

Location based Probabilistic Load Forecasting of EV Charging Sites: Deep Transfer Learning with Multi-Quantile Temporal Convolutional Network

Mohammad Wazed Ali

Intelligent Embedded Systems (IES), University of Kassel, School of CIT, Technical University of Munich, Dept. of Computer Science and Engineering, Intelligent Embedded Systems (IES), University of Kassel, Kassel, Germany

Asif bin Mustafa

Munich, Germany

Md. Aukerul Moin Shuvo

Rajshahi University of Engg. & Technology, Rajshahi, Bangladesh

Bernhard Sick

Kassel, Germany

Abstract—Electrification of vehicles is a potential way of reducing fossil fuel usage and thus lessening environmental pollution. Electric Vehicles (EVs) of various types for different transport modes (including air, water, and land) are evolving. Moreover, different EV user groups (commuters, commercial or domestic users, drivers) may use different charging infrastructures (public, private, home, and workplace) at various times. Therefore, usage patterns and energy demand are very stochastic. Characterizing and forecasting the charging demand of these diverse EV usage profiles is essential in preventing power outages. Previously developed data-driven load models are limited to specific use cases and locations. None of these models are simultaneously adaptive enough to transfer knowledge of day-ahead forecasting among EV charging sites of diverse locations, trained with limited data, and cost-effective. This article presents a location-based load forecasting of EV charging sites using a deep Multi-Quantile Temporal Convolutional Network (MQ-TCN) to overcome the limitations of earlier models. We conducted our experiments on data from four charging sites, namely Caltech, JPL, Office-1, and NREL, which have diverse EV user types like students, full-time and part-time employees, random visitors, etc. With a Prediction Interval Coverage Probability (PICP) score of 93.62%, our proposed deep MQ-TCN model exhibited a remarkable 28.93% improvement over the XGBoost model for a day-ahead load forecasting at the JPL charging site. By transferring knowledge with the inductive Transfer Learning (TL) approach, the MQ-TCN model achieved a 96.88% PICP score for the load forecasting task at the NRELsite using only two weeks of data.

Index Terms—Electric Vehicle, Load Forecasting, Charging Sites, Deep Transfer Learning, TCN, Quantile Regression.

I. INTRODUCTION

To meet the Paris Agreement demand adopted in 2015, electric vehicles (EVs), combined with a renewable energy mix, are a great candidate to replace fossil fuel-based internal combustion engines and reduce global warming, greenhouse gas emissions, and climate change [1][2]. In recent years, electric car sales have experienced exponential growth, with around 14 million new electric cars sold alone in 2023 and expected to reach 17 million in 2024 [3]. However, this tremendous eagerness to transform the transport sector with EVs and increase the total share of the grid's renewable energy mix brings many unique challenges, such as power outages due to excessive energy demand. Very often, there is no fixed charging behavior, and any energy demand pattern exists in advance for EVs, which may lead to uncontrolled charging and

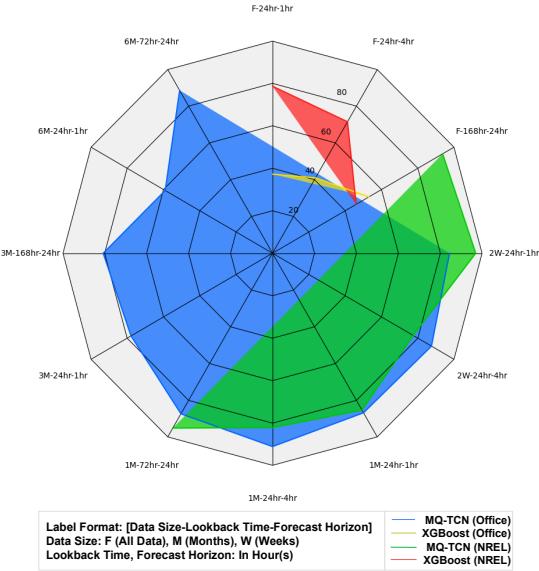


Fig. 1: Load Forecasting at JPL and NREL Electric Vehicle Charging Sites using Deep MQ-TCN model with Inductive Transfer Learning.

grid instability [4]. In addition, integrated renewable energy sources depend highly on weather, location, and other climate-related factors [5].

Therefore, stakeholders like EV charging station operators, energy traders, and energy producers need to know the aggregated energy demand of a charging site in advance at a nearby time or for special occasions such as public holidays [6]. Understanding EV users' aggregated energy demand at nearby time points (1-hour, 4-hour, 12-hour, 24-hour forecast horizon) of a particular charging site will give stakeholders a significant advantage in incentivizing pricing strategies or charging EVs intelligently [7].

Many past research works based on statistical or machine learning-based specialized load models failed due to the lack of adaptability and inability to model complex interactions among heterogeneous temporal features and irregular energy usage patterns [8][9]. Some models are very sensitive to outliers [10], and some can not handle large datasets well

[11][12]. In recent years, researchers have started to use deep learning (DL) models. However, much of this DL-based research was completed earlier with a large amount of data and high computational resources [13], point forecasts were performed only without uncertainty estimates, and knowledge transfer did not occur [14][15]. In addition, many of these models consider only numerical and temporal features and ignore informative high cardinal categorical features [16]. Recently, some research ideas have been discussed and analyzed to overcome the limitations of earlier models without in-depth experimental evaluations [5][17]. In our research, we not only analyzed but also focused on in-depth experimental evaluation of our ideas to overcome the following challenges:

Firstly, to collect an abundant amount of data in the electromobility (e-mobility) domain is not easy as it contains sensitive personal information and due to its economic value [18]. **Secondly**, training an individual DL-based load model from scratch for geographically separated charging sites is costly and time-consuming. **Thirdly**, building an efficient DL model with abundant data requires massive computation power and thus generates high temperatures, which oppose the objectives of the Paris Agreement [19].

Therefore, this research considers inductive Transfer Learning (TL) from a knowledge-transferring perspective among geographically separated charging sites and ensures self-adaptive properties in the models. Rather than developing DL models for point forecasting, we consider a quantile regression-oriented probabilistic approach to capture uncertainty in combination with Temporal Convolutional Network (TCN) [20]. In this research, we proposed Multi-Quantile TCN (MQ-TCN), an architecture described briefly in Section II novel in the e-mobility domain for load forecasting of charging sites incorporating quantile regression techniques and TCN capabilities. To evaluate MQ-TCN architecture using real-world datasets for load forecasting tasks, we are motivated to explore the following research questions in this research:

- 1) How to develop a multivariate multi-step load forecasting model of charging stations using a deep MQ-TCN for the e-mobility domain?
- 2) How can we effectively transfer day-ahead load forecasting knowledge using a deep inductive TL approach among geographically separated EV charging sites with low computational resources?
- 3) How can we handle uncertainty?
- 4) How does the performance of day-ahead load forecasting in TL settings change with lookback period, forecast horizon, and training data size?

This research used historical charging session data from four real-world EV charging sites: Caltech, JPL, Office-1, and NREL [21]. Here, the Caltech, JPL, and Office-1 data is obtained from the primary dataset, 'ACN' [22]. Experimental evaluation of these datasets using proposed MQ-TCN architecture in comparison to the XGBoost [23] and Deep Auto-Regressive recurrent networks (DeepAR) [24] models leads to the following contributions:

- **We are the first** to propose utilizing the inductive TL approach in our knowledge using a multi-step quantile regression method with deep learning architecture like multivariate TCN to forecast day-ahead load of geographically separated EV charging sites under extremely limited computational resources (data, computing power, etc.) for the e-mobility domain.
- **Without the TL approach**, our proposed MQ-TCN architecture achieved an impressive Prediction Interval Coverage Probability (PICP) score of 93.62% as shown in Figure 1, and this resulted in an improvement of around 28.93% for one day-ahead of load forecasting than usual XGBoost model at the JPL site in the same settings.
- **With the inductive TL approach**, with the pretrained MQ-TCN model, using only 2 weeks of data, it reached a PICP score of 96.88% as illustrated in Figure 1 for hourly forecasting at NREL site. This resulted in an improvement of 18.23% compared to the XGBoost model trained on 6 months of data without TL. In similar TL settings at the Office-1 site, MQ-TCN achieved 91.04% for 4 hours-ahead forecasting and 87.30% for one day-ahead forecasting task using only 1 month of data.
- **To capture the uncertainty** of the model's generated load forecasts of EV charging sites, our utilized PICP and Winkler Score (WS) metrics efficiently estimate uncertainties jointly by quantifying accurate forecasting within efficient prediction intervals.
- **By investigating the impact** of data size, lookback period, and forecasting horizon time via experimental evaluation. It has been observed that transferring knowledge with an inductive TL approach significantly reduces learnable parameter size without losing forecasting accuracy. While applying the inductive TL approach from JPL to the Office-1 site with the same settings and using the proposed MQ-TCN model, a 72% reduction in learnable parameter size is observed for data sizes of 2 weeks compared to 1 month. However, this reduction only compensates for 3.4% of the PICP score. Moreover, while varying the lookback period and the length of the forecast horizon, no explicit observation has been found related to the forecasting accuracy.

