

一种融合稀疏注意力机制和BlockDiagonal矩阵的xLSTM-Informer深度学习模型

王鹏超, 黎锁平*, 周永强

兰州理工大学理学院, 甘肃 兰州

收稿日期: 2026年3月20日; 录用日期: 2026年4月30日; 发布日期: 2026年5月13日

摘要

在新能源高渗透率背景下, 电力现货价格频繁波动, 传统模型在处理这类非线性、非平稳时序数据时存在性能瓶颈。本文提出xLSTM-Informer模型, 融合xLSTM、稀疏注意力机制(ProbSparse Attention)及BlockDiagonal结构压缩方法, 构建出适应强波动市场特征的预测架构。其中xLSTM顺序堆叠了一个xLSTM乘性门控单元(mLSTM)和一个结构化LSTM (sLSTM)模块, 并且将BlockDiagonal结构压缩方法运用到mLSTM中, 更好地提取时序局部特征的注意力, 旨在处理长序列建模时信息传递受限、参数冗余以及时序特征表达能力不足等问题。模型在输入数据归一化过程中联合采用层归一化与组归一化策略, 增强对尺度异构数据的鲁棒性。实验采用甘肃省2024年上半年电力现货市场数据, 在多种天气场景(晴天、大风、阴天、无风)下进行测试。结果表明, xLSTM-Informer在五项主要评价指标(RMSE, MAE, MAPE, TIC, R^2)上均优于CNN、CNN-LSTM、Informer和iTransformer等基准模型。其中在典型晴天场景下, RMSE仅为8.863, MAPE为4.556%, R^2 高达0.987, 展现出对极端市场变化的强适应能力和稳定性。

关键词

xLSTM-Informer, 稀疏注意力机制, BlockDiagonal, 电力现货价格

A Deep Learning Model of xLSTM-Informer Fusing Sparse Attention Mechanism and BlockDiagonal Matrices

Pengchao Wang, Suoping Li*, Yongqiang Zhou

School of Science, Lanzhou University of Technology, Lanzhou Gansu

Received: March 20, 2026; accepted: April 30, 2026; published: May 13, 2026

*通讯作者。

文章引用: 王鹏超, 黎锁平, 周永强. 一种融合稀疏注意力机制和 BlockDiagonal 矩阵的 xLSTM-Informer 深度学习模型[J]. 人工智能与机器人研究, 2026, 15(3): 685-700. DOI: 10.12677/airr.2026.153065

Abstract

Against the backdrop of high new energy penetration, electricity spot prices exhibit frequent fluctuations, and traditional models suffer from performance bottlenecks when modeling such nonlinear and non-stationary time series. This paper proposes the xLSTM-Informer model, which integrates xLSTM, ProbSparse Attention, and a BlockDiagonal matrix compression method to construct a prediction framework adapted to highly volatile market characteristics. Specifically, xLSTM sequentially stacks a multiplicative LSTM (mLSTM) unit and a structured LSTM (sLSTM) module, with the BlockDiagonal compression method incorporated into the mLSTM to better capture attention over local time-series features. This design addresses the limitations of long-sequence modeling, including constrained information propagation, parameter redundancy, and insufficient representation of temporal patterns. To enhance robustness against scale-heterogeneous data, the model combines layer normalization and group normalization during input data standardization. Experiments are conducted using electricity spot market data from Gansu Province in the first half of 2024, evaluated across multiple weather scenarios (sunny, windy, cloudy, calm). The results demonstrate that xLSTM-Informer outperforms baseline models such as CNN, CNN-LSTM, Informer, and iTransformer across five key evaluation metrics (RMSE, MAE, MAPE, TIC, R^2). In the typical sunny scenario, the model achieves an RMSE of 8.863, MAPE of 4.556%, and R^2 of 0.987, verifying its strong adaptability and stability under extreme market variations.

Keywords

xLSTM-Informer, Sparse Attention, BlockDiagonal, Electricity Spot Price

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

电力现货价格的演变受到天气、负荷、发电结构及政策信号等多源异构因素影响，呈现出显著的非线性、非平稳与突变性特征[1][2]。传统统计模型受限于线性假设与静态机制，难以应对复杂时序依赖关系[3]。在此背景下，深度学习被广泛引入电力市场预测中。例如，LSTM(长短期记忆网络)[4]、RNN(循环神经网络)、CNN(卷积神经网络)等模型在电力负荷预测和电价预测中均表现出良好性能[5]-[8]。然而，LSTM受限于串行计算结构，难以处理超长输入序列；而Transformer(转换器)虽具有全局建模能力，但其 $O(n^2)$ 的注意力计算复杂度限制了实际应用的规模扩展性[7][8]。

已有研究尝试通过多种改进路径提升预测能力。例如，基于深度学习的电力价格预测已在国内外得到广泛探讨[9]，部分研究进一步结合注意力机制[10]或混合结构模型[7][8]来增强对非线性与非平稳性的刻画能力。但总体来看，现有方法仍存在对长序列建模不足、对局部突变响应不敏感以及计算效率偏低等瓶颈。

针对上述问题，本文提出一种融合多种结构优势的深度时序预测架构——xLSTM-Informer(长短期稀疏注意力预测模型)，旨在实现对电力现货价格的高精度、强鲁棒预测。该模型整合了多尺度因果卷积、稀疏注意力机制、双路径门控记忆结构(mLSTM与sLSTM)、BlockDiagonal参数压缩以及层归一化与组归一化联合策略，构建了一个兼顾全局依赖建模与局部突变感知的混合架构。本文的研究工作基于对现有文献的系统分析与模型创新，为新能源高渗透率背景下的电力市场预测提供新的方法论支持。

2. 相关工作

近年来,随着新能源渗透率不断提升,电力现货价格预测逐渐成为能源系统智能化中的核心任务之一,研究者们围绕建模结构、特征融合与时序学习等方向提出了多种解决方案。

在模型结构方面,文献[6]采用卷积神经网络(CNN)提取电价序列的局部特征,增强了模型对短时波动的响应能力。然而,CNN固有的局部感受野限制了其对全局趋势和长期依赖的建模能力。为弥补这一缺陷,文献[11]与[12]中采用长短期记忆网络(LSTM),利用其门控机制建模价格序列中的长期依赖关系。该方法在处理周期性和季节性数据上表现稳定,但在超长时间序列下仍面临计算效率低与训练资源消耗高的问题。

自注意力机制的引入显著拓展了时序建模能力。文献[13]提出基于 Transformer 的电价多步预测模型,通过全局注意力机制捕捉关键时刻的高阶依赖,提高了对异常点与价格拐点的识别能力。Transformer 在建模能力上表现优异,但其自身 $O(n^2)$ 的计算复杂度使其在长序列场景中不具备部署优势[14]。

