Predicting peak day and peak hour of electricity demand with ensemble machine learning

Tao Fu, Huifen Zhou, Xu Ma, Z. Jason Hou* and Di Wu*

Pacific Northwest National Laboratory, Richland, WA, United States

Battery energy storage systems can be used for peak demand reduction in power systems, leading to significant economic benefits. Two practical challenges are 1) accurately determining the peak load days and hours and 2) quantifying and reducing uncertainties associated with the forecast in probabilistic risk measures for dispatch decision-making. In this study, we develop a supervised machine learning approach to generate 1) the probability of the next operation day containing the peak hour of the month and 2) the probability of an hour to be the peak hour of the day. Guidance is provided on preparation and augmentation of data as well as selection of machine learning models and decision-making thresholds. The proposed approach is applied to the Duke Energy Progress system and successfully captures 69 peak days out of 72 testing months with a 3% exceedance probability threshold. On 90% of the peak days, the actual peak hour is among the 2 h with the highest probabilities.

KEYWORDS

batteries, electricity demand, energy storage, ensemble machine learning, load forecast

1 Introduction

Many cooperatives, municipally owned utilities, and other types of load serving entities (LSE) purchase power from electricity markets or through power purchase contracts. A capacity charge is paid based on the coincident demand during system peak hours. Effectively reducing the peak demand leads to significant economic and environment benefits, as well as improved power grid security and stability (Dai et al., 2021). Battery energy storage systems (BESS) are promising for peak demand reduction because of their flexibility and instantaneous response capability. An LSE does not know exactly when the peak hour will occur. Simply discharging BESS on all high-load days helps capture the peak hour and reduce the coincident demand, but causes unnecessary battery degradation and energy losses associated with charging/discharging. In addition, due to limited energy capacity, BESS may not be able to discharge at the rated power in all high-load hours. Therefore, advanced peak day and peak hour forecast methods are critical to maximizing benefits from BESS for peak demand reduction.

Predictions of the monthly peak hour can be derived from load forecast that spans the entire month. Traditional methods, such as autoregressive integrated moving average (ARIMA), forecast load with univariate historical load. More

advanced nonlinear machine learning (ML) methods have also been proposed for load forecasting with multivariate predictors, including weather, calendar, and economics (Hong and Fan, 2016). Examples of advanced nonlinear ML methods are K-nearest neighbors (KNN) (El-Attar et al., 2009), fuzzy regression models (Hong and Wang, 2014), support vector machine (SVM) (Niu et al., 2010), gradient boosting machine (GBM) (Massaoudi et al., 2021), random forest (RF) (Cheng et al., 2012; Huang et al., 2016), and artificial neural networks (ANN) (El Desouky and Elkateb, 2000; Ringwood et al., 2001; Saini, 2008). These approaches can also be combined to improve load forecasting. For example, El Desouky and Elkateb (2000) developed a hybrid ANN with ARIMA for load forecasting. El-Attar et al. (2009) proposed a multivariate load forecasting approach by combining support vector regression with a KNN local prediction framework. Berrisch et al. (2022) combined generalized additive models and deep ANN to predict high-resolution minimum and maximum peak load given only lower resolution data and weather information. Mao et al. (2009) proposed a self-organizing fuzzy neural network (SOFNN) for short-term load forecasting with a bilevel optimization method to find the best pre-training parameters. A monthly peak hour can be identified based on the value of the hourly load forecast. Such a method requires the hourly load forecast toward the end of the month. There are significant uncertainties associated with long-term hourly load forecast. A major challenge of predicting peak hours based on load forecast is how to model and quantify uncertainties, considering varying weather conditions (e.g., temperature, humidity, and wind speed) and nonlinear relationships between weather and load (El Desouky and Elkateb, 2000).
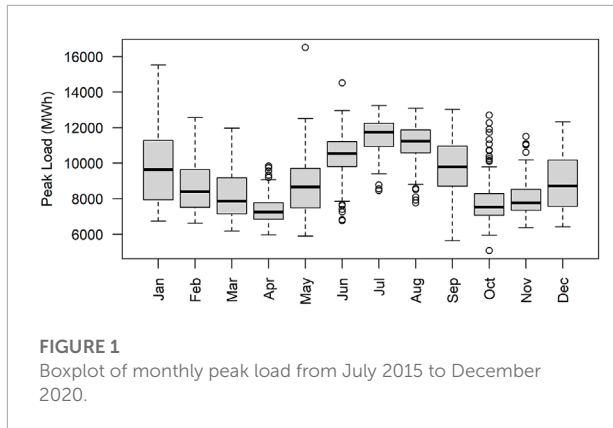
Efforts have also been made to directly predict peak hours. For example, Goodwin and Yazidi (2016) proposed an SVM and Gaussian mixture classification approach to directly estimate the peak hours for the next 7 days based on historical load. Leveraging short-term load forecasts, Jiang et al. (2014) adopted a probabilistic approach for estimating the probability of the next day containing the highest hourly demand of a year. A Naive Bayesian classification model was proposed for classifying whether an hour is a 5CP (top 5 coincident peaks of a year) hour (Ryu et al., 2016). Liu and Brown (2019a) proposed a convolutional neural network (CNN) classification model to predict 24 h ahead whether a day is a 5CP day. For the peak hour model, they built separate models for summer and winter using different classification methods: Naive Bayes, SVM, RF, AdaBoost, CNN, long short-term memory (LSTM), and stacked autoencoder (Liu and Brown, 2019b). More recently, Saxena et al. (2019) developed a hybrid classification model combining ARIMA, logistic regression (LR), and ANN for peak day prediction.

Both load forecast and direct peak prediction ML models require enough data for training and validation. In practice, however, complete historical records of load and weather are not always available, which warrants the data augmentation effort. In addition, these aforementioned approaches either ignored trailing or leading effects of factors or skipped dimension reduction and physical interpretation of the factor contributions. To summarize, despite the progress to date, additional research and development are needed to better support BESS dispatch decision-making for peak demand reduction, including 1) quantifying uncertainties associated with peak day and peak hour predictions, 2) addressing the data inadequacy issue *via* ML data augmentation, 3) evaluating choices of peak day probability thresholds for decision-making considering both prediction accuracy and precision, 4) including all physical and temporal factors to fully capture their correlations and causalities with system load peaking behaviors, e.g., trailing and leading effects, and 5) integrating ensemble ML model selection techniques to deal with factor collinearity, mixed data types, and overfitting, as well as trustworthy, explainable ML prediction.