In this article, Section II discusses related research in e-mobility. Section III formulates problem definition, and Section IV gives an overview of the proposed MQ-TCN model and discusses TL settings applied in this research. Section V presents details of the experimental settings, evaluation measures with results, and analysis. Section VI concludes this research article and describes future work.

II. RELATED WORK

E-mobility has become a prominent field in recent years, and forecasting the energy demand of charging sites has become an essential area of research. In this paper [14], EV charging demand forecasting models were developed for a day-ahead horizon and 15-minute resolution. The models

incorporated calendar and weather data, and the LSTM model was used. The authors of [25] evaluated the use of deep neural networks (DNNs) and tree-based machine learning models for forecasting energy demand at public EV charging stations and conclude that tree-based models outperform DNNs in accuracy and error metrics. The authors of [25] also studied how traffic, crowd distribution, weather, and charging station distribution affect charging demand [26]. In the deep learning model family, transformers perform better in e-mobility energy demand forecasting [13]. Probabilistic forecasting is an exciting research field that predicts the charging demand for e-mobility. Machine Theory of Mind Based Quantile Forecast Network (MBQFN) framework has proven to perform better than existing DeepAR, ARX-GARCH, DLQR, Persistence, and T-CKDE [27]. In an investigation of [28], Bayesian deep learning and quantile regression methods enhanced load forecasting accuracy in EV charging stations. This approach outperforms traditional methods in dealing with uncertainties. Numerous research were completed earlier to apply to electricity load forecasting using convolutional neural networks (CNNs) [29], [30], [31]. Recently, some research has been conducted on the transfer of knowledge in e-mobility, where data is limited in the target domain. The results indicated that using CNN-BiLSTM has improved accuracy in predicting EV charging demand and network voltage profiles while reducing computational requirements for power grid operations, but it is unable to handle uncertainty [15]. In [32], XGBoost was used alongside other algorithms on the ACN dataset for both classification and regression tasks with good scores at AUC, F1 Score, R², and RMSE metrics. It lacked adaptability and could not handle uncertainty. In [33], XGBoost predicted single-step charging demand forecasting at public stations.

III. PROBLEM DEFINITION AND FORMULATION

To formulate multi-step load forecasting task formally, let's denote a heterogeneous multivariate time series dataset of d ($d \geq 1$) number of non-predictive variables with length T as $X = \{x_t\}_{t=1}^T = \{x_1, x_2, \dots, x_T\}^T \in \mathbb{R}^{T \times d}$ and target variable as $Y = \{y_t\}_{t=1}^T = \{y_1, y_2, \dots, y_T\}^T \in \mathbb{R}^{T \times 1}$. Here, $x_t \in \mathbb{R}^d$ ($1 \leq t \leq T$) is a vector with d non-predictive variable at time step t whereas $y_t \in \mathbb{R}$ ($1 \leq t \leq T$) symbolizes the target variable at the same time step t . Based on the task definition and forecasting horizon δ ($\delta \geq 1$), an estimate of the target variable is the output of the load prediction model after $T + \delta$ denotes as $\{\hat{y}_t\}_{t=T+1}^{T+\delta} = \{\hat{y}_{T+1}, \hat{y}_{T+2}, \dots, \hat{y}_{T+\delta}\}$. In this article, we consider a deep neural network variant as a non-linear mapping function $F(\bullet)$ to forecast future value $\{\hat{y}_t\}_{t=T+1}^{T+\delta}$ with δ time steps ahead from non-predictive input $\{x_t\}_{t=1}^T$, and target series $\{y_t\}_{t=1}^T$ and represent as $\{\hat{y}_t\}_{t=T+1}^{T+\delta} = F(\bullet) = F(\{x_t\}_{t=1}^T, \{y_t\}_{t=1}^T)$.

This research article also focuses on transferring knowledge of load forecasting tasks among EV charging sites with similar characteristics using the deep TL approach. To transfer knowledge from source (S) tasks $\{\mathcal{T}_{sd_m}\}_{m=1}^M$ in where $M \in \mathbb{N}^+$ and source domain $\{D_{s_m}, \mathcal{T}_{sd_m}\}$ to a target (T) task \mathcal{T}_{td_m} ,

an objective function $f_{td}(\cdot)$ needs to be learned with limited training samples $\{x_{it_d}, y_{it_d}\}$ in target domain. In our research, both source and target tasks $(\mathcal{T}_{sd_m}, \mathcal{T}_{td_m})$ in TL settings are represented as multi-step load forecasting with the deep neural network as non-linear objective function ($f_{td}(\cdot)$) to transfer knowledge from source EV sites having abundant resources (data, computing power etc.) to resource constraint EV charging sites.

IV. METHODOLOGY

In this section, we define the related terms, methods and then describe our proposed architecture.

A. Multivariate MQ-TCN for Multi-Step Forecasting

TCN is a variant of CNNs that uses parallel computations consisting of three main components: causal convolution, dilated convolution, and a residual connection. One of the most important features of this system is its ability to process inputs of any length and produce outputs of the same length. This is achieved by using a 1D fully convolutional network (FCN) structure where each hidden layer matches the input layer in length. Also, zero-padding is added to maintain consistency across layers. Causal convolutions ensure that the model only focuses on historical data to prevent data leakage in the future [34]. Traditional causal convolutions have a limited reach that increases linearly with network depth, which is insufficient for long sequences or historical data. To address this limitation, TCN uses dilated causal convolutions that skip inputs at defined intervals, thus significantly expanding the receptive field's size [35].

For a sequence input $x_1, x_2, \dots, x_T \in \mathbb{R}$ and a filter $f_{filter} : \{0, 1, \dots, k-1\}$, the dilated convolution function G on element s of the sequence is expressed as:

$$G(s) = (x *_{d_l} f_{filter})(s) = \sum_{i=0}^{k-1} f_{filter}(i) \cdot x_{s-d_l \cdot i} \quad (1)$$

where k denotes the dimension of the filter, d_l is the dilation factor which determines the interval at which the input values are sampled, $*_{d_l}$ represents the dilated convolution operator, and $x_{s-d_l \cdot i}$ indicates a backward step through the sequence indexed from s .