为解决 Transformer 在长序列上的计算瓶颈,文献[15]提出了 Informer 架构,采用概率稀疏注意力机制(ProbSparse Attention),在保留建模能力的同时将计算复杂度降至 $O(n \log n)$,该模型在多个时间序列任务中表现出良好的精度与效率平衡。文献[16]进一步将 Informer 用于电力负荷预测,验证了其在电力时序建模中的可行性。

除了建模架构的优化,数据归一化和特征融合策略也是提升预测鲁棒性的关键。文献[17]指出,在高波动的新能源市场中,结合层归一化与组归一化能够有效提升模型的训练稳定性与泛化能力,尤其适用于多维输入变量存在尺度异质性的情境。

在结构压缩与部署效率上,文献[14]提出了块对角参数化策略,用于压缩 Transformer 中的全连接层权重矩阵,有效减少模型参数量和显存占用。这一方法与本文提出的 BlockDiagonal 机制在设计理念上高度一致,均致力于在保证建模能力的同时降低部署资源门槛。

综上所述,大多数模型在面对新能源出力突变、价格剧烈波动等极端情形时仍存在适应性差、响应滞后等问题。因此,我们提出融合注意力机制、门控结构与归一化策略的 xLSTM-Informer 模型。

3. 系统模型

3.1. xLSTM-Informer 模型架构

为应对电力现货市场中价格剧烈波动、局部突变频发和长期依赖建模难等挑战,本文提出了一种新型深度学习模型——xLSTM-Informer。如图 1 所示,我们的深度学习模型第一个阶段利用 xLSTM 模块,它在捕捉序列中的短期动态信息的同时,增强了模型对多维输入之间结构化关系的建模能力。此外,xLSTM 的门控更新机制在保留历史记忆与非线性表达的基础上,有效减轻了梯度消失与爆炸的问题,提升了对长序列依赖的表达效率。第二阶段利用融合稀疏注意力机制的 Informer 模块,大幅降低了时间复杂度,并通过解码器架构对未来时间步进行精准预测。并且这一阶段在处理超长时间序列时展现出良好的扩展性和稳定性。

3.2. 改进的 xLSTM 架构

为提升对多变量时间序列建模的表达能力和泛化能力,本文改进后的 xLSTM 的架构如图 2 所示,该结构顺序堆叠了一个 mLSTM 模块和一个 sLSTM 模块,旨在处理长序列建模时信息传递受限、参数冗余以及时序特征表达能力不足等问题。传统的 RNN 或标准 LSTM 往往在建模长程依赖和跨通道交互方面存在记忆退化和表示瓶颈,而 Transformer 类模型虽然具有优秀的全局建模能力,但在小数据或稀疏信号

下容易过拟合。xLSTM 模块融合了局部卷积感知、门控状态更新与注意力引导的表达机制，并借助参数共享的 BlockDiagonal 结构，在多个子空间中实现高效信息变换，并且兼顾时序数据的短期细节建模与长期依赖捕捉。

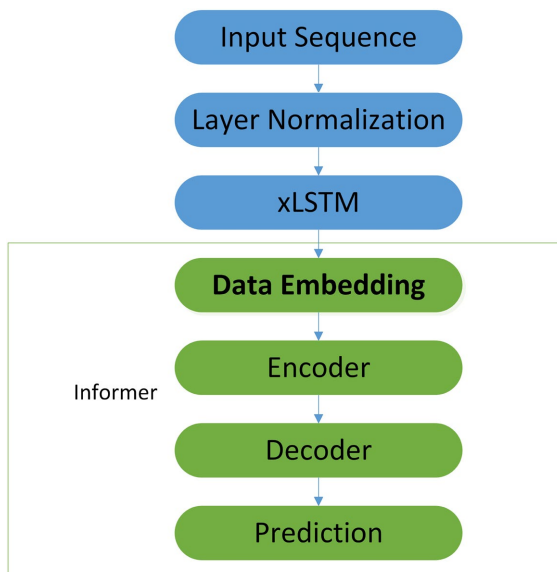


Figure 1. xLSTM-Informer model

图 1. xLSTM-Informer 模型

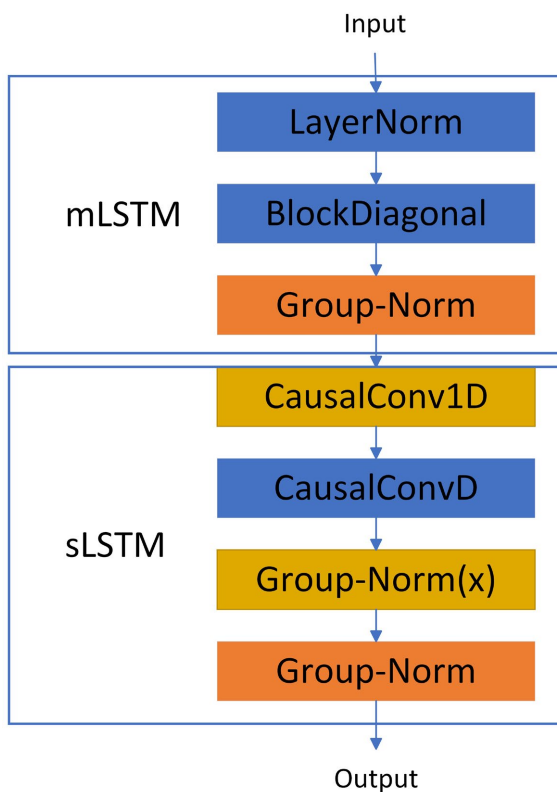


Figure 2. Architecture of the improved xLSTM

图 2. 改进的 xLSTM 架构

3.2.1. mLSTM 子模块

设在 xLSTM-Informer 编码器阶段中输入的多变量时间序列为 $X = [x_1, x_2, \dots, x_T]$, $x_t \in \mathbb{R}^D$, 其中 T 表示时间步的长度, D 表示每个时间步的特征维度, 变量 x_t 是第 t 个时间点的输入向量。

接下来对每个时间步进行归一化与投影处理, 可消除特征间分布偏移, 增强训练稳定性:

$$x_t^{(1)} = \text{LayerNorm}(x_t), \quad x_t^{up} = W_{up} x_t^{(1)} \in \mathbb{R}^{D'} \quad (1)$$

