Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
General short-term load forecasting based on multi-task temporal convolutional network in COVID-19

Zhenhao Zhang, b, Jiefeng Liu, a, *, Senshen Pang, a, Mingchen Shi, c, Hui Hwang Goh, a, Yiyi Zhang, a, Dongdong Zhang" 

a Guangxi Key Laboratory of Power System Optimization and Energy Technology, Guangxi University, Nanning, Guangxi 530004, China
b State Grid Chongqing Electric Power Company, Chongqing, Chongqing 401121, China
c State Grid Lishui Liandu Power Supply Company, Lishui, Zhejiang 323000, China

ARTICLE INFO
Keywords:
New coronavirus pandemic
Short-term load forecasting
Multi-task learning
Temporal convolutional network
Population mobility

ABSTRACT
The spread of the global COVID-19 epidemic has resulted in significant shifts in electricity consumption compared to regular days. It is unknown if standard single-task, single-indicator load forecasting algorithms can accurately reflect COVID-19 load patterns. Power practitioners urgently want a simple, efficient, and accurate solution for anticipating reliable load. In this paper, we first propose a unique collaborative TCN-LSTM-MTL short-term load forecasting model based on mobility data, temporal convolutional networks, and multi-task learning. The addition of the parameter sharing layers and the structure with residual convolution improves the data input diversity of the forecasting model and enables the model to obtain a wider time series receptive field. Then, to demonstrate the usefulness of the mobility optimized TCN-LSTM-MTL, tests were conducted in three levels and twelve base regions using 19 different benchmark models. It is capable of controlling predicting mistakes to within 1 % in the majority of tasks. Finally, to rigorously explain the model, the Shapley additive explanations (SHAP) visual model interpretation technology based on game theory is introduced. It examines the TCN-LSTM-MTL model’s internal mechanism at various time periods and establishes the validity of the mobility indicators as well as the asynchronous relationship between indicator significance and real contribution.

1. Introduction
The International Energy Agency (IEA) [1] reports that the new crown pneumonia (COVID-19) epidemic has damaged practically every industry in the world, including the power grid [1]. As the virus evolved, important strains developed one after the other, directly impacting the load patterns of practically all countries [2]. To halt the spread of the illness, numerous countries and areas have implemented lax or stringent policies and long-term measures, hence invalidating the similar day assumption [3]. Additionally, the COVID-19’s effect on power load is distinct from that of extreme weather [4] and geological disasters, indicating that forecasting methods based purely on weather indications will fail repeatedly. As a result, the efficacy of conventional, even optimized, procedures in the face of such a public crisis is unclear and dubious. The uncertainty and economy in energy scheduling [5] have always been the hotspot of the current research, and is still worth studying during the COVID-19 period.

According to the duration of the prediction, power system load forecasting can be classified as ultra-short-term [6], short-term [7], medium-term [8], and long-term [9]. Among these, due to its critical significance in power system operation planning, short-term load forecasting has currently emerged as a contemporary research hotspot. Since computers’ computing capacity has increased, neural networks (NNs) have gained considerable attention due to their superior adaptation to non-linear features. Support vector machines (SVM) and their derivatives have been frequently utilized to create GWO-SVM [10], PSO-SVM [11], and other non-sequential networks. There are also more non-linear methods that are extensively utilized, including extreme gradient boosting (XGBoost) [12], random forest (RF) [13], and other. However, none of the aforementioned models address the issue of depending solely on one indicator (historical load) or multiple conventional indicators (historical load, weather, holidays, etc.), which has significant limits.

The field of temporal neural networks has achieved numerous ground-breaking discoveries in the last few years. LSTM and GRU models are RNN network variations that perform exceptionally well in...
long-term load forecasting applications. They were combined with other optimization techniques and weather variables to create load models such as VMD-LSTM [14], CNN-LSTM [15], and CNN-GRU [16]. However, LSTM performs poorly and lacks sufficient robustness when extremely long sequences are supplied [17]. The introduction of the attention mechanism [18] suggests a way forward. The Transformer [19], Informer [17], LSTM-Attention [20] models are well suited to predicting applications involving longer time periods. However, certain models (for example, the Transformer model) have extremely complicated structures and a high computational complexity, which increases the likelihood of memory overflow. To address this issue, a number of versions based on temporal convolutional networks (TCNs) have garnered considerable attention in recent years. It is innocent and deceptively light. It has been combined with a light gradient boosting machine (LightGBM) and a convolutional neural network to create TCN-LightGBM [21], and CNN-TCN [22]. They have all improved significantly in recent years in terms of accuracy and computing performance.

Recently, transfer learning and multi-task learning (MTL) have recently been favored by scholars in the sector of electric power [20] for natural language processing (NLP) and image identification. Transfer learning requires careful consideration of both the timing of transfer learning [23] and the indicators to be employed for transfer learning [24]. Negative transfer is more likely to occur when indicators have a low association with the target tasks [25]. However, weakly connected indications can be deleted prior to MTL trainings using correlation analysis. It has been employed by certain scholars in collaborative forecasting of electric, heating, cooling, and air demands, as well as in tasks related to holidays [26] and communities [27]. They have all produced better results than single-task learning. Regrettably, whether they can contribute to the dramatic changes in the load pattern generated by COVID-19 remains an open question. In the past six months, methods based on transfer learning and reinforcement learning have been proposed to fight the uncertainty of COVID-19, mainly discussing the small sample problem [28] and hyperparameter optimization problem [29]. There are also literatures focusing on the modeling of load forecasting problems in specific regions (e.g., Ontario [30]). However, the models proposed in the above literature neglected to learn the multi-regional universality of load forecasting, and lacked generality.

