RESEARCH ARTICLE

Multi-Step Peak Power Forecasting With Constrained Conditional Transformer for a Large-Scale Manufacturing Plant

NAHYEON KIM[1], ARAVINDH BALARAMAN [1], KEUNWOO LEE[2], AND TAEWOON KIM [1], (Member, IEEE)
[1]School of Computer Science and Engineering, Pusan National University, Busan 46241, Republic of Korea
[2]Samsung Heavy Industries Company Ltd., Geoje 46241, Republic of Korea

Corresponding author: Taewoon Kim (taewoon@pusan.ac.kr)

ABSTRACT Despite of its importance and potential, the research on the peak power forecasting has received little attention. The decrease of the peak power not only reduces operational expense, but also avoids outages especially during the peak demand season. Thus, peak power forecasting, which is the key enabler for such advantages, can bring significant gains especially to a large-scale, energy-intensive manufacturing plant. This paper proposes a high-precision multi-step forecasting method to predict both the the peak power series and time of day the peak occurs. The proposed approach first predicts the peak power for a certain timespan (e.g., a day) by solving a regression problem with multiple features, including daily workload and weather forecast. Then, it generates hourly peak power series for the same timespan to increase the prediction accuracy and to identify the peak hour. In contrast to the daily workload plans, hourly plans rarely exist in practice and thus, hourly peak power forecasting is an auto-regression problem which becomes challenging as prediction timespan increases. In this work, a Constrained and Conditional Transformer (C2Transformer) is proposed for accurate multi-step peak power forecasting. The proposed model takes in the past hourly peak power series of length k along with a single peak predicted over a timespan of length n. Conditioning on the predicted long-term peak, the model generates n hourly power peaks. Also, the proposed C2Transformer has an additional constraint which is minimizing the difference between the predicted peak among n hours and the maximum value among the generated hourly peaks. Through extensive evaluations on a real data set from multiple sources, the proposed C2Transformer has shown superior performance to the widely used deep learning models.

INDEX TERMS Peak power forecasting, multi-step prediction, deep learning, transformer, manufacturing plants.

I. INTRODUCTION

Global economic development has led to a rise in electricity usage. As per the Energy Information Administration (EIA), energy consumption has been on the rise and is projected to continue this upward trajectory until 2050 (see Fig. 1 [1]). While the ongoing growth in energy consumption may be

The associate editor coordinating the review of this manuscript and approving it for publication was Rajeeb Dey .

considered an inevitable outcome of development at a certain stage, both industry and academia have voiced concerns about its adverse effects, notably its contribution to carbon emissions. Furthermore, energy consumption constitutes a significant portion of operating expenses (OpEx) for businesses. Consequently, substantial efforts have been made to mitigate power consumption.

To effectively reduce the power consumption while maintaining the desired production capacity, it is crucial
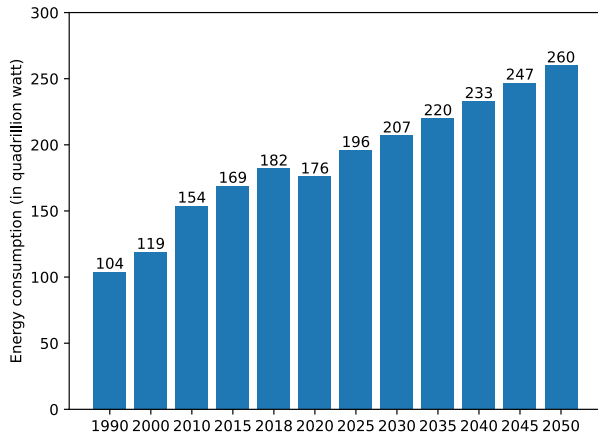
FIGURE 1. The energy consumption forecast worldwide in quadrillion watt units.

to accurately forecast power consumption [2]. There have been numerous studies conducted to improve the accuracy of power consumption forecasting and to enhance its accuracy using methods such as statistical techniques [3], [4], machine learning [5], [6], [7], and deep learning [8], [9], [10]. However, little attention has been given for predicting peak power consumption, despite its considerable importance in managing operating expenses (OpEx). For instance, certain electric power companies factor in the time of energy use when calculating bills for commercial customers [11]. This means that energy consumed during peak demand hours incurs higher costs. For another example, in South Korea, the electricity bill rate for the industrial sector is determined based on the previous year's peak power consumption. Therefore, accurately identifying when instantaneous power consumption reaches its daily peak is of significant importance in reducing OpEx.

In general, predicting and reducing (peak) power consumption is significantly challenging, and due to its substantial impact on business and nature, it has been widely studied in the literature. This includes the energy prediction and consumption in manufacturing industries [12] which is a main topic of the present study. Power consumption has also been considered with importance in the following domains: smart grid [13], Internet of Things [14], wireless communications [15], data center [16] to name a few. In other words, further development of the power prediction technology is still in demand, and has great potentials in wide range of domains. It should also be noted that data acquisition tools [17] and communication protocols [18] are the technologies that lay a foundation for reliable data collection.

In this context, this paper presents an innovative and highly accurate solution for forecasting peak power, aiming to predict both the hourly peaks and the specific time of day when the daily peak occurs. The key contributions of this paper can be summarized as follows:

- Despite its significance, forecasting peak power consumption has garnered less concentration when

compared to predicting total power usage. This paper introduces a framework designed to achieve precise multi-step peak power forecasting. The proposed framework incorporates several components, including improvements to the Transformer model, two distinct data set preparation strategies, and a two-level model ensemble.

- To enhance the precision of peak power prediction, this paper introduces a method that merges multi-variable regression and auto-regression. The proposed technique leverages multi-variable regression to forecast peak power over the long term and employs it to generate multi-step, short-term peak power series predictions.

- We have proposed two data preparation approaches, namely periodic sampling and continuous sampling, with the aim of improving prediction accuracy while simultaneously expanding the data-set size.

- In this paper, we introduce a novel approach known as Constrained and Conditional Transformer (C2Transformer), which serves as an enhancement to the Transformer model. This C2Transformer is designed to achieve highly accurate multi-step hourly peak power forecasting. The generated predictions for the short-term, denoted as n-step predictions, are subsequently employed to make longer-term peak power predictions and determine when these longer-term peaks are expected to occur. As an example, C2Transformer can generate hourly peak power series, which are then utilized to calculate the daily peak power levels and pinpoint the specific time of day when these peaks are likely to manifest. The conditional input, derived from a long-term peak prediction generated by an external regression model, plays a pivotal role in the decoder logic of the Transformer. It is utilized to enhance the accuracy of time-series predictions. Additionally, this same conditional input serves as an essential constraint within the proposed model. Specifically, it ensures that the maximum value among the generated peak series aligns with the external input, thereby maintaining consistency with the external peak prediction.

- The proposed model extends its capabilities to predict the timing of the long-term peak occurrence. While the model's primary output consists of hourly peak predictions, it recognizes that for plant operators, longer-term peaks like the daily peak hold greater significance. As a result, the proposed model can generate precise multi-step hourly peak predictions, which can, in turn, be used to derive accurate daily peak values and determine the exact time at which the daily peak is expected to transpire.

- To further enhance the performance of the trained model, we introduce a two-level ensemble approach. In the lower-level ensemble, models trained using the same type of data-set but with varying hyper-parameters are integrated using a max-combine rule to achieve precise peak predictions. In the upper-level ensemble, the results

obtained from the lower-level ensemble are combined using a mean-combine rule to enhance the model's overall generalization performance.

- To assess the efficacy of our proposed model, we conducted comprehensive evaluation and comparative studies, pitting it against commonly used methods for time-series prediction. These methods include ARIMA, 1D-CNN, LSTM, ConvLSTM, and the Vanilla Transformer model.

The remainder of this paper is structured as follows. Section II summarizes the previous studies related to the power forecasting. Section III presents the proposed constrained and conditional Transformer (C2Transformer) for peak power forecasting. The effectiveness of the proposed method is validated in Section IV, and Section V presents the conclusion of this study.

II. RELATED WORK

There have been several studies on power consumption prediction with various approaches. Power consumption prediction is usually cast as a time-series forecasting problem, hence traditional statistical method for has been adopted widely. Krishna et al. [3] proposed to use AutoRegressive Moving Average model (ARMA) [19] and AutoRegressive Integrated Moving Average (ARIMA) [20] model for predicting half-hourly power consumption. One of the findings therein is the inappropriateness of using ARMA because power consumption data is typically non-stationary. Consequently, the ARIMA model, with first-order differencing to make the data weakly stationary, demonstrated better performance when compared to ARMA. Alberg et al. [4] proposed a sliding window-based forecasting algorithm utilizing ARIMA or Seasonal ARIMA (SARIMA). The authors have demonstrated that the SARIMA-based sliding window forecasting algorithm is more effective. However, these models assume that the data is stationary and seasonal. Thus, traditional statistical approaches such as ARIMA and SARIMA may not be suitable for such data that entails little stationarity or seasonality.