其中 $x_t^{(1)}$ 是归一化后的输入, W_{up} 是左投影权重矩阵, x_t^{up} 是投影后的高维表示。

为了保持时间方向的一致性, 引入一维因果卷积:

$$x_t^{conv} = \text{CausalConv1D}(x_t^{up}) \quad (2)$$

其中 x_t^{conv} 是卷积后的时序局部特征。

3.2.2. 基于注意力的 BlockDiagonal 映射

为了提取时序局部特征的注意力表示, 通过 block-wise 投影生成注意力分量:

$$q_t = \text{BlockDiag}_q(x_t^{conv}) \quad (3)$$

$$k_t = \text{BlockDiag}_k(x_t^{conv}) \quad (4)$$

$$v_t = \text{BlockDiag}_v(x_t^{up}) \quad (5)$$

其中 q_t 是 query 向量, 用于从矩阵记忆中检索信息; k_t 是 key 向量, 经过缩放的线性变换结果, 用于计算记忆的相关性; v_t 是 value 向量, 代表待存储的信息。它们基于 BlockDiagonal 的子空间线性映射形成多头表示能力, BlockDiagonal 算法过程如表 1 所示。

Table 1. Algorithm of BlockDiagonal forward propagation

表 1. BlockDiagonal 前向传播过程算法

算法 1 BlockDiagonal 前向传播过程
输入: 特征张量 $x \in \mathbb{R}^{B \times D}$;
输出: y
模块初始化:
将输出维度 out_features 均分为 num_blocks 组;
每组输出维度为 block_out_features=out_features/num_blocks。
For: $i = 1, \dots, \text{num_blocks}$:
执行第 i 个子变换 $x_i = \text{Linear}(x)$;
将所有 x_i 在最后一维上拼接:
$y = \text{cat}([x_1, x_2, \dots, x_{\text{num_blocks}}], \text{dim} = -1)$ 。

为控制状态更新的比例, 引入指数形式的门控函数:

细胞更新状态公式:

$$C_t = f_t C_{t-1} + i_t v_t k_t \quad (6)$$

$$n_t = f_t n_{t-1} + i_t k_t \quad (7)$$

$$h_t = o_t \odot \tilde{h}_t, \quad \tilde{h}_t = g(C_t, q_t, n_t) = \frac{C_t q_t}{\max\{n_t^\top q_t, 1\}} \quad (8)$$

输入门公式:

$$i_t = \exp(\tilde{i}_t), \tilde{i}_t = w_i x_t^{\text{conv}} + b_i \quad (9)$$

遗忘门公式:

$$f_t = \sigma(\tilde{f}_t) \text{ OR } \exp(\tilde{f}_t), \tilde{f}_t = w_f x_t^{\text{conv}} + b_f \quad (10)$$

输出门公式:

$$o_t = \sigma(\tilde{o}_t), \tilde{o}_t = w_o x_t^{\text{up}} + b_o \quad (11)$$

其中 W_q 、 W_k 、 W_v 分别为查询向量、键向量、值向量的权重矩阵; ω_i 、 ω_f 、 ω_o 分别为输入门、遗忘门、输出门的权重向量; b_q 、 b_k 、 b_v 、 b_i 、 b_f 、 b_o 是偏置项; \odot 为逐元素相乘操作, 用于输入门与检索结果的结合。

输出通过 GroupNorm 正则化, 并融合跳跃连接:

$$\tilde{h}_t^{\text{mLSTM}} = \text{GroupNorm}(h_t W_{\text{skip}} x_t^{\text{conv}}) \quad (12)$$

其中 W_{skip} 是权重矩阵, $\tilde{h}_t^{\text{mLSTM}}$ 是 mLSTM 模块的最终输出表示。其中 GroupNorm 算法设计见表 2。

Table 2. Forward propagation of GroupNorm

表 2. GroupNorm 前向传播过程

算法 2 GroupNorm 前向传播过程
输入 $h_t \in \mathbb{R}^{B \times D}$;
对最后一维重塑为 $\mathbb{R}^{B \times N \times M}$, 其中 N 是分组数, 需满足 $D = N \times M$;
计算组内的特征均值 μ 和特征方差 σ^2 ;
执行标准化操作:
$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}}$
为每组引入可训练的缩放参数 γ_i 和偏移参数 β_i ;
执行缩放和平移操作:
$x = \hat{x} \cdot \gamma_i + \beta_i$
返回输出 x 。

3.2.3. sLSTM 子模块

LayerNorm 与 CausalConv 将 mLSTM 输出归一化并卷积以提取时序模式:

$$x_t^{(2)} = \text{LayerNorm}(\tilde{h}_t^{\text{mLSTM}}) \quad (13)$$

$$x_t^{\text{conv}} = \text{CausalConv1D}(x_t^{(2)}) \quad (14)$$

sLSTM 引入结构化门控机制, 通过选择性更新单元状态, 增强对输入中短期动态的建模能力。其计算模型如公式(15)~(22)所示:

细胞更新状态公式:

$$m_t = \max(f_t + m_{t-1}, i_t) \quad (15)$$

$$c_t = f_t c_{t-1} + i_t z_t \quad (16)$$

$$z_t = \varphi\left(W_z^T x_t^{(2)} + r_z h_{t-1} + b_z\right) \quad (17)$$

$$n_t = f_t n_{t-1} + i_t \quad (18)$$

$$h_t = o_t \tilde{h}_t, \tilde{h}_t = c_t / n_t \quad (19)$$

输入门公式:

$$i_t = \exp\left(W_i^T x_t^{\text{conv}} + r_i h_{t-1} + b_i\right) \quad (20)$$

遗忘门公式:

$$f_t = \exp\left(W_f^T x_t^{\text{conv}} + r_f h_{t-1} + b_f\right) \quad (21)$$

输出门公式:

$$o_t = \sigma\left(W_o^T x_t^{(2)} + r_o h_{t-1} + b_o\right) \quad (22)$$

其中 m_t 是门控归一化的最大激活, c_t 是时刻 t 的细胞状态, 存储长期记忆; f_t 为遗忘门, 控制前一时刻 c_{t-1} 的保留程度; i_t 是输入门, 控制新信息流入细胞状态的程度; o_t 为输出门, 控制细胞状态 c_t 流向隐藏状态 h_t 的程度; n_t 是归一化状态, 用于稳定细胞状态更新; h_t 是时刻 t 的隐藏状态, 用于循环连接和输出; W_z 、 W_i 、 W_f 、 W_o 分别为输入、输入门、遗忘门、输出门的权重矩阵; r_z 、 r_i 、 r_f 、 r_o 分别为输入、输入门、遗忘门、输出门的循环权重矩阵; b_z 、 b_i 、 b_f 、 b_o 是偏置项; σ 和 \exp 是激活函数。

3.2.4. 门控投影与残差输出

为了增强表示能力并引入非线性门控机制以捕捉更复杂的特征关系, 采用双投影与激活函数形成门控残差结构:

$$u_t = W_{\text{up}}^{(1)} h_t \quad (23)$$

$$v_t = \text{GELU}\left(W_{\text{up}}^{(2)} h_t\right) \quad (24)$$

$$r_t = W_{\text{down}}\left(u_t v_t\right) \quad (25)$$

$$\tilde{h}_t^{\text{final}} = r_t + x_t^{(2)} \quad (26)$$

其中 u_t 是对当前隐藏状态 h_t 进行线性投影的“信息部分”; v_t 是对 h_t 进行另一组线性变换并通过 GELU 激活后形成的“门控部分”; r_t 是通过下投影 W_{down} 还原到原维度; $\tilde{h}_t^{\text{final}}$ 是最终的输出, 含有残差连接。