The COVID-19 has had a huge impact on the validity and robustness of the power load forecasting model. This topic is the main stream of current scholars’ research. The purpose of this study is to propose practical solutions to issues such as reliance on a single indicator, insufficient use of related data, model inflexibility, and poor interpretability, while also taking into account the threat posed by COVID-19. Specifically, we have made the following innovations and contributions:

• To solve the problem of robustness of long-term input in training and forecasting, and make full use of valuable information in other parallel tasks, the TCN-LSTM-MLT collaborative forecasting model based on TCN and multi-task learning is proposed. The parameter sharing layers are introduced to improve the anti-overfitting ability, accuracy, and robustness under a long time series.

• To reduce the difficulty in capturing power load patterns, this paper innovatively introduces the population mobility data released by Apple [31] and Google [32] in COVID-19.

• To fully prove the model’s versatility in different periods and regions in COVID-19, this paper uses actual public data from 2018 to 2021 to conduct comparative experiments in 12 regions or countries in the world.

• To solve the interpretability problem of the load forecasting model, The Shapley additive explanation (SHAP) [33,34] with a solid game theory basis is innovatively used to visualize the influence of each indicator on the power load at each point. Besides, this paper provides an in-depth analysis of the relationship between the actual contribution and the correlation of the indicators by using SHAP.

2. Load trends analysis

Travel restrictions and other containment measures had a substantial impact on the economy, social output, and living during COVID-19, resulting in an entirely distinct load shift trend. From January 1st, 2019 through August 15th, 2021, Fig. 1 depicts the load-changing pattern in four major cities in the United States (New York, Columbus, Philadelphia, and Washington, D.C.) and four European nations (Switzerland, France, Italy, and Germany).

In the first row of Fig. 1, the graphs for 2019 and 2020 for four cities in the United States are compared using the same numerical coloring standard. The four cities’ spring 2019 electricity load figures are much higher than those for 2020. Among these, the power loads of New York City and Columbus exhibit notable red surges in 2019, but nearly vanish in 2020. The similar adjustments have occurred in Philadelphia’s power load, and the peak change in the power load has grown smoother. They all accurately depict COVID-19’s major impact on load change patterns in various cities. By 2021, almost all cities had recovered their electricity loads relative to 2020, but they were still lower than in 2019, which is consistent with the social fact that the gradual resumption of work and production in various locations has occurred in COVID-19 through normalized prevention and control. However, over these three years, all cities experienced summer load surges, indicating that the impact of COVID-19 may not be as significant as the summer temperature.

Apart from the fact that the United States would face identical conditions during the COVID-19 outbreak and pandemic in the spring of 2020, European countries (the second row of Fig. 1) will have some minor differences. To begin, the increases in power load values in the first few months of 2021 were clearly longer than those in 2019, which might be attributed to the severe ice and snow weather in early 2021. Additionally, the spring, summer, and fall electricity load values for European countries in 2021 are much lower than in 2019, and the length is longer than for the four cities. This also demonstrates the extent to which COVID-19 has impacted European countries.

As a result, COVID-19’s impact on numerous countries and cities is long-lasting, far-reaching, and difficult to the load. Additionally, there are differences in the patterns of power load change at various scales and levels of power systems that must be considered in various categories.

3. Methodology proposal

3.1. Population mobility

Population mobility is a term that refers to the varied short-term and periodic migrations of the population within regions that can be derived from a variety of sources. Large Internet businesses (such as Google, Apple, and Huawei, among others) can collect the geographic location data from customers’ smart devices and then clean and cluster it.

Population mobility is influenced by a variety of factors, including: time, space, policy, and weather, all of which affect the load. Therefore, we believe that incorporating population mobility into the load forecasting process has the potential to increase the accuracy of load forecasting. The generalized expression of incorporating mobility data can be expressed by (1):

$$FL = f(t, w, m, hl, o)$$

(1)
where the descriptions of the features are as follows:

- $f(\cdot)$ represents the non-linear mapping relationship between the forecasting values and the input indicator data.
- $FL$ is the forecasting value, and its length is determined by different types of load forecasting tasks.
- $t \in [0, 23]$ is the 24 hours of a day.
- $w$ represents a series of weather indicators, which can be indicators that directly affect power consumption, such as temperature, precipitation, and light intensity, etc.
- $m$ represents a series of indicators of population mobility, which can be the traffic flow of a region or country, the flow of people in a specific area, etc.
- $hl$ represents historical load data. The length can be determined according to actual needs.
- $o$ represents other indicators, mainly indicators that may help improve forecasting accuracy. They can be quantified or non-quantified, such as policies, user habits, etc.

### 3.2. Temporal convolutional network (TCN)

Convolutional neural networks (CNNs) have been widely employed throughout the world. However, due to the convolution kernel’s size constraint, it is typically regarded as unsuitable for simulating timing problems. This perspective has shifted with the advent of the temporal convolutional network (TCN). TCN is a CNN network version created specifically for timing problems, utilizing the advantages and structure of the CNN network ResNet. The TCN’s basic structure is depicted in Fig. 2.

TCN contains a number of essential parameters that need to be defined. The convolution kernel size $K_{size}$ represents the number of units from the previous layer that are used to perform convolution on a single unit from the following layer. A larger kernel size results in an increased receptive field $R_{field}$. The padding mode corresponds to the mode of convolution. Generally, we utilize causal convolution and set the padding mode to “casual” in this case. $N_{stack}$ represents the number of stacks of residual blocks to use. In contrast to the typical CNN structure, TCN employs dilated convolution, which enables the convolutional input to be an interval. Dilated convolution is mostly used in the time dimension in TCN, as shown in subplot A of Fig. 2. The sampling rate (hole coefficient) $d_i$ control the size of the interval. In TCN, dilated convolution is mainly performed in the time dimension. For instance, in the subplot A of Fig. 2, each number is sampled, implying a sampling rate of one. The intermediate layer’s “Dilations = 2” specifies that each pair of points in the input is treated as a single input sample. By adjusting the aforementioned parameters, we can calculate the $R_{field}$ as (2):

$$R_{field} = 1 + (K_{size} - 1) \times N_{stack} \times \sum d_i$$

In general, causal convolution, dilated convolution, and residual structure are the three fundamental components of TCN. When the network depth is constant, TCN has a larger receptive field than conventional causal convolution and is capable of adjusting the size of the receptive field flexibly. As a result, the TCN network is well-suited for the development of a dependable sequential load forecasting model.
3.3. Multi-task learning

Power systems are distributed, multi-regional, and multi-node systems. The capacity for power generation and consumption in a particular region has an indirect effect on the capacity for power generation and consumption in other related regions; the same is true for weather and population migrations.

