引文格式：江威,徐杨青,江强强,等. 融入深度位置编码与注意力机制的深基坑形变预测方法[J]. 地理空间信息, 2026, 24(3): 103-107. doi:10.3969/j.issn. 1672-4623.2026.03.021

融入深度位置编码与注意力机制的深基坑形变预测方法

江　威[1]，徐杨青[1]，江强强[1]，孙　杰[2*]，田承昕[2]
（1. 中煤科工集团武汉设计研究院有限公司，湖北 武汉 430064；2. 中国地质大学（武汉），湖北 武汉 430064）

摘　要：深基坑工程作为地铁车站、地下停车场、商业综合体等地下空间开发建设中的关键环节，其形变和施工风险控制问题日益突出。传统方法在基坑围护结构形变预测中依赖经验公式与数值模拟，存在难以准确刻画不同深度之间关系的局限。以武汉市马湖地铁站深基坑工程为例，结合围护结构侧向位移的时空特性，提出了一种基坑形变预测模型DE-SE-Net，通过引入深度位置编码，有效克服了离散特征难以被神经网络有效利用的问题，同时借助特征和时间步注意力机制，提升了模型对关键特征和关键时间步的敏感性。结果表明，该方法能提高围护结构形变预测精度，提升工程管理的科学性和前瞻性。

关键词：深度位置编码；形变预测；神经网络
中图分类号：P258　　**文献标志码：**B　　**文章编号：**1672-4623（2026）03-0103-05

Deformation Prediction Method of Deep Foundation Pit Integrating Deep Positional Encoding and Attention Mechanism

JIANG Wei[1], XU Yangqing[1], JIANG Qiangqiang[1], SUN Jie[2], TIAN Chengxin[2]
(1. CCTEG Wuhan Engineering Company, Wuhan 430064, China;
2. China University of Geosciences (Wuhan), Wuhan 430064, China)

Abstract: As a critical component in the development and construction of underground space projects such as metro station, underground parking lots, and commercial complexes, the deep foundation pit engineering is confronted with increasingly prominent challenges in deformation and construction risk control. Traditional approaches for predicting the deformation of retaining structures in deep foundation pit rely heavily on empirical formulas and numerical simulations, which often struggle to accurately capture the inter dependencies across different depths. Taking the deep foundation pit engineering of Mahu metro station in Wuhan City as the research case, we proposed a foundation pit deformation prediction model named DE-SE-Net by incorporating the spatio-temporal characteristics of lateral displacements of retaining structure. By incorporating deep positional encoding, the proposed method effectively addressed the challenge of utilizing discrete features in neural networks, and integrated feature-wise and temporal-step attention mechanisms to enhance the model's sensitivity to critical features and key temporal-steps. Experimental results demonstrate that the DE-SE-Net model achieves higher prediction accuracy for lateral displacements of retaining structures, thereby enabling more scientific and proactive engineering management.

Key words: deep positional encoding, deformation prediction, neural network

基坑形变不仅直接威胁到工程本体结构的安全，还可能导致邻近建筑物不均匀沉降、城市路面开裂或地下管线断裂损毁等问题，给周边建筑环境稳定性和公共安全体系完整性带来风险。随着基坑工程开挖深度与规模的增加，传统数值分析和经验公式等方法的局限性越来越突出[1]，因此学者们开始探索新方法来提升预测精度和方法的适应性。灰色理论及其改进模型[2]因适用于小样本和不完全信息场景，成为常用方法但缺乏系统性评价标准[3-7]；自回归积分滑动平均模型（ARIMA）与前馈神经网络也被广泛应用于工程形变预测中[8-11]。随着深度学习的发展，针对基坑形变监测数据显著的时间序列特征，循环神经网络（RNN）成为一种潜在的预测方法[12]；而长短期记忆网络（LSTM）因能解决RNN梯度消失问题在时序预测领域备受关注[13]，其拓展模型Bi-LSTM、CNN-LSTM也被尝试应用[14]。综合考虑时序特征，TCN-LSTM与Transformer

项目来源：中国煤炭科工集团有限公司科技创新面上项目（2022-2-MS006）。
第一作者简介：江威（1987—），硕士，高级工程师，从事深基坑与地铁隧道变形监测、工程测量、无人机航测等工作，E-mail：xkdjiangwei@qq.com。
通信作者：孙杰（1982—），博士，副教授，研究方向为遥感大数据处理与应用，E-mail：jiesun@cug.edu.cn。