3.3. 稀疏注意力机制

传统 Transformer 在处理长序列时面临注意力计算复杂度高 ($O(n^2)$) 的问题。xLSTM-Informer 采用 In-former 结构中提出的 ProbSparse Attention 机制, 显著降低了注意力计算开销至 $O(n \log n)$, 并通过重要性评分筛选关键时间节点, 有效过滤噪声干扰。

稀疏注意力机制的计算公式为:

$$\text{ProbSparseAttention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) * V \quad (27)$$

其中 QK^T 表示 Query 和 Key 的点积; $\sqrt{d_k}$ 为缩放因子, 防止数值过大导致 softmax 梯度消失。该机制在电力现货价格预测任务中表现出较强的语义选择性, 能够动态关注价格拐点、突变峰值等信息密集区域, 从而提升模型对极端市场情形的适应能力。

3.4. 归一化机制融合设计

归一化策略在提升深度时序模型在高波动场景下的训练稳定性方面具有关键作用。考虑到电力市场数据维度多、尺度差异大且波动剧烈，xLSTM-Informer 在不同阶段引入如下的两类归一化策略：

层归一化(Layer Normalization)：用于稳定跨时间步长的梯度传播过程，防止梯度爆炸与消失，表达如下：

$$\text{LayerNorm}(x) = \frac{x - \mu}{\sigma} \quad (28)$$

其中 μ 和 σ 分别是输入数据的均值和标准差。

组归一化(Group Normalization)：通过对通道分组归一化处理，提高模型对不同变量间尺度异质性的鲁棒性，表达式如下：

$$\text{GroupNorm}(x) = \frac{x - \mu_g}{\sigma_g} \quad (29)$$

其中 μ_g 和 σ_g 是对输入数据按组进行归一化时的均值和标准差。

xLSTM-Informer 模型融合了门控注意力与归一化机制，既提升了模型在高频扰动下的稳定性，也增强了跨步长、跨模态等复杂场景中的建模能力。例如，可通过外接图神经网络(GNN)引入空间拓扑特征，或结合自然语言嵌入器(如 BERT)处理政策文本类信息，从而提升模型的市场感知能力。

4. 数据与参数设置

为验证 xLSTM-Informer 模型在电力现货市场价格预测任务中的有效性，本文对数据预处理、模型训练、参数设置及评估指标等进行操作与分析。利用真实新能源历史电力市场数据进行建模训练，确保实验结果具备较强的工程应用参考价值。

4.1. 数据来源与描述

本研究选取我国西北地区甘肃省电力交易中心公开发布的现货市场数据作为研究对象，时间跨度为 2024 年 1 月 1 日至 2024 年 6 月 30 日，数据样本时间间隔为 15 分钟。表 3 展示了八类核心变量的描述性统计，其中电价单位是元/MWh，发电量单位是 MWh。由于我们的变量名太长，为了表格整体性的观感，这里我们用 $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$ 分别表示电价平均值、实时中长期参考价、火电发电量、光伏发电量、甘肃省用电量、风力发电量、水力发电量、电力现货价格。

Table 3. Descriptive statistics of eight core variables

表 3. 八类核心变量的描述性统计

	电价 平均值	实时中 长期参考价	火电 发电量	光伏 发电量	甘肃省 用电量	风力 发电量	水力 发电量	电力 现货价格
count	16,800	16,800	16,800	16,800	16,800	16,800	16,800	16,800
mean	232.7	239.7	11943.2	3794.8	19248.4	5195.9	3563.6	232.7
std	131.4	179.0	4368.7	4853.1	1250.7	3113.7	1718.1	203.3
min	40	39.99	4336.15	0	16340.13	62.55	883.68	40
25%	131.27	40.00	8360.80	8.72	18382.41	2685.11	2116.22	40
50%	200.89	240.00	11495.97	173.63	19230.02	4700.78	3183.59	180
75%	327.65	380.00	15350.73	8149.12	20119.64	7319.09	5110.37	400
max	561.77	650.00	21312.16	17467.30	22802.09	15650.72	7452.98	650

在数据划分方面, 将前 70% 的样本作为训练集, 后 20% 划为测试集, 剩余 10% 划为验证集。训练集中采用滑动窗口法生成样本, 每个样本包含过去 10 天的数据用于预测第 11 天的电价。

为了提高模型对高波动场景的识别能力, 训练过程中引入动态采样策略, 对价格波动率超过特定阈值的时间段提高采样概率。这种策略在电力负荷预测研究中已被证实可有效提升模型对极端情况的拟合能力。

4.2. 模型对比与消融实验设计

为全面评估所提出的 xLSTM-Informer 模型在电力现货价格预测任务中的有效性与优越性, 本文从模型对比实验与消融实验两个方面进行系统设计, 并结合多种评价指标对模型性能进行综合分析。

在模型对比实验方面, 选取以下典型预测模型作为对比对象:

- (1) CNN: 基于卷积神经网络的价格预测模型, 能够提取局部时间特征;
- (2) CNN-LSTM: 结合卷积网络与长短期记忆网络, 兼顾局部特征提取与时序依赖建模能力;
- (3) iTransformer: 针对时间序列任务优化的 Transformer 结构, 在特征维度建模方面具有优势。

上述模型涵盖了从传统深度学习结构到 Transformer 架构的多种代表性方法, 能够从不同角度反映所提模型的性能提升情况。

在消融实验方面, 为验证所提出模型中关键结构的有效性, 本文以原始 Informer 模型作为基准模型, 设计结构级消融实验。具体而言, 通过在 Informer 模型中引入 xLSTM 模块构建改进模型, 并在相同数据集与实验设置下对两种模型进行对比分析。该实验旨在评估 xLSTM 模块在增强时间序列长短期依赖建模能力方面的作用。通过比较两种模型在各项评价指标上的表现差异, 可以直观反映所提模型在预测精度与稳定性方面的改进效果。

4.3. 实验训练参数与模型评价指标

所有实验在 Intel Core i5 CPU 环境下进行, 软件平台为 PyTorch 3.6。表 4 展示了模型的主要参数设置。

Table 4. Parameter settings

表 4. 参数设置

输入序列长度	预测步长	批量大小	学习率	学习率调度	最大训练轮数	早停策略阈值
10	1	64	0.0001	patience = 25	100	15

为全面评估所提出的 xLSTM-Informer 模型在电力现货价格预测任务中的性能表现, 本文选取了五个常用的回归类评估指标: 均方根误差(RMSE)、平均绝对误差(MAE)、平均绝对百分比误差(MAPE)、泰尔不等系数(Theil Inequality Coefficient, TIC)以及决定系数(R²)。各指标涵盖了预测精度、误差幅度、相对偏差与拟合能力等多个维度, 能够从不同角度反映模型的预测质量。