In this paper, we propose an ensemble learning approach taking advantage of multiple ML techniques, including RF, GBM, and LR, for predicting peak day and peak hour to better support BESS dispatch decision-making. To quantify the prediction uncertainties, we develop, validate, compare, and select optimal ML models to generate the probability of the next day to be the monthly peak day (note that one and only one peak day is defined for each billing cycle, which is a month), and the probability of an hour to be the peak hour of a day. The proposed approach also features ML-based data augmentation to address the data inadequacy issues and a procedure to select exceedance thresholds for decision-making. We use a comprehensive set of predictors, including actual load in previous days of the month, day-ahead load forecast, weather forecast, and their derivatives, as well as temporal factors to help improve peak predictions. By applying bagging and boosting techniques, the ensemble tree-based ML method can help avoid overfitting and deal with unbalanced data of mixed types, e.g., categorical vs continuous (Domingos, 2012; Breiman, 2001; Friedman, 2001; Cieslak and Chawla, 2008;, 2012; Jaiswal and Samikannu, 2017). We also use naive load-based peak predictions to demonstrate the effectiveness of the proposed ML-based prediction framework.

2 Dataset

The proposed ML-based prediction framework is generic and can be applied to any system. The Duke Energy Progress (DEP) system is used for illustration and analysis presented in

FIGURE 1
Boxplot of monthly peak load from July 2015 to December 2020.

FIGURE 2
Actual hourly load 2015−2020. The dots represent the actual peak hours in **(A)** January and **(B)** July.

FIGURE 3
Daily peak load vs daily maximum and minimum temperature in **(A)** January and **(B)** July.

this paper. The input dataset includes historical load and weather data, which are described as follows.

2.1 DEP load

The DEP system consists of two balancing authorities: DEP East and DEP West. The data are obtained from the Energy Information Administration (EIA, 2021), including both day-ahead forecasts and actual demand with an hourly resolution. Complete records are only available for about 6 years starting from July 2015. Data augmentation is proposed and used to generate 21 years of data for training and testing purposes, as described in Section 3.3.

A boxplot of monthly peak load is provided in Figure 1. As can be seen, the highest monthly peak loads typically occur in mid-winter (e.g., January) and mid-summer (e.g., July). Figure 2 plots the system load vs time for January and July. The peak day can occur at the beginning, in the middle, or at the end of a month, and there is no obvious pattern, which increases the difficulty for peak day prediction.

2.2 Weather data

The weather datasets are obtained from the National Oceanic and Atmospheric Administration website (NOAA, 2021). In this study, the weather data at the Raleigh weather station are used as it is the closest weather station to the city of Raleigh, which is the geographic center of DEP and the second biggest city in North Carolina. The raw weather data include dry-bulb air temperature, dew point temperature, wind speed, and visibility, with temporal resolutions between 15 min to an hour from 2000 to 2020. To match the temporal resolution and coverage of the load data, the weather data are mapped to the nearest exact hours to generate hourly data. In addition, the hourly humidity H is calculated
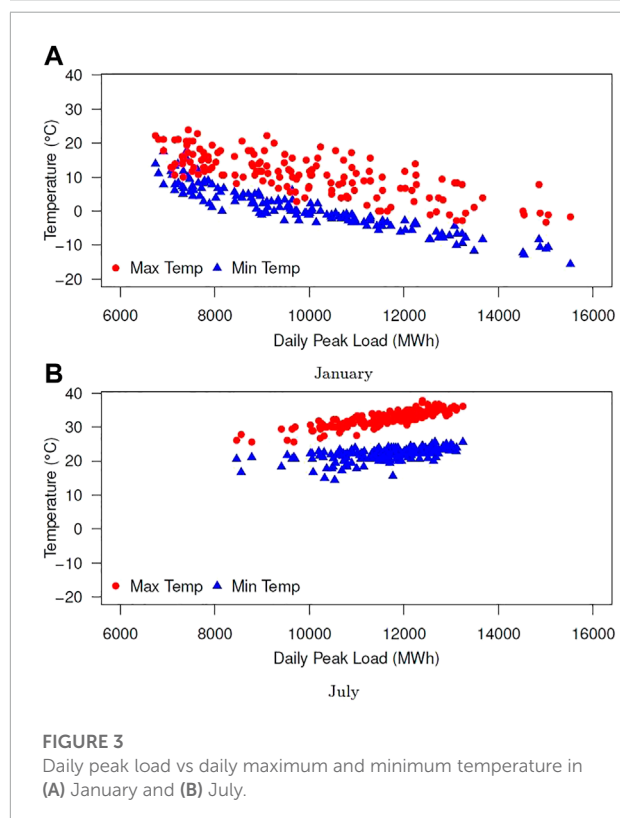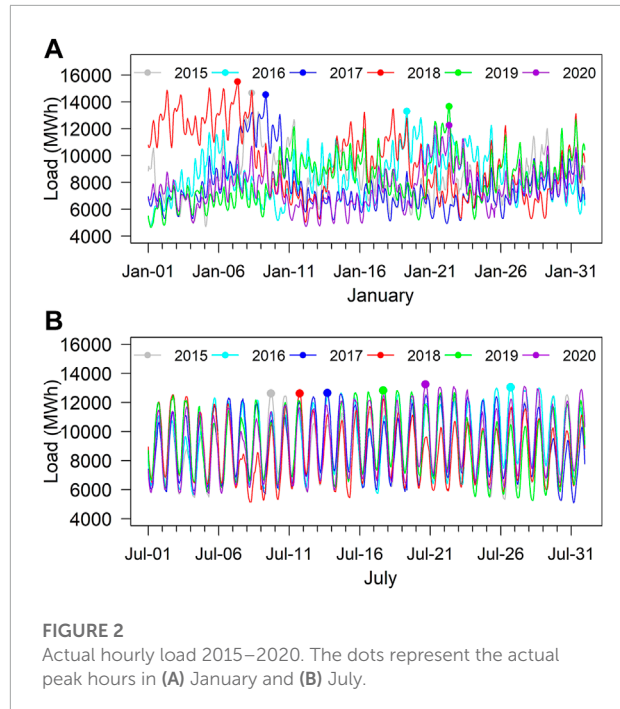
according to Bosen (1958):

$$H = \exp\left(\frac{17.625 \times DT}{243.04 + DT} - \frac{17.625 \times T}{243.04 + T} \right) \times 100\% , \qquad (1)$$

where T is the dry-bulb air temperature (°C), DT is the dew point temperature (°C), and $\exp(\cdot)$ is the exponential function.