也被应用于形变预测[15]。目前，卷积神经网络、RNN和注意力机制等神经网络模型已被广泛应用于形变预测，但在基坑形变预测中仍存在不足：围护结构的侧向位移与测点深度密切相关，且相邻深度的形变曲线通常呈现相似性，传统数值型深度特征难以准确刻画不同深度之间的关系，而神经网络模型也难以有效提取该信息，导致模型无法充分学习深度间的关联性，进而影响预测结果的准确性。鉴于此，本文提出了一种融合深度位置编码和SE注意力机制的形变预测模型（DE-SE-Net），旨在提高基坑形变预测精度，构建基于预测的主动安全防控体系。

1 数据来源

本文采用的数据来源于武汉市轨道交通8号线二期马湖站建设项目，车站全长552.3 m；场地地形平坦，地势起伏不大，坡降较缓，地面高程在19.35~26.08 m之间；地貌单元属于长江Ⅲ级阶地的剥蚀堆积平原区（三级阶地湖积区）；分布地层自上而下为填土及第四系全新统一般黏性土、第四系上更新统冲、洪积物、第四系残积物、东湖群砂岩、砂质泥岩。基坑开挖建设期间，按照规范布设各类形变监测点和受力监测点，并按照规定频率进行监测记录。

深基坑围护结构的水平位移是基坑工程安全监测与设计中的核心指标，直接关系到基坑稳定性、周边建（构）筑物安全和地下管线保护。影响深基坑围护结构水平位移的因素包括地质与水文条件、基坑自身几何与设计参数、施工因素、外部荷载与环境因素等。从水文地质条件、基坑分仓几何分布、施工周报、现场监测数据和周边环境等实际可掌握的数据出发，本文共获取较重要的14个特征因子，即地表沉降（mm）、建筑物沉降（mm）、管线沉降（mm）、结构沉降（mm）、结构水平位移（mm）、支撑轴力（kN）、土压力（MPa）、地连墙钢筋应力（kN）、分层土体类别（1素填土、2黏土、3强风化泥质砂岩、4中风化泥质砂岩）、围护结构水平位移深度分布（mm）、基坑暴露时间（d）、底板是否浇筑（0未浇筑、1浇筑完成）、地下水位（mm）和已开挖长度（m）。围护结构的侧向位移是衡量基坑稳定性的重要指标，其变化趋势能反映基坑开挖过程中围护结构的受力特性。本次实验以武汉市轨道交通8号线马湖地铁站深基坑监测项目中围护结构侧向位移监测点CX47为研究对象，将数据按照深度分组，获取影响范围内的特征数据；再将每条数据按照8∶2的比例划分训练集和测试集，其中训练集包含1 955条数据，测试集包含552条数据。监测项目通常采用测斜管进行监测，每米设置一个测点，以捕捉不同深度的变化趋势。由图1可知，不同深度的位移变化特征存在明显差异：浅层区域受地表荷载、土体松弛和降雨等因素影响较大，形变波动较显著；深层区域受周围土体的约束作用较强，变形幅度较小且较稳定；中部区域变形最显著，形成典型的“大肚子曲线”。

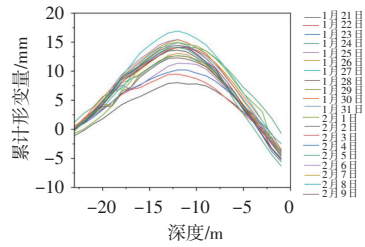


图1 围护结构侧向位移监测点CX47部分周期监测数据

2 研究方法

2.1 模型设计

本文提出的DE-SE-Net模型融合了深度位置编码和SE注意力机制，整体结构由TCN模块、特征强化模块、GRU模块、深度位置编码模块和时间步注意力模块组成（图2），通过多层结构的协同作用，实现对基坑围护结构侧向位移的精准预测。

