



**HAL**  
open science

# Residential Short-Term Load Forecasting via Meta Learning and Domain Augmentation

Di Wu, Can Cui, Benoit Boulet

► **To cite this version:**

Di Wu, Can Cui, Benoit Boulet. Residential Short-Term Load Forecasting via Meta Learning and Domain Augmentation. IFIP International Workshop on Artificial Intelligence for Knowledge Management (AI4KMES), Aug 2021, Montreal, QC, Canada. pp.184-196, 10.1007/978-3-030-96592-1\_14. hal-04120825

**HAL Id: hal-04120825**

**<https://inria.hal.science/hal-04120825v1>**

Submitted on 7 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

# Residential Short-Term Load Forecasting via Meta Learning and Domain Augmentation

Di Wu<sup>1</sup>, Can Cui<sup>2</sup>, and Benoit Boulet<sup>1</sup>

<sup>1</sup> McGill University, Montreal, Quebec, H3A 0E9, Canada

<sup>2</sup> Google LLC, New York, NY, 10044, USA

{di.wu5@mail,benoit.boulet}@mcgill.ca, tracycui@google.com

**Abstract.** With the increasing adoption of electric devices and renewable energy generation, electric load forecasting, especially short-term load forecasting (STLF), has recently attracted more attention. Accurate short-term load forecasting is of significant importance for the safe and efficient operation of power grids. Deep learning-based models have achieved impressive success on several applications, including short-term load forecasting. Yet, most deep learning models do require a large amount of training data. However, in the real world, it may be very difficult or even impossible to collect enough data to train a reliable machine learning model. This makes it hard to adopt deep models for several real-world scenarios. Thus, it will be very helpful if deep learning models can be learned to tackle tasks with limited amount of training data and unseen tasks. In this work, we propose to use the meta-learning framework to train a long short-term memory-based model for short-term residential load forecasting. Specifically, by minimizing the task-level loss (loss over several tasks), the model is trained to perform well on different tasks. We also use domain randomization techniques to further augment the training tasks, which may further improve the generalization ability of the proposed model. Our model is evaluated on real-world data sets and compared against some classic forecasting models.

**Keywords:** Electric load forecasting, meta learning

## 1 Introduction

Electric load forecasting aims to forecast the future load consumption based on a set of historical records as well as other external factors [1]. It has been shown that accurate electric load forecasting is of significant importance for the efficient operation of modern power grids. The forecasting accuracy can help significantly reduce the total operation cost [2, 3]. As shown [4] in 1% increase of forecasting error would cause a £10 million increase in operating cost per year for the UK power systems. Depending on the forecasting horizon, electric load forecasting ranges from short-term forecasting (minutes or hours ahead) to long-term forecasting (years ahead) [5]. Short-term forecasting, main focus of this work, is mainly used to support real-time energy dispatch.

Electric load forecasting could be very challenging due to multiple uncertain factors. Specifically, there are uncertain factors for both the energy demand side and the energy generation side. For example, in recent years, different types of electrical appliances are integrated into the power grids [2], such as electric vehicles [6–12]. With the high penetration increase of electric vehicles, there will be a corresponding charging consumption. This electric consumption will be highly affected by human behaviors. On the other side the renewable energy generation has also been increasing very quickly in recent years. As shown in [13], the renewable energy generation has been increasing almost exponentially in the past ten years. Due to these reasons, a single residential unit may be more challenging than an industrial building. This can be more challenging when only limited data is available. In our work, we mainly focus on tackling the electric load forecasting for single houses.

Electric load forecasting has been an important research topic for the past few years. In general, the electric load forecasting algorithms can be generally categorized into two groups: statistical [14–16] and machine learning methods [17–20, 5, 21, 22]. The benefit of autoregressive integrated moving average (ARIMA) model was showcased in [15, 16]. Different types of machine learning-based methods have also been utilized for electric load forecasting including support vector regression (SVR) [18, 19], general additive models [21], neural networks (NNs) [23, 24], and random forest [25, 26].