(1) 均方根误差(RMSE): 衡量预测值与真实值之间整体误差的均方根, 反映预测的总体精度, 对较大误差更为敏感:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (30)$$

其中 \hat{y}_i 为第 i 个预测值, y_i 为对应的真实值, n 为样本数量。RMSE 越小, 说明模型整体误差越低。

(2) 平均绝对误差(MAE): 衡量的是预测值与实际值之间的平均绝对差值, 其对异常值不敏感, 具有

良好的可解释性:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (31)$$

MAE 适合衡量模型在一般场景下的稳定误差表现, 与 RMSE 可互为补充。

(3) 平均绝对百分比误差(MAPE): 用于衡量预测误差的相对偏差程度, 能更直观反映预测值相对于实际值的偏离百分比:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (32)$$

MAPE 对实际值为零的样本较为敏感, 因此适用于实际值不为零的预测任务。MAPE 越小, 说明模型相对误差越小, 具有更高的实用性。

(4) 泰尔不等系数(TIC): 综合考虑了预测值与真实值的趋势一致性, 适用于时间序列预测场景。其值区间为[0, 1], 越接近 0 表示预测效果越好:

$$\text{TIC} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n \hat{y}_i^2} + \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2}} \quad (33)$$

TIC 能够较好地评价模型是否捕捉到了电价走势的变化趋势, 尤其适用于周期性波动显著的场景。

(5) 决定系数(R²): 用于衡量预测结果与真实值之间的拟合优度, 其值区间为[0, 1], 越接近 1 表示模型越能解释目标变量的变异性:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2} \quad (34)$$

其中 \bar{y} 为实际值的均值。R² 越高, 说明模型拟合能力越强, 是衡量整体预测性能的核心指标之一。

上述五个评估指标各有侧重: RMSE 与 MAE 从误差绝对值出发, MAPE 强调误差的相对比例, TIC 衡量趋势拟合能力, 而 R² 则反映全局拟合优度。综合这些指标, 可以全面评估模型在精度、稳定性与趋势感知能力等方面的表现。

5. 实验结果与分析

图 3 展示了甘肃电力现货价格时间序列的特征分解, 包含四个关键组成部分: 原始数据(Raw Data)、趋势(Trend)、季节性(Seasonality)和残差(Residual), 时间跨度为 2024 年 1 月至 2024 年 6 月。

原始时间序列呈现显著的非平稳性和波动性, 价格范围在 400~650 元/MWh 之间。月内周期(约 30 天): 与新能源出力月度波动相关, 峰值出现在每月中旬(新能源大发期), 谷值在月末(电网调峰需求增加)。

趋势分量通过 STL 分解(Seasonal-Trend Decomposition Using LOESS)提取, 呈现先降后升的 U 型曲线。1~3 月趋势值从 285 元/MWh 下降至 235 元/MWh(降幅 17.5%), 4~6 月回升至 305 元/MWh(升幅 29.8%)。根据这些信息我们可以更好地处理电力现货价格的特征, 以便于预测。

如表 5 所示, xLSTM-informer 模型在 RMSE (36.959)和 MAE (9.312)指标上表现最优, 显著低于其他模型。xLSTM-informer 的 MAPE 仅为 7.349%, 远低于 iTransformer (55.514%)和 CNN (76.333%), xLSTM-informer 的 R² 值为 0.922, 接近 1。综合各项指标, xLSTM-informer 模型在预测精度、拟合优度和预测一

致性方面均表现出显著优势。

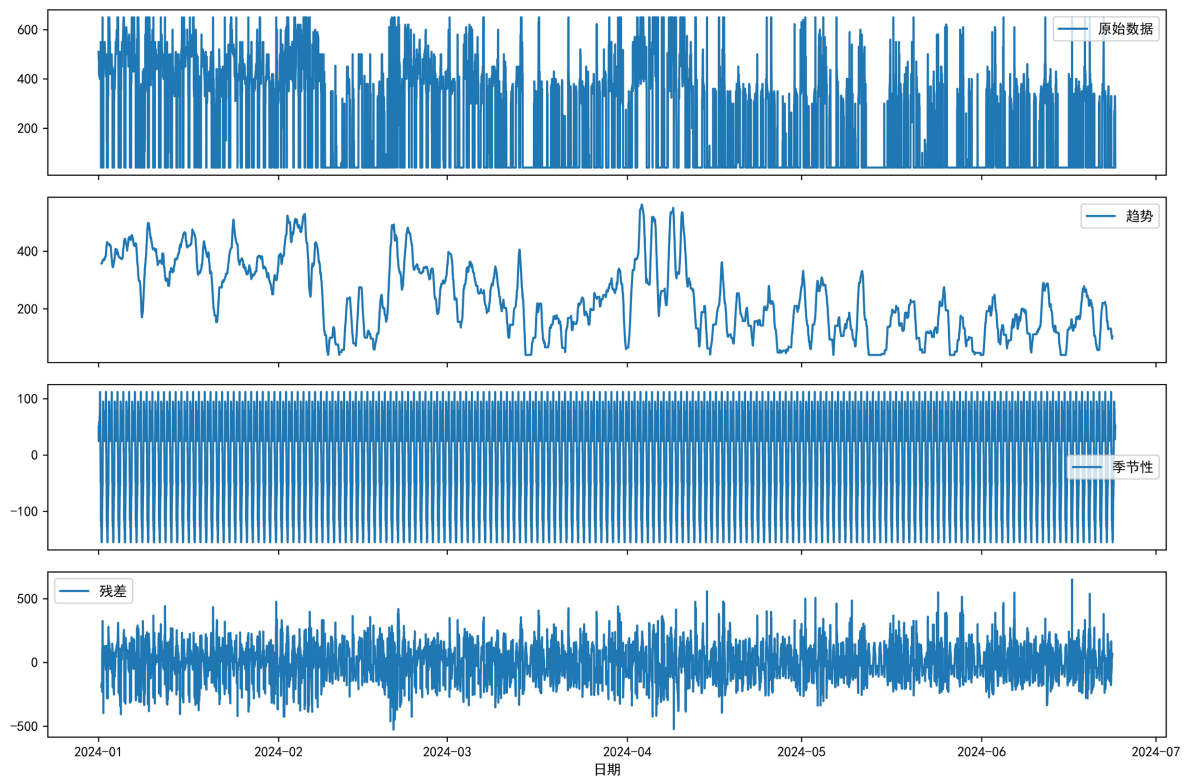


Figure 3. Time-series feature decomposition of electricity spot price

图 3. 电力现货价格时间序列特征分解

Table 5. Performance metrics of five models

表 5. 五种模型各类指标结果

指标	RMSE	MAE	MAPE (%)	TIC	R ²
xLSTM-informer	36.959	9.312	7.349	0.091	0.922
CNN	90.736	51.720	76.333	0.224	0.638
CNN-LSTM	70.538	30.054	36.805	0.175	0.781
iTransformer	91.805	43.531	55.514	0.227	0.629
Informer	54.408	25.554	32.815	0.135	0.870