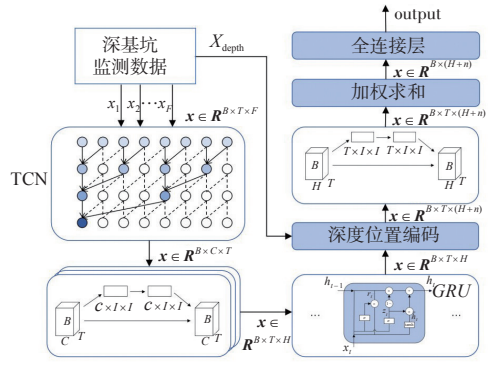


图2 DE-SE-Net模型架构

将数据输入TCN模块前，需对原始的围护结构侧向位移数据进行预处理（缺失值填充和异常值剔除），以确保输入数据的完整性和有效性。此外，为减少不同深度间的尺度差异，按深度对形变值进行分组，并分别进行归一化处理，使模型能更准确地学习不同深度下的形变特征。模型输入由监测数据和深度信息两部分构成，监测数据输入TCN模块，深度信息通过深度位置编码模块输入模型。监测数据包含多个周边监

测点的数值型变量，如支撑轴力、冠梁沉降、冠梁水平位移等，其特征向量表示为 $\boldsymbol{X}_{\text{monitoring}} \in \boldsymbol{R}^{B\times T\times F}$，其中 B 为批次大小，T 为时间维度，F 为特征维度。将监测数据的时序序列输入TCN模块进行时序特征提取。TCN通过扩展卷积与残差链接机制，可在较少时间步数内提取复杂时序特征并建模长时依赖关系。扩展卷积通过在卷积核元素间设置空洞扩大模型感受野，使每层能跨越部分时间步捕获更长跨度的依赖关系。

特征强化模块引入SE注意力机制以增强输入特征的表达能力：TCN模块提取的时序特征矩阵为 $\boldsymbol{X}\in \boldsymbol{R}^{B\times C\times T}$，$C$ 为通道数。模块的核心目标是为各通道赋予权重，使模型能自适应聚焦关键特征。

GRU模块用于进一步处理特征强化模块输出的时序特征。GRU通过更新门和重置门控制信息流动，决定哪些历史信息应当保留，哪些应当遗忘。GRU通过处理每个时间步的特征，输出隐藏状态 h。假设GRU隐藏状态维度为 H，则经过GRU处理后，维度的转换表达式为：$\boldsymbol{X}_{\text{in}} \in \boldsymbol{R}^{B\times C\times T} \rightarrow \boldsymbol{X}_{\text{mid}} \boldsymbol{R}^{B\times T\times C} \rightarrow \boldsymbol{X}_{\text{out}} \in \boldsymbol{R}^{B\times T\times H}$，其中 $\boldsymbol{X}_{\text{in}}$ 为特征强化模块输出的特征矩阵，$\boldsymbol{X}_{\text{mid}}$ 为经过维度变换后的特征矩阵，$\boldsymbol{X}_{\text{out}}$ 为经过GRU处理后的特征矩阵。

深度信息通过位置编码将原始的离散深度特征转换为低维连续向量，使神经网络能够学习不同深度之间的关联性。假设深度特征为 x_{depth}，则通过位置编码将每个深度映射为一个 n 维向量，得到深度嵌入矩阵，即 $\boldsymbol{X}_{\text{depth}} = f\left(x_{\text{depth}}\right) \in \boldsymbol{R}^{B\times n}$，其中 $f\left(x_{\text{depth}}\right)$ 为深度特征的编码映射函数，n 为编码维度。将 $\boldsymbol{X}_{\text{depth}}$ 扩展至每个时间步 T，则有：$\boldsymbol{X}_{\text{depth_expanded}} \in \boldsymbol{R}^{B\times T\times n}$。将 $\boldsymbol{X}_{\text{out}}$ 和 $\boldsymbol{X}_{\text{depth_expanded}}$ 按维度拼接，得到拼接后的特征矩阵，即 $\boldsymbol{X}_{\text{concat}} \in \boldsymbol{R}^{B\times T\times (H+n)}$。

