

## Research Article

# An Improved Load Forecasting Method Based on the Transfer Learning Structure under Cyber-Threat Condition

Luo Zhao <sup>1</sup>, Xinan Zhang,<sup>2</sup> Yifu Chen,<sup>3</sup> Xiuyan Peng,<sup>1</sup> and Yankai Cao <sup>4</sup>

<sup>1</sup>College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China

<sup>2</sup>Department of Electrical, Electronic and Computer Engineering, The University of Western Australia, Perth 6009, Australia

<sup>3</sup>Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison 53706, USA

<sup>4</sup>Department of Chemical and Biological Engineering, University of British Columbia, Vancouver V6T 1Z3, Canada

Correspondence should be addressed to Luo Zhao; zhaoluo425@163.com and Yankai Cao; yankai.cao@ubc.ca

Received 6 June 2022; Revised 5 July 2022; Accepted 11 July 2022; Published 24 August 2022

Academic Editor: Aboul Ella Hassanien

Copyright © 2022 Luo Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Smart grid is regarded as an evolutionary regime of existing power grids. It integrates artificial intelligence and communication technologies to fundamentally improve the efficiency and reliability of power systems. One serious challenge for the smart grid is its vulnerability to cyber threats. In the event of a cyber attack, grid data may be missing; subsequently, load forecast and power planning that rely on these data cannot be processed by generation centers. To address this issue, this paper proposes a transfer learning-based framework for smart grid scheduling that is less reliant on local data while capable of delivering schedules with low operating cost. Specifically, the proposed framework contains (1) a power forecasting model based on transfer learning which can provide high quality load prediction with limited training data, (2) a novel adaptive time series prediction method with modeling time series from a covariate shift perspective that aims to train the forecasting model with a strong generalization capability, and (3) a day-ahead optimal economic power scheduling model considering a shared energy storage station.

## 1. Introduction

In recent years, the emergence of renewables and big data has prompted a reform of the electrical network. As a consequence, the concept of a smart grid becomes increasingly popular [1]. A smart grid is defined as the next generation electrical grid with power-flow control, self-healing, and energy reliability using digital communications. Compared to the conventional power system, the smart grid is designed to integrate millions of smart sensors and advanced computing technologies into the whole grid [2]; it can efficiently realize real-time automatic control, intelligent regulation, online analysis and decision making, cooperative interaction, and other advanced functions of the power grid. One feature of the smart grid is the high share of renewable generation, which poses a threat to its reliability due to the renewable intermittency. One widely adopted solution to this problem is to employ the energy storage system (ESS) in microgrids. Since wind and photovoltaics power are

nondispatchable parts, the dispatching strategy of ESS becomes an important component in the smart grid with renewable generation.

Most existing energy storage planning approaches rely on historical data or highly precise generation/load forecast [3, 4]. Current power forecasting methods can be broadly categorized into two types: methods based on statistical analysis [5] and methods based on artificial intelligence (AI) algorithms [6]. For methods in the first category, the Bayesian theory, multivariable linear regression, and autoregressive moving average (ARIMA) are often employed [7, 8]. These methods are better suited for fitting data that have periodic features; the high percentage of renewable energy sources considerably increases the randomness of power variations, making methods in this category less suitable for applications in smart grids. Methods based on AI algorithms, on the other hand, are theoretically favorable in predicting an output for systems with high nonlinearity and complex dynamic properties. In particular,

a recurrent neural network (RNN) [9] helps handle non-linear problems and capture more dynamic relationships between the input and the forecasted output compared to the artificial neural network (ANN) [10], and the long short-term memory (LSTM) unit proposed in [11] further improves its performance in the prediction of time series data.

A noteworthy concern is that all the abovementioned methods require sufficient training data. The heavy reliance on information networks leads to higher cyber risks for smart grids [12]. In the event of a cyber-attack, the local database could be tampered and lost, which leads to serious consequences [13]. For instance, Ukraine's electricity supply system was hacked in 2017 through an attack on the data aggregator, which is the central node containing data from data collection base stations [14]. This event caused massive power outages, paralyzed connected nodes, and prevented the control center from accessing customer load data in time for power control and scheduling operations. It is of great interest to reduce the grid control centers' heavy reliance on local users' data, so that the day-ahead or long-term power dispatch would not be blocked by insufficient data. The primary issue that needs to be addressed is load forecasting based on inadequate local data.

Transfer learning (TL) [15, 16] is a suitable tool for addressing the challenges discussed above. It is of great value to introduce transfer learning in power systems to efficiently utilize resources from different regions, discover the commonality of different datasets, and establish transfer learning-based forecasting methods. The key idea of transfer learning is to use the existing experience to solve similar tasks, exploit similarities between data and models, and apply the trained content to new tasks. Specifically, transfer learning allows the use of knowledge in the dataset with complete labels (i.e., the source domain) to solve problems in the dataset with missing labels (i.e., the target domain), using a trained model with good generalization capability [17]. The research in [18] presents the outstanding contribution of transfer learning in the field of image processing. Innovative breakthroughs have also been made in the field of classification and target detection [19] in recent years. Lu et al. [20] proposed a general transfer learning-based framework for load forecasting with limited data. The influence of adopting different kernel functions in transfer learning for fault diagnosis is studied by Li et al. in [21]. Scholars in [22] investigate the superiority of transfer learning in extracting features and aim to predict the wind speed in different environments. Yin et al. [23] proposed a hybrid transfer learning-based wind power forecasting model. Unfortunately, the potential relationship between statistical properties in time series and transfer learning is ignored in these works.

One critical issue for developing transfer learning-based forecasting methods is how to train a forecasting model with strong generalization capabilities. Many published forecasting methods for smart grids are based on the assumption that historical data follow the same distribution. The scholars in [24] made great improvements in load forecast based on machine learning in certain areas, but the generalization ability of their proposed method is not very promising since

the differences between distributions of data are not considered. In a typical grid, especially those with high penetration of the renewable system that introduces high stochasticity, the distribution of the data in the temporary structure changes over time. Consider the illustrative graph in Figure 1, the probability distribution of  $P_x$  varies for different intervals, and the temporal covariate shift phenomenon happens after adding a new segment of data, where  $P_a \neq P_b \neq P_c \neq P_{test}$ . Here, the aforementioned issue is embodied in two aspects. First, how to build an adaptive prediction model to weaken the effect of covariate shift and accommodate the diversity of sample data. Second, how to develop a probability distribution algorithm to minimize the divergence between the distribution for different intervals.

Load prediction provides basic data for generation planning, day-ahead market offers, and intraday market trading, and it is important for the economic dispatch of power system. In the wind/photovoltaic/energy storage complementary microgrid, the generation plan of energy storage is the only dispatchable part. Researchers in [25] proposed a general method for the capacity and power of energy storage batteries and constructed a capacity allocation scheme for energy storage batteries, researchers in [26] introduced the control and communication technology, and operation principle of cloud energy storage based on the Irish power system. At present, the research on shared energy storage is in its initial stage, and the existing work takes shared energy storage systems as the main research object to analyze the business model and profitability of shared energy storage systems, while in-depth research on the charging and discharging behavior and economic benefits of users' participation in the shared energy storage system is limited. This paper introduces shared energy storage plants among different user groups and establishes an optimal scheduling model with the objective to minimize daily operation cost of user groups.

The main contributions of this paper are summarized as follows:

- (1) To address the challenge that the generation center fails to develop power planning for grid operation due to inadequate local user data, we propose a power forecasting method based on transfer learning, where data from the source domain can provide valuable reference information. Additionally, case studies provide a detailed analysis of how to choose the appropriate source domain, and the effect of negative transfer on model performance is analyzed.
- (2) This work proposes to model the time series of load prediction from the covariate shift perspective. To train a forecasting model with strong generalization capability with transfer learning, we generate a combination mode where the probability distributions of time series vary for different intervals, and an optimal split method is proposed to ensure that the segments being divided are the most dissimilar ones. The temporal distribution matching algorithms are proposed to minimize the divergence between the distribution for different intervals. Dynamic

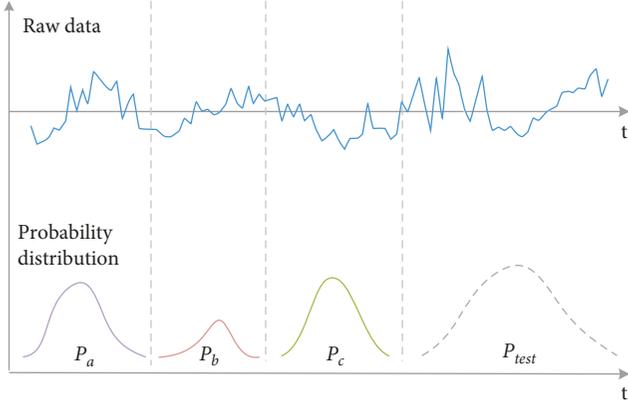


FIGURE 1: The temporal covariate shift phenomenon.

programming (DP) is applied to optimize the optimal division points. The case study shows that the maximum improvement of the proposed forecasting method is up to 52.8% in mean absolute percentage error (MAPE) compared to other transfer learning-based methods, and up to 64.4% compared to the traditional method.

