IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 12 May 2024, accepted 2 July 2024, date of publication 8 July 2024, date of current version 26 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3424688

RESEARCH ARTICLE

Short-Term Electricity Demand Forecasting for DanceSport Activities

KEYIN LIU[1], HAO LI[2], AND SONG YANG[3,4]
[1]Chengdu Sports University, Chengdu, Sichuan 610041, China
[2]College of Computer Science, Sichuan University (SCU), Chengdu 610065, China
[3]Huawei Technologies Company Ltd., Shenzhen 518057, China
[4]Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Corresponding author: Song Yang (y14681@sina.com)

ABSTRACT This paper introduces a novel hybrid deep learning-based approach for short-term electricity demand forecasting in dance sport activities. Traditional deep learning methods often overlook important spatial dependencies and key features like trend and seasonal patterns. To address these limitations, we propose a model that combines Transformer for temporal feature extraction and Graph Neural Networks for spatial feature extraction, enabling prediction based on spatial-temporal features. Additionally, we employ the decomposition techniques to extract seasonal and trend features from dance sports data. By integrating early fusion (feature-level fusion) and late fusion (score-level fusion) strategies, our model achieves superior performance, outperforming baseline methods by over 4% on benchmark datasets. Additionally, we conduct the ablation study to comprehensively analyze the impact of each module on prediction accuracy, providing valuable insights into the contribution of spatial, temporal, seasonal and trend features to the overall forecasting performance.

INDEX TERMS Short-term demand forecasting, graph neural networks, DanceSport, hybrid fusion.

I. INTRODUCTION

DanceSport, as a competitive form of ballroom dance, can be usually divided into two subcategories with ten different dance styles, including International Standard Dance, such as Waltz, Viennese Waltz, Foxtrot, Quick Step and Tango, and International Latin Dance, e.g., Rumba, Cha Cha, Jive, Samba and Paso Doble [1]. In the past two decades, DanceSport has emerged as a widely embraced physical activity, captivating enthusiasts worldwide and fostering the establishment of numerous training institutions [2], [3]. These training facilities, often strategically located in residential areas, cater to the growing demand of DanceSport activities. However, unlike conventional training institutes, DanceSport facilities exhibit a distinctive characteristic: during training sessions, the instantaneous electricity demand will experience a significant surge, nearly doubling the typical power consumption, which demands the proactive energy reservoir planning from power grid companies. The heightened demand primarily stems from the power-intensive nature of video analytics equipment, air conditioning systems employed in instructional programs, and various auxiliary devices such as microwaves.

The pronounced electricity consumption patterns in DanceSport activities present a unique challenge, as traditional electricity load forecasting models may not adequately capture the dynamic and intensified nature of power usage during training sessions. Recognizing the need for tailored short-term electricity demand forecasting in this context, this paper aims at building an DanceSport-specific electricity demand forecasting model. Such a model would enable electricity providers to anticipate and make preparations for peak demand instances, facilitating targeted energy reservoir planning for specific locations. The proactive approach not only ensures the efficient allocation of power resources but also addresses the specific energy dynamics associated with DanceSport activities. By delving into the intricacies of electricity consumption within this specialized domain,

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

the developed predictive model is able to contribute to power distribution strategies and the overall sustainability of DanceSport training facilities [4].

While Section II will present a brief literature review on short-term time-series data forecasting models, here, we would like to highlight that existing methods have not effectively addressed the inherent characteristics of DanceSport, leading to the suboptimal forecasting performance. DanceSport exhibits salient short-term correlation features, particularly during warm-up sessions conducted by users and instructors before training. While the power consumption may not experience a significant surge during this phase, the integration of key electrical features associated with warm-up activities, such as initiating dance music playback and displaying video footage of users engaging in pre-training exercises, has the potential to accurately predict the short-term electricity demand. Additionally, DanceSport courses exhibit distinct periodic characteristics, with lighter training sessions like Latin dance occurring on workdays and more intensified DanceSport activities, such as street dance, Ballet, taking place during weekends. These recognizable patterns contribute significantly to the predictive accuracy of the electricity surge, providing invaluable insights for tailored forecasting in the context of DanceSport.

In view of the aforementioned characteristics, this paper introduces a robust short-term electricity demand forecasting model specifically designed for DanceSport activities. Unlike traditional approaches, our model utilizes a hybrid deep learning-based framework which combines early fusion at the feature level with the late fusion at the score level. We employ a Transformer-based architecture to capture the temporal dependence of DanceSport activities using the multi-head attention mechanism. Furthermore, we integrate GNN to extract spatial features which characterize the spatial dependence in DanceSport electricity consumption, accounting for activities such as users drinking water or changing clothes during DanceSport warm-up activities.

Furthermore, we employ the time series data decomposition technique to extract seasonal and trend features. These features are then combined through feature fusion, leveraging the outputs of both Transformer and GNN models, which are subsequently fed into a hybrid Deep Neural Network (DNN) for electricity demand prediction. Our approach, which incorporates spatial, temporal, and periodic information, surpasses conventional time-series prediction models, showcasing superior accuracy in forecasting DanceSport electricity consumption. Figure 1 demonstrates the flowchart of our proposed method.

The paper's contributions can be summarized as follows: (1) A novel hybrid deep learning-based forecasting framework is proposed for short-term electricity demand in DanceSport activities, leveraging Transformer architecture with a multi-head attention mechanism, graph neural networks, and time series data decomposition to comprehensively capture temporal, spatial, seasonal and trend features. (2) The inclusion of GNN specifically addresses spatial dependence, modeling user behaviors during warm-up, such as water consumption and clothing changes. (3) Demonstrated through quantitative performance comparison, the proposed model outperforms state-of-the-art algorithms, showcasing the effectiveness of the multi-faceted strategy in accurately predicting DanceSport electricity consumption patterns.