The receptive capacity of a TCN is inherently linked to the network's depth, represented by n , and the dilation factor, denoted as d_l . It is crucial to stabilize the TCN, particularly when deepening it, to overcome common issues of deep networks such as performance degradation and vanishing gradients. In this solution, a residual module is used, leveraging residual connections to enhance and simplify the network training process [36]. Such a module and its connection are illustrated in Figures 2b and 2c. To approximate a given function $H(x)$, we can reformulate it using another function $G(x)$ defined as $G(x) := H(x) - x$. This allows us to recast the original learning target as $G(x) + x$, based on the concept of residual connections. This adjustment simplifies the optimization process compared to optimizing the unaltered

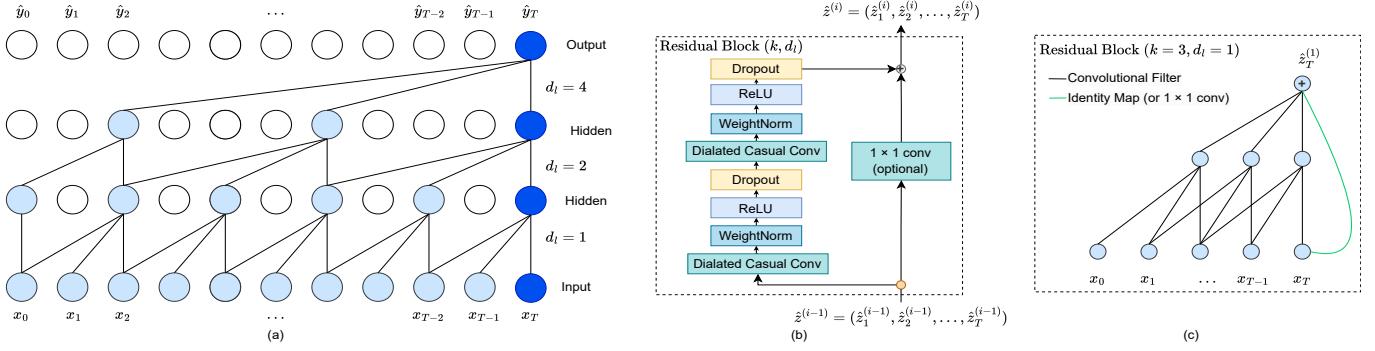


Fig. 2: Components of a TCN (a) A dilated causal convolution applied convolutional filters with dilation factors $d_l = 1, 2$, and 4 , and a filter size of $k = 3$; (b) In a TCN residual block, if the input and output have different sizes, a 1×1 convolution is included to adjust for the difference; (c) Illustration of Residual Connection within a TCN [34].

function $H(x)$ directly. The outcomes are then routed back as inputs to the block, producing an output o as illustrated by:

$$o = \text{Activation}(x + G(x)) \quad (2)$$

This mechanism permits the network layers to adjust the identity mapping incrementally rather than relearning the entire transformation from scratch. This is substantially beneficial for deep neural networks' depth-wise learning capacity.

Instead of single-point forecasting, we propose a TCN architecture tailored for quantile regression [37] in e-mobility demand forecasting in Figure 3. The TCN processes a sequence of inputs $x_{T-p}, x_{T-p+1}, \dots, x_T$, where p denotes the number of past observations and d represents the full length of the input vector. The network's output serves as an intermediate forecast value, which is then input into subsequent fully connected (FC) layers. Then, each FC layer in the model specializes in predicting a distinct quantile of the forecast distribution. Specifically, the output for the first quantile is given by $\hat{y}_{T+1}^{q^0}, \hat{y}_{T+2}^{q^0}, \dots, \hat{y}_{T+\delta}^{q^0}$; for the second quantile by $\hat{y}_{T+1}^{q^1}, \hat{y}_{T+2}^{q^1}, \dots, \hat{y}_{T+\delta}^{q^1}$; and so on, up to the n -th quantile. The

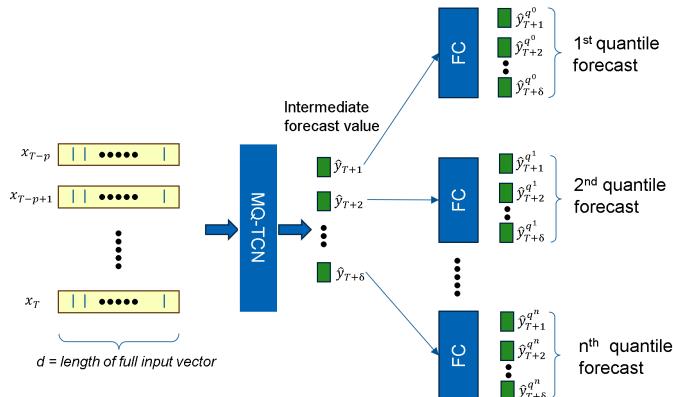


Fig. 3: Multi-Quantile Forecasting using TCN.

TCN model is trained using the Pinball Loss function L_{PB} ,

which is designed to optimally penalize the forecast errors for each quantile level q as follows:

$$L_{PB}(q, y_{T+i}, \hat{y}_{T+i}^q) = \begin{cases} (1-q) \cdot (y_{T+i} - \hat{y}_{T+i}^q) & \text{if } \hat{y}_{T+i}^q \geq y_{T+i}, \\ q \cdot (y_{T+i} - \hat{y}_{T+i}^q) & \text{if } \hat{y}_{T+i}^q < y_{T+i}, \end{cases} \quad (3)$$

for each forecast horizon $i = 1, 2, \dots, \delta$ and corresponding quantile level q . This function ensures that the quantile forecasts are not only centered but also capture the inherent variability of the future values, providing a comprehensive probabilistic forecast which is crucial for decision-making in dynamic and demand-responsive environments such as e-mobility services.

B. Transfer Forecasting Knowledge with Inductive TL

TL transfers knowledge from one EV load forecasting model trained on abundant data to forecast load on geographically separated charging sites with limited data. Based on past studies and scientific literature [38], we define TL and its related terms as follows:

If x_t^d is a sample data point at time t with d non-predictive features, a domain D can be characterized by (i) feature space, χ^{d+1} (including non-predictive and target feature space) and (ii) marginal probability distribution $P(X)$ with $X = \{x \mid x_i \in \chi, i = 1, \dots, d\}$. Domain D in TL can be denoted as $D = \{\chi, P(X)\}$. A task, \mathcal{T} , can be defined via label space \mathcal{Y} and objective function $f(\cdot)$ can be denoted as, $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$ in where $f(\cdot)$ can be only learned from training data ($\{x_i, y_i\} \mid x_i \in X$ and $y_i \in Y$). A generic definition of inductive TL is as follows:

Definition (Inductive Transfer Learning): Given a source domain D_{sd} and a learning task \mathcal{T}_{sd} , and a target domain D_{td} and learning task \mathcal{T}_{td} , the goal of inductive TL aims to assist improve the learning of the target predictive function $f_{td}(\cdot)$ in D_{td} using the knowledge in D_{sd} and \mathcal{T}_{sd} , where $\mathcal{T}_{sd} \neq \mathcal{T}_{td}$.

In accordance with the above formal definition, in our research experiment, our target load forecasting function is

a non-linear deep learning model as described in the previous section under the name of MQ-TCN. The source and target domain feature spaces can be denoted as $\chi_{sd} \in \mathbb{R}^{(T \times d_{sd})}$ and $\chi_{td} \in \mathbb{R}^{(T \times d_{td})}$, respectively, where $(T \times d)$ is the optimal length of the input sequence values of individual domains, known as the lookback window used to train MQ-TCN model. The marginal probability distributions of the source and target variables (future energy consumption) are different: $P_{sd}(X) \neq P_{td}(Y)$. Both the source and target domain data in our research come from geographically separated EV charging sites and are not identical ($D_{sd} \neq D_{td}$) in terms of data distribution. The energy consumption profile of the EV user often varies from area to area, even within the same country, due to different lifestyles, economic status, and other factors. The source (\mathcal{Y}_{sd}) and target (\mathcal{Y}_{td}) label spaces in the case of energy forecasting model depend on the length of the forecast horizons. As the predictive function $f(\cdot)$ is responsible for transferring knowledge in TL settings, but $f(\cdot)$ is learned by providing training data from geographically separated EV charging sites. Therefore, source and target forecasting functions are also not equal: $f_{sd}(\cdot) \neq f_{td}(\cdot)$. In our research, we investigate both homogeneous TL ($\chi_{sd} = \chi_{td}$; $\mathcal{T}_{sd}(Y) = \mathcal{T}_{td}(Y)$) and heterogeneous TL ($\chi_{sd} \neq \chi_{td}$; $\mathcal{T}_{sd}(Y) \neq \mathcal{T}_{td}(Y)$) in where source and target features space can be both (same or different) and also for a constant source label space, forecasting horizons are variable (both same and different) for target task (\mathcal{T}_{td}) in various experimental settings.

We have applied a ‘head replacement’ approach in our inductive TL experiments, as our TL based experiments are homogeneous and heterogeneous. In heterogeneous TL experiments, the label spaces of the source and target domains are different. However, this approach enables flexibility to make experiments with target domain (geographically separated charging infrastructure) load forecasting tasks of various time horizons ($\mathcal{T}_{td} \in \mathbb{R}^{1,4,24}$).