We address this issue in this work by utilizing multi-task learning (MTL), as shown in Fig. 3. Multi-task learning is a form of collaborative learning that has been widely used in a variety of industries. We synthesis the forecasting tasks of many surrounding locations for concurrent training during the load forecasting process. When a load forecasting model is trained in a single region, the internal parameters of the model are modified not only by the training of the region’s own data, but also by information transferred from other regions’ training tasks, as in (3):

$$FL_i = M(t, d, type, FO_i, FT_i) \quad i \in \{1, 2, ..., n\}$$

where $FL_i$, $FO_i$, $FT_i$ represent the $i$th forecasting task, the own knowledge of the $i$th task, and the knowledge transferred from other related tasks, respectively. $n$ is the number of forecasting tasks that need to be trained.

3.4. Mobility-optimized TCN-LSTM-MTL model

We examine two distinct sorts of tasks in the COVID-19 in this paper. The first is called single-task learning. The target area’s data is input through numerous hidden layers and then produces the predicting results directly without any additional information input. The other form is multi-task learning, in which we divide our time between tasks related to the goal area and tasks related to knowledge transfer. Fig. 4 illustrate the structure of the mobility-optimized TCN-LSTM-MTL to perform short-term load forecasting. We artificially separate each region’s load forecasting tasks into target and auxiliary tasks, and the region’s enormous volumes of data are normalized and sent through numerous dense layers to lower the number of neural nodes. Then, all data is combined and merged using the merge layer. After this procedure is complete, several sharing layers made of multiple TCN-LSTM structures assist in parameter sharing utilizing knowledge gained from previous jobs and subsequently in data separation via the data separation layer in order to continue the following process. Finally, data sorting and reduction in dimension are carried out.

It is worth mentioning that each task in multi-task learning is
completely parallel, with no target tasks and auxiliary activities such as transfer learning. Readers can artificially establish target tasks based on their preferences.

3.5. Shapley additive explanations for load forecasting

The load forecasting model, like many other neural network models, is a black-box model. It is critical to evaluate models in order to determine the model’s validity and to investigate the method by which the features exert their influence.

In this research, we adopt a visual way to explain the real influence of various load-related indicators on the output of the TCN-LSTM-MTL model using Shapley additive explanations (SHAP) [34]. This process can be streamlined, as illustrated in Fig. 5. Using Shapley values, the SHAP technology ties efficient credit allocation to local explanations.

The SHAP method’s core parameter is the Shapley value. Because multiple factors affect the power load, if each power-related feature is abstracted as a team member and the load forecasting results are taken as the total revenue, the Shapley values can be used to quantify the impact of each team member on the total revenue (increasing or decreasing the result). For a sample \( x_i = \{ x_{1i}, x_{2i}, \ldots, x_{ji}, \ldots, x_{ni} \} \) of historical data \( X \) with \( n \) samples, the Shapley value \( \phi_j \) of the \( j \)-th factor \( x_j \) can be defined as (4):

\[
\phi_j = \sum_{S \subseteq \{x_i\} \setminus \{x_j\}} \frac{|S|! (n - |S| - 1)!}{n!} \left( \text{val}(S \cup \{x_j\}) - \text{val}(S) \right)
\]

Fig. 4. The schematic diagram of the short-term load forecasting process using mobility optimized TCN-LSTM-MTL model.

Fig. 5. The schematic diagram of the SHAP technology.
where $S$ is the subset of the remaining features after the value of the $j$-th feature $x_j$ in $x_i$ is discharged; $|S|$ is the number of features of subset $S$; $val(S)$ is a characteristic function, which represents the degree of influence of the features in $S$ on the load forecasting results through “cooperation” process, and its value can be calculated through the output values of load forecasting model $\hat{F}$ using the features which are not included in the subset $S$, as shown in (5).

$$val(S) = \int \hat{F}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \, dx - \frac{1}{n} \sum_{i=1}^{n} \hat{F}(x_i) \quad x \notin S$$  \hspace{1cm} (5)$$

(5) requires performing an integral summation operation on each load-related feature in the sample $x$ that is not included in $S$. As a result, we can quickly get the equation for SHAP from (6):

$$y_t = y_{\text{base}} + f(x_1^i) + f(x_2^i) + \ldots + f(x_n^i)$$  \hspace{1cm} (6)$$

where $y_{\text{base}}$ represents the mean value of the predicted load values, $f(x_i^i)$ represents the Shapley value (contribution value) of the first load-related feature in the $i$-th sample to the load forecasting result $y_t$. If $f(x_i^i) > 0$, it means that the first feature positively affects the predicted load result $y_t$ and increases $y_t$ from $y_{\text{base}}$; otherwise, it has a negative effect and makes it decrease.

After training the TCN-LSTM-MTL model, we will use the SHAP in order to gain a better understanding of the direction and impact of each load-related indicator.

4. **Short-term load forecasting case studies**

4.1. **The test’s fundamental environment**

(1) **Operation Platform**

This paper’s scripts are hosted on Google Colab. The Google Colab platform runs Chromium OS 12 on NVIDIA P100 GPUs, Intel Xeon processors, and the TensorFlow 2.7 deep learning framework.