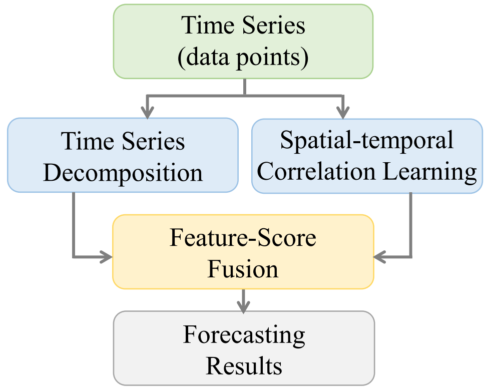

FIGURE 1. A flowchart of our proposed method.

The remainder of the paper is organized as follows: In Section II, a brief literature review is presented, focusing on short-term load forecasting methods as well as feature fusion for predictive modeling. Section III introduces the background knowledge of Transformer and graph neural networks (GNN), providing the necessary contexts for the subsequent methodology presentation. The proposed hybrid deep learning-based forecasting model, integrating Transformer, GNN, and time-series data decomposition for feature extraction, is presented in Section IV. To validate the model's effectiveness, comparative experiments with canonical baseline algorithms are conducted on benchmark datasets in Section IV, complemented with an ablation study to discern individual contributions of each component within the model. We end the paper with conclusion and future work in Section VI.

II. LITERATURE REVIEW

In this section, we first present a brief overview of methodologies used for short-term load forecasting (STLF), and then introduce fusion methods for predictive modelling, in that the proposed model in this paper is essentially a fusion model for short-term load forecasting.

A. SHORT-TERM LOAD FORECASTING (STLF)

Short-term load forecasting (STLF) holds paramount importance in various domains and has thus attracted increasing attentions from both academic researchers and industrial entrepreneurs over the past several decades [5], [6], [7], [8], [9].

Early studies predominantly rely on the auto-regressive moving average (ARMA) model and its variants [10], [11], [12], [13] as foundational frameworks for short-term load forecasting. These models aimed to predict load based on linear characteristics within historical load records. However, the inherent complexity of dynamic load patterns renders the linearity assumption of these forecasting models limiting in terms of the prediction accuracy. Consequently, the introduction of deep learning-based methods (DLMs) become imperative, seeking to capture intricate features in diverse manners. The transition from classical statistical methods to DLMs marks a shift towards more sophisticated modeling techniques capable of accommodating the intricate dynamics inherent in short-term load forecasting.

In the realm of short-term load forecasting (STLF), modern Deep Learning Models (DLMs) like DRN [14], RNN [6], [15], LSTM [16], and CNN [17] have prominently advanced temporal feature extraction. However, to the best of our knowledge, DLMs often focus solely on temporal dynamics and overlook the crucial spatial dependencies within electricity consumption data [18], [19], [20]. Additionally, these models often lack explicit mechanisms for capturing the periodical patterns, potentially compromising their accuracy in predicting nuanced load variations over specific time intervals.

B. FUSION METHODS FOR PREDICTIVE MODELING

Fusion methods have emerged as effective strategies to mitigate the limitations of individual models, making them promising candidates for constructing robust hybrid models for prediction. These approaches are particularly valuable for addressing the inherent challenges associated with complex and dynamic datasets. In the academic domain, two primary perspectives on fusion methods have gained prominence: early fusion and late fusion.

Early fusion operates at the feature level, merging information before it reaches the predictive model. The approach integrates diverse features from multiple sources, enhancing the model's ability to capture comprehensive patterns in the data. Subsequently, the fused features are fed into Deep Learning Models (DLMs) for prediction. The advantage lies in the combination of diverse information, allowing the model to capitalize on the strengths of each source. Notable examples include early fusion techniques employed in multi-modal applications, where features from different data modalities are integrated for more holistic predictions [21], [22], [23]. On the other hand, late fusion involves individual models to generating prediction independently. These predictions are later integrated into a unified hybrid model during the later stages of the modeling process, providing a deeper and more comprehensive understanding of data patterns. Late fusion, particularly observed in ensemble methods like stacking and bagging, proves highly effective when managing diverse data sources or models with distinct advantages [24], [25].

The proposed model in this paper employs a hybrid fusion approach comprising both feature-level early fusion and score-level late fusion, combining Transformer, GNNs, and time-series data decomposition, for enhanced forecasting capabilities in short-term DanceSport electricity demand prediction.

III. BACKGROUNDS: TRANSFORMER AND GRAPH NEURAL NETWORKS

Since our proposed methodology leverages the Transformer-based multi-head attention model for DanceSport's temporal feature extraction and integrates Graph Neural Networks (GNNs) for the spatial feature extraction, this section provides the essential background knowledge of Transformer and GNNs.

A. TRANSFORMER AND MULTI-HEAD ATTENTION

The Transformer architecture, introduced by Vaswani et al. [26], has become a cornerstone in natural language processing and sequential data modeling. Its innovative design replaces recurrent or convolutional structures with a self-attention mechanism that allows for capturing long-range temporal dependencies efficiently. Transformer consists of an encoder-decoder structure, and in our context, we primarily focus on the encoder part for temporal feature extraction.