V. EXPERIMENTAL SETUP AND EVALUATION

This section give details about datasets, experimental setup and evaluation measure used in our research.

A. Data

Table I provides detailed statistics about ACN and NREL datasets. ACN contains historical EV charging session information from the California Institute of Technology (Caltech), Jet Propulsion Laboratory (JPL), and a small office workplace (Office-1) [22]. On the other hand, the NREL dataset represented historical charging session data of the National Renewable Energy Laboratory (NREL). The table shows Caltech has 54 charging stations with 13 different data features. JPL has slightly fewer stations than Caltech but more charging sessions. With only 8 stations, the Office-1 location has significantly fewer sessions than the other locations. Compared to the other three charging locations, NREL has the most stations, 141, and the highest number of sessions over the most extended period of approximately 59 months [21]. This site includes 16 data features, indicating a broader data

collection and features than other sites. Figures 4a, 4b, and 4c

TABLE I: E-Mobility Dataset Statistics

Charging Sites	# of Charging Stations	# of Charging Sessions	Duration (Months)	# of Features
Caltech	54	31424	40	13
JPL	50	33638	35	13
Office-1	8	1683	30	13
NREL	141	40979	59	16

provide a comprehensive view of energy consumption trends at Caltech, JPL, Office-1, and NREL locations. In Figure 4a, Caltech shows peak energy usage during daytime hours, JPL in the early afternoon, the Office-1 at midday, and NREL unusually peaks in the early morning. Figure 4b illustrates varying holiday energy consumption with notable spikes at NREL during Susan B. Anthony Day and Martin Luther King Jr. Day, while JPL peaks on Veterans Day. On the other hand, the Office-1 consumes very little energy during holidays. Lastly, Figure 4c displays annual energy usage, highlighting NREL’s significant variability and the Office-1’s consistent consumption across years, allowing for detailed analysis of temporal energy patterns across sites.

We prepared our datasets by addressing missing values where feasible to maintain consistent features across all locations. Anomalies were removed using Facebook Prophet [39], and domain knowledge. Hourly data samples were standardized for uniformity. We utilized sine-cosine transformations to capture the cyclic nature of temporal data. Numerical features were normalized using a Min-Max scaler method. Using an encoder-based dimensionality reduction technique, we translated a high-dimensional categorical feature such as ‘StationID’ from 186 unique values to only 30 unique values of lower dimension. For low-dimensional categorical features, we applied different encoding strategies based on the number of unique categories: for features with 2 to 5 unique values, we used one-hot encoding. For features with 5 to 10 unique values, we used embedding layers in our DNNs to improve model performance. To avoid disruptions caused by the COVID-19 pandemic, we curated the data to include records only until March 2020.

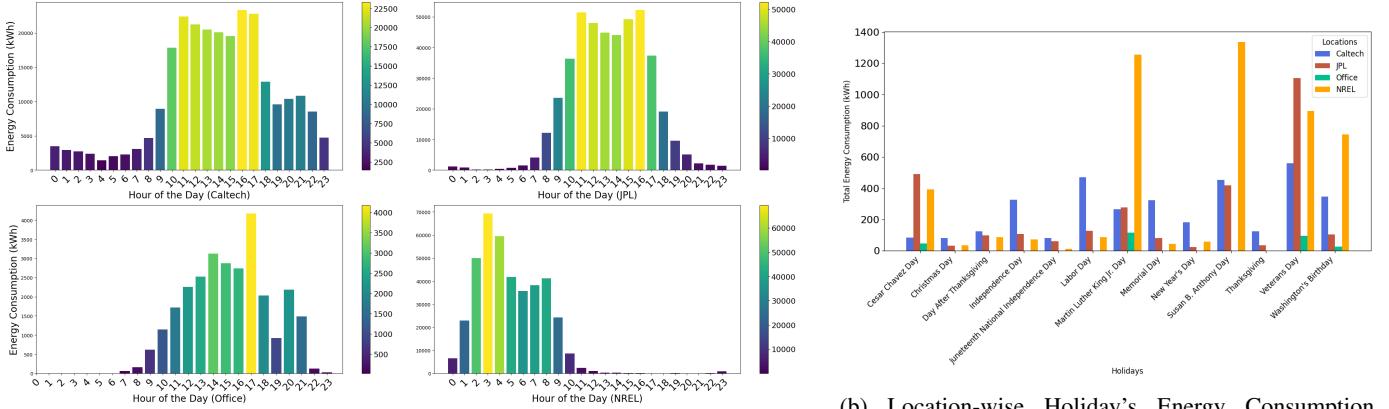
B. Evaluation Metrics

In our experiments, we evaluated model performance using several evaluation metrics: PICP, Pinball Loss, WS Score, and Normalized Deviation (ND). Pinball Loss is described in Equation 3. PICP, WS, and ND are defined as:

$$\text{PICP} = \frac{1}{N} \sum_{i=1}^N \varepsilon_i \quad (4)$$

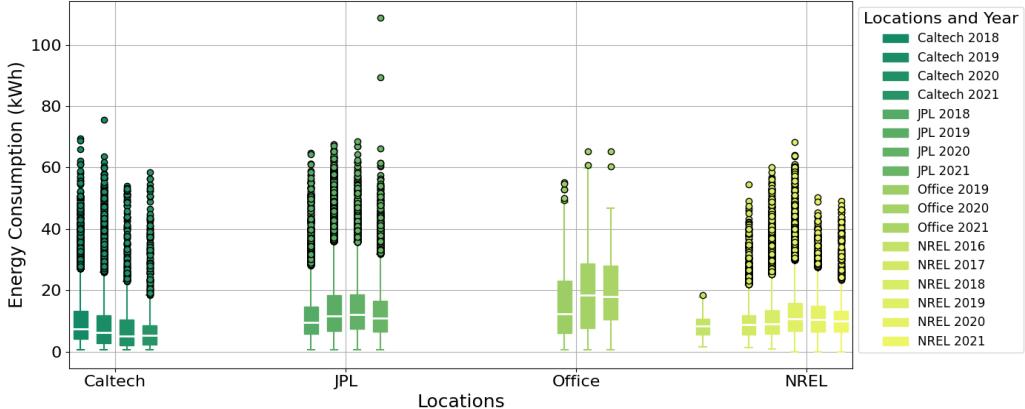
where the variable ε_i is defined as follows:

$$\varepsilon_i = \begin{cases} 1, & \text{if } y_{T+i} \in [L_{T+i,q}, U_{T+i,q}], \\ 0, & \text{if } y_{T+i} \notin [L_{T+i,q}, U_{T+i,q}]. \end{cases} \quad (5)$$



(a) Location-wise Hourly Energy Consumption (KWh).

(b) Location-wise Holiday's Energy Consumption (KWh).



(c) Location-wise Annual Energy Consumption (KWh) Boxplots.

Fig. 4: Energy Consumption Trends and Comparisons for All Locations

TABLE II: Forecasting Performance Comparison without Transfer Learning (Source Domain).

Caltech							JPL				
Model	PICP	PinBall Loss	WS	ND	Lookback Time	Forecast Horizon	Model	PICP	PinBall Loss	WS	ND
MQ-TCN	80.19	3.38	38.90	0.3523	168	24	MQ-TCN	93.62	2.28	21.55	0.2226
XGBoost	84.93	3.47	37.04	0.3602	n/a	n/a	XGBoost	64.69	1.82	19.58	0.1770
DeepAR	78.96	3.13	33.41	0.3254	168	48	DeepAR	n/a	n/a	n/a	0.1770
MQ-TCN	82.79	3.39	38.94	0.3649	336	24	MQ-TCN	75.50	2.35	25.04	0.2266
MQ-TCN	80.90	3.04	32.37	0.3261	336	48	MQ-TCN	87.85	2.64	20.93	0.2573
MQ-TCN	80.90	3.04	32.37	0.3261	336	48	MQ-TCN	88.32	1.89	18.16	0.1814

$$WS = \sum_{i=1}^N \begin{cases} \gamma, & L_{T+i,q} \leq y_{T+i} \leq U_{T+i,q} \\ \gamma + \frac{2(L_{T+i,q} - y_{T+i})}{\alpha}, & y_{T+i} < L_{T+i,q} \\ \gamma + \frac{2(y_{T+i} - U_{T+i,q})}{\alpha}, & y_{T+i} > U_{T+i,q} \end{cases}$$

(6)

$$ND = \frac{\sum_{i=1}^N |y_{T+i} - \hat{y}_{T+i}|}{\sum_{i=1}^N |y_{T+i}|}$$

(7)

Here, the variable ε_i is a Boolean variable that indicates the coverage behavior of prediction intervals (PIs). y_{T+i} denotes

the actual value at time $T + i$, \hat{y}_{T+i} denotes the forecasted value at the same timestamp, and N is the number of periods over which the forecast is evaluated. γ is the interval width and $\alpha = 0.15$ in our settings.