为了能进一步证明本文模型的优势，我们提取了八个时间段的数据用于各个模型预测结果的指标对比，这八个时间段的数据根据天气情况分为四类：无风(6.8~6.9, 6.18~6.19)，大风(5.8~5.29, 6.25)、阴天(5.28~5.29, 6.20~6.21)、晴天(6.21~6.22, 6.22~6.23)。在这四种变动的天气下，发电量极其不稳定，使得价格波动明显，所以，如果我们的模型可以在这些情况下能够精准预测，那就说明我们模型的鲁棒性更好。

图 4 展示了五种模型在无风与大风两种天气条件下四个典型时段的预测效果，其中：(a)、(b)分别对应无风时段 6.8~6.9 和 6.18~6.19，(c)、(d)分别对应大风时段 5.28~5.29 和 6.25。图中给出了各模型在各时段的预测值与实际值的拟合曲线，用于直观对比其预测性能。表 6 对各模型在 RMSE、MAE、MAPE (%)、

TIC 和 R^2 等指标下的具体表现进行了进一步量化分析。

xLSTM-informer 在大风的两个时段均保持 $R^2 > 0.85$ 、 $TIC < 0.15$ ，且误差指标显著低于其他模型，证明其对强时序波动、非平稳数据的建模能力最佳，可能归因其独特的时序分解结构或自适应注意力机制。