(2) **Data Sources**

Table 1 lists the sources of historical load data, population mobility data, and meteorological data used in our testing. The International Unit System is used to express all fine-grained load data.

Historical load statistics from nation-level sources (Switzerland, France, Germany, and Italy), ISO-level sources (APS, CAISO, NEVP, and PACE), and metropolitan-level sources (New York, Columbus, Philadelphia, and Washington D.C.) are used. Meanwhile, because the load forecasting model is a time series model, we downgraded and flattened the time to 32 fine-grained dimensions for representation, with 24 dimensions representing the 24 hours ($H_1$, $H_2$, ..., $H_{24}$) of a day, 7 dimensions representing the 7 days (Sun, Mon, ..., Sat) of a week, and one dimension representing holidays.

(3) **Benchmark Model Settings**

To enable a more accurate comparison of our model’s effects with those of other traditional models during COVID-19, we established a range of benchmark testing methodologies, as detailed in Table 2.

Five dashes divide the benchmark model into six categories. The first, second, and third categories are established and mature models of single-task learning based on RNN, LSTM, and TCN. The fourth class of models is those that employ multi-task learning on the pure COVID-19 datasets, such as LSTM-MTL and our proposed TCN-LSTM-MTL models. The fifth kind incorporates a model based on the mechanism of attention. The final category includes all other conventional non-sequential models, such as XGBoost, RFR, and so on. T, W, L, and M denote fine-grained time (through flattening), weather, and historical load, respectively, for feature types. Additionally, to account for the effect of differing time periods on the load projection findings, we designated the period from January 1st, 2018 to December 31st, 2019 as the “old” data and the period from February 15th, 2020 to August 15th as the “new” data.

Additional parameters used in our tests are included in Table 3. Each dataset is segmented into training, validation, and test sets. The forecasting horizon for the “new” dataset is from 0 a.m. on August 9th to 11 p.m. on August 15th, 2020. There are 168 load values to anticipate; the forecasting horizon for the “old” dataset is from 0 a.m. on December 25th to 11 p.m. on December 31st of 2019, which is also 168 steps.

---

**Table 2**

| Model name        | Task type | Feature | Data type |
|-------------------|-----------|---------|-----------|
| NN-Old            | Single    | T, W, L | Old       |
| NN-Retrain-New    | Single    | T, W, L | New       |
| NN-Mob-Single     | Single    | T, W, L, M | New    |
| LSTM-Old          | Single    | T, W, L | Old       |
| LSTM-Retrain-New  | Single    | T, W, L | New       |
| LSTM-Mob-Single   | Single    | T, W, L, M | New    |
| TCN-Old           | Single    | T, W, L | Old       |
| TCN-Retrain-New   | Single    | T, W, L | New       |
| TCN-Mob-Single    | Single    | T, W, L, M | New    |
| TCN-LSTM-MTL      | Multiple  | T, W, L, M | New    |
| LSTM-Attention    | Single    | T, W, L, M | New    |
| Other non-sequential models | Single | T, W, L, M | New     |

**Table 3**

| Content                | Setting               |
|------------------------|-----------------------|
| The time span of datasets | 730 days (old), 183 days (new) |
| Training set           | 716 days (old), 169 days (new) |
| Validation set         | 7 days after the training set |
| Test set               | The last 7 days of datasets |
| Optimizer              | Adam                  |
| Learning rate          | 0.001                 |
| Loss function          | Mean absolute error (MAE) |
| Data normalization method | maximum and minimum normalization |
| Correlation check method | Spearmann’s rank correlation coefficient |
| Results evaluation method | mean absolute percentage error (MAPE) and Theil inequality coefficient (TIC) |
4.2. Results analysis

(1) The Correlation of Mobility Data

Table 4 shows the Spearman correlation coefficients $S$ between electrical load and mobility in New York City during COVID-19. As a point of comparison, we use the correlation coefficient value for the temperature indication.

Five of the nine mobility indicators have a higher correlation than the temperature indicator. The driving indicator has a value of 0.846, which is 36.45% more than the temperature indicator. Additionally, we observe that the residential indicator’s correlation values are negative, indicating that the behavior of remaining at home is negatively correlated with the power load, which corresponds to the real scenario during the COVID-19 era.

As a result, the population mobility indicators may have values that correspond to variations in the power load during COVID-19.

(2) Calculation Time Analysis

We will conduct performance tests on several benchmark models in this section during COVID-19. The computation time for single-task learning is shown in Table 5.

The calculation time for the LSTM-Mob-Single model based on the LSTM model in four cities ranges between 49.1 s and 54.1 s, while the time for the TCN-LSTM-Mob-Single model based on the TCN-LSTM model in four cities ranges between 68.5 s and 70.9 s, indicating that the difference in model training time between regions is negligible when the dataset is the same type and size and the model structure is the However, the model with the TCN structure requires an additional 29.8

### Table 4
The Spearman Correlation Coefficient Results.

|                    | Temperature | Driving | Transit | Walking | Retail and recreation |
|--------------------|-------------|---------|---------|---------|-----------------------|
| $S$                | 0.620       | 0.846   | 0.724   | 0.691   | 0.684                 |
| Grocery and pharmacy | 0.530       | 0.594   | 0.638   | –0.539  | 0.461                 |

### Table 5
Training Time Of Single Task Learning.

|                     | New York | Columbus | Washington DC | Philadelphia |
|---------------------|----------|----------|---------------|--------------|
| LSTM-Mob-Single     | 52.2 s   | 54.1 s   | 55.2 s        | 49.1 s       |
| TCN-LSTM-Mob-Single | 68.5 s   | 70.2 s   | 72.8 s        | 70.9 s       |

### Table 6
The Training Time Of Multiple Task Learning.

| Parallel task | 1          | 2          | 3          | 4          | 5          |
|---------------|------------|------------|------------|------------|------------|
| LSTM-MTL     | 52.2 s     | 72.8 s     | 86.4 s     | 104.7 s    | 118.6 s    |
| Speedup rate | 0%         | 46.0%      | 86.9%      | 101.1%     | 120.6%     |
| TCN-LSTM-MTL | 68.5 s     | 94.1 s     | 114.8 s    | 132.7 s    | 152.0 s    |
| Speedup rate | 0%         | 47.4%      | 84.2%      | 112.8%     | 132.2%     |

Throughout COVID-19, we will evaluate the accuracy of numerous benchmark models. We designated New York City as the target task and Washington D.C., Columbus, and Philadelphia as auxiliary tasks.