- (3) With the accurate load prediction obtained from (1) and (2), a novel shared energy storage station (ESS) concept is proposed to form a framework for optimal scheduling based on the transfer learning method. The case study shows that the proposed framework can reduce the overall operating cost of the micro-grid and maximize the benefit of grid operation, thus addressing the issue of energy curtailment [27] as well as the high cost of energy storage.

This paper is organized as follows: Section 2 presents the structure of the transfer learning-based forecasting algorithm with the limited data set, and modeling time series from a covariate shift perspective, which is critical to determining the generalization ability of forecasting models. Section 3 proposes a framework of optimal dispatch planning for the distributed microgrid, a shared energy storage station (ESS) is formed as a solution for multisource power grid scheduling, and the dispatchability of the cyber-attacked

area is analyzed when using the proposed optimal economic dispatching method. The performance of transfer learning in addressing fragmented test data of the target domain is then developed. Moreover, using data obtained from the proposed forecasting method, the economic analysis of the proposed energy storage station is shown in Section 4. Finally, the main findings are included in Section 5.

## 2. Methodology of the TCS-Transfer Learning Model

**2.1. Transfer Learning-Based Structure.** In most machine learning tasks, the training and test sets come from the same feature space and are subject to the same probability distribution. When the actual conditions are not satisfied, it takes a lot of resources to collect data from the target domain to retrain a model. In particular, in a feature-rich smart grid, retraining the model becomes inefficient and time-consuming, which can seriously disrupt the schedule of the generation center, transfer learning [28] is a new solution to this problem.

Maximum mean discrepancy (MMD) is applied in this paper to define a more specific formula for the TL problem:

$$f^* = \arg \min_{f \in H} \frac{1}{N_s} \sum_i^{N_s} l(f(x_i), y_i) + \xi \text{MMD}(D_s, D_t), \quad (1)$$

where  $H$  represents the vector space that satisfies the objective function,  $N_s$  means the number of samples of the source domain  $D_s$ , and  $D_t$  means the source domain and target domain.  $N_t$  represents the number of samples of the target domain, and  $l(x, y)$  denotes the loss function suitable for different structures, and  $x$  and  $y$  represent the sample and label, respectively. Argmin denotes the value of the variable that minimizes the objective function.  $\xi$  represents the metric coefficient. This paper selects root mean squared error (RMSE) as the measurement of error. It is considered that MMD is one of the most widely used distance metrics in TL among many statistical measures, which is an effective way to measure the correlation between any two different domains  $D_i$  and  $D_j$ , and it can be formulated using the following equations:

$$\text{MMD}[\mathbb{F}, D_i, D_j] = \left[ \frac{1}{m(m-1)} \sum_{i \neq j}^m k(x_i, x_j) + \frac{1}{n(n-1)} k(y_i, y_j) - \frac{2}{mn} \sum_{i,j=1}^{m,n} k(x_i, y_j) \right]^{1/2}, \quad (2)$$

$$k(x, x') = e^{-\|x-x'\|^2/2\sigma^2}, \quad (3)$$

where  $D_i$  and  $D_j$  represent any two different domains, and their sample sizes are  $m$  and  $n$ .  $k(\cdot, \cdot)$  represents the Gaussian kernel function. Radial basis kernel function  $k(x, x')$  is used for mapping to the higher dimensional spaces.  $x'$  is the center point of the kernel function,  $\sigma$  denoted the

expectation, which controls the range of action of the Gaussian kernel function; the larger its value, the larger the local range of the influence of the Gaussian kernel function.  $\mathbb{F}$  denotes the unit ball in the regenerative nuclear Hilbert space (RNHS).

It is worth noting that not all source domains are suitable to be selected, Krizhevsky et al. [29] provides an in-depth analysis of the role of pretraining models for transfer tasks, a pretraining model can be used as benchmark models for the task of the target domain, researchers in [30, 31] also demonstrate that for datasets with different distributions, pretraining in an appropriate source domain can greatly improve the accuracy of the results, and source domain that has a weak correlation with the target domain may cause a negative transfer phenomenon. Therefore, it is essential to analyze the discrepancy between the  $\mathcal{D}_s$  and  $\mathcal{D}_t$ ; this paper measures the difference by calculating the MMD value, and the source domain can be selected for pretraining if the MMD is below the preset threshold value.

### 2.2. Problem Formulation of Temporal Covariate Shift (TCS).

The variate  $\{x_i\}$  of the time series is assumed to follow the same probability distribution in most existing prediction methods, it may have achieved satisfactory results on specific datasets, such as the load of a stand-alone device in a traditional predict scene, which is relatively uncomplicated in its diversity. However, this assumption is not realizable in the actual application due to the huge amount of data and features, the variation of data distributions with the time changing cannot be ignored. Thus, our problem can be formulated as follows: split a given time series  $\mathbb{S}$  with  $m$  labeled dataset into  $k$  segments with the most dissimilar distribution  $\{\mathbb{S}_1 \dots \mathbb{S}_k\}$ .

$$\begin{cases} \mathbb{S} = \{\mathbb{S}_1, \dots, \mathbb{S}_p \dots \mathbb{S}_q \dots, \mathbb{S}_k\}, \\ \sum \mathbb{S} = m. \end{cases} \quad (4)$$

With reference to the definition of covariate drift [32] in the classification field, the temporal covariate shift can be presented such that the whole intervals in the same period  $i$  follow the same probability distribution  $P_{\mathbb{S}_{p,q}}(x, y)$ . The distribution will be different when the time period changes  $P_{\mathbb{S}_p}(x) \neq P_{\mathbb{S}_q}(x)$  ( $p \neq q \in (1, k)$ ). To train a prediction model with excellent generalization performance under temporal covariate shift, the main issue is to capture the common knowledge shared among different periods of  $\mathbb{S}_{p,q}$ .

According to the principle of maximum entropy [33, 34], finding intervals that are the most distinct from each other can help maximize the capacity of shared information within a time series under temporal covariate change issues. It is fair to make distributions of each interval as diverse and feasible to maximize the entropy of the overall distributions of an array. This enables for more generic and adaptable future data modeling. Worst-case training using the original sequence enables the model to cope with the stochastic nature of the unknown data. Figure 2 shows the structure of the proposed predicting method based on transfer learning, and the primary task is to split the time series into  $k$  segments. The splitting problem of the time series in (5) is solved by the greedy algorithm.

$$\max_{m_1, \dots, m_k} \frac{1}{k} \sum_{1 \leq \lambda_i \neq \lambda_j \leq k} MMD(\mathbb{S}_p, \mathbb{S}_q), \quad (5)$$

$$s.t. \forall i, \quad \Delta_1 \leq |\lambda[j] - \lambda[i]| \leq \Delta_2,$$

where array  $\lambda[1 \dots m]$  represents a set of time series,  $m$  represents the length of it,  $MMD(\cdot, \cdot)$  means the distribution-based distance function to measure the distance between distributions of any two segments  $\mathbb{S}_p, \mathbb{S}_q$ ,  $\lambda[c], c \in [i, j]$  denotes the coordinate position of the cut at each time.  $\lambda[i]$  and  $\lambda[j]$  are the left and right endpoints of the segment to be segmented, respectively.  $\lambda[j]_{-1}$  is the right endpoint of the previous interval and  $\lambda[i]$  is the start point of the next interval.  $\Delta_1$  and  $\Delta_2$  represent predefined constants to prevent trivial solutions.