To overcome the limitations and to make more accurate predictions, various machine learning methods have been proposed. Zhou et al. [5] hypothesized that weather conditions and power usage are closely related. Consequently, the authors clustered the data based on weather conditions and compared three widely used algorithms: Back Propagation (BP), Radial Basis Function (RBF), and Support Vector Regression (SVR). The evaluation results revealed that SVR achieved the smallest errors in terms of Root Mean Square Error (RSME) and Mean Absolute Percentage Error (MAPE). Additionally, the clustering-based approach resulted in more accurate predictions than the counterparts without clustering. R. et al. [6] conducted evaluations using ARIMA and Extreme Gradient Boosting (XGBoost) [7] to predict the energy consumption where hourly energy consumption and the related features are available. According to the authors' findings, XGBoost outperformed ARIMA in terms of RMSE.

Furthermore, the authors identified the best-performing hyper-parameter set for XGBoost.

Recently, deep learning-based approaches have been widely used for power forecasting due to their outstanding performance and the availability of large dataset. Kim et al. [8] compared hourly power usage prediction performance between Long Short-Term Memory (LSTM) [9] and the Double Seasonal Holt-Winter algorithm [21]. According to the authors, LSTM performed better in terms of RMSE, while the Double Seasonal Holt-Winter algorithm failed to capture sudden changes in power usage patterns. Kim and Cho [10] proposed to enhance LSTM by combining it with CNN[1]-LSTM for predicting hourly power consumption. The authors pointed out that the CNN layer is used to select influential features between variables affecting power consumption, while the LSTM layer is used for learning and forecasting the underlying time-series power consumption patterns. Evaluation results showed that the CNN-LSTM model achieved the least RMSE compared to Conditional Restricted Boltzmann Machine (CRBM) [22], FCRBM [23], and Seq2Seq [24]. However, the authors' proposed model therein is designed for individual household electric data, which may pose challenges when applying it to large-scale power plant scenarios. Recently, emerging deep learning methods such as Generative Adversarial Network (GAN) [25] and Transformer have been applied [26] to forecast the multi-step power consumption. Tian et al. [27] presented a parallel prediction scheme with GANs for forecasting power consumption patterns at 30-minute intervals. Initially, a GAN model generates parallel data based on the original data. Then, data mixed with original data and parallel data are used to train the prediction models such as Back-Propagation Neural Network (BPNN) [28], Extreme Learning Machine (ELM) [29], and SVR. According to their evaluations, prediction models trained with mixed data from GAN achieved smaller MAE than those trained with original data alone. Rao et al. [30] proposed using Transformer with multi-head attention and position encoding mechanisms for total power predictions at 15-minute intervals. Evaluation results showed that Transformer achieved smaller MAPE than LSTM, BP, and ARIMA, with a shorter training time.

The primary goal of the aforementioned studies is to learn and forecast the total power consumption. While understanding and forecasting the total power consumption is essential, peak power prediction also deserves attention for the following reasons. Reducing the peak power can be an effective strategy to reduce the OpEx. For example, in South Korea, in the case of the industrial sector customers, electric rates for the coming year are determined by the peak power of the current year. That is, even without reducing the total power use, lowering the peak power can reduce the electric rates. In addition, it can effectively prevent power outages especially during the peak seasons such as summer when nationwide power usage surges.

---

[1]Convolutional Neural Network.

Companies can redistribute their workload based on peak power prediction to prevent outages and reduce electric charges. However, such an intelligent redistribution cannot be implemented without any peak power prediction scheme which has received less attention compared to the prediction of the total power consumption.

There are a few studies that paid attention to the peak power forecasting. Liu et al. [31] compared the peak prediction performance of the several classical methods, including Naive Bayes, SVMs and Random Forests, to the deep learning models, e.g., CNN and LSTM. These models predict the daily peak power demand for the next 24 hours using total power demand, weather and date. Among the several methods considered therein, LSTM achieved the best performance in terms of precision, recall, and accuracy. Despite its noteworthy performance, the LSTM model in their paper primarily focuses on predicting peak time accurately. In addition to predicting accurate peak time, predicting the peak power levels with high precision is also important, and in this work we propose to predict both the peak power and the peak hour at the same time accurately.

Zhang et al. [32] attempted to simultaneously predict long-term energy consumption and peak power demand with a single model. To achieve this goal, the authors proposed sequential-XGBoost, which uses the XGBoost algorithm with different configurations in a sequential manner. Sequential-XGBoost predicts monthly energy consumption and peak power demand for the next 1-3 years timespan. According to the authors, the proposed model achieved the lowest MAE compared to the widely used methods such as ARIMA and LSTM. However, such long-term monthly predictions may not immediately help redistribute power usage based on predictions. Therefore, in this paper we proposes a hourly and daily-basis peak power prediction method so that the plant operator can make short-term operation plan and to quickly respond to the power peak to be happening within a few hours from now.

The aforementioned studies have achieved outstanding performance under the assumed configurations, but they have some inherent limitations at the same time. The research presented in this paper seeks to address these limitations and to make the following enhancements. Firstly, There is a limited number of studies aiming to predict peak power levels, despite the importance and potential of peak forecasting. This paper proposes multi-step peak power forecasting, which is crucial for reducing OpEx and preventing power outages that is especially important during the peak seasons. Secondly, Many of the previous studies have treated the time series prediction problem as an auto-regression problem. However, depending on the available dataset, a better approach can be taken to achieve higher accuracy. In this paper, we propose to combine both multi-regression and auto-regression to further enhance the prediction accuracy. In a nutshell, the prediction from the former is passed to the latter as a conditional input to improve accuracy. Thirdly, We propose two data sampling and ensemble strategies to enhance prediction accuracy,

increase the dataset size, and increase the generalization performance of the trained model. As a result, we propose an enhancement to the state-of-the-art Transformer model to improve multi-step time series forecasting accuracy by introducing a conditional input and an additional constraint; details will be introduced in the following Section III.

III. PROPOSED IDEA
In this study, the primary goal is to solve the multi-step hourly peak power forecasting problem for large-scale manufacturing plants.

Considering a large-scale manufacturing site consisting of multiple factories and buildings makes the power prediction problem becomes much more complex. However, the resulting solution can be scalable and robust. A manufacturing business typically owns and operates a fixed number of factories, docks and equipment–at least, these numbers do not change frequently. While one might assume a certain power consumption rate for each equipment on the shop floor to make the problem simple. However, such assumptions are not practical, and they still do not render the problem deterministic. To accurately forecast the peak power of a manufacturing site, numerous elements must be taken into account. In practice, it is nearly impossible to consider every element affecting the power usage. Additionally, there are always a large degree of uncertainties influencing the power usage such as irregular shipbuilding orders, sudden workload drops due to deteriorating weather conditions, and more. Thus, the problem of peak power forecasting for a large-scale manufacturing is a complex and challenging to solve while yielding a scalable and robust solution.

We assume that the following data set is available: 1) daily workload and weather data, 2) daily peak power, and 3) hourly peak power. We also assume that the daily workload/weather data is available for the desired period of time, and it is the same to the daily peak power data. However, the hourly peak power time-series data is available only for the past 12 months; which is the case for the electric power generation company, Korea Electric Power Corporation (KEPCO), in South Korea.

Considering the multiple variables available, the daily peak power prediction can be cast as a multi-variable regression problem by letting daily workload and weather be the input features. However, hourly peak power forecasting for the forecast horizon of n (e.g., $n = 24$) hours is an auto-regression problem due to the lack of available features.

Multi-variable regression for a single-step prediction is a well-known problem and can be solved by various techniques [33], and one of the state-of-the-art and leading technologies is XGBoost [34]. XGBoost is a scalable tree-boosting system that is widely applied to various machine learning problems, such as regression and classification [7]. XGBoost is constructed by additionally stacking decision trees and greedily ensemble them. This system enables XGBoost to achieve high accuracy. Also, additional advanced features of XGBoost, such as scalability and parallel learning, enhance

the computational speed. Furthermore, regularized learning objectives and sub-sampling can effectively prevent the over-fitting problem in XGBoost.