TABLE 1 R-squared of daily peak load vs daily minimum and maximum temperature.

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
T_min	0.86	0.72	0.58	0.04	0.57	0.39	0.30	0.33	0.36	0.38	0.40	0.74
T_max	0.58	0.50	0.29	0.11	0.80	0.80	0.73	0.77	0.77	0.50	0.26	0.56

3 Methodology

3.1 Calculation of peak day and peak hour probabilities

3.1.1 Peak day model

There is a strong and nonlinear correspondence between power system load and weather (Sobhani et al., 2020). **Figure 3** plots the DEP daily peak load vs minimum and maximum temperature on each day in January and July from July 2015 to December 2020. The R-squared values between the daily minimum/maximum temperatures and the daily peak load for all months are listed in **Table 1**. Two key observations are highlighted as follows:

- There is a strong negative correlation (with $R^2 > 0.7$) between the daily minimum temperature and load in winter months (December to February), and a strong positive correlation between the daily maximum temperature and load in summer months (May to September).
- The correlation between daily peak load and daily minimum and maximum temperature is relatively weak ($R^2 \leq 0.5$) in spring and fall (March, April, May, October, and November), which suggests that temperature is a less effective predictor of peak load during these months.

Based on these observations, both direct and indirect models are developed and tested in this study. The selected predictors are the same for both models, including load and weather variables that are derived from the hourly data and a weekday/weekend indicator, as listed in **Table 2**. Because of the lack of historical weather forecast data, actual weather data are used for the ML model development.

- The direct model directly predicts the probability of the next day to be the peak day of the month. In this model, a binary response variable takes a value of 1 if a day is the monthly peak day and 0 otherwise. Ideally, the predictors should also include the load and weather forecast toward the end of the month to better capture how future load may affect whether the next day is the peak day. However, long-term load and weather forecasts are generally unavailable and therefore only forecasts of up to 7 days are used as predictors. As a result, the direct model actually links partial month predictors to the full month peak day indicator.

TABLE 2 Predictors for the peak day model.

Predictor	Description
load_max	Maximum load from the day-ahead forecast
T_min*	Minimum temperature from the day-ahead forecast
T_max*	Maximum temperature from the day-ahead forecast
weekdayIdx	0 or 1 indicating whether the next day is weekday or weekend
prev_month_max	Maximum actual load of the month to date
prev_MAX	Maximum actual load on previous day
T_min_day_2_to_7*	Minimum forecast temperature within the next 2–7 days
T_max_day_2_to_7*	Maximum forecast temperature within the next 2–7 days

- The indirect model outputs the probability of the next day to be the up-to-date peak day, which is defined as the peak day of a time window starting from the beginning of the month and ending at day 7 into the future or the end of the month.

Therefore, the model links predictors to the peak day indicator at the exactly matched time window. Then, the obtained up-to-date peak day probability (P_{date}) is converted to the monthly peak day probability (P_{month}) by applying a multiplier (P_{mul}) that reflects the chance of the monthly peak day occurring within the up-to-date time window, as expressed in **Eq. 2**:

$$P_{month} = P_{date} \, P_{mul}. \tag{2}$$

The multipliers corresponding to different time windows can be generated by examining the distribution of historical peak days. It is found that the chance is generally proportional to the length of the time windows in a given month. In other words, each day in a month has an equal probability to be the monthly peak day. Therefore, the multiplier can be defined as

$$P_{mul} = \min\left[1, \frac{n+6}{N}\right], \tag{3}$$

where N is the number of days of the current month and n is the day of the month for the next day. Unlike the direct model, which completely ignores the impacts of the load beyond 7 days, the indirect model takes future load into account in a stochastic manner.

3.1.2 Peak hour model

The day-ahead load forecast based peak hour prediction is subject to great uncertainties. Therefore, we introduce supplementary predictors in the ML peak hour model to help improve the prediction. The response variable is a binary variable that indicates whether an hour is the peak hour of the day. The predictors include both hourly load forecast and weather data for the operating day, as well as a weekday or weekend binary indicator, as listed in **Table 3**. We have the following considerations for including the major predictors:

- For each hour t, we use not only the temperature and humidity for the hour, as they directly impact the load components, e.g., heating and cooling, but also the temperature from $t − 3$ to $t + 3$ to explore the trailing or leading effects between temperature and load.
- Similarly, the forecast loads from hour $t − 3$ to $t + 3$ are included in the predictors because the peaks tend to cluster in groups.
- The rank of load forecast for hour t is included to distinguish the peak hour from other hours, particularly those with comparable loading levels.
- The maximums of load forecast before and after hour t are included to describe the relative position of the load in hour t with respect to highs in the past and future hours of the day. In comparison to load forecast of each hour beyond hours $t − 3$ to $t + 3$, the use of these two predictors helps reduce modeling complexity and avoid overfitting, while effectively capturing the overall impacts of the other hours on the operating day.
- In addition, as adjacent days tend to have more similar peaking behaviors, attributes from the previous day, including the rank of actual load for hour t, the forecast, and the actual load for hour t, are also used as predictors.

3.2 ML models

Given the high dimensionality of the predictors and the expected non-linear relationships with response variables, we adopt the ensemble tree-based ML method: RF and GBM, also to avoid overfitting and deal with unbalanced data of mixed types, e.g., categorical vs continuous. The models are developed in Python with main packages from the sklearn and xgb libraries.

RF is a tree-based ensemble learning method that can handle categorical variables, continuous variables, or a combination of both (Breiman, 2001). RF constructs a number of ensemble trees, with each tree trained by a randomly selected subset of input data using a bootstrap aggregating technique. At each node of a tree, instead of choosing the best split among all predictors, a

TABLE 3 Predictors for the peak hour model.