Fig. 6 and Tables 7 illustrate the forecasted findings for New York City from August 9th to August 15th, 2020. Without mobility data, the RNN model cannot track changes in the load pattern during COVID-19. This condition is also present in the LSTM, TCN, and other non-sequential models, where the forecasting errors of the SVR and BR models all exceed 5%. More importantly, the forecasting errors of the French RNN model reach 43.684 percent, which is completely unacceptable. However, when mobility data optimization is incorporated into the model training process, considerable changes occur. The LSTM and TCN models with the “new” dataset containing mobility data are capable of controlling the error to less than 2%; this is also true for the RNN model with mobility optimization (NN-Mob-Single). The preceding experiments demonstrate that during the COVID-19 era, regardless of whether it is a classic RNN model, an LSTM model, or a developing TCN model, as long as it is trained using historical load, weather, and other indicators, it will be unable to successfully predict the load. This demonstrates the effectiveness of mobility data in optimizing load forecasts.

In comparison to combined TCN and LSTM model (TCN-LSTM), the maximum error of the mobility optimized TCN-LSTM model (TCN-LSTM-Mob-Single) occurs in the ISO-level task of APS (1.213 %, which is only 34 % of that of TCN-Mob-Single). This demonstrates how the combination of TCN and LSTM can help improve load forecasting accuracy. However, the error of LSTM-Attention exceeds 2 % (up to 3.693 %) in metropolitan-level tasks and exceeds 3 % (up to 4.109 %) in ISO-level tasks, which represent average accuracy. Additionally, the TCN-LSTM-Mob-Single has a lower TIC value than the LSTM-Attention in the majority of activities. As a result, we can conclude that the TCN-LSTM-Mob-Single model has improved accuracy, numerical stability, robustness of muscular strength, and generalizability.

By examining the findings of multi-task-based models (TCN-LSTM-MTL and LSTM-MTL), it is clear that the training and forecasting accuracy of the multi-task-based models is much higher than that of the single-task-based LSTM-Mob-Single and TCN-Mob-Single models. Their curves are more similar to the real load, demonstrating the significant benefit of multi-task learning in high-precision load forecasting. With the aid of multi-task learning, even standard LSTM models can attain optimal forecasting accuracy. However, among the 19 benchmark models, the TCN-LSTM-MTL model has the highest forecasting accuracy (shown in navy blue and bold). The highest forecasting accuracy was achieved by eight tasks. According to portion C (including portion C.1) of Fig. 6, the TCN-LSTM-MTL model curve’s forecasting results are closer to the actual value than the LSTM-MTL model curve’s forecasting results. This proves conclusively that the addition of TCN accurately captures the specific load patterns and adequately explains the success of the TCN-LSTM-MTL model we suggested.

We proceed to compare the effects of various activities at various
Fig. 6. The forecasting curves and error probability distribution group diagrams of the New York’s load forecasting task during the period from August 9th to 15th of 2020 in COVID-19. (A) The change curves and error probability distribution diagrams of LSTM-Old, LSTM-Retrain-New, LSTM-Mob-Single, LSTM-MTL and actual load; (B) The change curves and error probability distribution diagrams of TCN-Old, TCN-Retrain-New, TCN-Mob-Single, TCN-LSTM-Mob-Single, TCN-LSTM-MTL and actual load; (C) The change curves and error probability distribution diagrams of TCN-LSTM-Mob-Single, TCN-LSTM-MTL, LSTM-MTL and LSTM-Attention and actual load.
scale levels of regions using horizontal observations, using the TCN-LSTM-MTL model (which has the highest accuracy) as an example. Except for APS’s forecasting duty, practically all regions’ forecasting mistakes may be kept below 1 %. (the lowest is only 0.442 %, which is 5.6 % higher than the highest accuracy of LSTM-MTL, 8.8 % higher than the highest accuracy of TCN-LSTM-Mob-Single, and 28.2 % higher than the highest accuracy of LSTM-Mob-Single). APS’s forecasting error is 1.121 %, which is likewise a respectable outcome. These findings reveal unequivocally that the TCN-LSTM-MTL model is sufficiently versatile and capable of accurately adapting to forecasting tasks at various sizes and geographies. The ISO-level forecasting results had the largest average error (0.831 %), followed by the metropolitan level (0.806 %) and the national level (0.536 %). This could be because the ISO-level activities all have regular load change law curves, and it’s difficult to find a fine-grained representation of the entire ISO. In the case of national-level tasks, the start and shutdown of a single huge load will have a negligible effect on the national-level load, resulting in a consistent time series of load variations.

In general, both the mobility optimized TCN-LSTM-MTL model and the LSTM-MTL model can be employed as a realistic and universal synergistic load forecasting model during the COVID-19. Through these two multi-task learning models, enterprises managing multi-regional power loads can achieve high-precision load forecasting. The TCN-LSTM-Mob-Single model is an excellent option for small power providers.

