAN P. KARTHIGAIKUMAR PAUL A., SATHEESKUMARAN S., KUMAR R. ECG Signal Preprocessing and SVM Classifier-Based Abnormality Detection in Remote Healthcare Applications. IEEE Access, 2018, 6, pp. 9767–9773, doi: 10.1109/ACCESS.2018.2794346.

[15] WANG K., XU C., ZHANG Y., GUO S., ZOMAYA A.Y. Robust Big Data Analytics for Electricity Price Forecasting in the Smart Grid. IEEE Transactions on Big Data, 2019, 5(1), pp. 34–45, doi: 10.1109/TBDATA.2017.2723563.

[16] YUAN Y., DEHGHANPOUR K., BU F., WANG Z. A Multi-Timescale Data-Driven Approach to Enhance Distribution System Observability. IEEE Transactions on Power Systems, 2018, pp. 0885–8950, doi: 10.1109/TPWRS.2019.2893821.