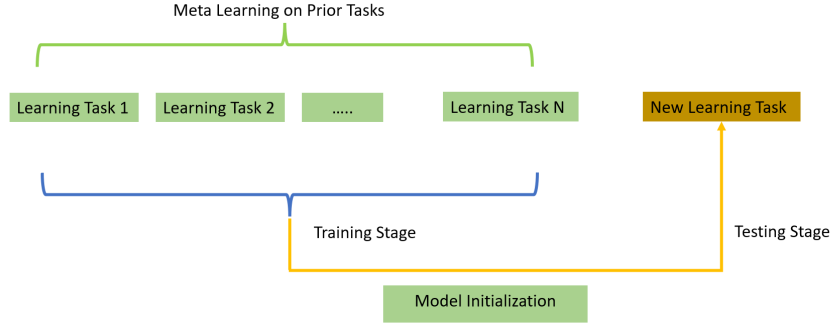
Most of the current works assume that we have a large amount of training data and stable data distribution [27]. In the real world, we may need to deal with a new house with a limited amount of data or even no labeling data. However, most of the current machine learning-based models would require a large amount of data to learn a reliable forecasting model. Without enough data, there will be a significant performance degradation for the model performance. Thus it will be very beneficial if we can learn a robust forecasting model that is robust to the changes of forecasting tasks. In this work, we propose to use meta-learning [28–32] to deal with this challenge. Specifically, we mainly use meta learning and data augmentation to learn a good model initialization based on which the model can adapt fast on a new task.

The remainder of this paper is organized as follows. The technical background of this paper is presented in Section II. The short-term load forecasting method is presented in Section III. Real-world data sets based evaluation results are presented in Section IV. Finally, the conclusion and future work are presented in Section V.

## 2 Background

### 2.1 Meta Learning

Deep learning models have achieved impressive success for different types of applications in recent years. However, most of these applications would require a huge amount of training data which can be very challenging in real world. Meta-learning [33, 34], also known as learning to learn, aims to empower the model



**Fig. 1.** Meta Learning Concept

to learn fast with a limited amount of data. This can be very helpful for machine learning models, especially for deep learning based models which usually would require a large amount of training data. As shown in Figure 1, different from other classic learning paradigms, meta learning focuses on improving the task level generalization. After finishing the learning on multiple tasks, the meta learner can perform better, i.e., perform better after experiencing the same number of samples from the new task.

Different from traditional machine learning paradigms, the meta-learner is trained on a distribution of similar tasks, with the goal of learning a strategy can generalize to related but unseen tasks from a similar task distribution. Depending on how the the meta learner is trained, there mainly three types of meta learning method, i.e., metric based methods [35], [36], [37], optimization based methods [28], and context based methods [38]. Optimization based method aims to learn a good initialization which can enable the meta learner to perform fast adaptation once presented with a new task.

## 2.2 Time Series Forecasting

The short-term load forecasting can be treated as a time series problem. Time series forecasting aims to predict the future values of time series data given the observed history. Different types of features have shown to be effective for electric load forecasting. In our work, we aims to utilize the three types of features including lagged electric load, lagged temperature information, and weekday/weekend information for the one hour ahead electric load forecasting. One-hour ahead electric load forecasting can be very helpful for the real-time energy dispatching.