4.3. Variable timing sequence length analysis

To properly grasp the model’s robustness, we will evaluate several benchmark models under varied temporal sequence training and forecasting situations. We merely alter the amount of time series data that the model receives. Fig. 7 illustrates the results of variable sequence tests. The LSTM-MTL and TCN-LSTM-MTL models, which include multi-task learning, outperformed other single-task learning models in forecasting. TCN-LSTM-MTL outperforms the LSTM-MTL model in the processing of longer time sequences. However, the difference between the two models is negligible when the sequence input is less than or equal to 7 days. When the sequence length is 1, 3, or 4 days, the LSTM-MTL model is marginally more accurate than the TCN-LSTM-MTL model. This demonstrates that the LSTM-MTL model is more suitable for training models with short-time inputs. When a longer sequence input is used, the TCN-LSTM-MTL model is preferable. In comparison, it is not difficult to observe that most single-task models, with the exception of the LSTM-Attention model, have more errors in any sequence input than Table 7

| MAPE(%) (TIC100) | NN-Old | NN-Retrain-New | NN-Mob-Single | LSTM-Old | LSTM-Retrain-New | LSTM-Mob-Single | TCN-Old | TCN-Retrain-New | TCN-Mob-Single | LSTM-LSTM-MTL | LSTM-MTL | LSTM-Attention |
|------------------|--------|----------------|--------------|---------|-----------------|----------------|---------|----------------|--------------|---------------|---------|----------------|
| NY               | 0.056  | (3.945)        | 9.441        | 8.277   | 8.887           | 9.174          | 9.789   | 9.977          | 10.079       | 10.121        | 10.121  |
| COL              | 0.056  | (3.945)        | 9.441        | 8.277   | 8.887           | 9.174          | 9.789   | 9.977          | 10.079       | 10.121        | 10.121  |
| WDC              | 0.056  | (3.945)        | 9.441        | 8.277   | 8.887           | 9.174          | 9.789   | 9.977          | 10.079       | 10.121        | 10.121  |
| PHIL             | 0.056  | (3.945)        | 9.441        | 8.277   | 8.887           | 9.174          | 9.789   | 9.977          | 10.079       | 10.121        | 10.121  |
| CAISO            | 0.056  | (3.945)        | 9.441        | 8.277   | 8.887           | 9.174          | 9.789   | 9.977          | 10.079       | 10.121        | 10.121  |
| APS              | 0.056  | (3.945)        | 9.441        | 8.277   | 8.887           | 9.174          | 9.789   | 9.977          | 10.079       | 10.121        | 10.121  |
| NEVP             | 0.056  | (3.945)        | 9.441        | 8.277   | 8.887           | 9.174          | 9.789   | 9.977          | 10.079       | 10.121        | 10.121  |
| PAC              | 0.056  | (3.945)        | 9.441        | 8.277   | 8.887           | 9.174          | 9.789   | 9.977          | 10.079       | 10.121        | 10.121  |
| SWIT             | 0.056  | (3.945)        | 9.441        | 8.277   | 8.887           | 9.174          | 9.789   | 9.977          | 10.079       | 10.121        | 10.121  |
| FRA              | 0.056  | (3.945)        | 9.441        | 8.277   | 8.887           | 9.174          | 9.789   | 9.977          | 10.079       | 10.121        | 10.121  |
| ITA              | 0.056  | (3.945)        | 9.441        | 8.277   | 8.887           | 9.174          | 9.789   | 9.977          | 10.079       | 10.121        | 10.121  |
| GER              | 0.056  | (3.945)        | 9.441        | 8.277   | 8.887           | 9.174          | 9.789   | 9.977          | 10.079       | 10.121        | 10.121  |

Note: 1) the name of cities and countries are abbreviated. Specifically, NY = New York, COL = Columbus, WDC = Washington D.C., PHIL = Philadelphia, SWIT = Switzerland, FRA = France, ITA = Italy, GER = Germany. 2) the color gradient mapping level of each cell is determined by the MAPE error result in the cell, and green, yellow, and red correspond to small, medium, and large MAPE error values, respectively.
multi-task learning. When the input sequence duration exceeded 7 days, the LSTM-Mob-Single model increased rapidly. Although it began to decrease after ten days, it quickly recovered and stabilized at the 16-day high. This demonstrates that while LSTM is known as successful for dealing with large sequences, it is inappropriate for training on comparatively longer sequences.

In the model testing process, the model error ranking of the long sequence input is identical to the model training process. Almost all models, however, exhibit wavy error variations as the duration of the input sequence increases. When the input sequence is less than or equal to six days, all models’ errors decrease dramatically with increasing sequence length. Except for the LSTM-MTL and TCN-LSTM-MTL models, the errors of benchmark single-task models have increased significantly when the sequence length increases. This demonstrates that both of these multi-task models are suitable for forecasting lengthy time series, with the TCN-LSTM-MTL model performing better with relatively short sequence input.

By and large, the accuracy of multi-task learning models used for training and forecasting changing time sequence lengths is significantly higher in COVID-19.

4.4. Model interpretation

The SHAP graphs of the TCN-LSTM-MTL model of New York City in COVID-19 are shown in Fig. 8 and Fig. 9, respectively.

The comprehensive influence of indicators is shown in Fig. 8, which displays the top twenty indicators with the greatest impact on electrical load from February 15th, 2020 to August 15th, 2021. Temperature indications have the largest influence, and the majority of indicators have an effect in both directions. Traditional weather indicators (temperature, cloud cover, pressure, and humidity) all contributed significantly to the electrical load. This demonstrates unequivocally that weather indicators are functional throughout COVID-19. Additionally, all of the mobility indicators that we introduced are listed in the top 15, demonstrating the mobility indicators’ significant optimization potential for power load during COVID-19. Meanwhile, comparing the ranking of Fig. 8 to the correlation ranking in Table 4 demonstrates that indicator correlation verification can actually be utilized to determine whether load-related indicators are possibly useful for load forecasting. However, a strong correlation does not necessarily imply a substantial contribution, which is asynchronous.