**2.3. Matching Process and Fine-Tune.** The proposed method is designed to acquire the common information shared by distinct intervals by comparing their probability distributions, this section presents the process of how to pretrain a model after obtaining the optimal split intervals in raw data. In comparison to approaches that simply depend on local or statistical knowledge, the pretraining model  $\mathcal{M}$  can produce a nice generalization on unknown datasets of the target domain. The pretraining issue formulated in Section 2.2 can be solved by the domain generalization (DG) [35] method, and the distribution matching loss of the network can be established as follows:

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^k \mathcal{L}_{\text{pre}}(y_i, \hat{y}_i) + \xi \sum_{1 < i < j < k} MMD(\mathbb{S}_i, \mathbb{S}_j), \quad (6)$$

where  $\mathcal{L}(\cdot, \cdot)$  denotes the MSE loss in the source domain, and  $\mathbb{S}_i$  and  $\mathbb{S}_j$  are any two different intervals of  $\lambda[1 \dots m]$ . In the proposed method, LSTM is employed as the main body of the network structure. Due to the special memory unit of LSTM, potential relationships between data in a time series can be preserved, which can provide high accuracy for the prediction results. Compared to the conventional recurrent neural network algorithm, which contains only one state  $h_t$ , the LSTM structure introduces cell states  $c_t$  to develop potential relationships in a long time series. The LSTM structure can be formulated as follows:

$$\begin{aligned} i_t &= \sigma(W_i * [h_{t-1}, x_t] + b), \\ \tilde{c}_t &= \tanh(W_c * [h_{t-1}, x_t] + b), \\ c_t &= f_t \odot c_{t-1} + i_t * \tilde{c}_t, \\ o_t &= \sigma(W_o [h_{t-1}, x_t] + b), \\ h_t &= o_t * \tanh(c_t), \end{aligned} \quad (7)$$

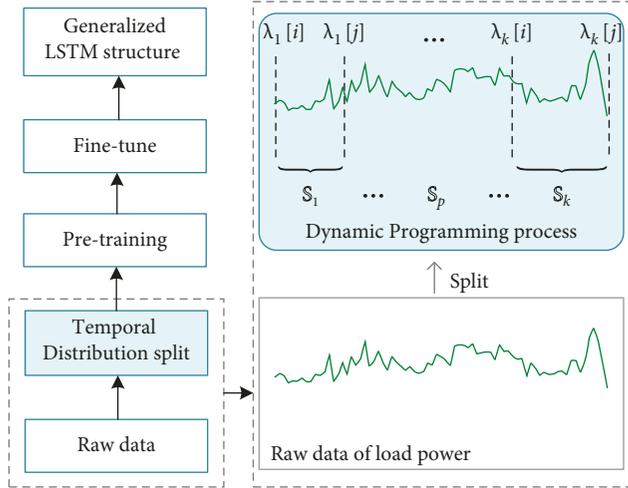


FIGURE 2: Schematics of the proposed predicting structure based on transfer learning.

where  $x_t$  is the input data,  $W_{i,c,o}$  represent the weight matrices,  $i_t$ ,  $f_t$ , and  $o_t$  are the input, forget, and output gates of the LSTM structure, respectively.  $b$  represents the bias value.  $c_{t-1}$  denotes the state of the memory cell. The candidate value  $\tilde{c}_t$  is generated by tanh layer.  $h_t$  presents the output value,  $h_{t-1}$  represents the output value of the previous unit, the sigmoid function is denoted as  $\sigma$ , and  $*$  represents the dot product.  $\odot$  is the element-wise product. Figure 3 shows the structure of LSTM; the type of data being inputted is described in Section 4.

Figure 4 shows the flowchart of the proposed transfer learning-based forecasting structure. The implementation steps of the whole network are as follows:

- (1) Start by collecting the raw data of power fluctuations from neighboring cities and calculating the probability distribution of the candidate datasets
- (2) Calculate the distance between the source and target domains according to eq. (2) and select the appropriate source domain as the input data of the pretrain model
- (3) For the selected source domain, it is first divided into  $k$  most dissimilar segments using the proposed dynamic programming-based method according to eqs. (4) and (5).
- (4) The  $k$  segments are considered as different domains. Eq. (6) is used as the new loss function in the model with LSTM as the main network. After obtaining a prediction network with strong generalization capability, the target domain with very less data is used to fine-tune by using Eq. (1)–(3).

The proposed temporal covariate shift issue focuses on an easily neglected problem in time series Dynamic programming solves the optimal spilt problem of the time series, considering the fragment as  $k$  independently distributed individuals. The generalization ability of the pretraining model using source domain data is greatly improved by taking the differences between these segments as an

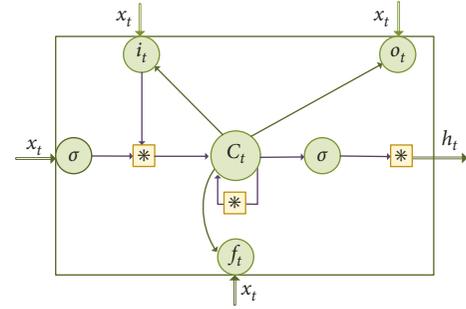


FIGURE 3: LSTM unit structure.

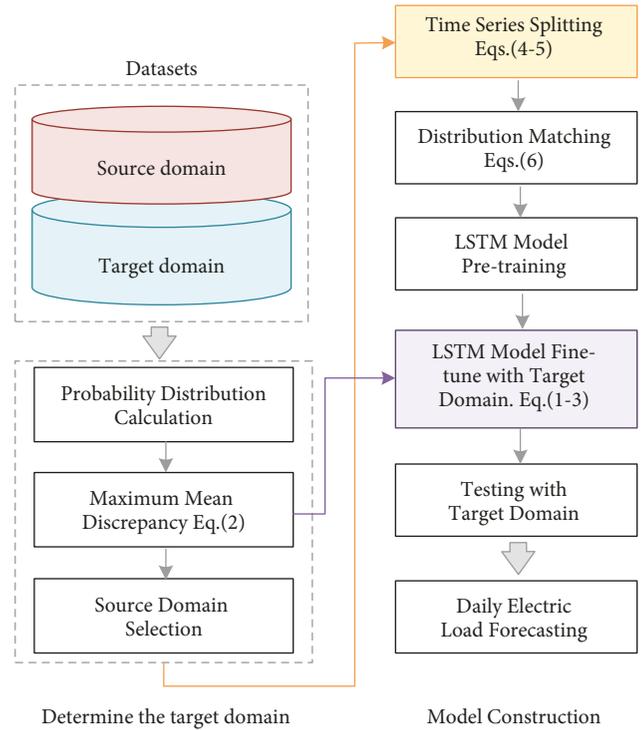


FIGURE 4: Power forecasting structure based on the TCS-transfer learning model.

additional consideration in the loss function. The deep LSTM network can be trained with less learning time while overcoming the problem of lacking sufficient local data as training datasets.

### 3. Optimal Dispatch Planning for Distributed Microgrid

**3.1. Smart Grid with Multiple Sources of Energy Supplies.** Smart grids are increasingly used in recent years for both residential and industrial purposes. In many microgrids, the hybrid wind-solar generation plant is employed owing to the complementary nature of wind and solar generation patterns. Therefore, the smart grid with renewable generations is studied in this paper as shown in Figure 5. From the perspective of the Internet of Things (IoT) network structure of the smart grid, the data is collected by a large number of

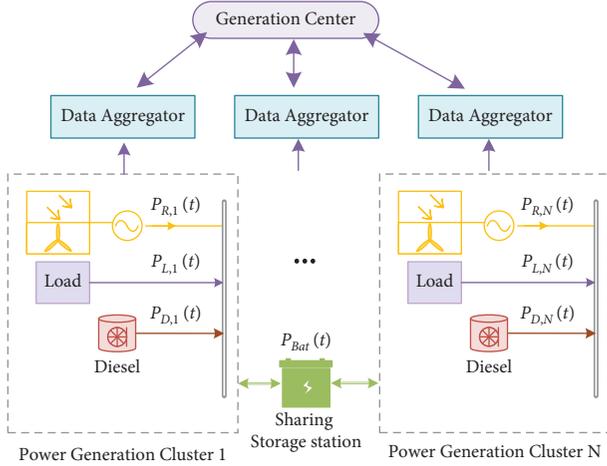


FIGURE 5: Conceptual framework of the multienergy complementary power generation system.

smart meters in the users' network, encrypted by the gateway, gathered by the data aggregator, and then transmitted to the control center, and this network data distribution characteristic is called the "funnel effect" [36]. In a smart grid with deeply integrated information systems, the form of fault propagation has more possibilities. When a node in the network fails, it will trigger the failure of the related nodes, and the existing scheme requires all users to work collaboratively. As long as one user fails, the whole system cannot operate normally. The transfer learning-based forecasting method proposed in this paper can substantially reduce the dependence of generation centers on the adequacy of local data.