Auto-regression multi-step forecasting is also a well-known problem, and there exist various solutions, including statistical approaches e.g., (S)ARIMA and ETS, machine learning-based approaches e.g., linear/polynomial regression, and deep learning-based models e.g., RNN (LSTM/GRU), one-dimensional CNN (1D-CNN), and ConvLSTM (Convolutional LSTM). In addition, Transformer which was initially proposed for building a language model has shown outstanding performance for auto-regression multi-step forecasting problems [35].

Transformer [26] is a transduction model composed of an encoder and a decoder, relying on multi-head self-attention without any recurrent or convolutional layers. Self-attention is used instead of additional layers to reduce computational complexity and the number of parameters, and to create tighter dependencies between inputs and outputs. Transformer has shown outstanding performance in language translation tasks and is also applicable to other tasks, including time series prediction. In addition to its performance, Transformer offers faster training speeds compared to the models that rely on recurrence or convolution. In general, multi-variable regression can achieve high accuracy if the available features are highly correlated with the output to predict. On the other hand, auto-regression model learns solely from the historical data of a single feature. Consequently, its accuracy may degrade, particularly when the number of prediction steps (i.e., the forecasting horizon) increases [36].

Considering the available dataset and the limitation of the auto-regression model, we propose an enhancement to the Transformer model that incorporates the prediction from the multi-variable regression model. The proposed multi-step hourly peak power prediction method operates as follows. Given the daily workload and the weather forecast, the daily peak power is predicted by a multi-variable model such as XGBoost for a certain date. Given the predicted daily peak power and k number of past peak power measured each hour, the enhanced Transformer model generates peak power predictions for n hours for the date. By providing the high-accuracy daily peak power as an input to the auto-regression Transformer model, the accuracy of the time-series forecasting can be enhanced. Fig. 2 illustrates the overall system composition of our proposed model.

The proposed Transformer model is called Constrained, Conditional Transformer (or C2Transformer), and it is specifically designed for enhancing the accuracy of the multi-step peak power forecasting. To be specific, the proposed enhancements to the Transformer model is twofold. We first provide the predicted daily peak power as a *conditional* input to the Transformer model to improve the accuracy of hourly peak power series forecasting. In addition, we add a *constraint* to the model, to ensure that the maximum

value in the generated time-series data matches the provided conditional input. The formal mathematical representation of the proposed enhancements is described as follows.

Let $(y_h)_{h=1,2,...,n}$ be the sequence of actual hourly peak power series for a specific date, and we assume $n = 24$ hours for simplicity in this section. Given the daily peak $p = \max\{y_h | \forall h\}$ for that date, the objective of this study is to 1) predict the hourly peak power series $\hat{\mathbf{y}} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)$ for the given timespan in such a way that the generated sequence is highly accurate (i.e., $y_h = \hat{y}_h, \forall h$), 2) the maximum value among the predictions closely matches the given p (i.e., $p = \max\{\hat{y}_h | \forall h\}$), and 3) the time of the peak matches between the actual and the predicted peak series (i.e., $\arg\max\{y_h | \forall h\} = \arg\max\{\hat{y}_h | \forall h\}$). To accomplish this goal, we propose to enhance the Transformer model by re-defining the objective of the model as follows. Given $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ and p, we solve the following equations to obtain the optimal $\hat{\mathbf{y}} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)$:

$$\min_{\hat{\mathbf{y}}}. \quad \alpha \frac{1}{n} \sum_{\forall h} (y_h - \hat{y}_h | p)^2$$
$$+ (1 - \alpha) \times (p - \max\{\hat{y}_h | \forall h\})^2, \quad (1)$$

where $h = 1, 2, \cdots, n$ and $\alpha \in [0, 1]$ is a design parameter assigning a weight to the first term in the objective function. It can be interpreted as an importance given to the corresponding term. The y_h represents the actual peak power at hour h for a specific date, while $\hat{y}_h$ represents the predicted peak power at hour h, given the daily peak power p for the same date. The first term calculates MSE of the predicted peak sequence. On the other hand, the second term calculates the squared error of the daily peak by computing the difference between the daily peak and the maximum value among the predicted hourly peak series. The problem addressed above includes the proposed enhancement to the vanilla Transformer whose objective function can be written as follows: $\min_{\hat{\mathbf{y}}} \frac{1}{n} \sum_{\forall h} (y_h - \hat{y}_h)^2$.

Furthermore, to enhance both learning efficiency and prediction accuracy, we have employed two sampling methods and ensemble approaches, as illustrated in Fig. 3. Assuming $n = 24$, the objective is to accurately forecast the peak power series for a day, covering the hours from 1 am to 24 pm, ensuring both the predicted sequence and its maximum value are accurate. To train the auto-regressive time series forecasting model, the available dataset consists of hourly peak power series for the past 12 month (or 365 days approximately), resulting in $365 \times 24 = 8760$ hourly peaks. We also have excluded the weekend data for prediction accuracy [37]. The overall of the proposed approach is shown in Fig. 4.

The hourly peak data set is first divided into 365 sets of 24 hourly peaks so that each set includes peaks from 1am to 24pm, which is called periodic sampling. By letting k, i.e., lookback window size as shown in Fig. 2, be the integer multiple q of 24 (i.e., $k = q \times 24, q \in \mathcal{Z}_{++}$), we get periodic samples for training and testing, where Z_{++} is the

$$lookback\ window\ size: k\ (hours),\ forecasting\ horizon = t+1, t+2, \ldots t+n\ (hours)\ for\ d$$

$\hat{y}_{t+1}, \hat{y}_{t+2}, \ldots, \hat{y}_{t+n}$ (hourly peak power prediction)

$$L_1^{mse} = \frac{1}{n} \sum_{h=t+1}^{t+n} (y_{t+h} - \hat{y}_{t+h})^2$$

Hidden space representations

Daily peak power:

$$\hat{y}_{max} = \max\{y_h | h = t+1, t+2, \ldots, t+h\}$$

$y_{t-(k-1)}, \cdots, y_t$ (hourly peak power series for the past k hours)

Encoder (Past data anaylsis)

Decoder (Hourly peak power forecasting)

$$L_2^{mse} = (y_{max} - \hat{y}_{max})^2$$

$\underline{x}_d$ (daily feature vector for day d)
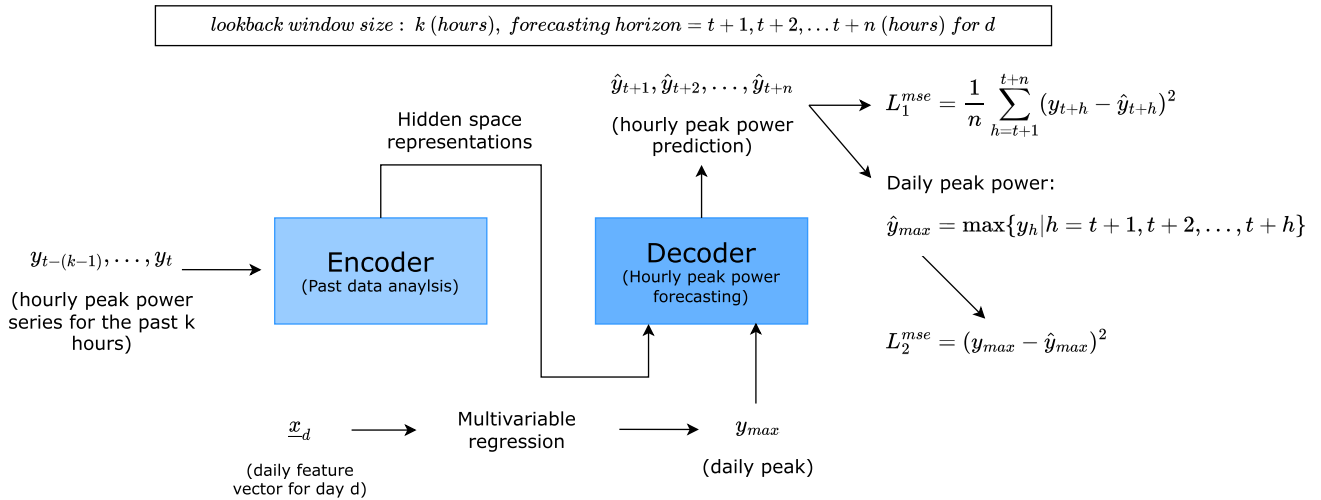
Multivariable regression

y_{max} (daily peak)