在时间步注意力模块中，利用时间步注意力机制对每个时间步的输出特征进行加权赋值，使模型能够更加专注于对预测任务贡献更大的时间步。该模块的输入为深度位置编码模块拼接后的特征矩阵，其核心是通过线性层来计算每个时间步的注意力权重。首先通过两个线性层进行处理，生成每个时间步的注意力权重 $\text{attn}_{\text{weights}} = \text{Softmax}\left(\text{Linear}_2\left(\text{ReLU}\left(\text{Linear}_1(x)\right)\right)\right)$，对这些权重进行Softmax操作，得到对时间步进行归一化的权重；再将 $\text{attn}_{\text{weights}}$ 与原始输入特征 x 进行逐元素相乘，得到加权后特征 $\text{weighted_}x \in \boldsymbol{R}^{B\times T\times C}$；最后对加权后的特征按时间步进行求和操作，得到最终的加权 weighted_out。对加权的时间步特征进行汇总，输出一个维度为 $\boldsymbol{R}^{B\times C}$ 的表示；经过全连接层，生成模型的预测结果。

2.2　实验设计

1）实验环境配置。实验选取Python作为主要开发语言，以PyTorch为深度学习神经网络训练的核心框架。实验运行环境基于Windows10操作系统，学习框架为PyTorch 2.5.1、编程语言为Python 3.9.13、CPU为Intel(R) Core(TM) i5-8300H CPU @ 2.30GHz、GPU为NVIDIA GeForce GTX 1050 Ti (4 GB)、CUDA 11.1。为实现灵活的环境管理，实验选取Anaconda作为环境管理工具。

2）对比模型。实验采用的对比模型包括：①ARIMA是统计学上常用的时序数据预测方法，由自回归和移动平均模型组成，通过对时间序列进行趋势差分（消除趋势和周期项），捕捉数据的内在规律；②Bi-LSTM是RNN的一种变体，将输入序列同时取前向和后向进行先处理，通过双向获取整个序列的上下文信息，可以更充分地了解整个序列的信息；③TCN-LSTM利用TCN的一维卷积和扩张卷积提取时间序列的局部特征并保留长期依赖关系，提取的特征再传入LSTM建模时间上的依赖关系，进一步挖掘全局的时序关系，最后送到输出层用于预测，相较于LSTM，其在特征提取和序列建模上表现更优，增加了特征的表达能力；④Transformer是基于注意力机制的模型，将输入序列通过多头自注意力机制计算位置向量的加权表示，得到序列的位置关系信息，加入位置编码，使模型记忆序列位置信息，包括编码器和解码器，前者提取输入序列的表示，后者产生输出序列。

3）评价指标。本文采用的评价指标包括均方根误差（RMSE）、平均绝对误差（MAE）和平均绝对百分比误差（MAPE）。RMSE通过计算预测值与真实值间误差的平方根来度量模型的预测性能，反映了模型的平均误差幅度。MAE常用于评估时间序列预测模型的预测性能，通过计算每个样本预测值和真实值间的绝对误差来衡量模型的预测准确性。MAPE通过计算每个样本预测值与真实值之间的百分比误差来评估模型的准确性。与RMSE和MAE不同，MAPE直接给出了预测误差占真实值的比例，因此在比较不同规模的预测值时，具有更好的解释性；但当真实值接近零

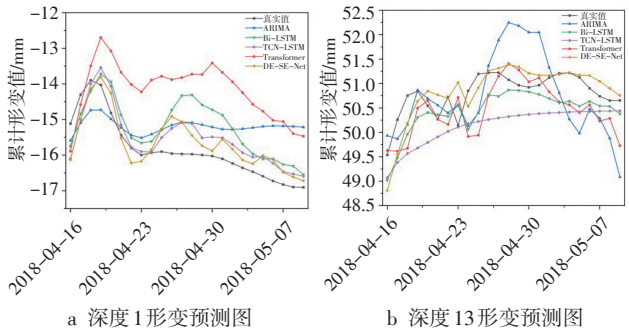
时，可能会出现较大的误差。

3 实验结果

将DE-SE-Net模型与4种对比模型进行比较，评价指标见表1，不同深度的形变预测效果见图3，结果表明：①DE-SE-Net通过深度位置编码、特征注意力和时间步注意力，能充分提取序列的长短期依赖关系和特征信息，并能学习到不同深度序列之间的形变关系，与TCN-LSTM模型相比，RMSE、MAE和MAPE分别降低了20%、22.22%和13.17%，取得了最佳预测效果；②作为经典的传统时序模型，ARIMA的RMSE、MAE和MAPE均表现最差，表明其在深基坑形变预测任务中的精度较低，特别是在面对复杂的非线性特征时，表现不如其他现代深度学习模型；③Bi-LSTM的预测性能较好，尤其是MAE、MAPE明显低于ARIMA，说明其能有效捕捉序列中的时间依赖性和变化趋势，进而提高预测精度；④Transformer略优于Bi-LSTM，且各指标均优于ARIMA，依靠编码器—解码器结构和自注意力机制能挖掘长时间依赖关系，在数据复杂程度较高时，其非线性特征提取能力在预测精度上具有一定的优势；⑤TCN-LSTM结合了TCN和LSTM模型，通过卷积层高效提取局部特征，并利用LSTM捕捉序列的时间依赖，其预测精度仅次于DE-SE-Net模型，展现了较强的预测性能。