C. Multivariate Multi-Step Probabilistic Load Forecasting

In this section, we described results and analysis in response to research question 1 by evaluating the proposed MQ-TCN model on only Caltech and JPL datasets.

In addition to our proposed MQ-TCN model, we also explored two well-known forecasting models, namely XG-

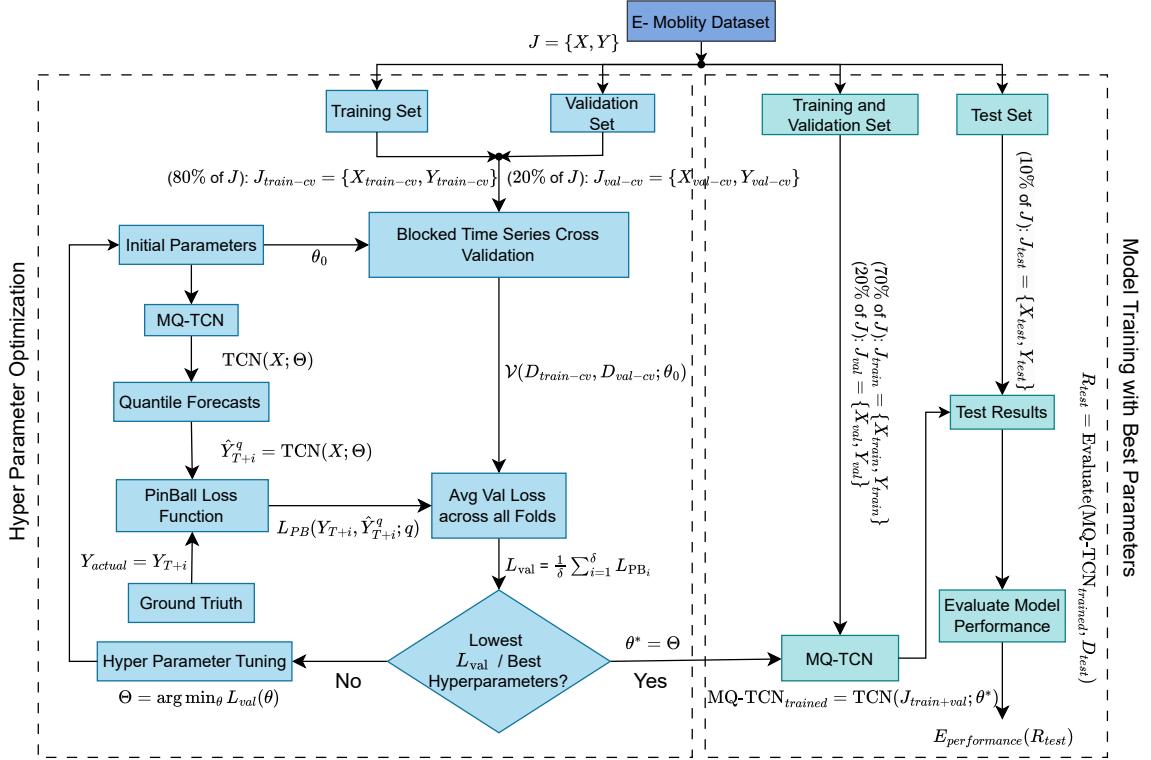


Fig. 5: Experimental Design for MQ-TCN Forecasting in E-Mobility Applications.

boost and DeepAR, as baselines. Past research works did not consider high cardinal categorical variables in conjunction with temporal and numerical features, particularly concerning multivariate, multi-step load forecasting of charging station using deep learning models [14] [25]. As our dataset is heterogeneous time-series nature, we have additional engineering challenges in adapting traditional XGBoost and DeepAR architectures. We modified the traditional XGBoost architecture, which is suitable only for point-based forecasting, to quantile regression based on probabilistic forecasting to compare with MQ-TCN. In addition, we needed to modify stride size in standard DeepAR architecture to fit our experimental heterogeneous time-series dataset. We divided our entire experimental datasets into training, validation, and test sets, as shown in Figure 5. To evaluate the model’s performance, we start by creating a test set. This set consists of the last 10% J_{test} of the data $J(X, Y)$ and is set aside to evaluate the model. The remaining data is then divided into several folds f_{fold} for time-series blocked cross-validation. Each fold contains 80% $J_{train-cv}$ of the data for training and 20% J_{val-cv} for validation. We calculate the average validation loss over all folds f_{fold} and use the loss function L_{PB} to identify the best hyperparameter set. With the best hyperparameters, we retrain the model using 90% $J_{train+val}$ of $J(X, Y)$ without shuffling. Finally, the MQ-TCN model is trained and evaluated using the previously separated J_{test} test set.

Eight load forecasting models were developed and trained

on Caltech and JPL datasets with varying settings for three quantile levels (0.05, 0.50, 0.90). These models were based on MQ-TCN, XGBoost, and DeepAR, and utilized various lookback periods (336 hours, 168 hours) and forecasting horizons (48 hours, 24 hours). Since Office-1 has a very limited number of samples, we excluded it here. For training XGBoost and DeepAR, we select and keep settings similar to the best performing MQ-TCN model based on the highest PICP score for comparison. A Bayesian optimization based method named Tree Prazen Sampler [40] was used to search for hyperparameters for all models and settings. We compared our proposed MQ-TCN and XGBoost models using PICP, Pinball Loss, and WS and compared MQ-TCN and DeepAR using ND. The optimization was based on ND in the case of DeepAR, and for our case, we considered 50% quantile predictions to make the comparison.

Results and Detailed Analysis: Experimental results displayed in Table II indicate a significant improvement in the performance of our MQ-TCN model, compared to the XGBoost baseline, across various lookback periods and forecasting horizons. Our MQ-TCN model achieved higher coverage rates, particularly for site Caltech, where the optimal model utilized 168 hours of historical data and forecasted for 48 hours. The XGBoost baseline had a coverage rate of 78.96%, while the MQ-TCN model achieved a PICP of 84.93%, which is an impressive 5.97% coverage enhancement. We observe improvements of 28.93% multi-step forecasting (24 hours) at

TABLE III: Results of Load Forecasting at Office-1 Site with Inductive TL (Target Domain).

Office-1							
Model	Data Size	Lookback Time	Forecast Horizon	PICP Loss	PinBall	WS	ND
MQ-TCN (JPL-Source)	Full Data	168	24	93.62	2.28	21.55	0.2226
XGBoost DeepAR		168	24	53.03 n/a	0.35 n/a	4.27 n/a	0.4430 1.0620
XGBoost DeepAR	Full Data	72	24	60.45 n/a	0.35 n/a	3.82 n/a	0.4305 1.0587
XGBoost DeepAR		24	4	41.02 n/a	0.33 n/a	3.48 n/a	0.4291 1.0810
XGBoost		24	1	37.24	0.33	3.62	0.4236
MQ-TCN	2 weeks	24	4	87.50 84.38	0.89 0.90	8.14 8.83	0.4641 0.4725
MQ-TCN	1 month	72	24	87.30 91.04 86.76	0.66 0.74 0.73	6.18 6.03 6.21	0.4398 0.6716 0.5372
MQ-TCN	3 months	168	24	80.71 86.12 81.02 77.88	0.40 0.61 0.49 0.49	3.77 5.07 5.14 5.44	0.4108 0.5862 0.4695 0.4509
MQ-TCN	6 months	168	24	89.23 88.55 84.14 59.08	0.45 0.42 0.45 0.44	4.67 4.25 4.37 4.44	0.4489 0.4262 0.4434 0.4353

site JPL with our MQ-TCN over XGBoost. However, DeepAR outperform MQ-TCN and XGBoost in terms of ND at the Caltech location. A sample load forecasting of 168 hours for the source sites Caltech and JPL are shown in Figure 7a and 7b.