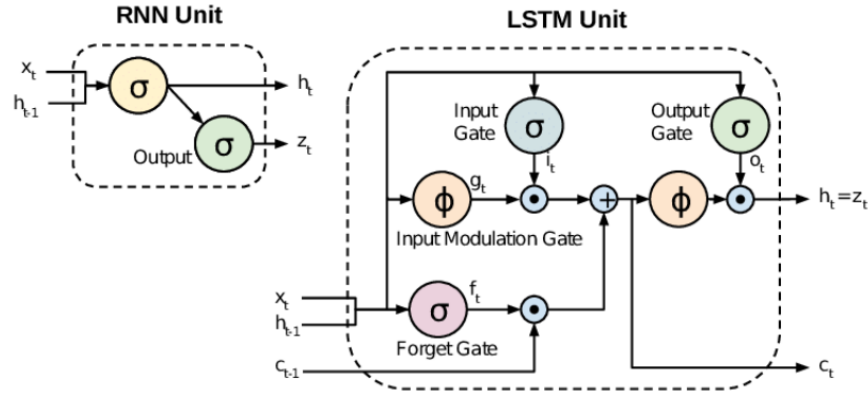


Fig. 2. LSTM Unit and RNN Unit

### 2.3 LSTM based Load Forecasting

Neural networks (NN), especially deep neural networks have shown impressive breakthroughs for different applications due to their strong representation power [39], [40]. Typical neural networks include feed forward neural networks, convolutional neural networks, and recurrent neural networks (RNNs). Feed forward neural network is the most basic type of NNs which consists of an input layer, one or more fully connected hidden layers, and a final output layer. Each hidden layer consists of several neurons realizing the non-linear transformation of outputs from previous layer.

Different from feed forward neural networks, RNNs are sequence based models and the information from previous timesteps can be used for the current timestep. This makes RNNs very suitable to model temporal correlated data and deal with time series problems. RNNs are mainly trained with backpropagation through time. However, there could be gradient vanishing and gradient exploding problems when training long-range dependencies with vanilla RNNs models [41]. The comparison of RNN unit and LSTM unit is presented in Figure 2<sup>3</sup>. As shown in this Figure, we can see that there are multiple additional cells for LSTM. Due to its excellent ability on sequential modelling, in this paper, LSTM based model is adopted for short-term residential load forecasting.

## 3 Methodology

Most of the current forecasting method would train a forecasting based on a fixed historical data sets. This may limit the generalization ability for In this paper, we propose to utilize meta learning to improve the model generalization for the

<sup>3</sup> Figure is from [42]

---

**Algorithm 1** Model-Agnostic Meta-Learning

---

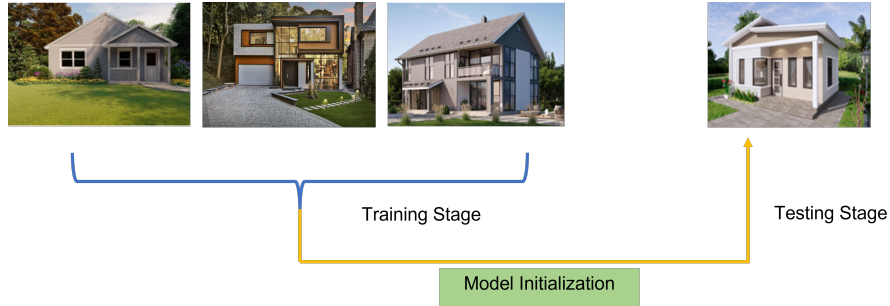
**Require:**  $p(\mathcal{T})$ : distribution over tasks  
**Require:**  $\alpha, \beta$ : step size hyperparameters

- 1: randomly initialize  $\theta$
- 2: **while** not done **do**
- 3:   Sample batch of tasks  $\mathcal{T}_i \sim p(\mathcal{T})$
- 4:   **for all**  $\mathcal{T}_i$  **do**
- 5:     Evaluate  $\nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta})$  with respect to  $K$  examples
- 6:     Compute adapted parameters with gradient descent:  $\theta'_i = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta})$
- 7:   **end for**
- 8:   Update  $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'_i})$
- 9: **end while**

---

**Fig. 3.** Overview for the model agnostic meta learning

electric load forecasting. Figure 4 presents an overview for the main objective of this paper. Our primal objective is to learning a good model initialization that can perform well on load forecasting for a new house.



**Fig. 4.** Application of meta learning for single house load forecasting

### 3.1 Model Agnostic Meta Learning

MAML is one of gradient based meta learning method and has shown to be effective for different types of applications such as traffic light control [43], image classification [44], and nature language processing [45]. MAML aims to learn a good model initialization based on multiple sampled tasks. As shown in Figure 3, there are mainly two loops of optimization. In the first loop, a large amount of tasks will be sampled and we will then learn a model for each sampled tasks based on the limited amount of data of that specific task. In the second loop, the model initialization  $\theta^*$  will be learned based on optimizing the meta loss.