FIGURE 2. Proposed hourly peak power forecasting framework, including the proposed enhancements to the Transformer model: i) providing a daily peak as a conditional input to the decoder, and ii) forcing the daily max to be as close as the maximum value in the predicted hourly peak series, where *k* and *n* is the lookback window size and the forecasting horizon, respectively, in the unit of hours.
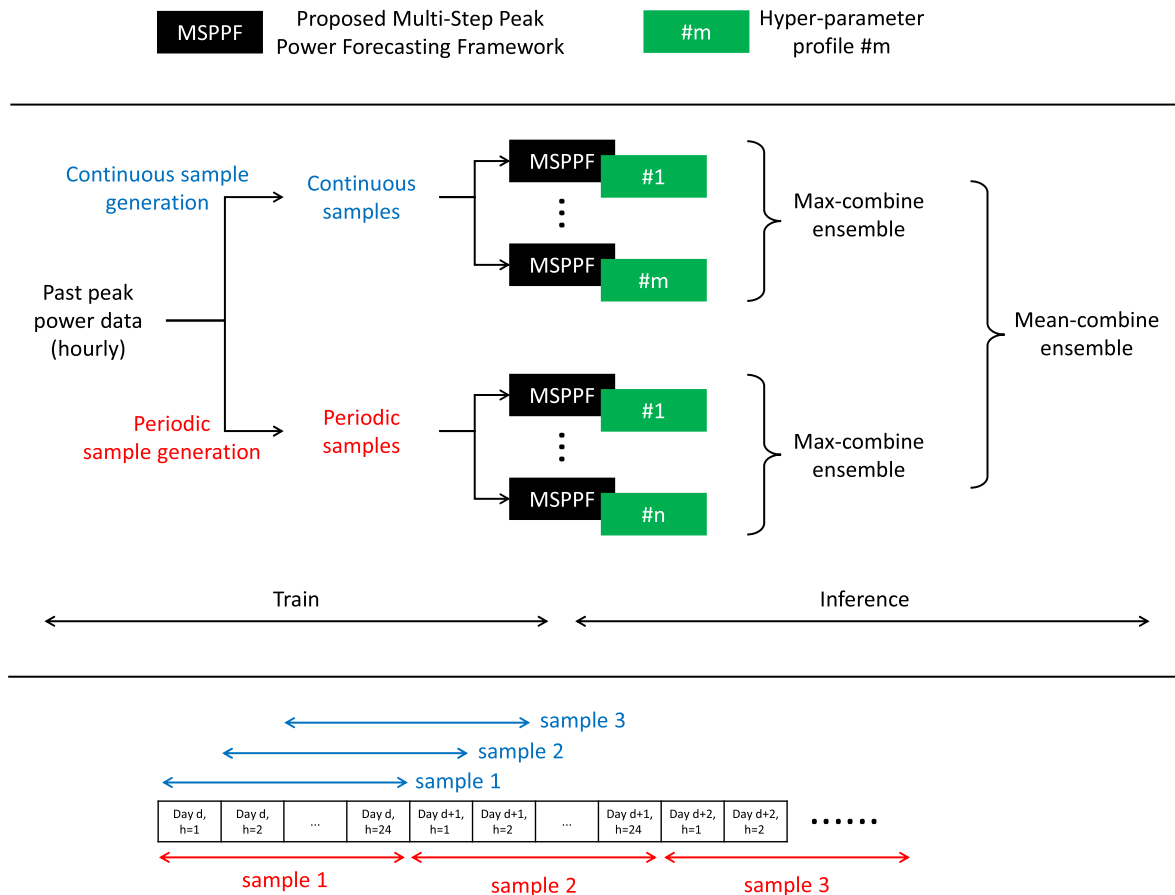

FIGURE 3. The proposed two sample construction methods, i.e., continuous sampling and periodic sampling, and two-level ensemble strategies, i.e., lower-level max-combine and upper-level mean-combine ensemble.

strongly positive integer set. To be specific, for given q, we get $365 - (q - 1)$ samples. This periodic sampling results in the samples, where each sample has q set of hours peaks

from 1 am to 24 pm. This approach prepares the dataset in the same period as that of the model output. Such data preparation method may enhance the accuracy of prediction,
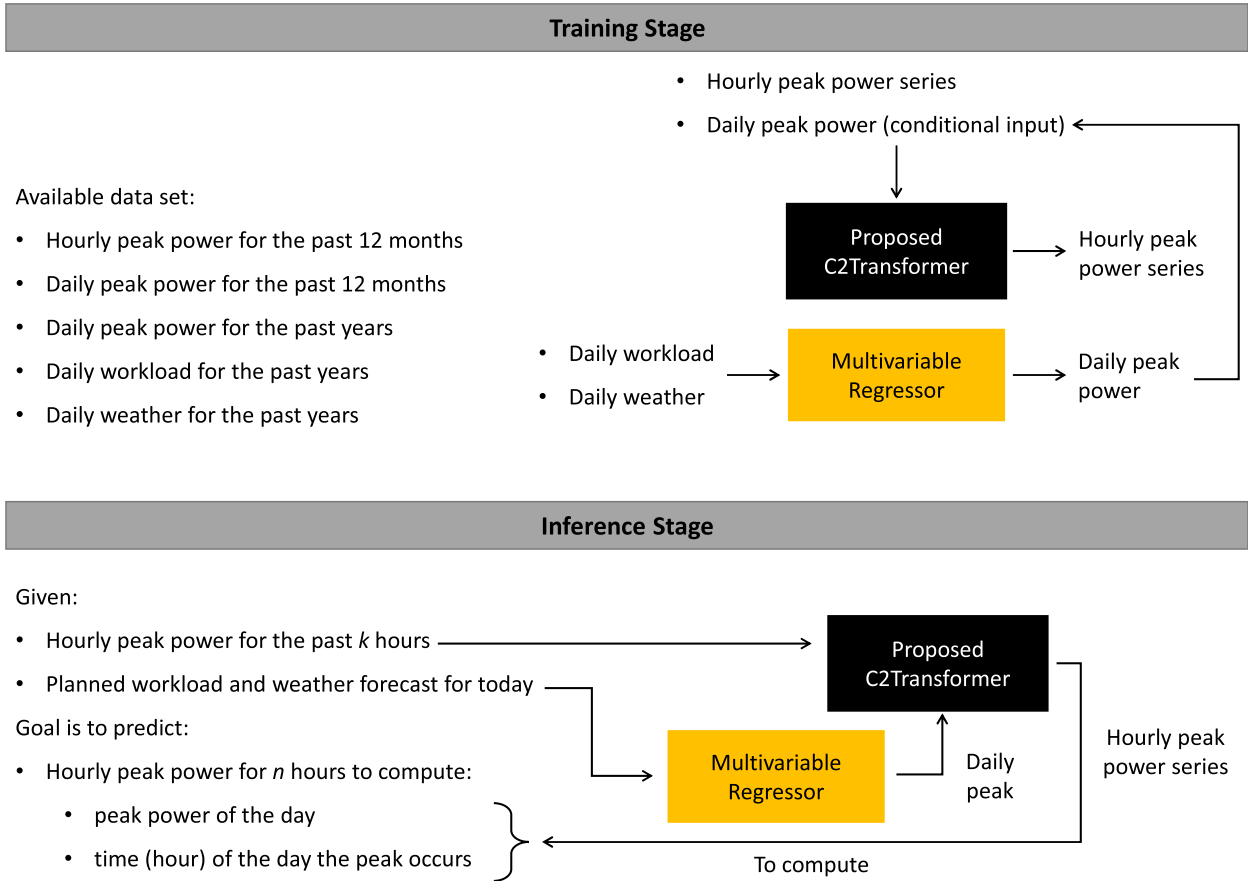
FIGURE 4. Illustration of the overall workflow of both training and inference stages focusing on the input dataset and predicted output.

but the number of samples becomes small which may result in under- or over-fitting [38].

For effective learning, we augmented the data set to increase the number of samples. The other data sample preparation approach we used is called continuous sampling, and it does not follow daily-based sample construction. To be specific, in continuous sampling, hourly peak data is not divided into sets of 1pm-24pm peaks. For a given lookback window size k, it draws k data from the beginning without requiring the hour of the first data to be 1am. Consequently, this approach yields a total of $8760 - (k-1)$ samples. To put it shortly, The periodic sampling approach prepares data set in a way to further enhance the prediction accuracy, while the continuous sampling increases the number of samples and it prevents under-/over-fitting.

The two datasets are then used to train the enhanced Transformer model. Additionally, we have used different sets of hyper-parameters, such as model complexity and learning rate, to make different models have different capability as marked by hyper-parameter profile #m in Fig. 3. From the empirical studies we have performed, models trained with different hyper-parameter profiles have different characteristics. For example, large-capacity models trained with large learning rates are good at capturing sudden changes in the power peaks, while low-capacity models with small

learning rates can capture the overall trend. To leverage such models all together, we propose two-level ensemble approaches. First, we apply a max-combine ensemble method among the predictions generated by the models trained with the same dataset, i.e., either periodic or continuous samples. In particular, the max-combine rule is applied among the models since the goal is to predict peak powers. Subsequently, the ensemble result from the models trained with continuous sampling and the one from those trained with periodic sampling are combined using a mean operation to enhance generalization performance.