Predictor	Description
load_forecast	Forecast load for hour t
T*	Forecast temperature for hour t
humidity*	Forecast humidity at hour t
weekendIdx	A binary indicator: 1 if the operating day is on a weekend and 0 otherwise
peak_prev_day	A binary indicator: 1 if hour t is the peak hour on the previous day, and 0 otherwise
T_m_1*	Forecast temperature for hour $t − 1$
T_m_2*	Forecast temperature for hour $t − 2$
T_m_3*	Forecast temperature for hour $t − 3$
T_p_1*	Forecast temperature for hour $t + 1$
T_p_2*	Forecast temperature for hour $t + 2$
T_p_3*	Forecast temperature for hour $t + 3$
load_m_1	Forecast load for hour $t − 1$
load_m_2	Forecast load for hour $t − 2$
load_m_3	Forecast load for hour $t − 3$
load_p_1	Forecast load for hour $t + 1$
load_p_2	Forecast load for hour $t + 2$
load_p_3	Forecast load for hour $t + 3$
prev_max_load	Maximum forecast load from hour 0 to hour $t − 1$
after_max_load	Maximum forecast load from hour $t + 1$ to 23
rank_load_forecast	Rank of forecast load for hour t on the operation day
rank_load_prevDay	Rank of actual load for hour t in on the previous day
load_prevDay	Actual load for hour t on the previous day
load_prevDay_forecast	Forecast load for hour t on the previous day

randomly sampled subset of predictors is selected, and the best split is chosen from the subset predictors. Each tree can grow to the maximum possible depth. For classification problems, the final prediction is made by the majority votes from all the trees in the ensemble. RF has been successfully applied to high-dimensionality systems. Examples of its application in power systems include Lahouar and Slama (2015); Liu et al. (2021). RF models can be developed and optimized by refining key model configuration parameters such as the number of trees and tree depths, in addition to bootstrap re-sampling of subsets and multi-fold cross-validation. A general RF-based classification algorithm is summarized in **Algorithm 1**. In addition to estimating the probability of classification, an RF model also provides measures of relative feature importance of predictors.

GBM is another tree-based ensemble learning method (Friedman, 2001). Unlike RF, in which each tree is trained independently, GBM builds one tree at a time and the newly-built tree is added to previous trees to improve the overall prediction. It can be viewed as an iterative numerical optimization process with a goal of finding an additive model that minimizes the loss function. The new tree at each step is fitted to the residual of previous trees, and adding a new tree improves regions where previous trees did not perform well. While the result of RF is voted by all the trees after the tree-building process, the result of GBM is optimized throughout the tree-building process. Compared to RF, GBM can be more computationally intensive and more sensitive to noise in the training data set due to its iterative nature.

```
1: Randomly split the total dataset into
   training and testing (e.g., 75% and 25%)
2: for i = 1 to n_tree do
3:    Randomly select a subset n out of N
      samples
4:    Randomly select a subset of m predictors
      out of M factors (x₁,x₂,...,xₘ)
5:    Grow a random forest tree Ti and prune it
      to the optimal depth n_depth
6:    Output Ti tree forecast Cᵢ(x₁,x₂,...,xₘ),
      which is binary)
7: end for
8: Generate RF prediction by the majority vote
   of all tree forecasts Cᵢ, ∀i = 1,...,n_tree
```

Algorithm 1. A General RF-based Classification Algorithm.

LR is a classical statistical ML method that models the relationship between a set of predictors and a categorical response variable using a logistic function (sigmoid function) (Stoltzfus, 2011). The predictors of an LR model can be categorical and/or continuous variables, and the output is the probability of the response variable being one of the categorical outcomes. Because there are no "hyper-parameters" that can be used to tune an LR model, it is ideally used as a baseline model in predictive analyses.

3.3 Data augmentation

The raw load data are only available after July 2015, resulting in less than 80 monthly peak days. Separate models by month are highly desirable to better capture varying load patterns in different months. Therefore, data augmentation is needed to generate sufficient data for model training and reliable testing (Chawla et al., 2002). An ANN model for day-ahead load forecast has been developed using load and weather attributes (Berscheid et al., 2018). The model is developed using Matlab's Deep Learning Toolbox. The developed model has been cross-validated with historical load and weather data sets for major U.S. Balancing Authorities. In this study, using the DEP weather and load data sets from July 2015 to December 2020, the developed ANN model was trained to predict hourly load (including both actual load and day-ahead forecast) with the following predictors: hour, temperatures for the current hour and the past 3 h, humidity at the hour, day of the week for the predicting day, and weekday/weekend index. After a hyperparameter search, an optimized ANN model containing two layers and 20 neurons was obtained to capture load patterns and the relationship between load and weather embedded in the raw load data. Using the ANN model, we produced hourly actual load and day-ahead hourly
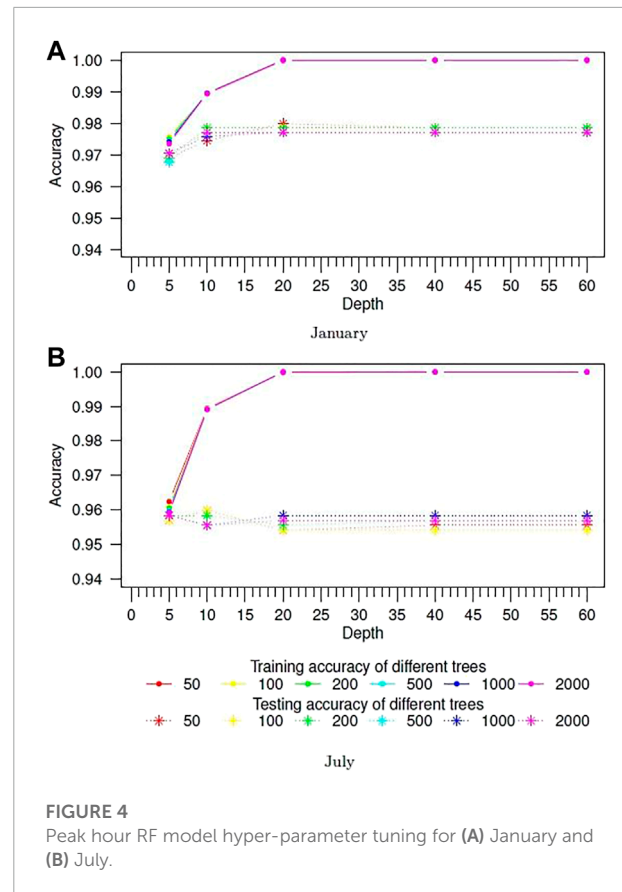

FIGURE 4
Peak hour RF model hyper-parameter tuning for **(A)** January and **(B)** July.

load forecast over a 15-year period with actual weather data from 2000 to 2015.