### 3.2 MAML based Electric Load Forecasting

In this work, we propose to use a MAML based forecasting framework to learn a robust forecasting model which will be able to deal with different forecasting tasks, i.e., forecasting for different houses including unseen houses. Figure 5 shows the overview of the two-stage forecasting framework. Specifically, in the first stage we will pretrain a base model  $\theta_0$  over multiple ( $N$ ) houses. In this stage, we assume that we have a large amount of the training data for these houses. In this paper, the base model is based on LSTM. It is worth noting that the proposed framework could also be applied for other base forecasting models. The second stage is normal MAML training but with the learned model parameters as initialization  $\theta_0$  rather than random initialization.  $\theta_0$  will be updated through training iterations.

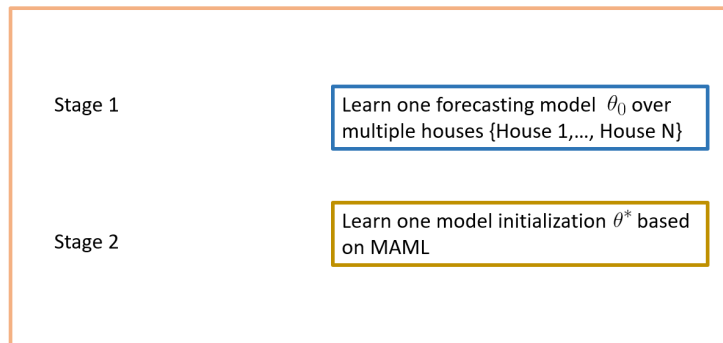


Fig. 5. Overview of the two stage forecasting method

## 4 Experiment Results

### 4.1 Dataset

In this paper, the electric load data sets from OpenEI are used to evaluate the effectiveness of proposed forecasting algorithms. In this paper, three types of features including lagged electric load (electricity consumed in the last four hours), lagged temperature information (temperature in the last four hours), and weekday/weekend information (1 for weekday and 0 for the weekend) are used for short-term electric load forecasting. To accelerate the training process, the feature normalization is implemented for all features. The used data set include one-year (2014) hourly load consumption data (8760 data points) for 72 houses in New York. We focus on the load consumption in winter times, particularly in the load consumption of December month.



## 4.2 Baselines and Evaluation Metrics

In this paper, mean average percentage error (MAPE) as shown in Eq. 1 and mean absolute error (MAE) as shown in Eq. 2 are used to evaluate the effectiveness of the proposed algorithms. As shown in these two equations,  $y'_i$  is the predicted load consumption and  $y_i$  is the real value for load consumption at time slot  $i$ .

$$MAPE = \frac{\sum_{i=1}^N \frac{|y_i - y'_i|}{y_i}}{N} \quad (1)$$

$$MAE = \frac{\sum_{i=1}^N |y_i - y'_i|}{N} \quad (2)$$

For benchmark results, four other frequently used forecasting methods are used as baseline models:

- Linear Regression
- SVR
- Feed-forward Neural Network (NN)
- LSTM

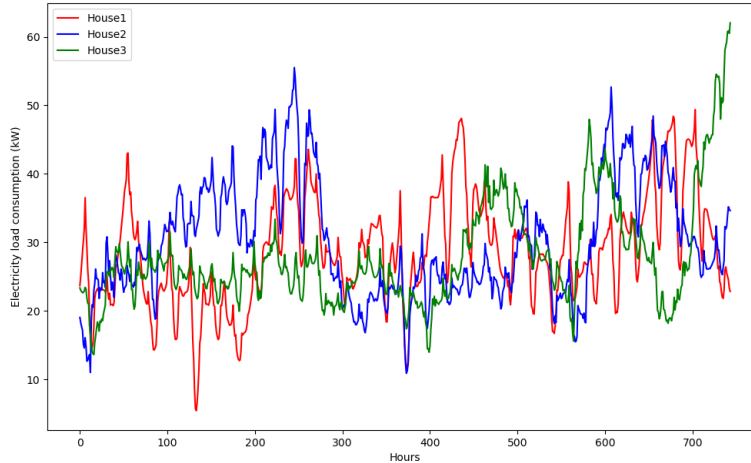
We compare the results of the baselines and proposed model for experiments described in Section 4.5 and Section 4.6.