IV. EVALUATION

To validate the effectiveness of the proposed C2Transformer in terms of prediction accuracy, we have carried out extensive evaluation and comparison studies. For comparison, we have implemented the following approaches that are widely adopted for multi-step power forecasting: ARIMA, LSTM, 1D CNN, ConvLSTM and vanilla Transformer. We have implemented the proposed C2Transformer along with others in Keras with TensorFlow backend. In addition, the proposed solution in this paper utilizes the XGBoost which is the state-of-the-art multi-variable regressor [39] as a multi-variable regressor to acquire the daily peak prediction to be used as a conditional input. All the evaluations are carried out

on a high-performance desktop computer with the following specifications: Intel© Core™ i7-12700K processor, NVIDIA RTX™ 3080 graphics card and 32GB RAM.

A. DATASET, PERFORMANCE METRIC AND PARAMETER CONFIGURATION

The dataset we have used for evaluation is a combination of multiple actual data from three different sources:

1. Daily workload record from Apr. 2016 to June 2023 acquired from a large-scale manufacturing site owned by one of the largest manufacturing companies in South Korea as well as worldwide. The particular site we have considered in this work is 4 million m^2-wide with three dry docks and five floating docks, and its berthing capacity is 24 vessels per year. The data is acquired by directly downloading it from the company's integrated Enterprise Resource Planning (ERP) and Manufacturing Execution System (MES).

2. Daily weather record from Apr. 2016 to June 2023; the daily average temperature and precipitation of the city where the manufacturing site of our interest is located. The data is acquired and downloaded from the publicly-accessible Open MET Data Portal, operated by Korea Meteorological Administration.[2]

3. Daily peak power history of the plant from Apr. 2016 to June 2023, and hourly peak power history of the plant for the past 12 months (i.e., July 2022 to June 2023). Both are acquired and downloaded from the publicly-accessible KEPCO's portal, and due to their policy, the hourly peaks are provided only for the recent 12 months.

The daily workload record is entered by the manager at scheduled intervals and stored in the company's integrated ERP-MES. Cumulative workload records for a given period (e.g., yearly reports for 2016-2022) can be downloaded from the same system, encompassing both planned and actual workloads. While the workload dataset is managed and accessible internally within the corporate entity, other datasets are controlled by external sources, such as MET Data Portal for the weather history and KEPCO for peak power measurements. The internal/external nature of the data can affect the granularity of the data. For instance, the raw workload records are collected and managed by the corporate, which allows easy computation for various periods such as weekly, monthly, quarterly, semiannually, annually, etc. Conversely, external data often lacks such fine granularity. In other words, one may not guarantee sufficient data granularity of the external dataset which can significantly impact the accuracy of the prediction model.

In addition to internal/external nature of the data, the fluctuations in the data can pose challenges when training a highly accurate model. Fluctuations in the data (or in the sensed readings, in general) may appear as a result from underlying physical status changes or additive undesired noise. In this work, the effect of the former is precisely

captured by the proposed approach which will be discussed shortly. Additionally, the effect of noise is minimized for several reasons. Firstly, the amount of workload and the peak power usage is assumed to be precisely measured by the corporation and KEPCO, respectively. Temperature and precipitation are the values sensed by the dedicated sensors, and they may occasionally be erroneous. However, the proposed approach in this work uses the mean value of each over a certain time horizon, and thus the effect of noise can be cancelled.

The overall illustration of the workflow for both training and inference stages focusing on the aforementioned dataset is depicted in Fig. 4. As a reminder, the hourly peak power series is fed to the proposed C2Transformer auto-regressor, which also takes a conditional input of the daily peak power predicted by the state-of-the-art multi-variable regressor, XGBoost. The proposed C2Transformer then generates hourly peak power series with high accuracy. The generated power series is then used to compute 1) the accurate peak power of the day by taking maximum value among the series, and 2) the time (hour) of the day the peak occurs.

In this paper, the evaluation and comparison studies are preceded by the following data pre-processing techniques. After pre-processing the hourly peak power data, there remains a total of 4344 hourly samples amounting to the 181 days of time span. The training and test dataset are divided into 171 days and 10 days of samples each. The pre-processing includes creating periodic dataset and continuous dataset, replacing the null values with zeros, excluding weekend data [37], unifying the date format among the data from multiple sources, etc. For multi-variable daily peak power prediction which is to be provided as a conditional input to the C2Transformer, a dataset composed of daily peak power series and daily workload/weather logs consisting of nine features are used, excluding the weekend data. Out of the nine features, seven are related to workload (e.g., processing, assembly, painting, etc.), and the remaining two are the average temperature and precipitation.

The multi-step peak power prediction errors are measured by the widely-used error functions such as MAE, MAPE, MSE and RMSE with their standard formulae being as follows, where y_i and $\hat{y}_i$ is the actual value (or called original value) and prediction, respectively.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{3}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

[2]https://data.kma.go.kr/resources/html/en/aowdp.html

TABLE 1. Configuration of the training parameters for deep learning models considered in this study.

	Epoch	Learning Rate	Batch Size	Optimizer
1D CNN	200	0.01	32	Adam
LSTM	200	0.005	32	Nadam
ConvLSTM	200	0.01	32	Adam
Vanilla Transformer	500	0.01	32	Adam
C2Transformer	500	0.01	32	Adam

For evaluation, we have used different values of lookback window sizes k and forecasting horizons n. Also, the configuration of the training parameters for different prediction models are shown in Table 1. To prevent over-fitting, both early-stopping and restoring-best-weights are enabled during training by using the callback functions provided by the Keras framework.

B. EVALUATION & COMPARISON RESULTS

First of all, we have trained a XGBoost multi-variable regressor for daily peak power prediction. The resulting daily peak power prediction is to be given to the proposed C2Transformer as a conditional input. To validate the model accuracy, the prediction results are shown in Fig. 5. The figure shows the normalized predictions after shuffling the data.[3] As it can be seen in the figure, the trained model can successfully capture the trend in daily peak changes. Although does not precisely detect sudden and unexpected drops happened around the day of 60, the trained model has achieved high generalization performance; i.e., MAE of 4225.58, MAPE of 6.19%, MSE of 29335757.66, and RMSE of 5416.25. In the following evaluation results, it has been shown that the achieved accuracy level of the daily peak prediction by XGBoost can contribute to enhancing the prediction accuracy of the hourly peak power by the proposed C2Transformer. Also, any further accuracy enhancement of the daily peak prediction model can help C2Transformer yield more accurate hourly peak prediction.

We have compared the accuracy of the hourly peak power forecasting between the widely used deep learning models and the proposed C2Transformer with different lookback window sizes $k = 120, 240, 360$ and forecasting horizon values $n = 24, 48, 72$. As a reminder, the unit of both k and n is hour, and thus, $k = 120$ and $n = 24$, for example, corresponds to the time span of five days and one day, respectively. The evaluation results are summarized in Table 2, Table 3, Table 4, and Table 5, where Table 2 and Table 3 are the experiment results using periodic data, and Table 4 and Table 5 are the results from the continuous data. For each evaluation configuration, the least

[3]Due to the Non Disclosure Agreement, the raw data cannot be presented as it is. Thus, Fig. 5, Fig. 6 and Fig. 7 show the normalized and shuffled results.
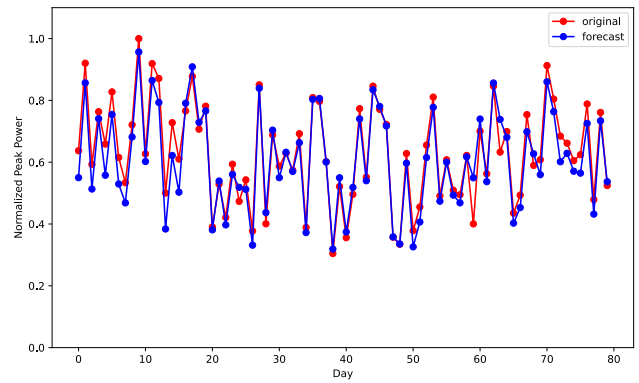

FIGURE 5. Normalized daily peak power prediction results by the trained multi-variable XGBoost model.

FIGURE 6. Normalized hourly peak power estimates of various models trained by the periodic data with the target configuration (i.e., lookback window size $k = 120$ and forecasting horizon $n = 24$).