To validate the ensemble ML model performance, a subset of 6 years data is randomly selected for model testing: 2001, 2006, 2008, 2011, 2019, and 2020. The remaining 15-year data are used for training. Note that the actual weather data are used for the ML model development because of the lack of historical weather forecast data.

4 Results

4.1 ML predictive model development

4.1.1 Model training

Ensemble ML models, including LR, RF, and GBM, are developed for peak hour predictions, with individual models for each month. For LR models, we use Akaike Information Criteria (Akaike, 1974) for the model selection. For RF and GBM models in each month, we evaluate the testing accuracy with respect to two hyper-parameters: number of trees and tree depths. **Figure 4** shows the training and testing accuracy with respect to different tree depths and different numbers of trees for both January and July RF models. The selected number of depths are 5, 10,
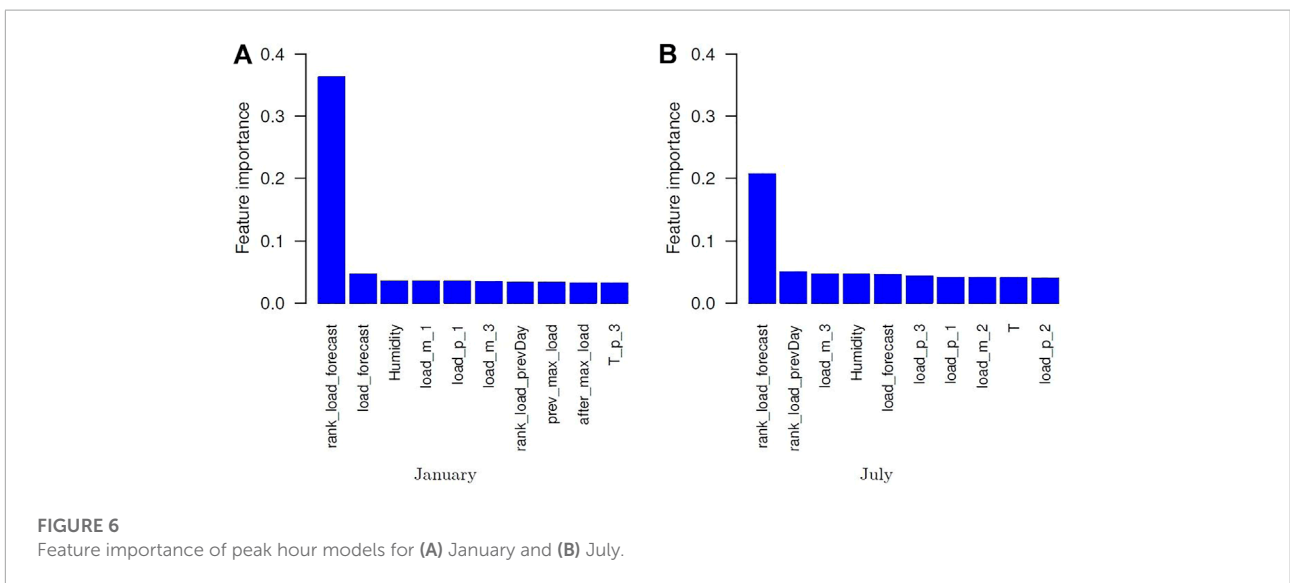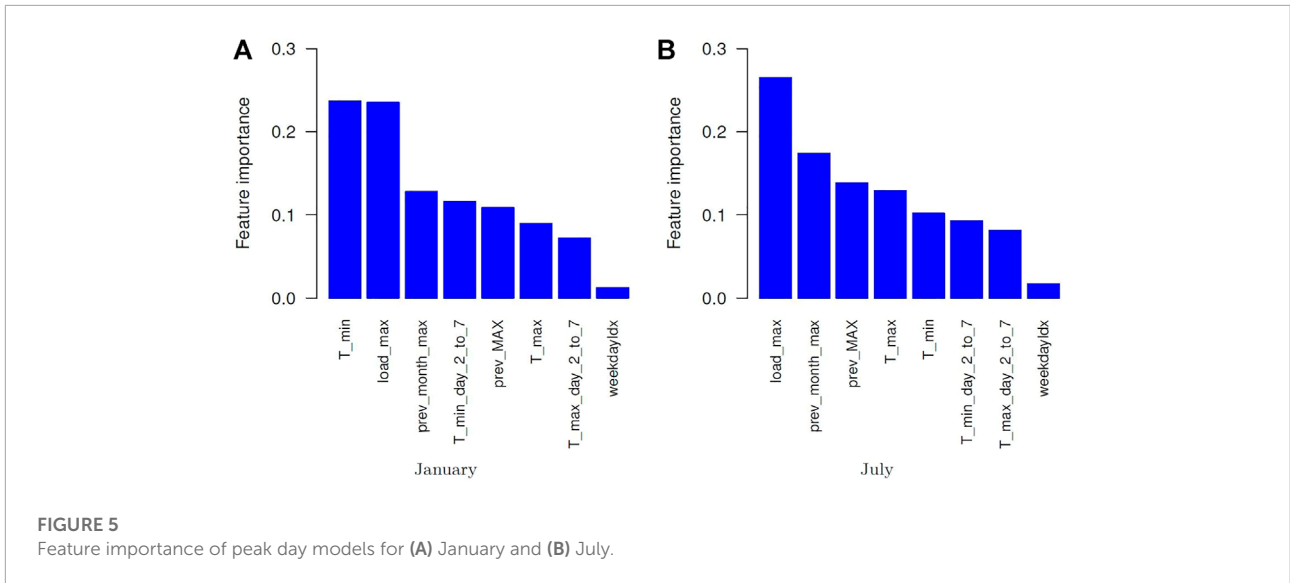
FIGURE 5
Feature importance of peak day models for **(A)** January and **(B)** July.

FIGURE 6
Feature importance of peak hour models for **(A)** January and **(B)** July.