## 4.3 Experiment Setup

As mentioned in section 4.1, the OpenEI dataset contains load consumption for 72 houses, in our experiment we pick 10 houses and treat each household as a class. The December load consumption of three randomly picked classes are shown in Fig. 6, as observed, the load consumption can vary a lot among different households. December load consumption of all 10 houses are combined, we train the models on December load consumption of other households (training classes) to predict the load for a particular household (test class).

## 4.4 Hyper Parameters

All baseline models are tuned with parameters with the best performance for the data set. Especially: For SVR, we choose RBF kernel,  $C = 1e3$ ,  $\epsilon=0.1$ ,  $\gamma=0.1$ ; For NN, we choose a 20 layer Multi-Layer Perceptron (MLP) with Rectified Linear Unit (ReLU) as the activation function,  $\alpha=0.001$ , 1000 maximum iterations; For LSTM, we build the LSTM model with an LSTM layer of 64 nodes, a dense layer with 128 nodes and use Adam as the optimizer. The data is broken into batches of 128 and we train the model for 1000 epochs. For MAML, the inner step size is 0.002, the outer step size is 0.005, the sample size is 20 for Section 4.5 and Section 4.6 and has selected values for Section 4.7.



**Fig. 6.** December Load Consumption of three houses

**Table 1.** MAPE (%) for short-term load forecasting on five houses

Method	House 1	House 2	House 3	House 4	House 5
Linear	56.07	60.12	57.89	65.12	95.54
SVR	55.47	58.86	56.91	58.75	94.09
NN	53.52	60.28	59.97	64.95	95.47
LSTM	46.77	50.97	47.62	51.55	81.22
MAML	<b>24.94</b>	<b>23.10</b>	<b>24.17</b>	<b>19.09</b>	<b>27.65</b>

#### 4.5 Evaluations on five houses

We evaluate the performance of the baseline and meta-learning models and reported the results on five houses. When training meta-learning model, we sample batch of tasks of 20 shots from training data then fine-tune and evaluate on test data. Table 1 and Table 2 respectively show the average forecasting MAPE(%) and MAE of different models on the five houses. As shown in the tables, the baseline models all have fairly large MAPE and MAE in general and perform particularly bad on House 5, while the meta-learning model has consistent performance across houses and shows on average a 55.9% improvement for MAPE and a 59.21% improvement for MAE over the best baseline model (LSTM). These results show that the model trained by meta-learning method generalize well to unseen data distributions while baseline models can fail badly when the test and training data has different distributions (House 5 as an example).

**Table 2.** MAE for short-term load forecasting on five houses

Method	House 1	House 2	House 3	House 4	House 5
Linear	15.13	15.39	17.01	18.12	15.61
SVR	15.11	15.26	16.77	16.57	15.55
NN	14.11	15.32	17.63	17.95	15.48
LSTM	12.60	13.21	14.23	14.55	13.44
MAML	<b>5.95</b>	<b>5.66</b>	<b>6.75</b>	<b>5.37</b>	<b>3.97</b>

**Table 3.** MAPE (%) for short-term load forecasting on five houses with noisy data

Method	House 1	House 2	House 3	House 4	House 5
Linear	48.15	52.04	51.61	52.64	83.34
SVR	45.18	50.18	48.94	46.03	81.63
NN	48.45	52.27	51.83	51.54	83.74
LSTM	42.00	48.63	45.88	44.82	79.76
MAML	<b>23.93</b>	<b>23.18</b>	<b>23.76</b>	<b>19.14</b>	<b>26.64</b>