Additionally, when examining the influence of indicators at different time points. The force group graphs in the B, C, D sections of Fig. 9 illustrate the visual interpretation results at multiple time points prior to and during the COVID-19 epidemic. To begin, Figures B and C depict the regular morning and nighttime loads in New York City in the absence of COVID-19. Whether in the morning or at night, the power load is high, and the transit, transit station, and driving indicators associated with commuting tend to increase the load. And, the workplace indicator associated with actual work behavior tends to decrease the load. However, the pattern of several indications of power load in COVID-19 has shifted dramatically. Figure D depicts the electricity load two months after February 21st, 2020, at the same time point (8 a.m.) as Figure B. The power load number has decreased significantly, and is now just 68.62 % of what it was two months ago. Additionally, those impacted by the travel ban, transit, and driving signs tend to significantly lower their power usage, which is the polar opposite of two months earlier.

Comparing Figures B and D reveals another intriguing phenomenon. Specifically, there is little variance in load patterns between weekdays and weekends during COVID-19, which is not typical.

4.5. Time and season validity verification

Due to the addition of COVID-19, the load forecasting methods needs to be vertically verified in different seasons and time to get enough industrial application feasibility. In this section, we will implement the time and seasons verification on the TCN-LSTM-MTL model we proposed to test the robustness of the model. We keep the parameters (model structure, model parameters, timing lengths, etc.) unchanged to control the number of variables. For example, if we need to carry out load forecasting tasks from July 9th to July 15th, 2021, we will use the...
We continued to obtain data from August 15th, 2020 to August 15th, 2021, and integrated with the original dataset with the time range from February 15th to August 15th, 2020. It is worth noting that the COVID-19 has experienced all stages that tend to normalize prevention and control in this time range, and multiple regions have experienced travel restrictions and national blockade multiple times during this period. Therefore, the test conditions of this section are stricter, and the model requires sufficient versatility and robustness to address the actual scenes of most COVID-19 fully.

The longitudinal test results of the times and seasons for the TCN-LSTM-MTL model are shown in Table 8. Almost all the load forecasting tasks of different levels have achieved good results, and the forecasting errors of most tasks are controlled by 1 % or less than 1 %. The largest error occurred in NEVP on the tasks of time range from December 9th to December 15th of 2020 with the maximum MAPE error value 1.701 %, and the corresponding 100 times TIC value 1.076. From the task of different levels, the average errors of the metropolitan, ISO level and national tasks is decremented, which is significantly different from Section 4.2 of June 9th to August 15th, 2020. This shows that the errors of the metropolitan and ISO load forecasting tasks are not necessarily sorted, and their errors will change over time and season. However, we can also find that the average error of the national-level forecasting task has been the lowest, which is the same as Section 4.2.

In general, the TCN-LSTM-MTL model has a good forecasting effect and robustness during different seasons and periods of COVID-19. This also fully proves the effectiveness of our proposed TCN-LSTM-MTL model.

5. Conclusion

The purpose of this research is to present a novel TCN-LSTM-MTL model for performing high-precision short-term load forecasting during COVID-19. The model is based on temporal convolutional networks, population mobility, and a multi-task learning algorithm. The following are the paper’s conclusions:

- By including a temporal convolutional network and multi-task learning, the forecasting model’s adaptability for long-sequence input training and forecasting was significantly improved. The TCN-LSTM-MTL model has been shown to significantly increase the model’s accuracy, robustness, and adaptability in COVID-19.
- The incorporation of population mobility data to COVID-19 successfully alleviated the difficulty associated with following the load pattern.
- The SHAP visualization model technology, which is based on game theory, elucidates the TCN-LSTM-MTL model’s operation at certain...
time points in COVID-19 and demonstrates the critical importance of population movement data. Additionally, the non-synchrony link between the correlation and contribution of characteristics is proven.

This paper’s work can be expanded by developing a load forecasting approach that is capable of dealing with harsh weather, geological disasters, economic crises, and other scenarios.

**CRediT authorship contribution statement**

Zhenhao Zhang: Writing – original draft, Writing – review & editing. Jiefeng Liu: Writing – original draft, Writing – review & editing. Senshen Pang: Visualization, Project administration. Mingchen Shi: Visualization, Project administration. Hui Hwang Goh: Investigation, Supervision. Yiyi Zhang: . Dongdong Zhang: Resources, Software, Data curation.

**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Data availability**

All data sources in the text are public data sources, they have been mentioned and cited in the text.

**References**

[1] Li Z, Ye H, Liao N, Wang R, Qiu Y, Wang Y. Impact of COVID-19 on electricity energy consumption: A quantitative analysis on electricity. Int J Electr Power Energy Syst 2022;140:108084. https://doi.org/10.1016/j.ijepes.2022.108084.

[2] Chen Y, Yang W, Zhang B. Using Mobility for Electrical Load Forecasting During the COVID-19 Pandemic. ArXiv E-Prints 2020.

[3] Liu J, Zhang Z, Fan X, Zhang Y, Wang J, Zhou Ke, et al. Power system load forecasting using mobility optimization and multi-task learning in COVID-19. Appl Energy 2021;310:118303.

[4] Wu D, Zheng X, Xu Y, Olsen D, Xia B, Singh C, et al. An open-source extendable model and corrective measure assessment of the 2021 texas power outage. Adv Appl Energ 2021;4:100056.

[5] Rouzbahani HM, Karimipour H, Lei L. A review on virtual power plant for energy management. Sustain Energ Technol Assessments 2021;47:101370. https://doi.org/10.1016/j.seta.2020.101370.

[6] Tan M, Yuan S, Li S, Su H, Li H, He FH. Ultra-Short-Term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning. IEEE Trans Power Syst 2020;35:2937–48. https://doi.org/10.1109/tpwrs.2019.2963109.