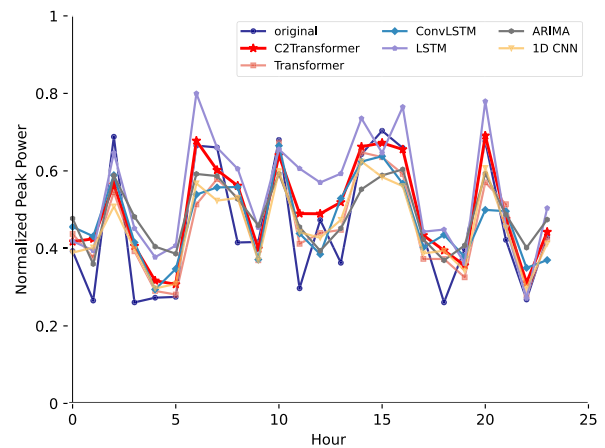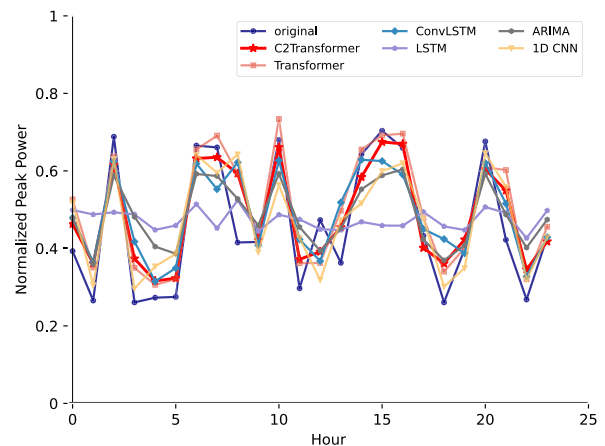**FIGURE 7.** Normalized hourly peak power level estimates of various models trained by the continuous data with the target configuration (i.e., lookback window size $k = 120$ and forecasting horizon $n = 24$).

MAE which is the primary performance measure in this study is highlighted in bold.

As it can be seen in Table 2, when the considered models are trained with periodic dataset, the proposed

TABLE 2. Multi-step forecasting performance of the considered deep learning models trained on periodic data with $n = 24$ where the least MAE is highlighted in bold.

k	1D CNN			LSTM			ConvLSTM			Transformer			C2Transformer		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
120	7968.97	0.196	9384.38	9807.27	0.270	12505.32	9155.02	0.240	10551.96	7040.57	0.173	8410.67	**6358.84**	0.180	8668.21
240	7447.02	0.193	9302.75	7413.99	0.190	9188.33	10856.87	0.320	14008.53	7536.09	0.221	10391.09	**6901.14**	0.191	8713.71
360	8734.69	0.196	10304.30	9759.18	0.199	11606.03	11618.89	0.341	14863.38	8245.21	0.227	10291.42	**7203.94**	0.198	9025.14

TABLE 3. Multi-step forecasting performance of the considered deep learning models trained on periodic data with $k = 120$ where the least MAE is highlighted in bold.

n	1D CNN			LSTM			ConvLSTM			Transformer			C2Transformer		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
24	7968.97	0.196	9384.38	9807.27	0.272	12505.32	9155.02	0.240	10551.96	7040.57	0.173	8410.67	**6358.84**	0.180	8668.21
48	7762.81	0.168	9784.41	8999.97	0.180	11520.73	10368.67	0.224	12790.66	7266.38	0.171	9467.01	**6561.18**	0.160	8400.70
72	6593.26	0.153	8527.23	7007.32	0.154	9218.75	10739.61	0.250	12742.63	6734.83	0.159	11203.10	**6450.38**	0.165	8342.88

TABLE 4. Multi-step forecasting performance of the considered deep learning models trained on continuous data with $n = 24$ where the least MAE is highlighted in bold.

k	1D CNN			LSTM			ConvLSTM			Transformer			C2Transformer		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
120	8031.22	0.197	9597.56	14484.45	0.363	15923.61	7896.01	0.212	9519.89	6405.09	0.169	8140.24	**6219.50**	0.168	7474.81
240	8473.21	0.214	9587.97	13091.58	0.332	14276.14	7864.46	0.216	9428.34	8353.20	0.217	9236.45	**6544.85**	0.178	7972.38
360	9916.41	0.201	11462.90	14422.14	0.305	18227.95	8915.17	0.217	10444.08	8988.08	0.244	10513.02	**7149.67**	0.206	8972.52

TABLE 5. Multi-step forecasting performance of the considered deep learning models trained on continuous data with $k = 120$ where the least MAE is highlighted in bold.

n	1D CNN			LSTM			ConvLSTM			Transformer			C2Transformer		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
24	8031.22	0.197	9597.56	14484.45	0.363	15923.61	7896.01	0.212	9519.89	6405.09	0.169	8140.24	**6219.50**	0.168	7474.81
48	8245.63	0.183	9925.22	12133.66	0.300	13800.37	9171.49	0.214	10693.03	6100.04	0.154	8012.49	**5678.86**	0.138	7307.02
72	7592.98	0.177	9079.27	10622.47	0.258	12316.49	10933.76	0.241	13388.66	6220.94	0.150	7940.07	**5577.49**	0.138	7350.56

TABLE 6. Prediction accuracy of the daily peak; trained by the periodic data with the target configuration (i.e., lookback window size $k = 120$ and forecasting horizon $n = 24$).

Approaches	MAE
ARIMA	9972.18
1D CNN	7920.71
LSTM	9617.98
ConvLSTM	3859.37
Vanilla Transformer	3093.18
C2Transformer	**1257.57**

TABLE 7. Prediction accuracy of the daily peak; trained by the continuous data with the target configuration (i.e., lookback window size $k = 120$ and forecasting horizon $n = 24$).

Approaches	MAE
ARIMA	9972.18
1D CNN	5814.09
LSTM	17962.79
ConvLSTM	7483.90
Vanilla Transformer	3033.35
C2Transformer	**2898.56**

TABLE 8. Prediction accuracy of the daily peak hour prediction trained by the periodic data with the target configuration (i.e., lookback window size $k = 120$ and forecasting horizon $n = 24$).

Approaches	Actual peak time (h)	Predicted peak time (h)	MAE
ARIMA	10	15	5
1D CNN	10	8	2
LSTM	10	14	4
ConvLSTM	10	**9**	**1**
Vanilla Transformer	10	**9**	**1**
C2Transformer	10	13	3

TABLE 9. Prediction accuracy of the daily peak hour prediction trained by the continuous data with the target configuration (i.e., lookback window size $k = 120$ and forecasting horizon $n = 24$).

Approaches	Actual peak time (h)	Predicted peak time (h)	MAE
ARIMA	10	15	5
1D CNN	10	13	3
LSTM	10	12	2
ConvLSTM	10	8	2
Vanilla Transformer	10	15	5
C2Transformer	10	**10**	**0**

C2Transformer has achieved the least MAE for all k. In particular, when $k = 120$, C2Transformer has shown the least performance with MAE being 6358.84. Among the rest deep learning models, the vanilla Transformer outperformed on average, showing the effectiveness of using the Transformer model for multi-step time-series forecasting. Also, the superiority of C2Transformer compared to the vanilla Transformer validates the effectiveness of the proposed enhancements, i.e., conditional input and constrained output. One interesting finding in this evaluation is that in the case of C2Transformer, increasing the value of k rather decreased the MAE/MAPE/RMSE performance, meaning that providing too much of the history data can degrade the multi-step prediction accuracy under the assumed configuration.

The following Table 3 also shows the performance of the considered deep learning models that are trained with periodic dataset. For the same lookback window of $k = 120$ (i.e., five days of time), the value of n was varied in this evaluation. As shown in the table, the proposed C2Transformer has yielded the least MAE error over all considered values of n. In other words, C2Transformer can successfully forecast peak power series over longer-term spans as well. One of the interesting findings drawn from this evaluation is that for the considered values of n, the increase in the forecasting horizon did not affect the MAE performance of C2Transformer much, validating the robustness of the proposed approach. It also was the same to 1D CNN, LSTM and vanilla Transformer, but ConvLSTM resulted in the opposite result, i.e., MAE is proportional to n in ConvLSTM.

The other two tables, Table 4 and Table 5, show the evaluation results of the models trained with the continuous dataset. For all different values of k and n, the proposed C2Transformer outperformed the rest. With $n = 24$, the C2Transformer yielded the least MAE error when $k = 120$, which was the case with periodic dataset as well. Surprisingly, as the value of n increases, C2Transformer recorded the better MAE performance, which shows the robustness of the C2Transformer with respect to the forecast horizon. Excluding the C2Transformer, the vanilla Transformer outperformed the rest as it did with the periodic data.