One pivotal component of the Transformer's encoder is the multi-head attention mechanism. Given an input sequence $\mathbf{X} = (x_1, x_2, \ldots, x_n)$, the multi-head attention mechanism operates by computing scaled dot-product attention in parallel across multiple attention heads. The attention scores A_i for each head i are calculated as follows:

$$A_i = \mathrm{softmax}\left(\frac{\mathbf{Q}_i \cdot \mathbf{K}_i^T}{\sqrt{d_k}}\right),$$

where $\mathbf{Q}_i$, $\mathbf{K}_i$, and $\mathbf{V}_i$ represent the query, key, and value projections for the i-th head, and d_k is the dimension of the key vectors. The final attention output for each head is obtained through weighted summation:

$$\mathrm{Head}_i = \mathrm{softmax}\left(\frac{\mathbf{Q}_i \cdot \mathbf{K}_i^T}{\sqrt{d_k}}\right) \cdot \mathbf{V}_i,$$

The outputs from all attention heads are then concatenated and linearly transformed to produce the multi-head attention output. This mechanism enables the model to attend to different aspects of the input sequence simultaneously, enhancing its ability to capture diverse temporal dependencies.

In recent years, GNNs have gained significant attention and witnessed continuous advancements. Various extensions and variations, e.g., Graph Transformer [27], Vision Transformer (ViT) [28], Anomaly Transformer [29], have been introduced to address specific challenges in different application domains.

B. GRAPH NEURAL NETWORKS (GNNS)

Graph Neural Networks (GNNs) have emerged as powerful tools for learning from graph-structured data [30]. Unlike traditional neural networks that operate on grid-like data, GNNs can effectively capture relationships and dependencies

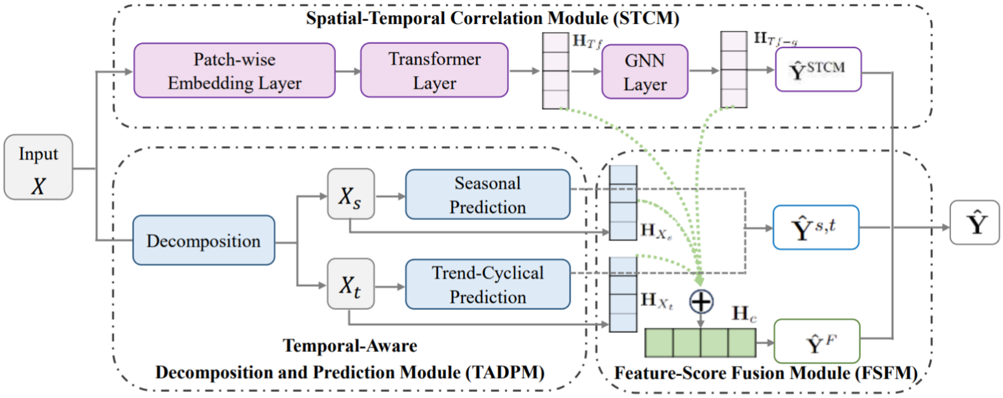

FIGURE 2. Architecture of the proposed multi-task learning model.

within non-Euclidean domains, making them well-suited for applications involving complex connectivity patterns.

At the core of GNNs is the aggregation of information from neighboring nodes in a graph. Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents the set of nodes and $\mathcal{E}$ denotes the set of edges. The information propagation in GNN can be formalized using the aggregation function:

$$h_v^{(l+1)} = \sigma\left(\sum_{u\in\mathcal{N}(v)} f\left(h_u^{(l)}, h_v^{(l)}, e_{uv}\right)\right).$$

Here, $h_v^{(l)}$ represents the node features at layer l for node v, $\mathcal{N}(v)$ denotes the neighbors of node v, f is a learnable aggregation function, and e_{uv} represents the edge features between nodes u and v. The function σ is a non-linear activation function.

One of the pioneering GNN architectures is the Graph Convolutional Network (GCN) proposed by Kipf and Welling [31]. GCN simplifies the aggregation function to:

$$h_v^{(l+1)} = \sigma\left(\sum_{u\in\mathcal{N}(v)} \frac{1}{\sqrt{\deg(u)\cdot\deg(v)}}\cdot h_u^{(l)}\right)$$

where $\deg(u)$ represents the degree of node u, and the aggregation weights are normalized by the geometric mean of the degrees of the connected nodes. GNNs have shown remarkable performance in tasks such as node classification, weakly supervised segmentation [32], and detection with missing data [33], demonstrating their versatility in modeling complex relationships in diverse domains.

IV. METHODOLOGY

In this section, we first give the problem definition. Following this, we delve into our proposed approach, a Multi-Task Learning Model (MTLM), tailored specifically for multi-variate time series forecasting problems, i.e., we investigate the electricity consumption prediction in this paper. Within MTLM, we elaborate on three key components: the Spatial-Temporal Correlation Module (STCM), the Temporal-Aware Decomposition and Prediction Module (TADPM), and the Feature-Score Fusion Module (FSFM).

A. PROBLEM DEFINITION

Given the historical time series $\mathbf{X} \in \mathbb{R}^{T\times N}$ from the past T time steps for N variables, where $\mathbf{x}^i \in \mathbf{X}$ denotes the i-th time series, i.e., $\mathbf{x}^i = \{x_1, \cdots, x_t, \cdots, x_T\}, i \in \{1, 2, \cdots, N\}$ and x_t is the electricity consumption at time step t. The goal of electricity demand forecasting is to predict the values $\hat{\mathbf{Y}} \in \mathbb{R}^{T_f\times N}$ for the future T_f time steps.

B. OVERVIEW

The overall framework of our proposed multi-task learning model (MTLM) is shown in Figure 2. MTLM mainly consists of three modules, namely spatial-temporal correlation module (STCM), temporal-aware decomposition and prediction module (TADPM) and feature-score fusion module (FSFM). In particular, STCM aims to learn the long-term temporal patterns and spatial correlations between time series based on a Transformer encoder and graph neural network layer, respectively. TADPM is adopted to capture both seasonal and trend-cyclical patterns in a decomposition way, which performs seasonal and trend predictions and fuses such predictions for the final time series forecasting. FSFM achieves the goal of multi-task learning by effectively utilizing the correlations and complementarities among multiple prediction tasks through fusion at both feature and score levels.