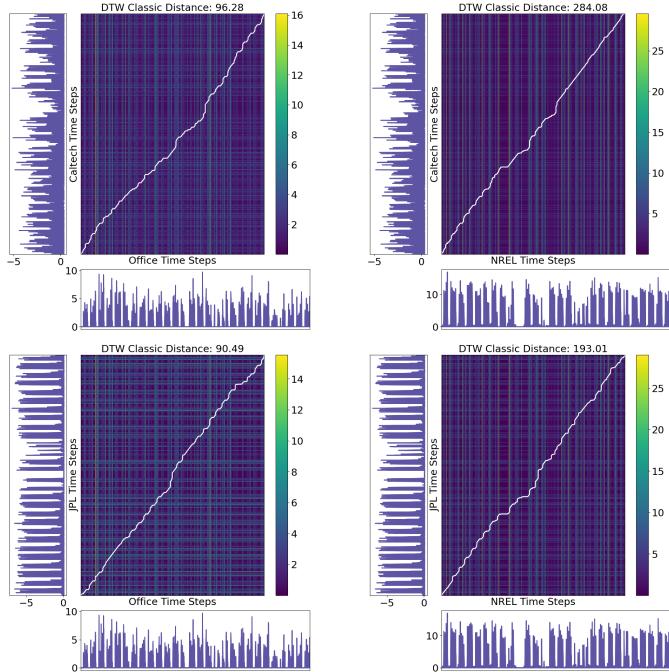


Fig. 6: Similarity Distance Measurement between Two EV Charging Sites Datasets using DTW method.

TABLE IV: Results of Load Forecasting at NREL Site with Inductive TL (Target Domain).

NREL							
Model	Data Size	Lookback Time	Forecast Horizon	PICP Loss	PinBall	WS	ND
MQ-TCN (JPL-Source)	Full Data	168	24	93.62	2.28	21.55	0.2226
XGBoost DeepAR		168	24	46.04 n/a	2.04 n/a	22.91 n/a	0.3112 1.0939
XGBoost DeepAR	Full Data	72	24	80.21 n/a	1.82 n/a	18.74 n/a	0.2944 1.0562
XGBoost DeepAR		24	4	71.49 n/a	1.73 n/a	17.98 n/a	0.2681 1.0520
XGBoost		24	1	78.65	1.74	17.49	0.2718
MQ-TCN	2 weeks	24	4	77.42 96.88	3.16 1.86	23.26 9.78	2.0426 1.2854
MQ-TCN	1 month	72	24	95.08 82.09 85.29	4.28 1.92 1.01	28.07 20.07 9.31	0.7647 0.5969 0.3166

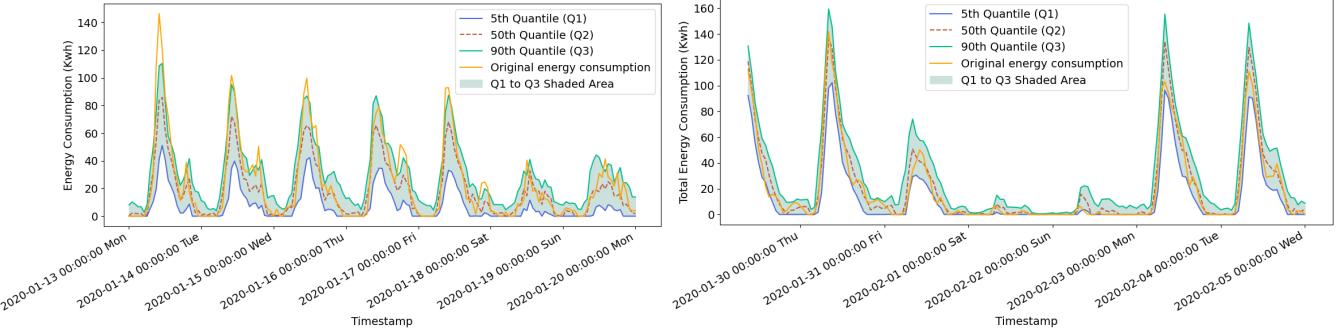
D. Location based Load Forecasting under Data Scarcity

In response to research question 2, we conducted experiments in two stages, taking into account data from four EV charging stations: Caltech, JPL, Office-1, and NREL. Among these four locations, Caltech and JPL sites have more data samples than Office-1 and NREL, as shown in the table. Office-1 and NREL need more data samples to build efficient load models. It would be highly advantageous to reuse or transfer knowledge from pre-trained models (trained on initially large amounts of data, such as Caltech and JPL) to low-resource (limited data) scenarios like at Office-1 and NREL sites to forecast load.

In the first stage, we analyzed EV users' demographics and charging infrastructure type (public, private, or workplace). In addition, we quantify the 'similarity' between any two charging sites out of four utilizing charging-related data with the 'Dynamic-Time-Warping (DTW)' method illustrated in Figure 6. The choice to prioritize TL from datasets with lower DTW scores, 'JPL-Office-1' and 'JPL-NREL', reflects their closer resemblance. Conversely, higher DTW scores were observed in 'Caltech-Office-1' and 'Caltech-NREL' pairs, indicating significant dissimilarities. Emphasizing datasets with inferior DTW scores enhances knowledge transfer efficacy, fostering improved model generalization and predictive accuracy [41].

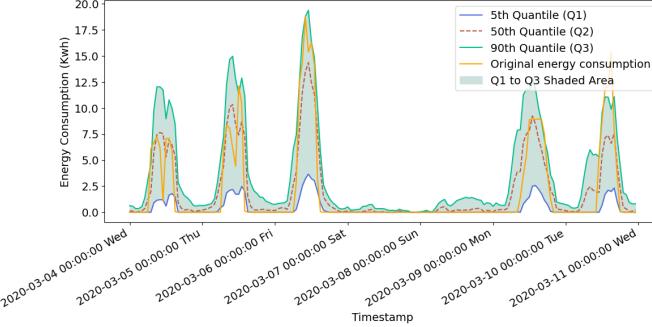
A sample load forecasting of 1 week (168 hours) and 1 day (24 hours) for the target site Office-1 and NREL is shown in Figure 7c and 7d. Also, Considering the optimal models for Office-1 and NREL based on the PICP, Figure 8a displays a histogram that quantifies the frequency of specific prediction error values. Unlike NREL's, it shows that Office-1's prediction errors are tightly clustered around zero. Figure 8b provides kernel density estimates of these prediction errors, highlighting their central tendencies and variances. It appears that the accuracy of Office-1's errors is higher and less variable as compared to NREL's errors, which are more dispersed.

In the second stage, we focus on optimizing knowledge

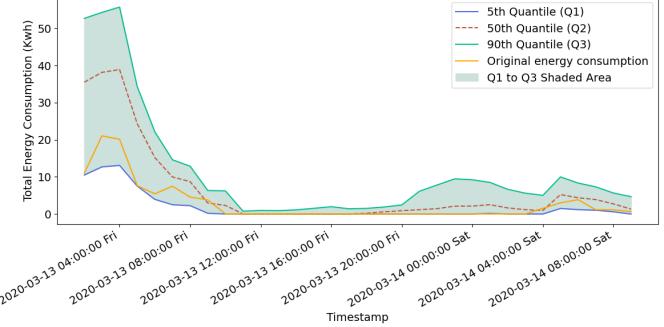


(a) Sample Energy Consumption in a Week (168 Hours) for the location Caltech.

(b) Sample Energy Consumption in a Week (168 Hours) for the location JPL.

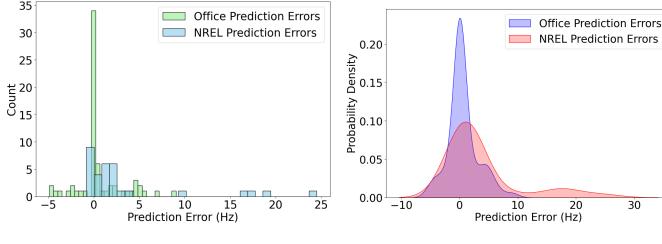


(c) Sample Energy Consumption in a Week (168 Hours) for the location Office-1 using TL from the location JPL.