TABLE 4 Comparison of BESS operation cycles and peak days captured between monthly peak day model and up-to-date peak day model.

Year	BESS operation cycles		Captured actual peak days	
	Direct model	Indirect model	Direct model	Indirect model
2001	64	78	12	12
2006	64	63	11	12
2008	64	72	11	12
2011	68	71	12	12
2019	77	87	12	10
2020	65	83	10	11
Average	67.2	75.7	11.3	11.5

20, 40, and 60, and the selected number of trees are: 50, 100, 200, 500, 1,000, and 2000. The results show that the testing accuracy converges when the tree depths are greater than 20 for both models. For the January model, the testing accuracy is comparable when the same number of depth is used regardless of the number of trees used. For the July model, 1,000 trees yield slightly higher accuracy than using other tree numbers when the tree depth is greater than 20.

4.1.2 Model comparison

For the peak hour models, each of the 3 ML ensembles is optimized with parameter search and cross-validation for reducing misfits and overfitting if possible. The performance of the optimized final models is compared in terms of the overall accuracy of capturing actual peak hours for each month in the 6 testing years. Overall, the RF models outperform both the LR and GBM models by 2%, In six monthly models—January, April, June, July, August, and September—the RF models perform better than the other two models by 5% and 3%, respectively. LR models perform slightly better in February, May, October, and December; while GBM models perform better in March and November. Because RF models perform the best overall, and in the focus months of January and July in particular, they are selected for further peak hour predictions and analyses. For the peak day prediction, RF models also perform better than or similar to GBM models, and the LR models do not converge due to their inability to handle imbalanced data sets. Thus, RF models are also selected for peak day prediction to be consistent with the model choice for peak hour prediction.

4.1.3 Feature importance

Feature importance (Saarela and Jauhiainen, 2021), which can be directly calculated from the RF model, measures the relative importance of each predictor in the fitted model. Figure 5 shows the ranked importance of the eight predictors in the January and July peak day models. For the January model, T_min (the minimum forecasted temperature) and load_max (the maximum forecasted load for the operation day) are the most significant predictors, with almost identical factors and feature importance values of 0.237 and 0.235, respectively. The other six variables are secondary. For the July model, load_max is more important than the other predictors. Load-related predictors (ranking from 1 to 3) are slightly more important than temperature-related predictors (ranking from 4 to 7).

For peak hour predictions, Figure 6 shows the top-ranked 10 of the 23 predictors with scaled importance in the January and July RF models. Rank_load_forecast (the rank of the load forecast of each hour) is the dominant predictor for both models, and the feature importance values are 0.362 and 0.207, respectively. Other import predictors include the load forecast for the January model and the rank load of the previous day for the July model.

TABLE 5 Comparison of predicted probabilities of actual peak days in 2020 between using the direct and indirect peak day models.

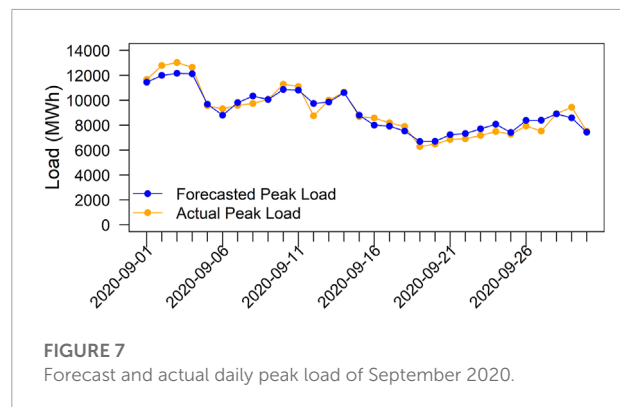| Actual Peak day	Maximum hourly Peak load (MWh)	Direct peak day model	Indirect peak day model
1/22	12,260	13.0%	25.7%
2/22	12,065	24.9%	27.1%
3/1	10,194	11.7%	4.5%
4/8	8,324	0.5%	3.8%
5/30	9,410	8.7%	37.5%
6/29	11,606	20.1%	28.0%
7/20	13,244	27.1%	33.7%
8/27	12,511	21.2%	30.3%
9/3	13,027	11.9%	2.9%
10/8	8,790	0.8%	8.6%
11/19	10,039	35.4%	66.8%
12/9	11,625	20.5%	16.4%

FIGURE 7
Forecast and actual daily peak load of September 2020.

To better understand the importance of predictors, one can refer to our principal component analysis and shrinkage discriminant analyses showing cross-dependence among predictors and response variables, as provided in **Supplementary Material**.

4.2 Verification of peak day and peak hour predictions

The above ML classification models yield parameter understanding and peak day/peak hour probabilities. In order to use such outputs for a typical BESS dispatch decision-making, one needs to consider that 1) the maximum number of charging/discharging cycles should be limited (e.g. less than 100 per year), and 2) the BESS can be charged/discharged up to 2 h per day, according to the recommendations from DEP. Therefore, thresholds are needed for identifying the peak days and peak hours. For example, peak day prediction can be set to exceed the probability threshold of 3% and peak hour prediction is the top 2 h with the highest peak probabilities in the operation day. The peak day exceedance probability threshold is initiated to be 3% with the assumption that each day of a month has an equal
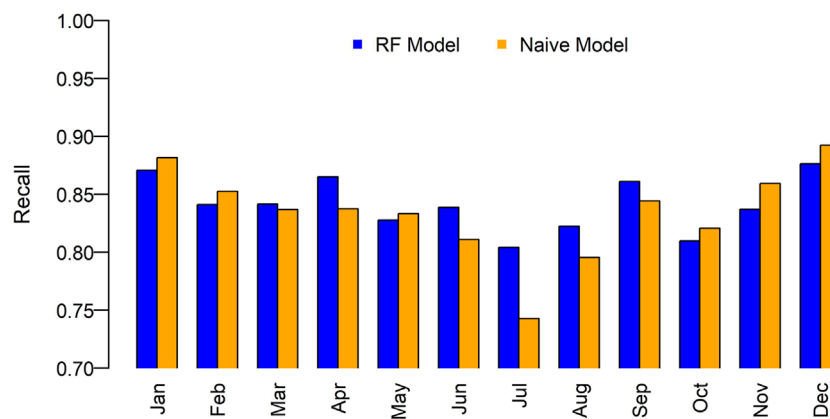
FIGURE 8
Comparison of prediction recall for each month between RF models and naive models.