C. SPATIAL-TEMPORAL CORRELATION MODULE

STCM contains three layers with patch-wise embedding, Transformer and graph neural network, where the patch-wise embedding layer encodes the time series to help capture more semantic relations between patches (i.e., segment) compared with modeling separate points. Then, these segment representations are fed into the Transformer layer for capturing the temporal dependencies. By doing this, it can significantly reduce the length of the input time series and make the representation of time series more efficient. Finally, a graph neural network layer takes the last segment representation together with a graph adjacent matrix as the input to further learn spatial and temporal correlations.

1) PATCH-WISE EMBEDDING LAYER

We split the time series $\mathbf{x}^i \in \mathbb{R}^T$ from variable $i \in \{1, 2, \cdots, N\}$ into P non-overlapping patches with length L following [34], i.e., $T = P \times L$. Hence, the k-th patch can be denoted as $\mathbf{x}^i_k \in \mathbb{R}^L$, which can be transformed into the latent space via the patch-wise embedding layer:

$$\mathbf{r}^i_k = f_e(\mathbf{x}^i_k), \tag{1}$$

where $f_e(\cdot)$ is a transformation function, such as linear projection or multilayer perceptron (MLP), $\mathbf{r}^i_k \in \mathbb{R}^d$ is the representation of segment $\mathbf{x}^i_k$ and d is the dimension. As such, the sequence of segment representations can be denoted as $\mathbf{r}^i = \{\mathbf{r}^i_1, \mathbf{r}^i_2, \ldots, \mathbf{r}^i_P\}$.

2) TRANSFORMER LAYER

The Transformer layer takes $\mathbf{r}^i$ as the input and generates a sequence of latent representations:

$$\mathbf{H}^i_{Tf} = f_{Tf}(\mathbf{r}^i), \tag{2}$$

where $f_{Tf}(\cdot)$ is a Transformer layer, note that it also can be a stacked Transformer layer. $\mathbf{H}^i \in \mathbb{R}^{P\times d}$ is the latent representations which encodes the temporal correlations within time series, i.e., $\mathbf{H}^i_{Tf} = \{\mathbf{h}^i_1, \mathbf{h}^i_2, \ldots, \mathbf{h}^i_P\}$.

3) GRAPH NEURAL NETWORK (GNN) LAYER

The learned patch-wise latent representations in the Transformer layer can be utilized in the downstream models (e.g., GNN-based framework), to further enhance the spatial-temporal pattern learning. Therefore, the representations $\mathbf{H}_{Tf}$ from all time series are fed into a GNN layer (e.g., Graph WaveNet [35]) as follows:

$$\mathbf{H}_{Tf-g} = \mathrm{MLP}(\mathbf{H}_{Tf}) + \mathbf{H}_g, \tag{3}$$

$$\mathbf{H}_g = f_g(\mathbf{H}_{Tf}), \tag{4}$$

where $\mathbf{H}_{Tf-g} \in \mathbb{R}^{N\times d}$ is the final representation of STCM, $\mathrm{MLP}(\cdot)$ transforms the $\mathbf{H}_{Tf}$ to the latent space of $\mathbf{H}_g$. $f_g(\cdot)$ denotes the Graph WaveNet. Thus STCM can predict future values $\hat{\mathbf{Y}}^{\mathrm{STCM}}$ with a regression layer and the loss is calculated by mean absolute error:

$$\mathcal{L}^{\mathrm{STCM}} = \mathcal{L}(\hat{\mathbf{Y}}^{\mathrm{STCM}}, \mathbf{Y}) = \frac{1}{T_f N}\sum_{t=1}^{T_f}\sum_{i=1}^{N} |\hat{y}^{\mathrm{STCM}}_{ti} - y_{ti}|. \tag{5}$$

D. TEMPORAL-AWARE DECOMPOSITION AND PREDICTION MODULE

The spatial-temporal correlation module, equipped with a Transformer layer and a GNN layer, shows strong capability in modeling long-term sequences and spatial correlations among time series. In the scenario of sports activities, electricity demand often exhibits seasonal and trend-cyclical characteristics, for instance, sports activities with periodical practice schedules and seasonal match arrangements. We thus design a temporal-aware decomposition and prediction module (TADPM) to further explore and exploit such seasonal and trend-cyclical patterns inspired by the recent study [36].

1) DECOMPOSITION

TADPM separates the time series $\mathbf{X} \in \mathbb{R}^{T\times N}$ into two parts following the decomposition method in [36]:

$$\mathbf{X}_t = \mathrm{AP}(Padding(X))_{kernel}, \tag{6}$$

$$\mathbf{X}_s = \mathbf{X} - \mathbf{X}_t, \tag{7}$$

where $\mathbf{X}_t, \mathbf{X}_s \in \mathbb{R}^{T\times N}$ represents the trend-cyclical and seasonal parts, respectively. $Padding(\cdot)$ keeps the series length unchanged, and $AP(\cdot)$ is adopted to separate different patterns with different *kernels*, thus $\mathbf{X}_t$ can be generated by the mean operation:

$$\mathbf{X}_t = \mathrm{mean}(\mathrm{AP}(Padding(X))_{kernel_1}, \cdots, AP(Padding(X))_{kernel_n}). \tag{8}$$

2) TREND-CYCLICAL PREDICTION

We adopt a linear regression to perform future trend-cyclical series prediction by taking $\mathbf{X}_t \in \mathbb{R}^{T\times N}$ as the input:

$$\hat{\mathbf{Y}}^t = f_{reg}(\mathbf{X}_t), \tag{9}$$

where $f_{reg}(\cdot)$ denotes the regression layer, and $\hat{\mathbf{Y}}^t \in \mathbb{R}^{T_f\times N}$ is the prediction results of trend part.