(d) Sample Energy Consumption in a Single Day (24 hours) for the location Office-1 using TL from the location JPL.

Fig. 7: Sample Forecasting of Source Domain (Caltech, JPL) and Target Domain (Office-1, NREL).



(a) Comparison of Prediction Error Histograms.

(b) Comparison of Prediction Error PDFs.

Fig. 8: Visual analysis of prediction errors between Office-1 and NREL sites (a) Histogram of prediction errors; (b) Kernel density estimate of prediction errors.

transfer between similar charging sites, as discussed earlier. We utilized the best pre-trained model for JPL's location and aimed to transfer knowledge within the exact usage pattern for locations: Office-1 and NREL. We used the pre-trained MQ-TCN model as a feature extractor, removing the last layer and adding TCN blocks on top of it. Additionally, we added fully connected layers to predict each quantile.

Results and Detailed Analysis: Tables III and IV present a detailed comparison of forecasting performance at Office-1 and NREL using inductive TL across various models, lookback periods, and forecast horizons. We compare our result with baseline DeepAR and XGBoost models trained with complete

data. Notably, with minimal data, our proposed MQ-TCN models exhibit robust performance across various data sizes, lookbacks, and forecasting horizons.

For the Office-1 site, XGBoost model, training with the entire dataset, we achieve a maximum PICP score of a maximum of 53.03% during multi-step forecasting and 37.24% across single-step forecasting scenarios. The performance of XGBoost generally shows a trend where shorter historical data periods and forecasting horizons correlate with poorer outcomes in terms of PICP and other metrics. XGBoost misses the very low forecasting values. DeepAR, trained with full data size for a one-day and four-hour forecast horizon, does not support single-step forecasting. Our proposed MQ-TCN outperforms DeepAR in terms of ND in all scenarios.

The MQ-TCN model, especially with a one-month dataset for Office-1 site, exhibits notable enhancements in PICP. It can reach as high as 91.04% when using one day of historical data for forecasting four hours ahead in a multi-step forecast. In contrast, it achieves a PICP score of 86.76% for single-step forecasting using the same dataset size. The mean PICP and ND scores of our proposed MQ-TCN are 83.36 and 0.4813, respectively. These scores are significantly better than the XGBoost's average PICP score of 47.935 and the DeepAR's average ND score of 1.067. Also, for NREL, we investigated how to transfer knowledge with a very limited amount of data, and we found that with a very limited amount of data, we achieved a PICP score of 96.88% for single-step and 95.08%

for multi-step forecasting, which also outperform the baseline models.

E. Uncertainty Handling and Impact of Variable Data Size, lookback Period and Forecast Horizon in TL

In this section, we executed experiments using various data sizes to assess the impact of data size, lookback period, and forecast horizon on the performance of the MQ-TCN model and to estimate uncertainty. We started with different data sizes, such as 6 months, 3 months, 1 month and 2 weeks for Office-1, and 1 month and 2 weeks for NREL location. We tested the lookback (168, 72, 24) and future steps (24, 4, 1) accordingly for multi-step and single-step prediction.

Results and Detailed Analysis: The bar chart in Figure 9 presents information on the accuracy of energy consumption predictions in two different locations, namely Office-1 and NREL. Table III shows that Office-1 has higher reliability in its energy consumption predictions, as reflected by a robust PICP score of 89.23%. This suggests that Office-1’s forecasts are generally more accurate, making its energy management strategy potentially more reliable and actionable. On the other

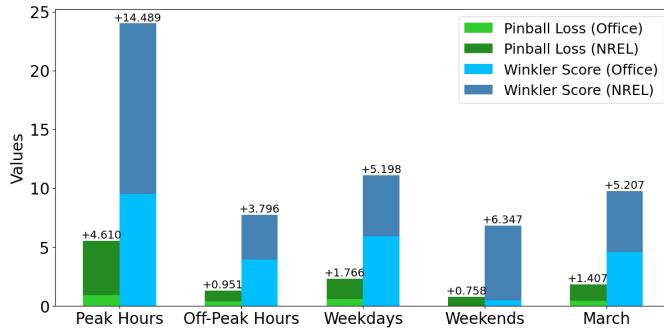


Fig. 9: Comparison of Predictive Precision and Interval Quality: Pinball Loss and Winkler Score for Office-1 and NREL

hand, NREL achieved an impressive PICP score of 96.88%, which might suggest superior forecasting but reveals a twist. NREL’s larger WS indicates that its PI is wider, suggesting that while NREL’s forecasts often capture the actual values, the range of these predictions is broad, pointing to less certainty about where the true values will fall within these ranges. The forecasting model for Office-1 is more precise, providing accurate predictions that make its energy management strategy more reliable.

We found no conclusive evidence of a relationship between lookback size, forecast horizon, and forecast accuracy. It was found that forecasting performance could be reduced or increased by increasing or decreasing the lookback period while keeping the forecast horizon constant. Similarly, with a fixed lookback period, increasing or decreasing the forecast horizon can increase or decrease forecast accuracy. As shown in Table III, at the Office-1 site, an increasing lookback period for a fixed day-ahead forecast horizon from 72 hours to 168 hours decreased forecast accuracy.

The scatter plot in Figure 10 compares locations (NREL,

Office-1, JPL, and Caltech) by data size, learnable parameters, and PICP score. NREL has PICP scores of 96.88 and 85.29, while the Office-1 ranges from 87.3 to 91.04. JPL scores 93.62, and Caltech scores 84.93. Notably, the target sites (Office-1 and NREL) have lower learnable parameters than the source sites (Caltech and JPL), with significantly higher learnable parameters. The plot demonstrates that even with very few learnable parameters in transfer learning (TL) settings, we can achieve comparable PICP scores. This reduction of the number of learnable parameters without comprising model’s efficiency to make efficient forecasts significantly reduce high computation cost (computation time, cost for data collection, and maintenance etc.) [42].

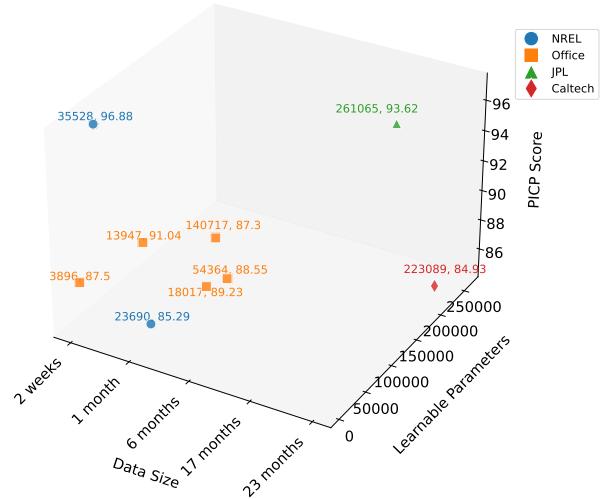


Fig. 10: Comparison of PICP Scores and Learnable Parameters across Different Data Sizes and Locations

VI. CONCLUSION AND FUTURE WORK

In our research, we successfully implemented and showed the applicability of our proposed MQ-TCN architecture for EV load forecasting tasks for multiple geographically separate EV charging sites under data scarcity. Based on the evaluation of the MQ-TCN model to transfer knowledge for load forecasting tasks among charging stations in TL settings, we described its effectiveness and adaptability. The MQ-TCN model provides load forecasting with minimal training data, precisely two weeks or 336 data instances, with high accuracy, reducing training time and, thus, computation cost. Though MQ-TCN achieved impressive results in transferring knowledge and showed adaptability to learn under data scarcity, it failed in some novel situations. We plan to explore this problem in the future, considering continual representation learning and the addition of meta-information.

ACKNOWLEDGMENT

This work results from the project SALM (01IS20098B) funded by BMBF (German Federal Ministry of Education and Research).