TABLE 6 Annual performance of BESS for 6 testing years using the proposed peak day and peak hour models.

Year	BESS operation Cycles	Number of peak days captured	Number of peak hours captured
2001	78	12	10
2006	63	12	12
2008	72	12	12
2011	71	12	11
2019	87	10	9
2020	83	11	11
Average	76	11.5	10.8

chance being the peak day of the month. The optimal choice of the threshold can be selected and verified by comparing the identified charging/discharging cycles.

4.2.1 Peak day prediction

RF models are developed for each month using both direct and indirect peak day definitions. After the next operation day is assigned a peak probability by the RF model, a decision to charge/discharge a BESS on this day can be made by comparing the predicted peak day probability with a predefined threshold. The BESS would only be operated if the predicted peak day probability is above the threshold. A starting value of such a threshold is 3%, with the assumption that any day within a month could be a peak day. As shown in **Table 4**, both the direct and indirect peak day models perform well in terms of the number of operation cycles and peak days captured per year for the 6 testing years. The BESS would be operated for a similar number of cycles on average, 67 and 76, which are much fewer than the maximum 100 cycles/year requirement. The average numbers of captured peak days are 11.3 and 11.5 per year, for the directed and modified peak day models, respectively.

Table 5 shows the predicted peak day probabilities for each month in 2020 using the direct and indirect peak day models. With 3% as the threshold, the direct peak day models miss the peak days in April and October, while the indirect peak day models only miss the peak day in September. As shown in **Figure 1**, compared with the historical peak loads, the maximum peak loads on 8 April 2020 (8,324 MWh), and 8 October 2020 (8,790 MWh), are at the lower end of the historical peak day loads, which causes the direct models to assign a less than 1% probability to both days and miss them. In comparison, the indirect models can still capture these days with assigned probabilities of 3.8% and 8.6%, respectively. For the peak day on 3 September 2020, the predicted peak day probability is only 2.9% from the indirect peak day model, although its peak load of 13,027 MWh is higher than the median historical peak day loads. The reason for such a low-probability estimate is because the neighboring days have comparable high load and the forecast peak load on 3 September 2020 (12,251 MWh), is lower than the actual peak load on 2 September 2020 (12,785 MWh), as shown in **Figure 7**, which shows the forecast and actual daily peak loads in September 2020.

4.2.2 Peak hour prediction

Practically, the BESS can be discharged for 2 h during a peak day; therefore, the 2 h with the highest probabilities from the peak hour model prediction in each day are selected for discharging. The peak hour prediction performance is verified in two ways: 1) comparing predictions from RF models with the naive model predictions using day-ahead load forecast; and 2) evaluating the number of captured peak hours following the proposed peak day identification (e.g., 3% exceedance probability) and peak hour prediction procedure.

Figure 8 shows the percentage of peak hours captured (true positives) relative to the total number of true positives and

false negatives, $TP/(TP + FN)$, in each month for the 6 testing years. The percentage metric, i.e., recall, is a good measure of classification model performance for imbalanced data like in this study. For both RF and naive models, the recall values are highest in December and January and lowest in July and August. The RF model has better overall performance than the naive model: of the 72 peak hours in the testing data, the RF model captures 67 (93.1%) while the naive model captures 63 (87.5%). In particular, the recall of the RF model for July is much higher than in the naive model: 80.4% vs. 74.3%.

Further evaluation of the entire prediction framework performance is based on the number of peak days and peak hours captured for all 6 testing years, using a 3% exceedance probability threshold for peak day prediction and a 2-h battery discharging time. The results are summarized in Table 6. The average number of peak hours captured is 10.8 per year for the 6 testing years. All 12 monthly peak hours in 2006 and 2008 are captured. For testing years 2011 and 2020, 11 of 12 peak hours are captured and the monthly peak hours in May 2011 and September 2020 are missed. In 2001, peak hours in May and August are missed. In 2019, peak hours in February, May, and September are missed. The number of BESS charging/discharging cycles is 76 per year on average, ranging from 63 to 87, which meets the 100 cycles or less per year requirement.

5 Conclusion

In this paper, we presented an advanced ensemble ML framework for predicting peak days and peak hours to better support BESS dispatch decision-making. The proposed approach features the probabilistic definition of peak day and peak hour, a comprehensive set of physical and temporal factors and predictors, nonlinear ensemble ML model implementation and selection, and data augmentation. With cross-validation and model comparison, the proposed ML framework has been proven to work effectively. The study also provided guidance on the model choices and favorable conditions for applying the proposed approach. The study generated an ML-enabled dataset including cleaned historical data and derived attributes as exploratory and response variables.

The ML models have been trained and validated during various time periods with different system behaviors, but one could expect the correspondence between temperature and peak consumption to be region-specific, although in general there is a positive correlation between temperature and demand in summer and a negative correlation between them in winter; therefore, we recommend the users to adopt our ML setup and prediction framework for any different region, but the

model parameters should be retrained and optimized. In locations where weather attributes (e.g., temperature) have little correlation with peak demand, e.g., in regions with oceanic or Mediterranean climate, the ML prediction performance might be weak. One area of future work is to develop methods to explicitly integrate mid-to long-term factors into forecast models. Another interesting research direction is to fully integrate state-of-the-art ML techniques such as ANN and LSTM to improve the efficiency and accuracy of peak predictions.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

TF and HZ conducted the data preparation and analytics, XM performed numerical simulations, ZH designed the framework, DW supervised the study, all co-authors contributed to the technical writing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fenrg.2022.944804/full#supplementary-material

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.*, 19 (6), 716–723. doi:10.1109/TAC.1974.1100705

Berrisch, J., Narajewski, M., and Ziel, F. (2022). High-resolution peak demand estimation using generalized additive models and deep neural networks. *arXiv* [Preprint]. Available at: https://arxiv.org/abs/2203.03342.

Berscheid, A., Makarov, Y., Hou, Z., Diao, R., Zhang, Y., Samaan, N., et al. (2018). "An open-source tool for automated power grid stress level prediction at balancing authorities," in 2018 IEEE/PES transmission and distribution conference and exposition (T&D) (IEEE).