Overall, the C2Transformer has achieved the best performance in terms of MAE regardless of the dataset. Although C2Transformer resulted in a slightly better performance when it is trained with continuous data than periodic, the difference on average is not significant. The superiority of the C2Transformer especially compared to the vanilla Transformer is due to the two key advancements we propose in this study: 1) the enhancements to the Transformer model by adding an additional output constraint and by providing a conditional input, and 2) two-level ensemble among the models trained with different hyper-parameters and dataset.

Except the C2Transformer, the vanilla Transformer has recorded the best MAE performance among the rest, showing the effectiveness of the Transformer model for multi-step time-series forecasting. On average, the LSTM model has recorded the largest MAE errors. Although it has effectively

TABLE 10. Best performance of the multi-step hourly peak series prediction of the considered approaches.

Approaches	MAE	MAPE	RMSE
ARIMA	9059.47	0.244	10087.44
1D CNN ($k = 240$, $n = 24$, periodic)	7447.02	0.193	9302.75
LSTM ($k = 120$, $n = 72$, periodic)	7007.32	0.154	9218.75
ConvLSTM ($k = 240$, $n = 24$, continuous)	7864.46	0.216	9428.34
Vanilla Transformer ($k = 120$, $n = 48$, continuous)	6100.04	0.154	8012.49
C2Transformer ($k = 120$, $n = 24$, ensemble)	**5858.16**	0.165	7606.68

TABLE 11. Best performance of the daily peak power prediction.

Approaches	MAE
ARIMA	9972.18
1D CNN ($k = 240$, $n = 24$, periodic)	3421.94
LSTM ($k = 120$, $n = 72$, periodic)	3818.30
ConvLSTM ($k = 240$, $n = 24$, continuous)	5702.57
Vanilla Transformer ($k = 120$, $n = 48$, continuous)	5595.31
C2Transformer ($k = 120$, $n = 24$, ensembled)	**2296.38**

TABLE 12. Best performance of daily peak time prediction of the considered approaches.

Approaches	Actual peak time (h)	Predicted peak time (h)	MAE
ARIMA	10	15	5
1D CNN ($k = 240$, $n = 24$, periodic)	10	13	3
LSTM ($k = 120$, $n = 72$, periodic)	10	13	3
ConvLSTM ($k = 240$, $n = 24$, continuous)	10	15	5
Vanilla Transformer ($k = 120$, $n = 48$, continuous)	10	15	5
C2Transformer ($k = 120$, $n = 24$, ensemble)	10	**10**	**0**

predicted the large-scale trend in peak power changes, LSTM failed to precisely predict the small-scale variations. Among the two deep learning models utilizing convolutional layers, i.e., 1D CNN and ConvLSTM, 1D CNN resulted in robust MAE performance against different values of k and n. On the other hand, ConvLSTM performed poorly especially when the values of n varies.

The following Fig. 6 and Fig. 7 show the time-series forecasting results of the considered deep learning models as well as ARIMA which is a widely-used statistical approach that are trained with periodic and continuous data, respectively, when lookback window size $k = 120$ and forecasting horizon $n = 24$. This particular pair of values ($k = 120, n = 24$) is called target configuration in this study, and will be used to compare the performance among the considered models in this study. As it can be clearly seen from both figures, C2Transformer can precisely capture the changes in the peak power and thus, it resulted in the smallest MAE performance. To be specific, C2Transformer trained with periodic data (i.e., Fig. 6) has shown less-smooth

predictions with effective capturing of sudden changes in hourly peaks. On the other hand, C2Transformer trained with continuous data (i.e., Fig. 7) has shown more precise fitting to the overall trend in peak changes by making smoother predictions, but with less effective capturing of sudden changes. In other words, the smoother model outperformed in capturing the large-scale changes, while the less-smoother model predicted the sudden changes with higher accuracy. This finding motivated us to further combine (i.e., the 2nd level ensemble; see Fig. 3) the model trained by the periodic data with the model trained by the continuous data so that the ensemble model can precisely capture the large- and small-scale changes of peaks simultaneous.

Considering the importance of predicting the daily peak which can be done by taking the maximum value among the predicted hourly peak series, we have carried out a comparison study for the daily peak accuracy among the considered methods in this study. The Table 6 and Table 7 show the MAE of the daily peaks predicted by the considered models trained by the periodic data and continuous data,

respectively. Regardless of the dataset whether it is periodic or continuous, the C2Transformer outperformed the rest by having the least error in predicting the daily peak. Especially when it is trained by the periodic dataset, the error becomes 1257 which is substantially less than any other approaches considered. The vanilla Transformer has shown the second best performance, validating its robustness against the dataset used for training. The ARIMA and LSTM resulted in high errors, implying their limited capacities in making multi-step predictions with the data having complex variations.

We also have carried out an evaluation on the prediction accuracy in forecasting the time (i.e., hour) of the day when the peak occurs. The evaluation and comparison results with the target configuration are summarized in Table 8 and Table 9. The former shows the results from the models trained with periodic data, while the latter with continuous data. On average, ARIMA and recorded a large MAE error in predicting the peak hour, while both ConvLSTM and C2Transformer resulted in the least MAE error. However, in what follows we will show that by performing an additional ensemble between the two C2Transformer models that are trained with periodic and continuous data each, C2Transformer can significantly enhance the accuracy of the peak time prediction.

Finally, we have compared the best-performing instances of the considered approaches, and the results are summarized in Table 10, Table 11 and Table 12. For each of the results included in the tables, we have chosen the values of k and n that yielded the least MAE for each approach. For example, 1D CNN recorded the least MAE when $k = 240$ and $n = 24$. On the other hand, C2Transformer resulted in the least MAE when $k = 120$ and $n = 24$ which is the same as the target configuration. As it can be seen from Table 10, C2Transformer resulted in the least MAE errors of 5858.16 among all approaches. Similarly, for the prediction of the peak time, C2Transformer recorded the least error (i.e., zero hour, see Table 12) with the MAE error of the daily peak value being 2296.38 as shown in Table 11. This performance enhancement validates one of the technical advancements proposed in this study, i.e., the two-level ensemble as shown in Fig. 3. The first-level ensemble among the models trained with either continuous or periodic samples enhance the accuracy of the trained model by carrying out max-combine ensemble. Then, the proposed approach combines the two ensemble'd models via mean-combine rule so that the generalization performance can be further enhanced. Such two-level ensemble with different combine approaches can effectively enhance the prediction performance of not only the peak series but also the time of the peak occurs.

V. CONCLUSION

In this study, we have introduced an innovative and effective framework for power forecasting, with a primary focus on predicting peak power usage in a large-scale plant. To enhance the accuracy of the multi-step peak predictions,

we have proposed a set of enhancements to the Transformer model, i.e., incorporating a conditional input and constrained output along with periodic/continuous sampling and two-level ensemble. This conditional input plays a crucial role in improving prediction accuracy, ensuring that the predicted daily peak aligns with the maximum value among the predicted hourly peak power levels. To generate the daily peak prediction used as an input for enhancing the proposed C2Transformer, the state-of-the-art XGBoost has been utilized in this study. This daily prediction is subsequently fed into the C2Transformer decoder module, enabling precise hourly peak power predictions over n hours of time span. Through comprehensive evaluations with the actual dataset, we have validated the effectiveness of the proposed approach. Furthermore, in comparison to the widely used deep learning models for time series predictions, we have shown that the proposed approach outperforms consistently.

REFERENCES

[1] EIA. (Oct. 6, 2021). *Energy Consumption Worldwide From 1990 to 2020, With a Forecast Until 2050* (in Quadrillion British Thermal Units)*. [Online]. Available: https://www.statista.com/statistics/263981/ worldwide-energy-consumption-forecast-between-1980-and-2030/

[2] M. A. Mat Daut, M. Y. Hassan, H. Abdullah, H. A. Rahman, M. P. Abdullah, and F. Hussin, "Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review," *Renew. Sustain. Energy Rev.*, vol. 70, pp. 1108–1118, Apr. 2017.

[3] V. Badrinath Krishna, R. K. Iyer, and W. H. Sanders, "ARIMA-based modeling and validation of consumption readings in power grids," in *Proc. Int. Conf. Crit. Inf. Infrastruct. Secur.*, Berlin, Germany, Oct. 2015, pp. 199–210.