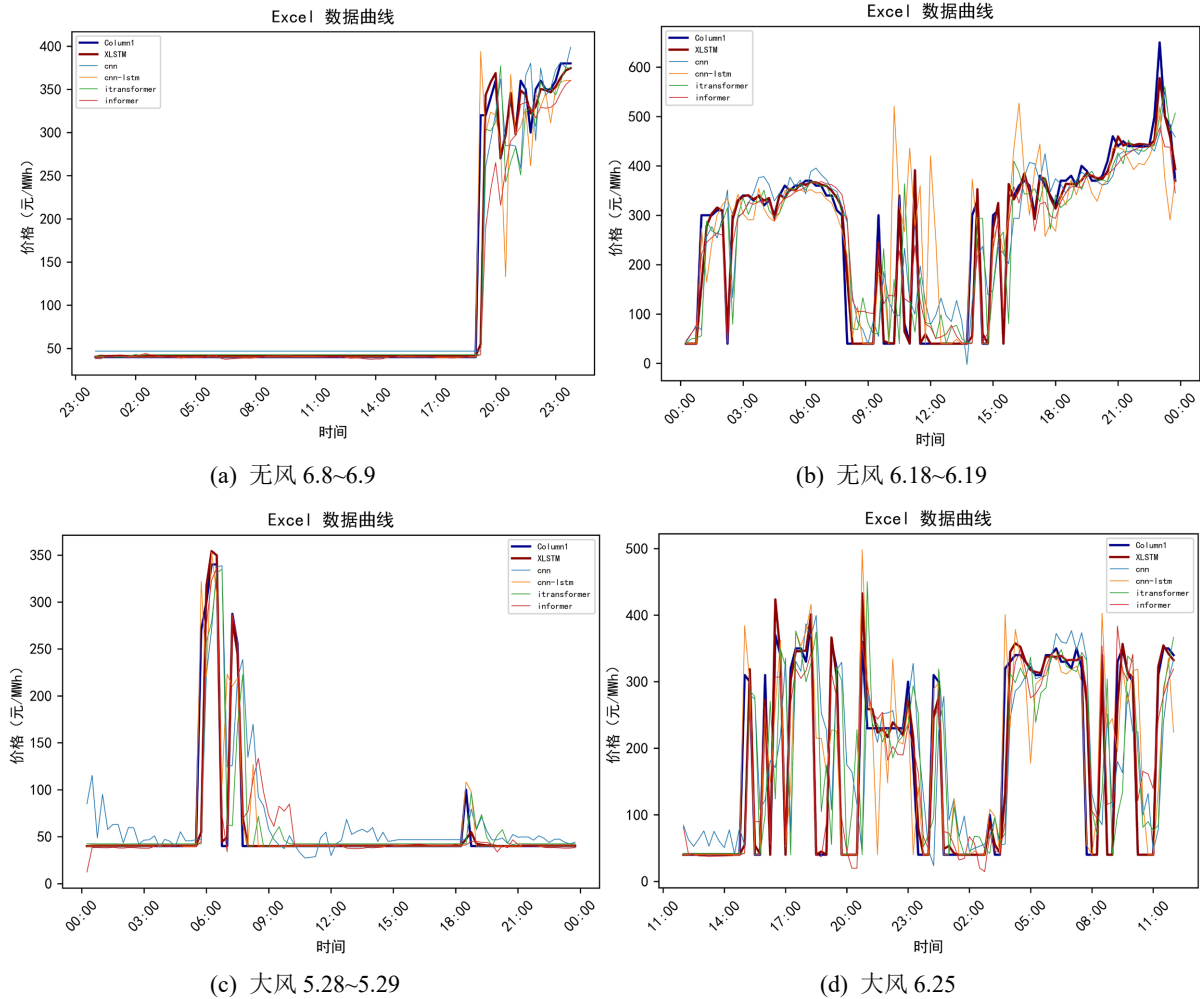


Figure 4. Fitting of predicted and actual values under windy and calm weather

图 4. 有风无风天气情况下预测值与实际值拟合

Table 6. Performance metrics of five models under windy and calm weather

表 6. 有风无风天气情况下五种模型各类指标结果

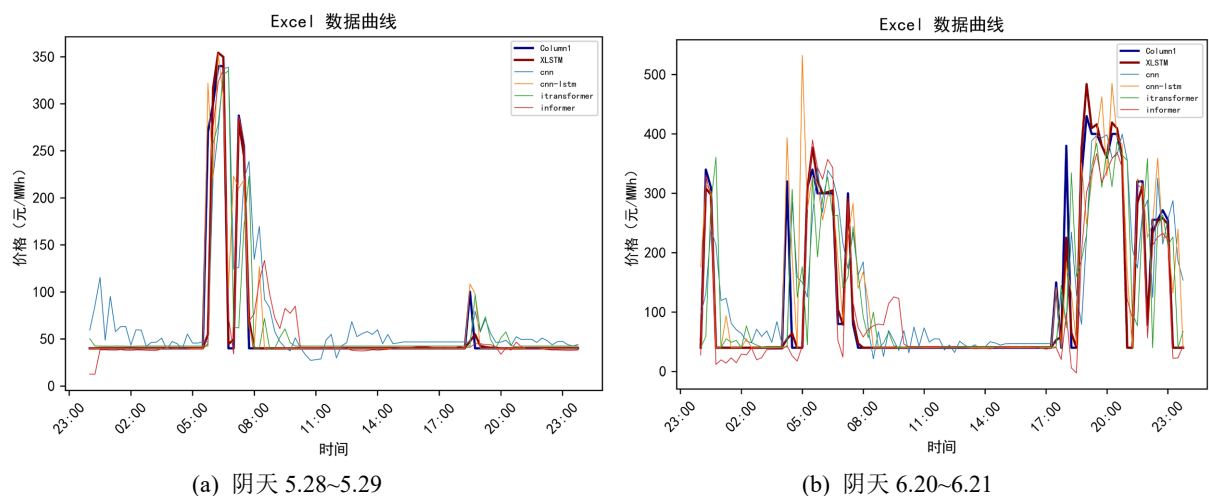
	指标	RMSE	MAE	MAPE (%)	TIC	R^2
无风 6.8~6.9	xLSTM-informer	26.655	4.809	2.085	0.084	0.954
	CNN	35.040	15.012	16.529	0.111	0.920
	CNN-LSTM	21.698	6.659	2.012	0.069	0.969
	iTransformer	35.001	11.441	7.761	0.113	0.920
	Informer	34.614	10.707	5.286	0.114	0.922

续表

无风 6.18~6.19	xLSTM-informer	32.201	13.080	8.584	0.052	0.958
	CNN	94.644	64.716	76.013	0.153	0.636
	CNN-LSTM	108.367	59.644	80.857	0.173	0.523
	iTransformer	104.535	63.192	68.242	0.170	0.556
	Informer	57.879	37.360	39.750	0.095	0.864
大风 5.28~5.29	xLSTM-informer	22.124	4.033	4.110	0.132	0.879
	CNN	54.206	25.970	51.174	0.309	0.274
	CNN-LSTM	26.489	7.523	12.087	0.155	0.827
	iTransformer	51.040	16.812	26.372	0.306	0.356
	Informer	31.263	10.730	19.993	0.184	0.758
大风 6.25	xLSTM-informer	40.923	14.500	15.276	0.089	0.914
	CNN	128.682	88.896	125.793	0.285	0.147
	CNN-LSTM	77.036	46.444	53.066	0.170	0.694
	iTransformer	126.902	79.013	102.766	0.282	0.171
	Informer	58.251	34.535	47.095	0.129	0.825

图 5 展示了五种模型在阴天与晴天两种天气条件下四个典型时段的预测效果, 其中: (a)、(b)分别对应阴天时段 5.28~5.29 和 6.20~6.21, (c)、(d)分别对应晴天时段 6.21~6.22 和 6.22~6.23。表 7 对各模型在 RMSE、MAE、MAPE (%)、TIC 和 R^2 等指标下的具体表现进行了进一步量化分析。

xLSTM-informer 在两个阴天时段均保持 $R^2 > 0.85$ 、 $TIC < 0.15$, 且误差指标显著优于其他模型, 晴天环境下 R^2 值高达 0.987, 表明本模型在晴天场景中能几乎完全捕捉数据的潜在模式。RMSE (8.863) 和 MAE (2.672) 均为最低, 且 TIC 值仅 0.043, 体现出预测极高的准确性。表明模型对中等波动强度的非平稳数据(如阴天的间歇性变化)具有稳定的建模能力, 这归因于 LSTM 与 Informer 的分层特征提取机制。



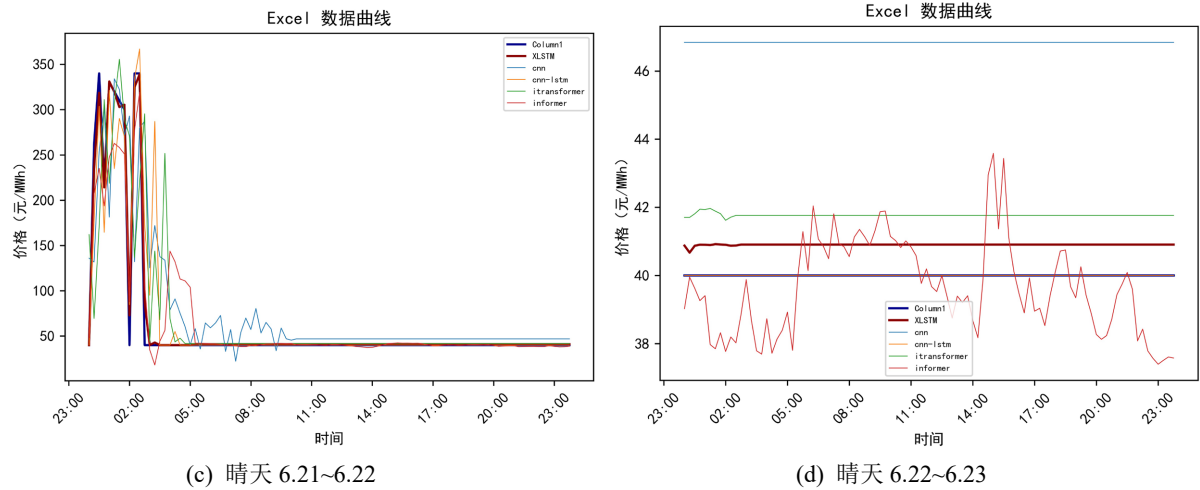


Figure 5. Fitting of predicted and actual values under cloudy and sunny weather
图 5. 阴天晴天情况下预测值与实际值拟合

Table 7. Performance metrics of five models under cloudy and sunny weather
表 7. 阴天晴天情况下五种模型各类指标结果

	指标	RMSE	MAE	MAPE (%)	TIC	R ²
阴天 5.28~5.29	xLSTM-informer	22.124	4.033	4.110	0.132	0.879
	CNN	54.206	25.970	51.174	0.309	0.274
	CNN-LSTM	26.489	7.523	12.087	0.155	0.827
	iTransformer	51.040	16.812	26.372	0.306	0.356
	Informer	31.263	10.730	19.993	0.184	0.758
阴天 6.20~6.21	xLSTM-informer	30.097	9.097	6.962	0.083	0.948
	CNN	104.158	67.318	99.936	0.288	0.375
	CNN-LSTM	77.715	33.753	50.276	0.208	0.652
	iTransformer	111.110	61.293	80.040	0.316	0.289
	Informer	52.693	28.163	31.043	0.152	0.840
晴天 6.21~6.22	xLSTM-informer	8.863	2.672	4.556	0.043	0.987
	CNN	54.889	26.966	49.799	0.264	0.518
	CNN-LSTM	33.289	8.940	14.340	0.163	0.823
	iTransformer	58.065	20.645	29.771	0.287	0.461
	Informer	27.461	11.776	17.046	0.143	0.879