REFERENCES

- [1] A. Banerji, K. Sharma, and R. L. R. Singh, "Integrating renewable energy and electric vehicle systems into power grid: Benefits and challenges," *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)*, pp. 1–6, 2021.
- [2] B. Singh, G. Majeau-Bettez, and A. H. Strømman, "Corrigendum to: Hawkins, t. r., b. singh, g. majeau-bettez, and a. h. strømman. 2012. comparative environmental life cycle assessment of conventional and electric vehicles. *Journal of industrial ecology* doi: 10.1111/j.1530-9290.2012.00532.x," *Journal of Industrial Ecology*, vol. 17, 2013.
- [3] International Energy Agency (IEA), "Global ev outlook 2024," <https://www.iea.org/reports/global-ev-outlook-2024>, 2024, 2024-05-10. [Online]. Available: <https://www.iea.org/reports/global-ev-outlook-2022>
- [4] K. Clement-Nyns, E. Haesen, and J. Driesen, "The impact of charging plug-in hybrid electric vehicles on a residential distribution grid," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 371–380, 2010.
- [5] M. W. Ali, "Deep adaptive knowledge-aware learning for e-mobility," in *Organic Computing - Doctoral Dissertation Colloquium 2022*, S. Tomforde and C. Krupitzer, Eds. kassel university press, 2022, p. 139–155.
- [6] B. Wang, D. Zhao, P. Dehghanian, Y. Tian, and T. Hong, "Aggregated electric vehicle load modeling in large-scale electric power systems," *IEEE Transactions on Industry Applications*, vol. 56, no. 5, pp. 5796–5810, 2020.
- [7] J. Vuelvas, F. Ruiz, and G. Gruosso, "Energy price forecasting for optimal managing of electric vehicle fleet," *IET Electrical Systems in Transportation*, vol. 10, no. 4, pp. 401–408, 2020.
- [8] Y. B. Khoo, C.-H. Wang, P. Paevere, and A. Higgins, "Statistical modeling of electric vehicle electricity consumption in the victorian ev trial, australia," *Transportation Research Part D: Transport and Environment*, vol. 32, pp. 263–277, 2014.
- [9] O. Frendo, N. Gaertner, and H. Stuckenschmidt, "Improving smart charging prioritization by predicting electric vehicle departure time," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, pp. 6646–6653, 2020.
- [10] M. Liang, W. Li, J. Yu, and L. Shi, "Kernel-based electric vehicle charging load modeling with improved latin hypercube sampling," in *2015 IEEE Power & Energy Society General Meeting*. IEEE, 2015, pp. 1–5.
- [11] E. Xydas, C. E. Marmaras, L. M. Cipcigan, A. S. Hassan, and N. Jenkins, "Forecasting electric vehicle charging demand using support vector machines," *2013 48th International Universities' Power Engineering Conference (UPEC)*, pp. 1–6, 2013.
- [12] M. Majidpour, C. Qiu, P. Chu, H. R. Pota, and R. Gadhi, "Forecasting the ev charging load based on customer profile or station measurement?" *Applied energy*, vol. 163, pp. 134–141, 2016.
- [13] S. Koohfar, W. Woldemariam, and A. Kumar, "Prediction of electric vehicles charging demand: A transformer-based deep learning approach," *Sustainability*, vol. 15, no. 3, 2023. [Online]. Available: <https://www.mdpi.com/2071-1050/15/3/2105>
- [14] G. Van Kriekinge, C. De Cauwer, N. Sapountzoglou, T. Coosemans, and M. Messagie, "Day-ahead forecast of electric vehicle charging demand with deep neural networks," *World Electric Vehicle Journal*, vol. 12, no. 4, 2021. [Online]. Available: <https://www.mdpi.com/2032-6653/12/4/178>
- [15] P. Banda, M. A. Bhuiyan, K. N. Hasan, and K. Zhang, "Assessment of hybrid transfer learning method for forecasting ev profile and system voltage using limited ev charging data," *Sustainable Energy, Grids and Networks*, vol. 36, p. 101191, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352467723001996>
- [16] K. Nath, S. N. Gowda, C. Zhang, R. S. Gowda, and R. Gadhi, "Short-term electric vehicle charging load forecasting using transfer and meta-learning," in *2024 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2024, pp. 1–5.
- [17] M. W. Ali, "Heterogeneous multi-source deep adaptive knowledge-aware learning for e-mobility," in *2022 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C)*, 2022, pp. 57–59.
- [18] C. C. García, R. G. i Ferret, M. C. Zambrano, M. S. Cardona, D. Gkatzoflias, P. Dilara, I. Drossinos, and A. V. Donati, "Data collection and reporting guidelines for european electro-mobility projects," 2014.
- [19] D. Gielen, F. Boshell, D. Saygin, M. D. Bazilian, N. Wagner, and R. Gorini, "The role of renewable energy in the global energy transformation," *Energy Strategy Reviews*, 2019.
- [20] C. Xu, Y. Sun, A. Du, and D. ce Gao, "Quantile regression based probabilistic forecasting of renewable energy generation and building electrical load: A state of the art review," *Journal of Building Engineering*, 2023.
- [21] C. Neuman, A. Meintz, and M. Jun, "Workplace charging data collection and behavior," 2021.
- [22] Z. J. Lee, T. Li, and S. H. Low, "ACN-Data: Analysis and Applications of an Open EV Charging Dataset," in *Proceedings of the Tenth International Conference on Future Energy Systems*, ser. e-Energy '19, Jun. 2019.
- [23] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou *et al.*, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, vol. 1, no. 4, pp. 1–4, 2015.
- [24] V. Flunkert, D. Salinas, and J. Gasthaus, "Deepar: Probabilistic forecasting with autoregressive recurrent networks," *ArXiv*, vol. abs/1704.04110, 2017.
- [25] H. Rathore, H. K. Meena, and P. Jain, "Forecasting of evs charging behavior using deep neural networks," in *2023 International Conference on Communication, Circuits, and Systems (IC3S)*, 2023, pp. 1–6.
- [26] H. J. Feng, L. C. Xi, Y. Z. Jun, Y. X. Ling, and H. Jun, "Review of electric vehicle charging demand forecasting based on multi-source data," in *2020 IEEE Sustainable Power and Energy Conference (iSPEC)*, 2020, pp. 139–146.
- [27] T. Hu, K. Liu, and H. Ma, "Probabilistic electric vehicle charging demand forecast based on deep learning and machine theory of mind," Jun. 2021, pp. 795–799.
- [28] D. Zhou, Z. Guo, Y. Xie, Y. Hu, D. Jiang, Y. Feng, and D. Liu, "Using bayesian deep learning for electric vehicle charging station load forecasting," *Energies*, vol. 15, no. 17, 2022. [Online]. Available: <https://www.mdpi.com/1996-1073/15/17/6195>
- [29] L. Cai, J. Gu, and Z. Jin, "Two-layer transfer-learning-based architecture for short-term load forecasting," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 1722–1732, 2020.
- [30] D. Zhou, S. Ma, J. Hao, D. Han, D. Huang, S. Yan, and T. Li, "An electricity load forecasting model for integrated energy system based on bigan and transfer learning," *Energy Reports*, vol. 6, pp. 3446–3461, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S235248472031711X>
- [31] M. Weber, M. Auch, C. Doblander, P. Mandl, and H.-A. Jacobsen, "Transfer learning with time series data: A systematic mapping study," *IEEE Access*, vol. 9, pp. 165 409–165 432, 2021.
- [32] A. Ostermann, Y. Fabel, K. Ouan, and H. Koo, "Forecasting charging point occupancy using supervised learning algorithms," *Energies*, vol. 15, no. 9, p. 3409, 2022.
- [33] A. Almaghrebi, F. Aljuheshi, M. Rafaie, K. James, and M. Alahmad, "Data-driven charging demand prediction at public charging stations using supervised machine learning regression methods," *Energies*, vol. 13, no. 16, p. 4231, 2020.
- [34] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.
- [35] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [37] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016.
- [38] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, pp. 43–76, 2019.
- [39] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, pp. 37 – 45, 2018.
- [40] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyperparameter optimization," *Advances in neural information processing systems*, vol. 24, 2011.
- [41] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Transfer learning for time series classification," in *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 1367–1376.
- [42] N. C. Thompson, K. Greenewald, K. Lee, and G. F. Manso, "The computational limits of deep learning," *arXiv preprint arXiv:2007.05558*, 2020.