3) SEASONAL PREDICTION

To capture the seasonal patterns from the seasonal part $\mathbf{X}_s$, we employ a well-designed module[1] in [36], thus the final prediction of the seasonal part can be expressed as:

$$\hat{\mathbf{Y}}^s = f_{MIC}(\mathbf{X}_s), \tag{10}$$

where $f_{MIC}(\cdot)$ indicates the multi-scale isometric convolution (MIC) layer, and $\hat{\mathbf{Y}}^s \in \mathbb{R}^{T_f\times N}$ is the seasonal prediction. Therefore, the final result is calculated by the mean operation of the trend-cyclical and seasonal predictions:

$$\hat{\mathbf{Y}}^{s,t} = mean(\hat{\mathbf{Y}}^s, \hat{\mathbf{Y}}^t). \tag{11}$$

E. FEATURE AND SCORE FUSION MODULE (FSFM)

Although both STCM and TADPM can predict future electricity demand separately, the prediction bias of individual modules inevitably affects the stability and robustness of overall predictions. This inspires us to design an ensemble strategy to combine the feature outputs and prediction results of multiple modules, so as to achieve more stable overall prediction performance. We thus propose a feature-score fusion module (FSFM), which fuses different features and the individual predictions from different modules. Specifically, FSFM concatenates latent representations (i.e., $\mathbf{H}_{Tf}$ and $\mathbf{H}_{Tf-g}$) from the Transformer layer and GNN

[1] For brevity, we omit its calculation process, interested readers can refer to MICN [36] for more details.

layer as well as both seasonal and trend parts (i.e., $\mathbf{X}_s$ and $\mathbf{X}_t$) as follows:

$$\mathbf{H}_c = \mathbf{H}_{Tf} \oplus \mathbf{H}_{Tf-g} \oplus \mathbf{H}_{X_s} \oplus \mathbf{H}_{X_t}, \tag{12}$$

where $\mathbf{H}_{X_s}$, $\mathbf{H}_{X_t}$ are the linear projection results with $\mathbf{W}_{s,t} \in \mathbb{R}^{T \times d}$, $\mathbf{H}_c$ is the feature concatenation result, which is fed into the regression layer $f_{reg}(\cdot)$ for prediction:

$$\hat{\mathbf{Y}}^F = f_{reg}(\mathbf{H}_c). \tag{13}$$

After obtaining the predictions from different modules, FSFM then performs score fusion in a simple mean operation for the final prediction as follows:

$$\hat{\mathbf{Y}} = \mathrm{mean}(\hat{\mathbf{Y}}^{STCM}, \hat{\mathbf{Y}}^{s,t}, \hat{\mathbf{Y}}^F), \tag{14}$$

where $\hat{\mathbf{Y}} \in \mathbb{R}^{T_f \times N}$ represents the prediction values for the future T_f time steps.

F. MODEL TRAINING

We first pre-train the STCM and TADPM, respectively, and then train the multi-task learning model based on the following loss function:

$$\mathcal{L} = \alpha_1 \mathcal{L}(\hat{\mathbf{Y}}^{\mathrm{STCM}}, \mathbf{Y}) + \alpha_2 \mathcal{L}(\hat{\mathbf{Y}}^{s,t}, \mathbf{Y}) + \alpha_3 \mathcal{L}(\hat{\mathbf{Y}}^F, \mathbf{Y}) + \alpha_4 \mathcal{L}(\hat{\mathbf{Y}}, \mathbf{Y}), \tag{15}$$

where $\alpha_i, i \in [1, 4]$ is the hyper-parameter to be learned for adjusting the importance of different prediction tasks.

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATASET

This paper utilizes the publicly available Almanac of Minutely Power dataset (AMPds2) dataset [37], which captures the residential energy consumption. AMPds2 is selected as it includes all three major types of energy usage: electricity, water, and natural gas. The dataset spans a significant time frame of two years and has been meticulously processed to fill in minor gaps algorithmically, ensuring a consistent recording frequency of one reading per minute.

Specifically focusing on electricity consumption, the dataset comprises readings from 21 physical meters and 2 software-calculated meters. Each meter provides comprehensive data across 11 measurement parameters: voltage, current, frequency, displacement power factor, apparent power factor, real power, real energy, reactive power, reactive energy, apparent power, and apparent energy. To ensure the integrity of comparisons, rigorous data cleansing procedures have been implemented on the AMPds2 dataset.

We exclusively focus on the Electricity_P dataset extracted from the comprehensive AMPds2, specifically targeting electrical power loads relevant to the discussed tasks. We carefully selected 15 meters from this dataset that align with the requirements of dancing sports activities. It's important to note that the electricity usage patterns for dancing sports are distinct from typical daily demands, as the devices used are tailored for training or competition in this specific sport. From the total of 23 meters available in the Electricity_P dataset, we handpicked and utilized 15 devices directly related to dancing sports. To consolidate the data, we calculated the combined load of these 15 meters, treating them collectively as another representative meter for our analysis.

B. EXPERIMENTAL SETUP

In this subsection, we compare the performance of our proposed methods with several baseline prediction models, which will be introduced in the following. Furthermore, we also perform the ablation study by removing part of the components and evaluate the prediction performance of the remaining modules.

1) EXPERIMENTAL DETAILS

For data pre-processing, we resampled the data to a half-hourly frequency. Subsequently, we segmented the data based on timestamp sequences, allocating 60% as the training set, and the two remaining 20% each as validation sets and testing sets, respectively.