6. 结束语

本文针对新能源渗透率不断上升、电力价格波动性日益增强背景下的现货市场预测问题，提出了一种集成多种结构优势的新型深度预测架构——xLSTM-Informer。该模型通过融合多尺度因果卷积、多头稀疏注意力机制、sLSTM 与 mLSTM 门控记忆路径、BlockDiagonal 参数压缩与双归一化策略，在准确性、稳定性与可部署性等多个关键维度实现显著突破。

在实证分析中，xLSTM-Informer 在预测精度上相较现有主流模型平均提升 10%以上，尤其在新能源

波动剧烈、价格突变频发的情景下表现出优异的鲁棒性与响应能力。模型的轻量化设计显著降低了部署门槛，使其具备在实际电网调度、市场交易平台乃至边缘设备中高频运行的工程可行性。消融实验进一步验证了模型中各子模块的协同作用，其中 xLSTM 结构在建模多变量时序的非线性交互中起到关键性支撑。

与此同时，xLSTM-Informer 架构的设计理念也为深度时序预测模型的演化提供了新方向：突破传统静态结构设计，转向融合记忆、注意力与正则控制的混合动态结构。随着智能电网、虚拟电厂与分布式储能的发展，对高精度、低延迟预测模型的需求将愈发强烈。

综上所述，xLSTM-Informer 不仅在理论结构上推进了深度时序建模技术的边界，也为构建高智能化、高韧性的未来电力系统提供了技术支撑与实践路径，其推广潜力与研究价值有望在多类能源预测任务中持续显现。

基金项目

本文得到国家自然科学基金项目(62467004)及“国网甘肃电科院 2024 年甘肃省调现货技术支持系统电力现货市场全景运营管理平台指标管理研究”项目(HX2025C51100001)资助。

参考文献

- [1] Pikus, M. and Waş, J. (2024) Predictive Modeling of Renewable Energy Purchase Prices Using Deep Learning Based on Polish Power Grid Data for Small Hybrid PV Microinstallations. *Energies*, **17**, Article 628. <https://doi.org/10.3390/en17030628>
- [2] Baskan, D.E., Meyer, D., Mieck, S., Faubel, L., Klöpper, B., Strem, N., *et al.* (2023) A Scenario-Based Model Comparison for Short-Term Day-Ahead Electricity Prices in Times of Economic and Political Tension. *Algorithms*, **16**, Article 177. <https://doi.org/10.3390/a16040177>
- [3] Weron, R. (2014) Electricity Price Forecasting: A Review of the State-of-the-Art with a Look into the Future. *International Journal of Forecasting*, **30**, 1030-1081. <https://doi.org/10.1016/j.ijforecast.2014.08.008>
- [4] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [5] Kong, W., Dong, Z.Y., Jia, Y., Hill, D.J., Xu, Y. and Zhang, Y. (2019) Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Transactions on Smart Grid*, **10**, 841-851. <https://doi.org/10.1109/tsg.2017.2753802>
- [6] Rafi, S.H., Nahid-Al-Masood, Deeba, S.R. and Hossain, E. (2021) A Short-Term Load Forecasting Method Using Integrated CNN and LSTM Network. *IEEE Access*, **9**, 32436-32448. <https://doi.org/10.1109/access.2021.3060654>
- [7] Zhong, B. (2023) Deep Learning Integration Optimization of Electric Energy Load Forecasting and Market Price Based on the ANN-LSTM-Transformer Method. *Frontiers in Energy Research*, **11**, Article 1292204. <https://doi.org/10.3389/fenrg.2023.1292204>
- [8] Yang, G., Du, S., Duan, Q. and Su, J. (2022) Short-term Price Forecasting Method in Electricity Spot Markets Based on Attention-LSTM-MTCN. *Journal of Electrical Engineering & Technology*, **17**, 1009-1018. <https://doi.org/10.1007/s42835-021-00973-5>
- [9] 武永江. 深度学习在电力现货价格预测中的应用研究[D]: [硕士学位论文]. 太原: 太原科技大学, 2021.
- [10] Laitos, V., Vontzos, G., Bargiotas, D., Daskalopulu, A. and Tsoukalas, L.H. (2024) Data-Driven Techniques for Short-Term Electricity Price Forecasting through Novel Deep Learning Approaches with Attention Mechanisms. *Energies*, **17**, Article 1625. <https://doi.org/10.3390/en17071625>
- [11] Kılıç, D.K., Nielsen, P. and Thibbotuwawa, A. (2024) Intraday Electricity Price Forecasting via LSTM and Trading Strategy for the Power Market: A Case Study of the West Denmark DK1 Grid Region. *Energies*, **17**, Article 2909. <https://doi.org/10.3390/en17122909>
- [12] Bâra, A., Oprea, S. and Băroiu, A. (2023) Forecasting the Spot Market Electricity Price with a Long Short-Term Memory Model Architecture in a Disruptive Economic and Geopolitical Context. *International Journal of Computational Intelligence Systems*, **16**, Article No. 130. <https://doi.org/10.1007/s44196-023-00309-3>
- [13] Shejul, K., Harikrishnan, R. and Kukker, A. (2024) Short-Term Electricity Price Forecasting Using the Empirical Mode

- Decomposed Hilbert-LSTM and Wavelet-LSTM Models. *Journal of Electrical and Computer Engineering*, **2024**, Article ID: 4575735. <https://doi.org/10.1155/jece/4575735>
- [14] Liu, B., Li, Z., Li, Z. and Chen, C. (2024) CL-Informer: Long Time Series Prediction Model Based on Continuous Wavelet Transform. *PLOS ONE*, **19**, e0303990. <https://doi.org/10.1371/journal.pone.0303990>
- [15] Tran, N.T and Xin, J. (2023) Fourier-Mixed Window Attention: Accelerating Informer for Long Sequence Time-Series Forecasting. arXiv: 2307.00493.
- [16] Xu, H., Peng, Q., Wang, Y. and Zhan, Z. (2023) Power-Load Forecasting Model Based on Informer and Its Application. *Energies*, **16**, Article 3086. <https://doi.org/10.3390/en16073086>
- [17] 程思远, 杨强, 石文娟, 等. 基于 LSTM + Self-Attention 的电力价格预测模型研究[C]//中国电机工程学会电力信息化专业委员会, 国家电网公司信息通信分公司. 2023 电力行业信息化年会论文集. 北京: 北京中电普华信息技术有限公司, 2023: 127-131.