[7] Lin J, Ma J, Zhu J, Cai Y. Short-term load forecasting using CNN-LSTM neural networks. Energy 2019;182:72–81. https://doi.org/10.1016/j.energy.2019.05.230.

[8] Ahmad T, Zhang H. Novel deep supervised ML models with feature selection for short-term load forecasting during periods with significant temperature variations in city of Burbank. Appl Soft Comput 2014;16:80–8. https://doi.org/10.1016/j.asoc.2013.12.001.

[9] Poghosyan A, Greetham DV, Haben S, Lee T. Long term individual load forecast under different electrical vehicles uptake scenarios. Appl Energy 2015;157:699–709. https://doi.org/10.1016/j.apenergy.2015.02.069.

[10] Barman M, Dev Choudhury NB. A similarity based hybrid GWO-SVM method of power system load forecasting for regional special event days in anomalous load situations in Assam. India Sustain Cities Soc 2020;61:102311. https://doi.org/10.1016/j.scs.2020.102311.

[11] Selakov A, Cvijetinovic D, Milovic D, Mellon S, Bekut D. Hybrid PSO-SVM method for short-term load forecasting during periods with significant temperature variations in city of Banja Luka. Appl Soft Comput 2020;85:105781. https://doi.org/10.1016/j.asoc.2020.105781.

[12] Wang Y, Sun S, Chen X, Zeng X, Kong Y, Chen J, et al. Short-term load forecasting of industrial customers based on SVMD and XGBoost. Int J Electr Power Energy Syst 2021;129:106830. https://doi.org/10.1016/j.ijepes.2021.106830.

[13] Fan G-F, Zhang L-Z, Yu M, Hong W-C, Dong S-Q. Applications of random forest in short-term load forecasting of industrial businesses, considering attention mechanism. Int J Electr Power Energy Syst 2022;137:108073. https://doi.org/10.1016/j.ijepes.2022.108073.

[14] Lu L, Wei Z, Zhang J, Zhang L, Tan Z, Tian Z A. VMD and LSTM based hybrid method of load forecasting for power grid security. IEEE Trans Ind Informatics 2022;18(0):6474–82.

[15] Kim T-Y, Cho S-B. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 2015;82:72–81. https://doi.org/10.1016/j.energy.2015.09.055.

[16] Afrasiabi M, Mohammad M, Rastegar M, Stanovic I, Afrasiabi S, Khazaei M. Deep-Based Conditional Probability Density Function Forecasting of Residential...
[17] Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting. AAAI 2021;35(12):11106–15.

[18] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Adv Neural Inf Process Syst 2017;30:5998–6008.

[19] Wolf T, Chaumond J, Debu T, Sankh V, Delangue C, Moi A, et al. Conf Empir. Methods. Nat Lang Process Syst Demontr 2020;2020:38–45.

[20] Qin J, Zhang Y, Fan S, Hu X, Huang Y, Lu Z, et al. Multi-task short-term reactive and active load forecasting method based on attention-LSTM model. Int J Electr Power Energy Syst 2022;135:107517.

[21] Wang Y, Chen J, Chen X, Zeng X, Kong Y, Sun S, et al. Short-Term Load Forecasting for Industrial Customers Based on TCN-LightGBM. IEEE Trans Power Syst 2021;36(3):1984–97.

[22] He S, Li C, Liu X, Chen X, Shahidehpour M, Chen T, et al. A per-unit curve rotated decoupling method for CNN-TCN based day-ahead load forecasting. IET Gener Transm Distrib 2021;15(19):2773–86.

[23] Pan SJ, Yang Q. A Survey on Transfer Learning. IEEE Trans Knowl Data Eng 2010;22:1345–59. https://doi.org/10.1109/TKDE.2009.191.

[24] Zhang Y, Luo G. Short term power load prediction with knowledge transfer. Inf Syst 2015;53:161–9. https://doi.org/10.1016/j.is.2015.01.005.

[25] Xu X, Meng Z. A hybrid transfer learning model for short-term electric load forecasting. Electr Eng 2020;102:1371–81. https://doi.org/10.1007/s00202-020-00930-x.

[26] Zeng P, Sheng C, Jin M. A learning framework based on weighted knowledge transfer for holiday load forecasting. J Mod Power Syst Clean Energy 2018;7:329–39. https://doi.org/10.1007/s40565-018-9435-z.

[27] Gilanifar M, Wang H, Sriman LMK, Ozguven EE, Arghandeh R. Multi-task Bayesian Spatiotemporal Gaussian Processes for Short-Term Load Forecasting. IEEE Trans Ind Electron 2020;67:5132–43. https://doi.org/10.1109/TIE.2019.2928275.

[28] Hu C, Zhang J, Yuan H, Gao T, Jiang H, Yan J, et al. Black swan event small-sample transfer learning (BEST-L) and its case study on electrical power prediction in COVID-19. Appl Energy 2022;309:118458.

[29] Zhang S, Chen Q, Zeng W, Guo S, Xu J. A novel transfer deep learning model with reinforcement-learning-based hyperparameter optimization for short-term load forecasting during the COVID-19 pandemic. J Intell Fuzzy Syst 2022;43(3):2869–82.

[30] Pirnia M, Elzaruge M, Beylunioflhu PC, Ahmed M, Nathwani J. Impact of COVID-19 on Ontario’s electricity market: Load, generation, emissions. Electr J 2022;35(4):107111.

[31] Apple. Mobility Data 2020:https://www.apple.com/covid19/mobility.

[32] Google. COVID-19 Community Mobility Reports 2020:https://www.google.com/covid19/mobility/.

[33] Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. Proc. 31st Int. Conf. Neural Inf. Process. Syst., Red Hook, NY, USA: Curran Associates Inc.; 2017. p. 4768–4777.

[34] Lundberg SM, Nair B, Vavilala M, Horibe M, Einses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018;2(10):749–60.