The generation plant of the microgrid is divided into  $N$  clusters, each having wind turbine and solar photovoltaic (PV) systems. The output power of the wind/PV system in the  $j$ -th cluster can be expressed as follows:

$$P_{G,j}(t) = P_{W,j}(t) + P_{PV,j}(t) + P_{D,j}(t), \quad j \in \{1, 2, \dots, N\}, \quad (8)$$

where  $P_{W,j}$ ,  $P_{PV,j}$ , and  $P_{D,j}$  represent the output power of wind turbines, photovoltaic systems, and diesel systems, respectively. In modern microgrids, generators are usually used as backup power sources, the term  $P_{D,j}(t)$  in (8) can be ignored if no power failure occurs. It can be further expressed in per unit (p.u.) value denoted by  $P_{*G,j}(t)$  as eq. (9), where  $P_{G,base}(t)$  denotes the base power in the DC bus,  $P_{G,j}(t)$  is the actual power value of the DC bus, and  $P_{Bat}(t)$  represents the power of the battery.

$$\begin{cases} P_{*G,j}(t) = \frac{P_{G,j}(t)}{P_{G,base}(t)}, \\ P_{PV,j}(t) + P_{W,j}(t) + P_{D,j}(t) + P_{Bat}(t) = P_{L,j}(t). \end{cases} \quad (9)$$

**3.2. Operation Mode and Market Rules of a Shared Station.** The concept of a shared energy storage plant is shown in Figure 5, in which the operator of an energy storage station

uses the financial advantage to establish a large shared energy storage plant among a group of customers, and unifies the operation and management of the energy storage station to provide shared energy storage services to multiple customers in the same distribution network area. The generation center forecasts the load based on historical electricity consumption data and plans to use shared energy storage plants for charging and discharging, which minimizes the economic cost of operation of the storage devices and saves the customer the investment cost of installing and maintaining the storage devices. Power market rules require the dispatch plan submitted to the grid operator 24 hours ahead.

Figure 6 depicts the framework of the proposed power dispatch schedule. The main purpose of this work is to solve the issue that the predicting process cannot operate due to insufficient local data, thus, the power data of the neighbor area is employed to train highly generalizable models, and the precise prediction model can be obtained after fine-tuning with few local data.

The generation center of the energy shared storage power station transmits the remaining power of users who need charging directly to users who need discharging according to the charging demand and discharging demand of each user at each time. The generation station of energy storage station can utilize the complementary nature of customers' power consumption behavior, i.e., the difference in power consumption behavior of the same customers at different times and different customers at the same times, and can maximize the investment in the least amount of energy storage to meet customers' demand for energy storage use. The specific optimization strategy of shared storage stations is detailed in Section 3.3.

**3.3. Optimal Economic Scheduling of Shared Energy Storage Station.** Having explained the forecasting method and the possible operational modes of the microgrid, the next task is to develop the dispatch strategy for the shared ESS to economically benefit users. The high investment cost of energy storage is the main reason that limits the application of energy storage technology on the demand side of the grid. This paper proposed the concept of a shared energy storage station, as shown in Figure 7, which is applied to the economic optimization scheduling of regional users, and the minimum daily operating cost of the user group is achieved by coordinating the charging and discharging power of the users. The energy storage system allows users to store electricity during the grid valley hours and release it during peak hours, thereby decreasing electricity costs and relieving the pressure on regulating the peak load. According to the charging demand and discharging demand of each user in each period, the generation center will deliver the remaining electric energy of the user who needs to discharge directly to the user who needs to charge. If the total charging and discharging demand of users in the same time period is discharged, the regulation center will decide whether the users' electricity needs to be purchased by the main grid or stored in the shared power station according to the electrovalence at that time.

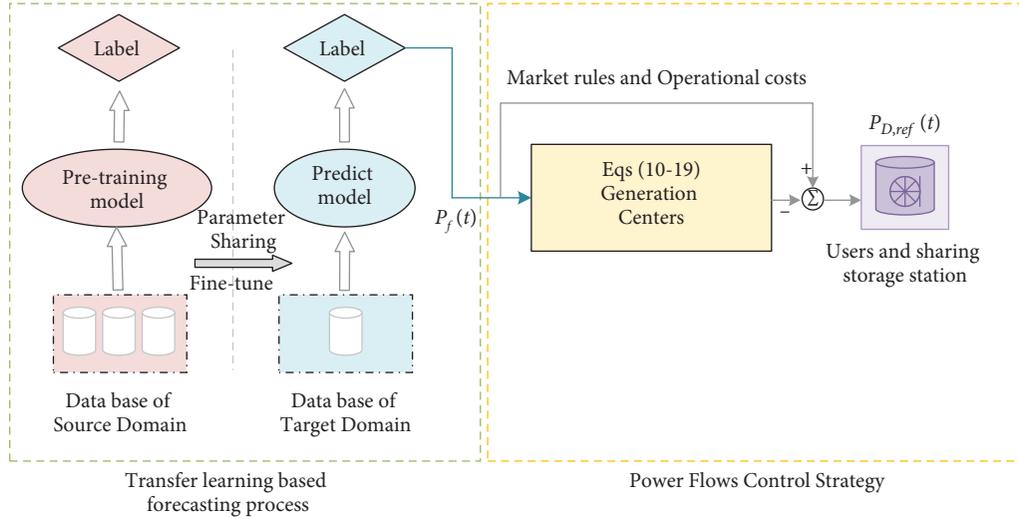


FIGURE 6: Schematics of the proposed optimal power planning procedure.

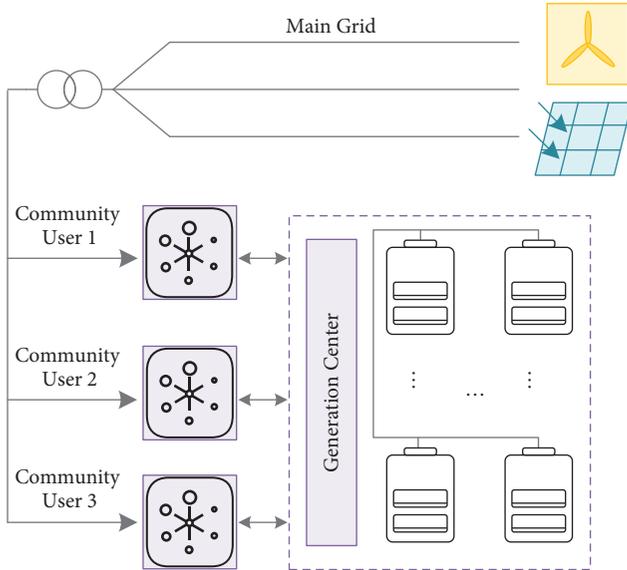


FIGURE 7: Schematic diagram of the shared energy storage station.

Based on the predicting method proposed in the previous sections, this part develops a scheme to obtain the minimum grid operating cost through proper charging and discharging of the shared energy storage station. By using the shared energy storage station, the user saves the investment costs for the installation and maintenance of energy storage devices. Users pay the service fee to the generation center in exchange for shared energy storage services. The service fee means the users pay to the generation center when they use the shared energy storage stations for charging and discharging, it is set as 0.16\$/kWh.

**3.3.1. Objective Function of the Optimization Model.** According to the market rules presented in Section 3.2, the user group connected to the shared energy storage station uses the typical daily operating cost optimization as the

objective function to determine the capacity, the maximum charging and discharging power of the energy storage station, and the charging and discharging power of the storage station for each time period of the user. The daily operating cost of the customer group includes the cost of electricity purchased from the grid and the service fee paid to the energy storage station.

$$\min C = C_g + C_s, \quad (10)$$

where  $C$  represents the daily cost of electricity for the user community;  $C_g$  and  $C_s$  denote the cost of electricity purchased by the customer from the grid and the service fee paid to the energy storage station, respectively.

$$\begin{cases} C_g = \sum_{i=1}^N \sum_{t=1}^T \rho(t) \cdot P_{G,i}(t) \cdot \Delta t, \\ C_s = \sum_{i=1}^N \sum_{t=1}^T \delta(t) \cdot [P_{E,D,i}(t) + P_{E,C,i}(t)] \cdot \Delta t, \end{cases} \quad (11)$$

where  $N$  represents the serial number of the user group connected to the same shared energy storage station, user groups of three areas are selected as case studies in this paper ( $i \in [1, 3]$ );  $\rho(t)$  (\$/kW · h) represents the price that the users purchase electricity from the grid.  $T$  denotes the scheduled time periods;  $P_{G,i}(t)$  indicates the power value purchased by the user  $i$  from the grid at a given time interval  $t$ ;  $\Delta t$  represents the unit time length of the power scheduled;  $\delta(t)$  is the service fee of the shared energy storage station;  $P_{E,D,i}(t)$  and  $P_{E,C,i}(t)$  are discharge power and charging power of energy storage station at time  $t$ , respectively.