In line with STEP's configuration, we set the input length to 772 (equivalent to 2 weeks) and divided it into patches of size 12. The prediction length matches one patch size, i.e., 12, and we assessed prediction performance at 3, 6, and 12 steps, with each step corresponding to half an hour. The encoder structure mirrors STEP's setup, involving pre-training a reconstruction task with a 70% masking rate to acquire the trained encoder. Similarly, the STCM structure follows STEP's model, while the TADPM structure is patterned after MICN, both maintaining a hidden dimension of 128 for latent representations.

Utilizing the same assessment methodology as in the prior study, we calculate the Mean Absolute Error (MAE) on z-score normalized data to standardize various variables onto a comparable scale. Additionally, we employ four metrics as assessment standards: MAE, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). Their definitions are outlined as follows:

$$\mathrm{MAE}_t = \frac{1}{n} \sum_{i=1}^{n} \left| y_t^{(i)} - \hat{y}_t^{(i)} \right|, \tag{16}$$

$$\mathrm{MSE}_t = \frac{1}{n} \sum_{i=1}^{n} \left(y_t^{(i)} - \hat{y}_t^{(i)} \right)^2, \tag{17}$$

$$\mathrm{RMSE}_t = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_t^{(i)} - \hat{y}_t^{(i)} \right)^2}, \tag{18}$$

$$\mathrm{MAPE}_t = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_t^{(i)} - \hat{y}_t^{(i)}}{y_t^{(i)}} \right|, \tag{19}$$

where $y_t^{(i)}$ and $\hat{y}_t^{(i)}$ represent the original and forecasted consumption time series, respectively, for all instances i at a specific time step t. It's important to note that, unlike MAE, MSE, and RMSE which are computed using normalized

TABLE 1. Electricity consumption short-term forecasting performance comparison.

Methods	Horizon 3				Horizon 6				Horizon 12			
	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE	MAPE
STEP	0.6232	1.1033	1.0504	0.2656	0.6425	1.1281	1.0621	0.2848	0.6603	1.2363	1.1119	0.2753
STID	0.6221	1.0877	1.0386	0.2867	0.6235	1.0973	1.0587	0.2801	**0.6233**	1.1135	1.0667	0.2761
MICN	0.6284	1.1074	1.0983	**0.2616**	0.6263	**1.0504**	1.0702	0.2712	0.6361	1.1131	1.1011	0.2668
Autoformer	0.6863	1.1126	1.0548	0.3383	0.6897	1.1150	1.0560	0.3388	0.6853	1.1548	1.0746	0.3553
LSTM	0.6889	1.2192	1.1042	0.3032	0.7155	1.3106	1.1448	0.3150	0.7380	1.2287	1.1085	0.3642
HA	0.8689	1.2981	1.1394	0.5146	0.8709	1.1416	1.1416	0.5159	0.8695	1.2988	1.1397	0.5144
ours	**0.6177**	**1.0693**	**1.0340**	0.2647	**0.6235**	1.0934	**1.0456**	**0.2669**	0.6238	**1.1098**	**1.0535**	**0.2644**

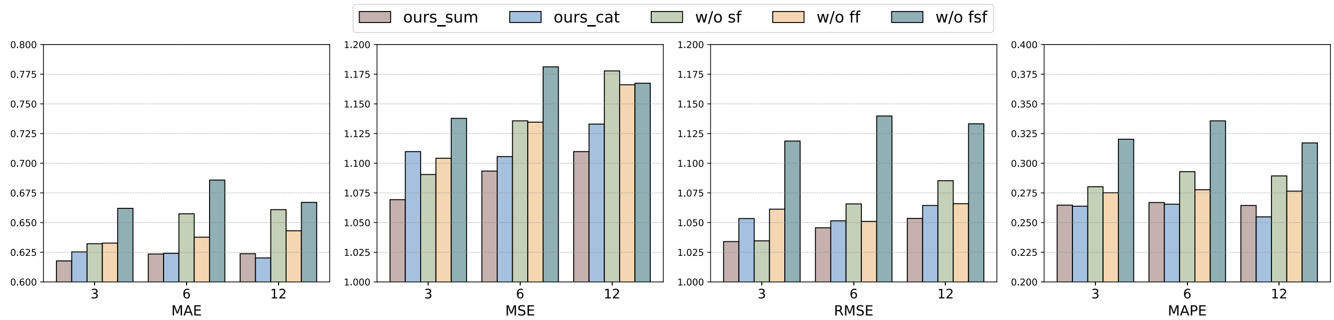

FIGURE 3. Ablation Study. sum: feature summation for FSFM; cat: feature concatenation for FSFM; sf: score fusion; ff: feature fusion; fsf: score and feature fusion.

original and predicted values, MAPE is calculated using de-normalized values.

2) BASELINE MODELS

This subsection details the implementation of several baseline models, namely STEP [34], STID [38], MICN [36], Autoformer [39], LSTM [40] and HA. It's important to note that we evaluate each deep learning baseline model alongside our proposed method using a standardized benchmark, BasicTS+. We maintain consistency with the experimental configurations of the baseline models by utilizing their official core code and hyperparameters for conducting experiments, with minor adjustments made only to certain settings such as learning rate and batch size.

STEP: STEP enhances traditional STGNNs by integrating a pre-training model that learns from extended historical data, significantly improving long-term dependency modeling and forecasting accuracy in multivariate time series.

STID: STID simplifies multivariate time series forecasting by embedding spatial and temporal identities into MLPs, enhancing accuracy and efficiency effectively.

MICN: MICN combines local features and global correlations to capture an overall view of the time series data by employing a multi-scale branch structure to model different latent patterns.

Autoformer: utoformer revolutionizes long-term time series forecasting with its novel decomposition architecture and Auto-Correlation mechanism, designed to effectively handle complex temporal patterns.

LSTM: LSTM networks utilize gates to manage information flow, adeptly handling long-term dependencies in time series data, thus excelling in contexts where historical context is key for prediction.