[4] D. Alberg and M. Last, "Short-term load forecasting in smart meters with sliding window-based ARIMA algorithms," *Vietnam J. Comput. Sci.*, vol. 5, nos. 3–4, pp. 241–249, Sep. 2018.

[5] X. Zhou, Z. Xu, X. Yu, and Y. Xia, "The analysis and prediction of power consumption in industry based on machine learning," in *Proc. Chin. Autom. Congr. (CAC)*, Hangzhou, China, Nov. 2019, pp. 5738–5742.

[6] R. Yashaswini and S. Vinay, "Improved energy consumption prediction using XGBoost with hyperparameter tuning," in *Proc. 4th Int. Conf. Emerg. Res. Electron., Comput. Sci. Technol. (ICERECT)*, New Delhi, India, Dec. 2022, pp. 1–3.

[7] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, Aug. 2016, pp. 785–794.

[8] J. Kim, J. K. Choi, Y. Heo, H. Seo, and J. Han, "Prediction of power consumption in the factory using long-short term memory," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2019, pp. 1211–1214.

[9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[10] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, Sep. 2019.

[11] D. Energy. (Oct. 4, 2019). *Energy Demand Charges Explained: What They Are and Why You Should Care*. Accessed: Sep. 27, 2023. [Online]. Available: https://sustainablesolutions.duke-energy.com/resources/energy-demand-charges-explained/

[12] J. Walther and M. Weigold, "A systematic review on predicting and forecasting the electrical energy consumption in the manufacturing industry," *Energies*, vol. 14, no. 4, p. 968, Feb. 2021.

[13] G. Hafeez, K. S. Alimgeer, A. B. Qazi, I. Khan, M. Usman, F. A. Khan, and Z. Wadud, "A hybrid approach for energy consumption forecasting with a new feature engineering and optimization framework in smart grid," *IEEE Access*, vol. 8, pp. 96210–96226, 2020.

[14] Y. L. Cheng, M. H. Lim, and K. H. Hui, "Impact of Internet of Things paradigm towards energy consumption prediction: A systematic literature review," *Sustain. Cities Soc.*, vol. 78, Mar. 2022, Art. no. 103624.

[15] D. Feng, C. Jiang, G. Lim, L. J. Cimini, G. Feng, and G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 167–178, 1st Quart., 2013.

[16] M. Dayarathna, Y. Wen, and R. Fan, "Data center energy consumption modeling: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 732–794, 1st Quart., 2016.

[17] G. Urbikain and L. N. L. De Lacalle, "MoniThor: A complete monitoring tool for machining data acquisition based on FPGA programming," *SoftwareX*, vol. 11, Jan. 2020, Art. no. 100387.

[18] E. Tapia, L. Sastoque-Pinilla, U. Lopez-Novoa, I. Bediaga, and N. L. De Lacalle, "Assessing industrial communication protocols to bridge the gap between machine tools and software monitoring," *Sensors*, vol. 23, no. 12, p. 5694, Jun. 2023.

[19] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, no. 3, pp. 278–305, Mar. 1991.

[20] A. Pankratz, *Forecasting With Univariate Box-Jenkins Models: Concepts and Cases*. Hoboken, NJ, USA: Wiley, 2009.

[21] J. W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," *J. Oper. Res. Soc.*, vol. 54, no. 8, pp. 799–805, Aug. 2003.

[22] V. Mnih, H. Larochelle, and G. E. Hinton, "Conditional restricted Boltzmann machines for structured output prediction," 2012, *arXiv:1202.3748*.

[23] G. W. Taylor, G. E. Hinton, and S. T. Roweis, "Two distributed-state models for generating high-dimensional time series," *J. Mach. Learn. Res.*, vol. 12, pp. 1025–1068, Mar. 2011.

[24] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada, Dec. 2014, pp. 3104–3112.

[25] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada, Dec. 2014, pp. 2672–2680.

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, Dec. 2017, pp. 5998–6008.

[27] C. Tian, C. Li, G. Zhang, and Y. Lv, "Data driven parallel prediction of building energy consumption using generative adversarial nets," *Energy Buildings*, vol. 186, pp. 230–243, Mar. 2019.

[28] J. Li, J.-H. Cheng, J.-Y. Shi, and F. Huang, "Brief introduction of back propagation (BP) neural network algorithm and its improvement," in *Advances in Computer Science and Information Engineering*, vol. 2. Cham, Switzerland: Springer, 2012, pp. 553–558.

[29] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231206000385

[30] Z. Rao and Y. Zhang, "Transformer-based power system energy prediction model," in *Proc. IEEE 5th Inf. Technol. Mechatronics Eng. Conf. (ITOEC)*, Chongqing, China, Jun. 2020, pp. 913–917.

[31] J. Liu and L. E. Brown, "Prediction of hour of coincident daily peak load," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Washington, DC, USA, Feb. 2019, pp. 1–5.

[32] T. Zhang, X. Zhang, O. Rubasinghe, Y. Liu, Y. Chow, H. Iu, and T. Fernando, "Long-term energy and peak power demand forecasting based on sequential-XGBoost," *IEEE Trans. Power Syst.*, early access, pp. 1–16, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10163761, doi: 10.1109/TPWRS.2023.3289400.

[33] A. Roberts and J. Roberts, *Multiple Regression: A Practical Introduction*. Thousand Oaks, CA, USA: Sage, 2020. [Online]. Available: https://books.google.co.kr/books?id=V_qbzQEACAAJ

[34] R. Bekkerman. (2015). *The Present and the Future of the KDD Cup Competition: An Outsider's Perspective*. Accessed: Sep. 27, 2023. [Online]. Available: https://www.linkedin.com/pulse/present-future-kdd-cup-competition-outsiders-ron-bekkerman

[35] N. Wu, B. Green, X. Ben, and S. O'Banion, "Deep transformer models for time series forecasting: The influenza prevalence case," 2020, *arXiv:2001.08317*.

[36] K. Goswami, A. Ganguly, and A. K. Sil, "Comparing univariate and multivariate methods for short term load forecasting," in *Proc. Int. Conf. Comput., Power Commun. Technol. (GUCON)*, Sep. 2018, pp. 972–976.

[37] Z. H. Osman, M. L. Awad, and T. K. Mahmoud, "Neural network based approach for short-term load forecasting," in *Proc. IEEE/PES Power Syst. Conf. Expo.*, Phoenix, AZ, USA, Mar. 2009, pp. 1–8.

[38] P. Thanapol, K. Lavangnananda, P. Bouvry, F. Pinel, and F. Leprévost, "Reducing overfitting and improving generalization in training convolutional neural network (CNN) under limited sample sizes in image recognition," in *Proc. 5th Int. Conf. Inf. Technol. (InCIT)*, ChonBuri, Thailand, Oct. 2020, pp. 300–305.

[39] XGBoost. *XGBoost Python Package*. Accessed: Apr. 11, 2023. [Online]. Available: https://xgboost.readthedocs.io/en/stable/python/

NAHYEON KIM received the B.S. degree from the School of Computer Science and Engineering, Pusan National University, Republic of Korea. She was an Exchange Student in computer science from the State University of New York at Oswego, for the 2022 Fall semester. Her research interests include artificial intelligence, deep learning, machine learning, big data analysis, and computer vision.

ARAVINDH BALARAMAN received the B.S. degree in computer applications from Thiruvalluvar University, Tamil Nadu, India, in 2017. He is currently pursuing the master's degree with the School of Computer Science and Engineering, Pusan National University, Republic of Korea. He has five years of industrial experience as a Cloud–DevOps developer. His research interests include artificial intelligence, deep learning, machine learning, cloud computing, and reinforcement learning.

KEUNWOO LEE received the B.S. degree in mechanical engineering from Inha University, Republic of Korea, in 2010. He is currently a Senior Research Engineer with Samsung Heavy Industries Company Ltd., Republic of Korea. His research interests include artificial intelligence, big data analysis, machine learning, and deep learning.

TAEWOON KIM (Member, IEEE) received the B.S. degree in computer science and engineering from Pusan National University, Republic of Korea, in 2008, the M.S. degree in information and mechatronics from the Gwangju Institute of Science and Technology, Republic of Korea, in 2010, and the Ph.D. degree in computer engineering from Iowa State University, Ames, IA, USA, in 2018. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Pusan National University. From 2018 to 2022, he was an Assistant Professor with the School of Software, Hallym University, Republic of Korea. From 2010 to 2013, he was a Research Engineer with the Telecommunications Technology Association, Republic of Korea. His research interests include modeling, optimization, and protocol design for wireless networking systems, such as WLAN, the IoT/sensor networks, HetNets and C-RANs, cloud/edge computing, and reinforcement learning-based autonomous agent control.