**3.3.2. Constraint Condition.** The constrains should be met by the proposed power planning model. They include the following: electrical power balance constraints [26] and operational constraints of the energy storage stations:

- (1) Power balance constrain of the whole grid:

$$P_G(t) + P_{PV,i}(t) + P_{W,i}(t) + P_{ESS}(t) - P_{L,i}(t) = 0, \quad (12)$$

$$P_{ESS} = P_{E,D,i}(t) - P_{E,C,i}(t) = \text{MAX}(P_{W,i}(t) + P_{PV,i}(t) - P_{L,i}(t)),$$

where  $P_G(t)$  is the power purchased from the main network;  $P_{PV,i}(t)$  represents the output power of the PV system of the  $i$ -th user in time interval  $t$ ;  $P_{W,i}(t)$  is the wind power of the  $i$ -th user in time interval  $t$ .  $P_{L,i}(t)$  denotes the load power of the  $i$ -th user, which is a predicted value obtained by the method proposed in the previous section;  $P_{ESS}$  is the power of the energy storage station, which satisfies the maximum power difference between generation and load in the grid for any period of time;  $P_{E,D,i}(t)$  and  $P_{E,C,i}(t)$  represent the power under discharging and the charging status of the energy storage system.

- (2) Charging and discharging power constraints for a shared energy storage station:

$$\begin{cases} 0 \leq P_{E,D,i} \leq P_{\max} \delta_{ESS\_D}^{\max}, \\ 0 \leq P_{E,C,i} \leq P_{\max} \delta_{ESS\_C}^{\max}, \\ \delta_{ESS\_D}^{\max} = \frac{SOC - SOC_{\min}}{SOC_l - SOC_{\min}}, \\ \delta_{ESS\_C}^{\max} = \frac{SOC_{\max} - SOC}{SOC_{\max} - SOC_h}, \end{cases} \quad (13)$$

$$\begin{cases} SOC = \left( SOC_0 - \frac{1}{Q \int_0^t \eta I dt} \right) \times 100\%, \\ SOC_{\min} \leq SOC \leq SOC_{\max}, \end{cases} \quad (14)$$

where  $P_{\max}$  is the rated maximum power of the charging and discharging power of the energy storage station;  $\delta_{ESS\_D}^{\max}$  and  $\delta_{ESS\_C}^{\max}$  are defined as the discharging and charging state factor, respectively, which ensure that the energy storage is not in an overcharged and discharged state;  $SOC_{\min}$  and  $SOC_{\max}$  represent the operating range of the energy storage;  $SOC_l$  and  $SOC_h$  represent the optimal working interval for energy storage;  $SOC$  denotes the state of the charge value of the energy storage;  $SOC_0$  is the initial state of charge;  $\eta$  is the charge and discharge efficiency;  $Q$  represents the electric charge quantity;  $I$  is the battery current.

- (3) Power balance constrain of the energy storage station:

$$\sum_{i=1}^N |P_{E,D,i}(t) - P_{E,C,i}(t)| = |P_D(t) - P_C(t)|, \quad (15)$$

$$\begin{cases} \delta_{ESS\_D}^{\max} + \delta_{ESS\_C}^{\max} \leq 1, \\ \delta_{ESS\_D}^{\max} \in \{0, 1\}, \delta_{ESS\_C}^{\max} \in \{0, 1\}. \end{cases} \quad (16)$$

The charging and discharging power of each user in a time period  $t$  needs to be balanced with the charging  $P_C(t)$  and discharging  $P_D(t)$  power of the energy storage station.

3.3.3. *Resolve Method.* The few nonlinear terms in the abovementioned constraints can significantly increase the difficulty and time during the solving process of numerically solving the aforementioned optimization problem. To overcome this challenge, Big-M [27] is adopted to linearize the nonlinear constraints in this work. The user scheduling model based on shared energy storage stations can be converted to a mixed integer linear programming problem. Thus, (15) can be reformulated as follows:

$$\begin{cases} 0 \leq P_{E,D,i} \leq P_{\max}, \\ 0 \leq P_{E,D,i} \leq \delta_{ESS\_D}^{\max} M, \\ 0 \leq P_{E,C,i} \leq P_{\max}, \\ 0 \leq P_{E,C,i} \leq \delta_{ESS\_C}^{\max} M, \\ \delta_{ESS\_D}^{\max} + \delta_{ESS\_C}^{\max} \leq 1, \\ \delta_{ESS\_D}^{\max} \in \{0, 1\}, \delta_{ESS\_C}^{\max} \in \{0, 1\}, \end{cases} \quad (17)$$

where the value of  $M$  is determined as  $10^8$  in this paper.

## 4. Case Study on the Smart Grid System

4.1. *Setup and Experimental Result of the Proposed Forecasting Method.* Field load data of three regions from Western Australia are utilized for the case studies, covering the time range of May 1, 2015, to July 1, 2021. A map of the regions of the presented case studies is shown in Figure 8, where Case 1 and Case 2 are two typical industrial type electricity consumption areas, and Case 3 is the residential user group. Each node represents a region of independent microgrids that do not interfere with each other and are powered primarily by renewable energy. It is noted that 80% of the historical data are used as the training set and the remaining 20% are used as a validation set. The train/valid sets are structured with a ratio of 8 : 2, the test set is from the real load data from users. When load forecasting is implemented, 168 steps are batched together to train the model and predict the net power in the next 24 steps, and the sampling interval is 1 hour and the forecasting horizon covers 24 hours. The input feature mainly consists of the population, temperature, and calendar data which contain the season, number of holidays, and weekdays. The features are summarized in Table 1.

Figure 9 shows the raw data collected from three cases presented in Figure 8. As shown by the yellow line, part of the data in the western mining smelter (Case 2) is absent because data loss occurred due to aggregator overload. The common deep learning-based predicting models are not universal for different data sources, and each region needs to

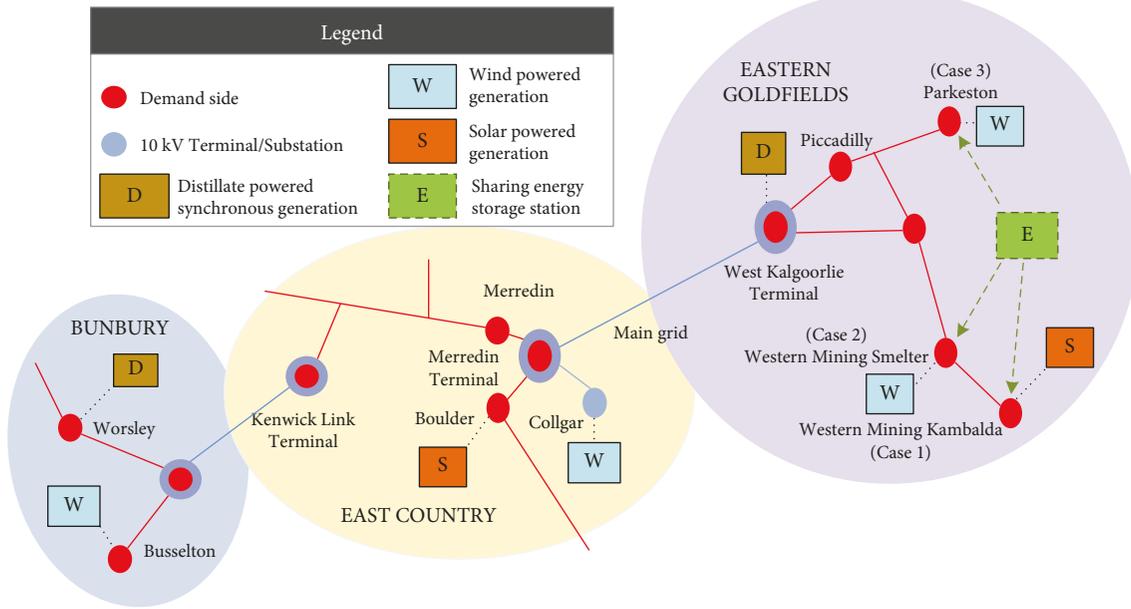


FIGURE 8: Geographical locations of the case study.