#### 4.6 Evaluations on five houses with noisy data

To further showcase the robustness of the proposed method, we compare its performance with other baselines on noisy data. Specifically, we add Gaussian noise of mean 0 and standard deviation 0.5 to all data. Table 3 and Table 4 respectively show the average forecasting MAPE(%) and MAE of different models on the five houses with noisy data. As shown in the tables, MAML still consistently has the best performance across houses and shows on average a 53.96% improvement for MAPE and a 56.07% improvement for MAE over the best baseline model (LSTM). Compared to Table 1 and Table 2, the performance of baseline models has improved because the Gaussian noise reduced the effect of training datas' distribution on the testing data, thus improved the model's ability to generalize to new data.

#### 4.7 MAML training with different number of shots

To evaluate how MAML performs when the number of selected samples (shots) in training step varies, we train MAML with different number of shots, more specifically 1-shot, 5-shot, 10-shot, 20-shot, 30-shot and 50-shot, and report the results in Table 5 and Table 6. As shown in the results, the performance remains steady across 1/5/10/20/30/50-shot training for all houses.

## 5 Conclusion

Accurate electric load forecasting is of significant importance for the safe and economic operation of power grids. Most current machine learning-based forecasting models assume that we have enough training data and that new data would follow similar distribution as training data. However, these assumptions

**Table 4.** MAE for short-term load forecasting on five houses with noisy data

Method	House 1	House 2	House 3	House 4	House 5
Linear	12.83	13.19	15.20	14.97	13.53
SVR	12.05	12.91	14.42	13.35	13.26
NN	12.81	13.21	15.20	14.70	13.60
LSTM	11.27	12.46	13.50	13.01	13.01
MAML	<b>5.90</b>	<b>5.63</b>	<b>6.63</b>	<b>5.40</b>	<b>4.10</b>

**Table 5.** MAPE (%) for short-term load forecasting on five houses with 1/5/10/20/30/50-shot MAML training

House	1-shot	5-shot	10-shot	20-shot	30-shot	50-shot
House 1	22.85	22.98	23.04	22.71	22.64	23.19
House 2	21.84	21.94	22.51	23.08	22.73	23.24
House 3	22.14	22.68	23.22	23.11	24.23	23.31
House 4	17.83	18.74	18.49	17.81	19.13	18.70
House 5	26.04	25.78	26.43	25.81	26.84	26.06

**Table 6.** MAE for short-term load forecasting on five houses with short-term load forecasting on five houses with 1/5/10/20/30/50-shot MAML training

House	1-shot	5-shot	10-shot	20-shot	30-shot	50-shot
House 1	5.64	5.67	5.60	5.67	5.69	5.78
House 2	5.29	5.36	5.40	5.37	5.46	5.50
House 3	6.22	6.26	6.37	6.38	6.50	6.44
House 4	5.10	5.23	5.23	5.04	5.30	5.29
House 5	3.90	4.00	3.88	3.89	3.96	3.98

may not be true for real-world applications. Thus it will be very beneficial if we can train a model with existing data that is robust and adapts quickly to new incoming data. In this work, we propose to use the meta-learning framework to learn a robust machine learning model for residential load forecasting and ran prototyping experiments to evaluate its performance. Experiment results are implemented on real-world data sets to showcase the effectiveness of the proposed method. For future work, we plan to further improve the proposed framework and design more comprehensive experiments for evaluation.

## References

1. H. Hahn, S. Meyer-Nieberg, and S. Pickl, “Electric load forecasting methods: Tools for decision making,” *European journal of operational research*, vol. 199, no. 3, pp. 902–907, 2009.
2. D. Wu, *Machine Learning Algorithms and Applications for Sustainable Smart Grid*. McGill University (Canada), 2018.
3. D. Wu, H. Zeng, C. Lu, and B. Boulet, “Two-stage energy management for office buildings with workplace ev charging and renewable energy,” *IEEE Transactions on Transportation Electrification*, vol. 3, no. 1, pp. 225–237, 2017.