HA: The Historical Average method, while simple, efficiently forecasts future values by averaging past data, proving especially effective in stable trend scenarios.

C. RESULTS AND ANALYSIS

We present the comparative results in Table 1. From Table 1, we can see that our proposed method outperforms the baseline models across multiple evaluation metrics. Specifically, our method exhibits the best prediction performance in terms of MAE, RMSE, and MAPE, across various prediction horizons, when compared to other models. For the MSE metric, only for Horizon 6, which is a 3-hour prediction, MICN achieves the best prediction performance, but outperforming our model by merely 4%.

The significant improvement can be attributed to the comprehensive nature of our model. By integrating spatial, temporal, seasonal, and trend features, our model captures a more nuanced understanding of the underlying data dynamics. This holistic approach enables more accurate predictions, especially in scenarios where factors like seasonality and trends play a crucial role. The enhanced predictive capabilities of our model has the potential to contribute significantly to power-grid companies. By providing a robust forecasting model, we contribute to the development of effective electricity distribution strategies. This, in turn, facilitates optimized resource allocation and operational cost reduction for power companies, making our model a valuable asset for the decision-making process in the energy sector.

D. ABLATION STUDY

In order to thoroughly assess the effectiveness of our proposed method, we conducted an ablation study in which we systematically removed key components and evaluated their impact on predictive performance. The ablation study results are presented in Fig. 3.

From the figure, we can see that each component plays a crucial role in enhancing the predictive accuracy of our hybrid deep learning-based model. We evaluate the difference of using summation or concatenation as the aggregation method inside GNN, and find that in most cases, summation achieves the better results, as indicated in 'ours_sum' and 'ours_cat' comparison in Fig. 3. Furthermore, we studied the effect of ignoring feature fusion (w/o ff), ignoring score fusion (w/o sf), ignoring both feature fusion and score fusion (w/o fsf) by merely relying the prediction results from Transformer+GNN, and report the respective performance in the figure. We can see that by integrating all the components together, we achieve the best performance in terms of the four evaluation metrics. Overall, the results of the ablation study reaffirm the effectiveness and robustness of our proposed hybrid deep learning-based model. By leveraging a combination of spatial, temporal, seasonal, and trend features, our model achieves superior predictive performance, making it a valuable tool for accurate electricity demand forecasting for DanceSport activities.

VI. CONCLUSION AND FUTURE WORK

This paper presents a hybrid deep learning-based model for short-term electricity consumption prediction of dance sport activities. The model utilizes multi-head attention for temporal feature extraction, and Graph neural networks for spatial feature distillation. Additionally, time-series data decomposition is used to explicitly extract the seasonal and trend feature of dance sports. With the spatial-temporal, seasonal and trend features, the hybrid model is able to yield more accurate electricity demands than canonical methods.

In the future, we plan to assess our model's robustness with disrupted electricity data and incorporate its results into power companies' strategies for improved residential electricity load balancing. Moreover, ongoing model refinement will include exploring new feature extraction techniques and integrating data from IoT devices to enhance predictive accuracy and reliability [41], with the goal of developing smarter and more adaptable forecasting models for sustainable energy management systems.

REFERENCES

[1] S. Dai, H. Xue, M. D. Giardina, and J. I. Newman, "Hybrid choreographic femininities: DanceSport, physical culture, and performative corporeality in contemporary China," *Qualitative Res. Sport, Exerc. Health*, vol. 16, no. 4, pp. 368–386, Jul. 2024.

[2] G. Tomescu, M.-I. Stănescu, M. Manos, L. Dina, and K.-A. Aivaz, "Using dancesport as an educational resource for improving institutionalized children's learning strategies," *Children*, vol. 10, no. 6, p. 1039, Jun. 2023.

[3] Z. G. Duan, J. Monseau, Y. He, and J. G. Qiu, "From cultural import to flourishing sport: A comprehensive history of dancesport in China (1864–2023)," *Int. J. Hist. Sport*, vol. 41, nos. 2–3, pp. 185–209, Feb. 2024.

[4] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019.

[5] M. Razghandi and D. Turgut, "Residential appliance-level load forecasting with deep learning," in *Proc. IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.

[6] H. Zhao, Y. Wu, L. Ma, and S. Pan, "Spatial and temporal attention-enabled transformer network for multivariate short-term residential load forecasting," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.

[7] S. S. Rangapuram, L. D. Werner, K. Benidis, P. Mercado, J. Gasthaus, and T. Januschowski, "End-to-end learning of coherent probabilistic forecasts for hierarchical time series," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 8832–8843.

[8] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. AAAI Conf. Artif. Intell.*, May 2021, vol. 35, no. 12, pp. 11106–11115.

[9] S. Liu, H. Yu, C. Liao, J. Li, W. Lin, A. X. Liu, and S. Dustdar, "Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–18.

[10] J.-F. Chen, W.-M. Wang, and C.-M. Huang, "Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting," *Electr. Power Syst. Res.*, vol. 34, no. 3, pp. 187–196, 1995.

[11] S.-J. Huang and K.-R. Shih, "Short-term load forecasting via ARMA model identification including non-Gaussian process considerations," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 673–679, May 2003.

[12] C.-M. Huang, C.-J. Huang, and M.-L. Wang, "A particle swarm optimization to identifying the ARMAX model for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 1126–1133, May 2005.

[13] B. A. Høverstad, A. Tidemann, H. Langseth, and P. Öztürk, "Short-term load forecasting with seasonal decomposition using evolution for parameter tuning," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1904–1913, Jul. 2015.

[14] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3943–3952, Jul. 2019.

[15] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—A novel pooling deep RNN," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Sep. 2018.