TABLE 1: Input feature of the forecasting model.

No.	Input feature	Description
1	No. of days	Integer
2	No. of holidays	Integer
3	No. of weekdays	Integer
4	Holiday length	Integer
5	Season	Binary
6	Population	Integer
7	Temperature	Integer

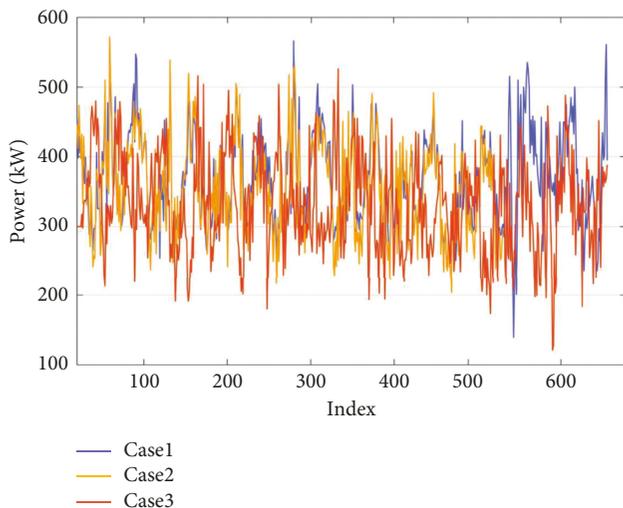


FIGURE 9: Raw data of three cases.

train a locally applicable predicting model based on its own database. According to the proposed transfer learning-based model, after comparing the MMD value of each database associated with the neighboring grid, the database of a microgrid in western mining Kambalda is chosen as the source domain.

To demonstrate the performance of the proposed structure, this work compares the proposed TCS-transfer learning model with four categories of methods listed as follows:

- (i) Traditional time series forecasting model LSTM with classical gradient descent: there are also BP [37], ANN, and CNN [38] belonging to this category, and the most applicable time series predicting model-LSTM is selected in this category
- (ii) The latest time series structure ELM [39] without backward propagation: the convergence time is substantially reduced and the training efficiency has been greatly improved. The first two categories are classical methods of time series forecasting.
- (iii) Variants of popular domain adaptation methods include MEDA-LSTM [40] and MMD-RNN [41], which are also based on the concept of transfer learning
- (iv) A branch of transfer learning is that a transformer with an attention mechanism [42, 43] has a stronger generalization capability than the classical method. This paper uses 6 encoder blocks of a transformer and 8 heads for self-attention.

The relevant parameters are listed in Table 2. The parameters that produce the best performance for each model are tuned by K-fold cross-validation. The following comparison discusses the effectiveness of the proposed TCS-transfer learning model mainly from two aspects: first, the power grid of the western mining smelter (Case 2) is with missing data, thus the traditional methods such as LSTM and ELM included in categories i and ii failed to forecast the load power due to the fragmented dataset in Case 2. For cases of missing data in the test set, transfer learning is the only method that can solve the problem. The first part of the

TABLE 2: Parameter set of different forecasting methods.

Model	Meaning	Value
TCS-transfer learning	Hidden layers	2
	Number of input layer nodes	4
	Number of hidden layer nodes	32
	Activation function	Sigmoid, tanh
	Learning rate	0.01
	Optimization function	ReLU
	Epochs of training	400
	Distance function	MMD
	Data preprocessing methods	Optimal splitting
	Hidden layers	2
MMD-RNN	Number of input layer nodes	4
	Number of hidden layer nodes	32
	Optimization function	ReLU
	Learning rate	0.01
	Epochs of training	500
	Distance function	MMD
	Data preprocessing methods	None
	Hidden layers	2
	Number of input layer nodes	4
	Number of hidden layer nodes	32
LSTM	Activation function	Sigmoid, tanh
	Optimization function	ReLU
	Learning rate	0.01
	Epochs of training	500

comparison in this section uses different transfer learning based approaches to predict the load power of Case 2. Second, this paper presents that the proposed method is also suitable for the time series forecast with a complete dataset. The comparison group of Case 1 and Case 3 contains both traditional and transfer learning based methods.

*4.2. Performance of Transfer Learning in Addressing Fragmented Test Data in the Target Domain.* This section mainly discusses the differences between the proposed TCS-transfer learning model and other transfer learning-based methods in dealing with load forecasting problems. In the first set of experiments in this paper, transformer, MMD-RNN, and MEDA-LSTM are used to compare the superiority of the proposed approach among transfer learning-based learning methods, the forecast load result is shown in Figure 10. Table 3 presents the prediction error of each method with RMSE and MAPE. These prediction errors are evaluated by

$$RMSE = \sqrt{\frac{\sum_{t=0}^n (\hat{x}_t - x_t)^2}{n}}, \quad (18)$$

$$MAPE = \frac{1}{n} \sum_{t=0}^n \frac{|\hat{x}_t - x_t|}{x_t} \times 100\%, \quad (19)$$

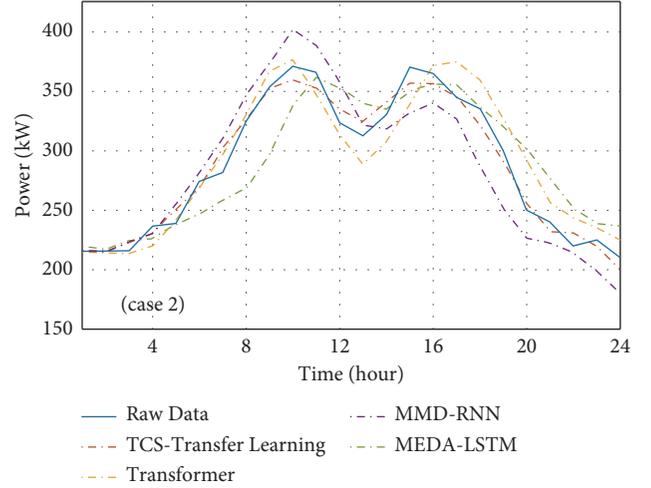


FIGURE 10: Accuracy comparison between TCS-transfer learning and other methods based on the database of Case 2.

TABLE 3: Comparison between MAPE and RMSE values obtained using different forecasting methods.

Methods	MAPE (%)	RMSE
Ref. [41] MMD-RNN	6.09	0.457
Ref. [40] MEDA-LSTM	4.94	0.395
Ref. [42] transformer	4.13	0.196
The proposed TCS-transfer learning	2.87	0.042

where  $\hat{x}_t$  denotes the predicted value of samples,  $x_t$  is the actual value of samples, and  $n$  represents the number of samples.

The convergence times of different forecasting methods are listed in Table 4; it takes 0.34s for convergence with regard to the proposed TCS-transfer learning structure, which is a 60.4% increase compared with the MMD-RNN. The prediction accuracy (MAPE) is improved by 52.8% over the RNN-based method, considering GPU cycles with Intel Core i9-12900K. Although the proposed method is slightly slower than the transformer in terms of convergence speed, it has a great improvement in prediction accuracy.

To evaluate the effect of different splitting methods on the training results, two additional methods of dividing the time series in Figure 11, i.e., splits A and B, were designed in this work. Split A denotes that the sequence is randomly divided into  $k$  segments; split B is the method where all intervals are with similar distributions, which is the opposite of our proposed split method. The proposed split method has the objective of minimizing the cost function in (5). The “distance” on the Y-axis means the distribution distance MMD with the green line and the RMSE denote the error with the blue line. As a result, it is critical that we divide the periods according to the worst case, where the distributions are the most varied.

*4.3. Results from the Case Study on Power Dispatch Strategy.* Figure 12 shows the results of power dispatch solved according to eqs. (10)–(19), whose location are illustrated in

TABLE 4: Convergence time (s) of different forecasting methods.