Bosen, J. F. (1958). An approximation formula to compute relative humidity from dry bulb and dew point temperatures. *Mon. Weather Rev.* 86, 486. doi:10.1175/1520-0493(1958)086<0486:aaftcr>2.0.co;2

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi:10.1023/a:1010933404324

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi:10.1613/jair.953

Cheng, Y.-Y., Chan, P. P., and Qiu, Z.-W. (2012). Random forest based ensemble system for short term load forecasting. Proceedings of the IEEE International Conference on Machine Learning and Cybernetics 1, 52.

Cieslak, D. A., and Chawla, N. V. (2008). "Learning decision trees for unbalanced data," in *Joint European conference on machine learning and knowledge discovery in databases* (Antwerp, Belgium: Springer), 241–256.

Dai, S., Meng, F., Dai, H., Wang, Q., and Chen, X. (2021). Electrical peak demand forecasting-A review. *arXiv* [Preprint]. Available at: https://arxiv.org/abs/2108.01393.

Domingos, P. (2012). A few useful things to know about machine learning. *Commun. ACM* 55, 78–87. doi:10.1145/2347736.2347755

EIA (2021). U.S. electric system operating data. Available: https://www.eia.gov/opendata/qb.php?category=2123635 (Accessed 1025, 2021).

El Desouky, A., and Elkateb, M. (2000). Hybrid adaptive techniques for electric-load forecast using ann and arima. *IEE Proc. Gener. Transm. Distrib.* 147, 213–217. doi:10.1049/ip-gtd:20000521

El-Attar, E., Goulermas, J., and Wu, Q. (2009). "Forecasting electric daily peak load based on local prediction," in 2009 IEEE Power & Energy Society General Meeting (Calgary, Canada: IEEE), 1–6.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statistics*, 1189–1232.

Goodwin, M., and Yazidi, A. (2016). A pattern recognition approach for peak prediction of electrical consumption. *Integr. Comput. Aided. Eng.* 23, 101–113. doi:10.3233/ica-160510

Hong, T., and Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* 32, 914–938. doi:10.1016/j.ijforecast.2015.11.011

Hong, T., and Wang, P. (2014). Fuzzy interaction regression for short term load forecasting. *Fuzzy Optim. Decis. Mak.* 13, 91–103. doi:10.1007/s10700-013-9166-9

Huang, N., Lu, G., and Xu, D. (2016). A permutation importance-based feature selection method for short-term electricity load forecasting using random forest. *Energies* 9, 767. doi:10.3390/en9100767

Jaiswal, J. K., and Samikannu, R. (2017). "Application of random forest algorithm on feature subset selection and classification and regression," in Proceedings of the IEEE World Congress on Computing and Communication Technologies (WCCCT).

Jiang, Y. H., Levman, R., Golab, L., and Nathwani, J. (2014). "Predicting peak-demand days in the Ontario peak reduction program for large consumers," in Proceedings of the 5th International Conference on Future Energy Systems.

Lahouar, A., and Slama, J. B. H. (2015). Day-ahead load forecast using random forest and expert input selection. *Energy Convers. Manag.* 103, 1040–1051. doi:10.1016/j.enconman.2015.07.041

Liu, F., Dong, T., Hou, T., and Liu, Y. (2021). A hybrid short-term load forecasting model based on improved fuzzy c-means clustering, random forest and deep neural networks. *IEEE Access* 9, 59754–59765. doi:10.1109/access.2021.3063123

Liu, J., and Brown, L. E. (2019a). "Effect of forecast accuracy on day ahead prediction of coincident peak days," in Proceedings of IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia), 661.

Liu, J., and Brown, L. E. (2019b). "Prediction of hour of coincident daily peak load," in 2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT) (Washington, DC: IEEE), 1–5.

Mao, H., Zeng, X.-J., Leng, G., Zhai, Y.-J., and Keane, J. A. (2009). Short-term and midterm load forecasting using a bilevel optimization model. *IEEE Trans. Power Syst.* 24, 1080–1090. doi:10.1109/tpwrs.2009.2016609

Massaoudi, M., Refaat, S. S., Chihi, I., Trabelsi, M., Oueslati, F. S., and Abu-Rub, H. (2021). A novel stacked generalization ensemble-based hybrid lgbm-xgb-mlp model for short-term load forecasting. *Energy* 214, 118874. doi:10.1016/j.energy.2020.118874

Niu, D., Wang, Y., and Wu, D. D. (2010). Power load forecasting using support vector machine and ant colony optimization. *Expert Syst. Appl.* 37, 2531–2539. doi:10.1016/j.eswa.2009.08.019

NOAA, (2021). Local climatological data station details. Available: https://www.ncdc.noaa.gov/cdo-web/datasets/LCD/stations/WBAN:13722/detail (Accessed 1025, 2021).

Ringwood, J. V., Bofelli, D., and Murray, F. T. (2001). Forecasting electricity demand on short, medium and long time scales using neural networks. *J. Intelligent Robotic Syst.* 31, 129–147. doi:10.1023/a:1012046824237

Ryu, B., Makanju, T., Lasek, A., An, X., and Cercone, N. (2016). "A Naive Bayesian classification model for determining peak energy demand in Ontario," in *Smart city 360°* (Toronto, Canada: Springer), 517–529.

Saarela, M., and Jauhiainen, S. (2021). Comparison of feature importance measures as explanations for classification models. *SN Appl. Sci.* 3, 272–312. doi:10.1007/s42452-021-04148-9

Saini, L. M. (2008). Peak load forecasting using Bayesian regularization, resilient and adaptive backpropagation learning based artificial neural networks. *Electr. Power Syst. Res.* 78, 1302–1310. doi:10.1016/j.epsr.2007.11.003

Saxena, H., Aponte, O., and McConky, K. T. (2019). A hybrid machine learning model for forecasting a billing period's peak electric load days. *Int. J. Forecast.* 35, 1288–1303. doi:10.1016/j.ijforecast.2019.03.025

Sobhani, M., Hong, T., and Martin, C. (2020). Temperature anomaly detection for electric load forecasting. *Int. J. Forecast.* 36, 324–333. doi:10.1016/j.ijforecast.2019.04.022

Stoltzfus, J. C. (2011). Logistic regression: A brief primer. *Acad. Emerg. Med.* 18 (10), 1099–1104. doi:10.1111/j.1553-2712.2011.01185.x