4. D. Bunn and E. D. Farmer, "Comparative models for electrical load forecasting," 1985.
5. H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on power systems*, vol. 16, no. 1, pp. 44–55, 2001.
6. D. Hall and N. Lutsey, "Electric vehicle charging guide for cities," *Consulting Report*, 2020.
7. D. Wu, H. Zeng, and B. Boulet, "Neighborhood level network aware electric vehicle charging management with mixed control strategy," in *2014 IEEE International Electric Vehicle Conference (IEVC)*, pp. 1–7, IEEE, 2014.
8. D. Wu, G. Rabusseau, V. François-lavet, D. Precup, and B. Boulet, "Optimizing home energy management and electric vehicle charging with reinforcement learning," *Proceedings of the 16th Adaptive Learning Agents*, 2018.
9. Q. Dang, D. Wu, and B. Boulet, "Ev charging management with ann-based electricity price forecasting," in *2020 IEEE Transportation Electrification Conference & Expo (ITEC)*, pp. 626–630, IEEE, 2020.
10. Q. Dang, D. Wu, and B. Boulet, "An advanced framework for electric vehicles interaction with distribution grids based on q-learning," in *2019 IEEE Energy Conversion Congress and Exposition (ECCE)*, pp. 3491–3495, IEEE.
11. Q. Dang, D. Wu, and B. Boulet, "A q-learning based charging scheduling scheme for electric vehicles," in *2019 IEEE Transportation Electrification Conference and Expo (ITEC)*, pp. 1–5, IEEE, 2019.
12. Q. Dang, D. Wu, and B. Boulet, "Ev fleet batteries as distributed energy resources considering dynamic electricity pricing," in *2021 IEEE 12th International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*, pp. 1–6, IEEE, 2021.
13. T. S. Adebayo, M. F. Coelho, D. Ç. Onbaşıoğlu, H. Rjoub, M. N. Mata, P. V. Carvalho, J. X. Rita, and I. Adeshola, "Modeling the dynamic linkage between renewable energy consumption, globalization, and environmental degradation in south korea: does technological innovation matter?," *Energies*, vol. 14, no. 14, p. 4265, 2021.
14. J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, "Arma models to predict next-day electricity prices," *IEEE transactions on power systems*, vol. 18, no. 3, pp. 1014–1020, 2003.
15. H. He, T. Liu, R. Chen, Y. Xiao, and J. Yang, "High frequency short-term demand forecasting model for distribution power grid based on arima," in *IEEE CSAE*, vol. 3, (Zhangjiaji, China), pp. 293–297, 2012.
16. H. Matsila and P. Bokoro, "Load forecasting using statistical time series model in a medium voltage distribution network," in *IEEE IECON*, (Washington, DC), pp. 4974–4979, 2018.
17. D. Wu, B. Wang, D. Precup, and B. Boulet, "Multiple kernel learning-based transfer regression for electric load forecasting," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1183–1192, 2019.
18. J. Ye and L. Yang, "A comparative study of ensemble support vector regression methods for short-term load forecasting," in *IEEE ICSAI*, (Nanjing, China), pp. 139–143, 2018.
19. Y. Chen, P. Xu, Y. Chu, W. Li, Y. Wu, L. Ni, Y. Bao, and K. Wang, "Short-term electrical load forecasting using the support vector regression (SVR) model to calculate the demand response baseline for office buildings," *Applied Energy*, vol. 195, pp. 659–670, 2017.