Methods	Time (s)
Ref. [41] MMD-RNN	0.86
Ref. [40] MEDA-LSTM	0.68
The proposed TCS-transfer learning	0.36
Ref. [42] transformer	0.21

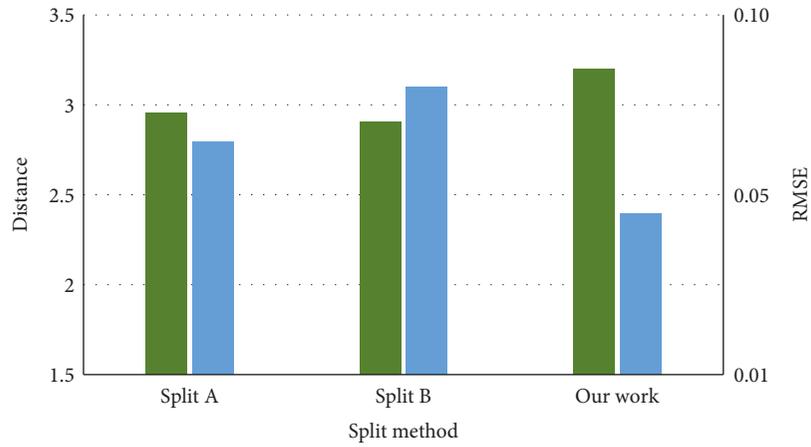
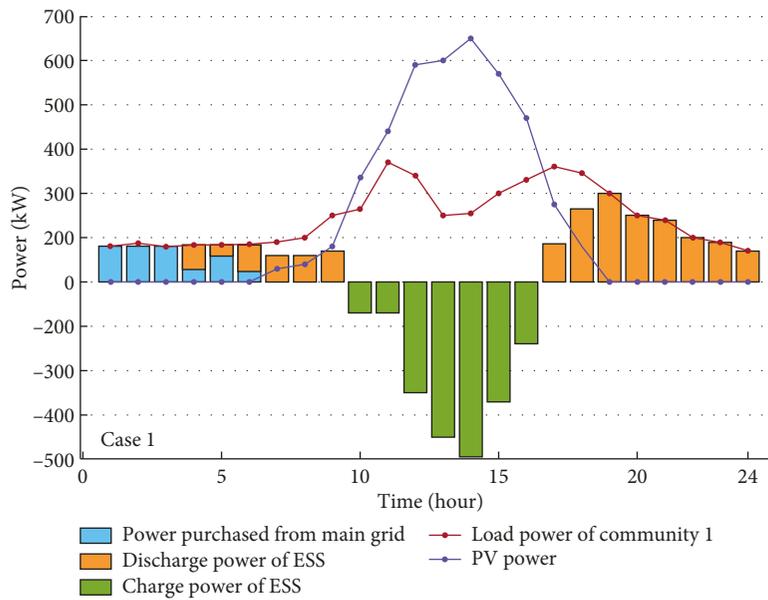


FIGURE 11: Different split methods.



(a)

FIGURE 12: Continued.



FIGURE 12: Power load balance curves of communities 1-3.

Figure 8. It can be observed from Case 1 in Figure 12 that the PV output power of the user group in the morning is less than the demand side power; the demanded electricity for this period is purchased from the grid as well as using part of the shared power station considering the optimal economics. When the PV power is higher than the demanded power, the surplus power within the community is stored by the storage power plant to avoid energy curtailment. At the period of 15:00-20:00, the demand of the community cannot be met by the PV system. While the electricity pricing is high in this time period, the undersupplied energy is provided by the

shared energy storage station. Table 5 provides the electricity rate obtained from the energy provider.

The configuration of the shared energy storage plant results in a capacity of 2,508 kWh and a maximum charge/discharge power of 637 kW. It can be observed from Figure 13 that at the period of 01:00-06:00, the electricity of users is purchased from the main grid and the power station does not provide electricity to the customer. During the 10:00-17:00, the energy storage station is in a charging state and the power rises from 382 kWh to the maximum value of 2,000 kWh, where the charging and discharging power value

TABLE 5: Electricity rates.

Level	Time frame	Electricity rates of the main grid [44] \$AUD/kWh
Peak	Commercial users 09:00–12:00; 14:00–19:00	0.253
	Residential users 07:00–09:00; 17:00–20:00	0.386
Average	Commercial users 12:00–14:00; 19:00–24:00	0.172
	Residential users 09:00–17:00; 20:00–24:00	0.322
Trough	Commercial users 24:00–09:00	0.073
	Residential users 24:00–07:00	0.086

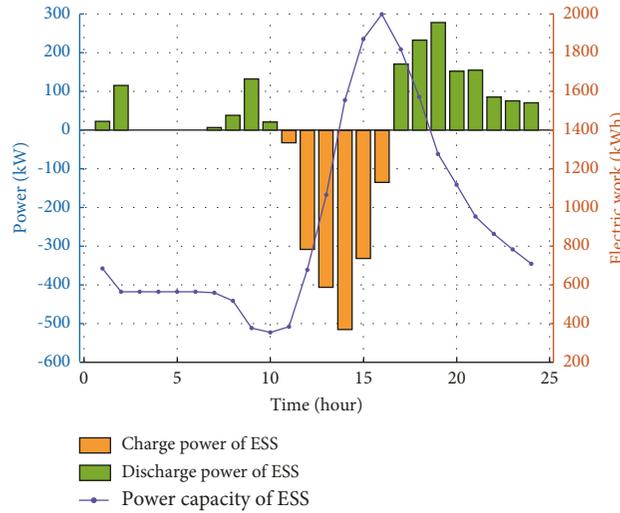


FIGURE 13: Output power and SOC curves of the shared energy storage station.

TABLE 6: User costs with using different energy storage configurations per day.

Optimization planning results of schemes 1 [45]				Optimization planning results of our work		
No. of user communities	Capacity of ESS (kW·h)	Maximum charge/discharge power (kW)	Operating costs (\$AUD)	Power purchase from the grid (kW·h)	Service fee of shared ESS (\$AUD)	Operating costs (\$AUD)
Community 1	2820	621	1416.5	583	702.5	759.2
Community 2	1005	163	566.3	16	593.4	595.4
Community 3	376	78	635.8	78	311.4	318.1
Total	4001	862	2618.6	677	1607	1672.7

Optimization planning results of schemes 1: independent configuration of the energy storage system within each user. Optimization planning results of our work: each user has access to a shared energy storage station.

of the energy storage plant is negative, which means that the energy storage plant is charging; if the charge/discharge power value is positive, it denotes that the energy storage plant is discharging.

As can be seen from Figures 12-13, the electric loads of communities 1–3 have reached a balanced state, and there is no phenomenon of energy curtailment, meanwhile, the energy storage station returns to the initial operation state after one cycle of operation, ensuring the normal operation of the next cycle of the energy storage station.

Table 6 compares the daily cost of each case considering different configuration methods of energy storage, the first method presents the independent energy storage system within each user. The total operating cost is AUD 2618.6, which is 36% more expensive than our proposed planning method, and the capacity of the energy storage required by

the customer is reduced by 37.3% due to the complementary nature of the customer’s power consumption behavior. Moreover, the energy curtailment is well addressed and the output power of renewables is fully utilized. One limitation with shared energy storage stations is that they are prone to cause harmonics after they are connected to the grid, which can compromise power quality; therefore, more management needs to be put into safe operation when using energy storage stations.

### 5. Conclusion

This paper proposed a framework for smart grid scheduling that is less reliant on local data while capable of delivering schedules with low operating costs. Specifically, the proposed framework contains the following: (1) a power

forecasting model based on deep transfer learning which can provide high-quality load prediction with limited training data; (2) a novel adaptive time series prediction method based on a neighboring area dataset that aims to train the forecasting model with strong generalization capability; (3) a day-ahead optimal economic power scheduling model considering the shared energy storage station. Results based on a case study with field load data in Western Australia showed that the maximum improvement of the proposed forecasting method is up to 52.8% in MAPE compared to other transfer learning-based methods, and up to 64.4% compared to the traditional method. The total operating cost after optimization according to the proposed method was reduced by 36.1%. These numbers indicate the proposed framework is a promising approach to solving power planning problems with incomplete datasets, in particular in addressing the cyber threats.

In the future, we plan to explore a deeper extension of TCS-transfer learning to a transformer for better performance. Moreover, this work only designed a centralized energy storage system. If multiple energy storage is needed, optimal coordination between these dispatch-oriented energy storage systems would be considered a promising area for future investigation.

## Data Availability

The features data input into the predicting model used to support the findings of this study have been deposited in the following repository: (1) Exemplary Energy Partners Company. (<http://www.exemplary.com.au/>), (2) Office Holidays (<https://www.officeholidays.com/countries/australia/2021>), and (3) Australia Net Migration Rate 1950-2022 (<https://www.macrotrends.net/countries/AUS/australia/net-migration>). The net power data used to support the findings of this study are currently under embargo, while the research findings are commercialized. Requests for data, 6/12 months after publication of this article, will be considered by the corresponding author

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the China Scholarship Council (No. 202006680038).