[16] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.

[17] H. J. Sadaei, P. C. de Lima e Silva, F. G. Guimarães, and M. H. Lee, "Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series," *Energy*, vol. 175, pp. 365–377, May 2019.

[18] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, and C. Zhang, "Connecting the dots: Multivariate time series forecasting with graph neural networks," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, Sep. 2020, pp. 753–763.

[19] Z. Yue, Y. Wang, J. Duan, T. Yang, C. Huang, Y. Tong, and B. Xu, "Ts2vec: Towards universal representation of time series," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 8, 2022, pp. 8980–8987.

[20] M. Aouad, H. Hajj, K. Shaban, R. A. Jabr, and W. El-Hajj, "A CNN-sequence-to-sequence network with attention for residential short-term load forecasting," *Electric Power Syst. Res.*, vol. 211, Oct. 2022, Art. no. 108152.

[21] A. Kedia, M. A. Zaidi, and H. Lee, "FiE: Building a global probability space by leveraging early fusion in encoder for open-domain question answering," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2022, pp. 4246–4260.

[22] M. Abdar, M. A. Fahami, S. Chakrabarti, A. Khosravi, P. Pławiak, U. R. Acharya, R. Tadeusiewicz, and S. Nahavandi, "BARF: A new direct and cross-based binary residual feature fusion with uncertainty-aware module for medical image classification," *Inf. Sci.*, vol. 577, pp. 353–378, Oct. 2021.

[23] L. Sun, Z. Lian, J. Tao, B. Liu, and M. Niu, "Multi-modal continuous dimensional emotion recognition using recurrent neural network and self-attention mechanism," in *Proc. Int. Conf. Multimodal Sentiment Anal. Real-Life Media Challenge Workshop*, 2020, pp. 27–34.

[24] Y. Yao and R. Mihalcea, "Modality-specific learning rates for effective multimodal additive late-fusion," in *Proc. Findings Assoc. Comput. Linguistics (ACL)*. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 1824–1834.
[25] C. Qiu, W. Wu, X. F. Zhang, and L. Wang, "Late fusion with triplet margin objective for multimodal ideology prediction and analysis," in *Proc. Conf. Empir. Methods Nat. Lang. Process.*, 2022, pp. 9720–9736.
[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. NIPS*, 2017, pp. 6000–6010.
[27] S. Yun, M. Jeong, R. Kim, J. Kang, and H. J. Kim, "Graph transformer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–15.
[28] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–12.
[29] J. Xu, H. Wu, J. Wang, and M. Long, "Anomaly transformer: Time series anomaly detection with association discrepancy," 2021, *arXiv:2110.02642*.
[30] C. Gao, Y. Zheng, N. Li, Y. Li, Y. Qin, J. Piao, Y. Quan, J. Chang, D. Jin, X. He, and Y. Li, "A survey of graph neural networks for recommender systems: Challenges, methods, and directions," *ACM Trans. Recommender Syst.*, vol. 1, no. 1, pp. 1–51, Mar. 2023.
[31] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent.*, 2017, pp. 1–10.
[32] Z. Qin, Y. Chen, G. Zhu, E. Zhou, Y. Zhou, Y. Zhou, and C. Zhu, "Enhanced pseudo-label generation with self-supervised training for weakly-supervised semantic segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, early access, 2024.
[33] X. Lin, Y. Zhou, Y. Liu, and C. Zhu, "A comprehensive review of image line segment detection and description: Taxonomies, comparisons, and challenges," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, 2024.
[34] Z. Shao, Z. Zhang, F. Wang, and Y. Xu, "Pre-training enhanced spatial–temporal graph neural network for multivariate time series forecasting," in *Proc. 28th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2022, pp. 1567–1577.
[35] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph WaveNet for deep spatial–temporal graph modeling," 2019, *arXiv:1906.00121*.
[36] H. Wang, J. Peng, F. Huang, J. Wang, J. Chen, and Y. Xiao, "MICN: Multi-scale local and global context modeling for long-term series forecasting," in *Proc. Int. Conf. Learn. Represent.*, 2022, pp. 1–16.
[37] S. Makonin, B. Ellert, I. V. Bajić, and F. Popowich, "Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014," *Scientific Data*, vol. 3, no. 1, pp. 1–12, Jun. 2016.
[38] Z. Shao, Z. Zhang, F. Wang, W. Wei, and Y. Xu, "Spatial–temporal identity: A simple yet effective baseline for multivariate time series forecasting," in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2022, pp. 4454–4458.
[39] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 1–16.
[40] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[41] H. Guo, Z. He, W. Sheng, Z. Cao, Y. Zhou, and W. Gao, "SEGAC: Sample efficient generalized actor critic for the stochastic on-time arrival problem," *IEEE Trans. Intell. Transp. Syst.*, early access, 2024.

KEYIN LIU received the master's degree from Chengdu Sport University, in 2015. She is currently pursuing the Ph.D. degree with the International College, Krirk University, majoring in music and dancology. She is a Lecturer with Chengdu Sport University. Her research interests include DanceSport education and choreography, as well as the application of technology in physical activities.

HAO LI received the bachelor's degree from the School of Computer Science, Shaanxi Normal University (SNNU), in 2020. He is currently pursuing the master's degree in computer science and technology with Sichuan University (SCU). His research interests include distributional reinforcement learning for reliable shortest path and reinforcement learning for safe autonomous navigation in dynamic environments.

SONG YANG received the Ph.D. degree in electronic science and technology from Beijing University of Posts and Telecommunications, China, in 2014. He is currently the CTO of the Cloud Security Division, Huawei Technologies Company Ltd., China. He is also a Graduate Supervisor with Chongqing University of Posts and Telecommunications, China. His current research interests include network security and cloud security.