20. Y. Dong, Z. Zhang, and W.-C. Hong, "A hybrid seasonal mechanism with a chaotic cuckoo search algorithm with a support vector regression model for electric load forecasting," *Energies*, vol. 11, no. 4, p. 1009, 2018.
21. D. Wu, B. Wang, D. Precup, and B. Boulet, "Boosting based multiple kernel learning and transfer regression for electricity load forecasting," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 39–51, Springer, 2017.
22. W. Lin, D. Wu, and B. Boulet, "Spatial-temporal residential short-term load forecasting via graph neural networks," *IEEE Transactions on Smart Grid*, 2021.
23. W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, 2019.
24. N. Kim, M. Kim, and J. K. Choi, "LSTM based short-term electricity consumption forecast with daily load profile sequences," in *IEEE GCCE*, (Las Vegas, NV), pp. 136–137, 2018.
25. J. Moon, Y. Kim, M. Son, and E. Hwang, "Hybrid short-term load forecasting scheme using random forest and multilayer perceptron," *Energies*, vol. 11, no. 12, p. 3283, 2018.
26. M. Son, J. Moon, S. Jung, and E. Hwang, "A short-term load forecasting scheme based on auto-encoder and random forest," in *International Conference on Applied Physics, System Science and Computers*, pp. 138–144, Springer, 2018.
27. T. R. Hoens, R. Polikar, and N. V. Chawla, "Learning from streaming data with concept drift and imbalance: an overview," *Progress in Artificial Intelligence*, vol. 1, no. 1, pp. 89–101, 2012.
28. C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International Conference on Machine Learning*, pp. 1126–1135, PMLR, 2017.
29. A. A. Rusu, D. Rao, J. Sygnowski, O. Vinyals, R. Pascanu, S. Osindero, and R. Hadsell, "Meta-learning with latent embedding optimization," *arXiv preprint arXiv:1807.05960*, 2018.
30. L. Zintgraf, K. Shiarli, V. Kurin, K. Hofmann, and S. Whiteson, "Fast context adaptation via meta-learning," in *International Conference on Machine Learning*, pp. 7693–7702, PMLR, 2019.
31. I. Olier, N. Sadawi, G. R. Bickerton, J. Vanschoren, C. Grosan, L. Soldatova, and R. D. King, "Meta-qsar: a large-scale application of meta-learning to drug design and discovery," *Machine Learning*, vol. 107, no. 1, pp. 285–311, 2018.
32. F. Mireshghallah, V. Shrivastava, M. Shokouhi, T. Berg-Kirkpatrick, R. Sim, and D. Dimitriadis, "Useridentifier: Implicit user representations for simple and effective personalized sentiment analysis," *arXiv preprint arXiv:2110.00135*, 2021.
33. T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, "Meta-learning in neural networks: A survey," *arXiv preprint arXiv:2004.05439*, 2020.
34. C. Giraud-Carrier, R. Vilalta, and P. Brazdil, "Introduction to the special issue on meta-learning," *Machine learning*, vol. 54, no. 3, pp. 187–193, 2004.
35. F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1199–1208, 2018.
36. J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," *arXiv preprint arXiv:1703.05175*, 2017.
37. A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "One-shot learning with memory-augmented neural networks," *arXiv preprint arXiv:1605.06065*, 2016.

38. A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, “Meta-learning with memory-augmented neural networks,” in *International conference on machine learning*, pp. 1842–1850, PMLR, 2016.
39. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
40. T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning,” *arXiv preprint arXiv:1509.02971*, 2015.
41. Y. Bengio, P. Simard, P. Frasconi, *et al.*, “Learning long-term dependencies with gradient descent is difficult,” *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.
42. A. Rassem, M. El-Beltagy, and M. Saleh, “Cross-country skiing gears classification using deep learning,” *arXiv preprint arXiv:1706.08924*, 2017.
43. X. Zang, H. Yao, G. Zheng, N. Xu, K. Xu, and Z. Li, “Metalight: Value-based meta-reinforcement learning for traffic signal control,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 1153–1160, 2020.
44. S. Khodadadeh, L. Bölöni, and M. Shah, “Unsupervised meta-learning for few-shot image classification,” *arXiv preprint arXiv:1811.11819*, 2018.
45. W. Yin, “Meta-learning for few-shot natural language processing: A survey,” *arXiv preprint arXiv:2007.09604*, 2020.