## References

- [1] Irena, "Planning for the renewable future," 2011, <https://www.irena.org/energytransition>.
- [2] M. Dabbaghjamesh, A. Kavousi-Fard, and Z. Y. Dong, "A novel distributed cloud-fog based framework for energy management of networked microgrids," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 2847–2862, 2020.
- [3] Z. Min, B. Wang, S. Guo, and J. Watada, "Multi-objective prediction intervals for wind power forecast based on deep neural networks," *Information Sciences*, vol. 550, pp. 207–220, 2021.
- [4] N. Victor and D. Lopez, "SI-LSTM: a Bi-directional LSTM with stochastic gradient descent optimization for sequence labeling tasks in big data," *International Journal of Grid and High Performance Computing*, vol. 12, no. 3, pp. 1–16, 2020.
- [5] M. S. Kivi, B. Blankely, M. Masters, and C. J. Bernacchi, "Development of a data-assimilation system to forecast agricultural systems: a case study of constraining soil water and soil nitrogen dynamics in the APSIM model," *Science of the Total Environment*, vol. 820, 2022.
- [6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [7] M. A. Alqahtani, "Cybersecurity awareness based on software and E-mail security with statistical analysis," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 6775980, 12 pages, 2022.
- [8] M. Cai, M. Pipattanasomporn, and S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques," *Applied Energy*, vol. 236, pp. 1078–1088, 2019.
- [9] M. S. Cui, "District heating load prediction algorithm based on bidirectional long short-term memory network model," *Energy*, vol. 254, no. 1, Article ID 124283, 2022.
- [10] X. Luo and X. Zhu, "Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge," *Energy*, vol. 225, Article ID 120240, 2021.
- [11] H. H. Dong, Y. Gao, Y. Fang, and M. Liu, "The short-term load forecasting for special days based on bagged regression trees in qingdao, China," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 3693294, 16 pages, 2021.
- [12] A. Sultana, A. Bardalai, and K. K. Sarma, "Salp swarm-artificial neural network based cyber-attack detection in smart grid," *Neural Processing Letters*, 2022.
- [13] Wikipedia, "Ukraine power grid hack," 2015, [https://en.wikipedia.org/wiki/Ukraine\\_power\\_grid\\_hack](https://en.wikipedia.org/wiki/Ukraine_power_grid_hack).
- [14] H. S. Lallie, K. Debattista, and J. Bal, "A review of attack graph and attack tree visual syntax in cyber security," *Computer Science Review*, vol. 35, Article ID 100219, 2020.
- [15] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
- [16] W. Cui, G. Zheng, Z. Shen, and S. Jiang, "Transfer learning for sequences via learning to collocate," in *Proceedings of the International Conference On Learning Representations*, New Orleans, LA, USA, May 2019.
- [17] M. N. Bisheh, X. Wang, S. I. Chang, and S. Lei, "Image-based characterization of laser scribing quality using transfer learning," *Journal of Intelligent Manufacturing*, 2022.
- [18] Y. Liu, S. Liu, J. Xu, X. Kong, L. Xie, and K. Chen, "Forest pest identification based on a new dataset and convolutional neural network model with enhancement strategy," *Computers and Electronics in Agriculture*, vol. 192, Article ID 106625, 2021.
- [19] J. Lines and A. J. Bagnall, "Time series classification with ensembles of elastic distance measures," *Data Mining and Knowledge Discovery*, vol. 29, no. 3, pp. 565–592, 2014.
- [20] Y. K. Lu, Z. Tian, R. Zhou, and W. Liu, "A general transfer learning-based framework for thermal load prediction in regional energy system," *Energy*, vol. 217, no. 15, Article ID 119322, 2021.
- [21] J. K. Li, M. Lin, Y. Li, and X. Wang, "Transfer learning network for nuclear power plant fault diagnosis with

- unlabeled data under varying operating conditions,” *Energy*, vol. 254, Article ID 124358, 2022.
- [22] T. Liang, Q. Zhao, Q. Lv, and H. Sun, “A novel wind speed prediction strategy based on Bi-LSTM, MOOFADA and transfer learning for centralized control centers,” *Energy*, vol. 230, no. 1, Article ID 120904, 2021.
- [23] H. Yin, Z. Ou, J. Fu, Y. Cai, S. Chen, and A. Meng, “A novel transfer learning approach for wind power prediction based on a serio-parallel deep learning architecture,” *Energy*, vol. 234, no. 1, Article ID 121271, 2021.
- [24] C. Li, X. Liu, Y. Cao et al., “A time-scale Adaptive dispatch method for renewable energy power supply systems on islands,” *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1069–1078, 2015.
- [25] C. Jiang and Z. Xia, “Application of a hybrid model of big data and BP network on fault diagnosis strategy for microgrid,” *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 1554422, 12 pages, 2022.
- [26] Y. Li, M. Vilathgamuwa, S. S. Choi et al., “Design of minimum cost degradation-conscious lithium-ionbattery energy storage system to achieve renewable power dispatchability,” *Applied Energy*, vol. 260, Article ID 114282, 2020.
- [27] B. S. Pali and S. Vadhera, “An innovative continuous power generation system comprising of wind energy along with pumped-hydro storage and open well,” *IEEE Transactions on Sustainable Energy*, vol. 11, no. 1, pp. 145–153, 2020.
- [28] H. Ismail Fawaz, “Transfer Learning for Time Series Classification,” in *Proceedings of the International Conference on Big Data*, pp. 1367–1376, Seattle, WA, USA, December 2018.
- [29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Proceedings of the Advances in Neural Information Processing Systems*, vol. 60, pp. 1097–1105, 2012.
- [30] M. Long, J. Wang, and M. I. Jordan, “Deep transfer learning with joint adaptation networks,” 2018, <https://arxiv.org/abs/1605.06636>.
- [31] X. Huang and S. Belongie, “Arbitrary style transfer in real-time with adaptive instance normalization,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1510–1519, Venice, Italy, October 2018.
- [32] Y. Du, J. Wang, W. Feng, S. Pan, and T. Qin, “AdaRNN: adaptive learning and forecasting for time series,” in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, Virtual Event, Australia, November 2021.
- [33] E. Tzeng, J. Hoffman, N. Zhang, and K. Saenko, “Deep domain confusion: maximizing for domain invariance,” 2018, <https://arxiv.org/abs/1412.3474>.
- [34] F. Y. Wang, J. Zhang, Q. Wei, X. Zheng, and L. Li, “PDP: parallel dynamic programming,” *IEEE-CAA Journal of Automatica Sinica*, vol. 4, pp. 1–5, 2017.
- [35] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM and other neural network architectures,” *Neural Networks*, vol. 18, no. 5–6, pp. 602–610, 2005.
- [36] W. P. Liu, T. Peng, R. Tang, and Y. Umeda, “An Internet of Things-enabled model-based approach to improving the energy efficiency of aluminum die casting processes,” vol. 202, Article ID 117716, 2020.
- [37] X. M. Gao, S. F. Yang, and S. B. Pan, “Optimal parameter selection for support vector machine based on artificial bee colony algorithm: a case study of grid-connected PV system power prediction,” *Computational Intelligence and Neuroscience*, vol. 2017, Article ID 7273017, 14 pages, 2017.
- [38] S. Q. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: towards real-time object detection with region proposal networks,” in *Proceedings of the 29th Annual Conference on Neural Information Processing Systems (NIPS)*, December 2015.
- [39] G. B. Huang, Q. Y. Zhu, and C. K. Siew, “Extreme learning machine: theory and applications,” *Neurocomputing*, vol. 70, pp. 489–501, 2006.
- [40] F. Z. Zhuang, Z. Qi, K. Duan et al., “A comprehensive survey on transfer learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.
- [41] Q. Zhou, W. Zhou, and S. Wang, “Semantic adaptation network for unsupervised domain adaptation,” *Neurocomputing*, vol. 454, pp. 313–323, 2021.
- [42] Z. Jin, J. Kim, H. Yeo, and S. Choi, “Transformer-based map-matching model with limited labeled data using transfer-learning approach,” *Transportation Research Part C: Emerging Technologies*, vol. 140, Article ID 103668, 2022.
- [43] A. Vecchiotti, S. Lee, and I. E. Grossmann, “Modeling of discrete/continuous optimization problems: characterization and formulation of disjunctions and their relaxations,” *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 433–448, 2003.
- [44] globalpetrolprices, “Retail energy price data,” 2022, <https://www.globalpetrolprices.com/>.
- [45] N. Lee, C. H. Nee, S. S. Yap et al., “Capacity sizing of embedded control battery-supercapacitor hybrid energy storage system,” *Energies*, vol. 15, no. 10, 2022.