表1 各模型实验结果评价指标对比

模型	RMSE/mm	MAE/mm	MAPE/%
ARIMA	1.29	1.20	7.05
Bi-LSTM	0.62	0.48	3.71
TCN-LSTM	0.50	0.36	2.05
Transformer	0.62	0.47	3.44
DE-SE-Net	0.40	0.28	1.78



a 深度1形变预测图 b 深度13形变预测图

图3 不同深度的形变预测效果

DE-SE-Net模型整体预测曲线最接近真实值，波动幅度小、误差控制好，对非线性变化响应良好；Bi-LSTM和TCN-LSTM模型表现尚可，但不如DE-SE-Net稳定；TCN-LSTM模型对突发变化敏感，但易震荡；Transformer模型初期严重失准，可能未充分训练或架构不适合该类任务；ARIMA作为传统统计模型，难以捕捉非线性动态变化。

基坑围护结构水平位移预测中前7个影响因素见图4。深度最显著，基坑开挖深度直接决定了围护结构的受力高度和侧向土压力大小，深度越大主动土压力累积越多，围护结构承受的弯矩与水平推力显著增加，深度还间接影响地下水位变化、支护体系刚度需求和变形模式，因此它是决定性几何参数，对水平位移有最强的控制作用。地下水位上升会增大孔隙水压力，降低土体有效应力，导致土体强度下降；同时增加浮力效应，使围护结构受到更大的侧向水压力。地表沉降和建筑沉降属于响应变量或反馈信号，虽非直接驱动力，但在监测预警系统中具有较高的关联性。支撑轴力与土压力的重要性较低，可能由于支撑轴力本身是结构响应，而非独立输入变量；土压力在模型中已被深度、土性、地下水等因素隐含表达。基坑暴露时间影响土体固结、蠕变和应力重分布过程，尤其在黏性土中长期暴露将导致缓慢变形；但其影响相对滞后且非线性，在短期预测中权重较低。上述结果表明，在基坑围护结构水平位移预测中，结构性参数（深度）和环境变量（地下水位）是最关键的驱动因子，而传统认为重要的“力学量”（如支撑轴力）反而因与其他变量高度相关而在模型中表现不突出。

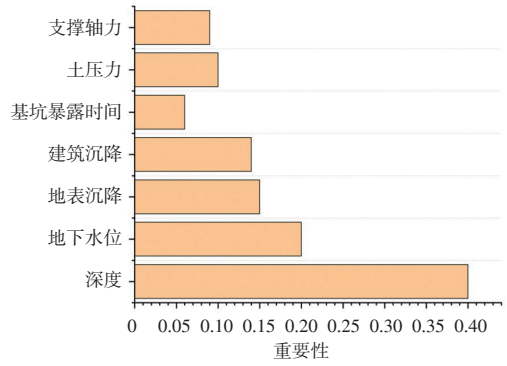


图4 影响因素的重要性

4 消融实验分析

DE-SE-Net模型中特征强化模块采用多头SE注意力机制，对不同通道的权重进行自适应分配；深度位置编码模块增强了模型对不同深度序列形变关系的理解；时间步注意力模块采用时间步注意力机制赋予时间步不同的权重。本文采用控制变

量方法，对DE-SE-Net模型进行消融实验，以分析各模块对模型性能的影响。参与消融实验的对比模型包括：①TCN-GRU-SE模型，在原有模型的基础上移除深度位置编码，仅采用形变数据和周边监测数据进行预测，以验证深度位置编码的作用；②DE-TCN-GRU-T模型，移除特征强化模块，将TCN模块的输出直接输入GRU模块，评估特征强化模块对模型性能的影响；③DE-TCN-GRU-F模型，移除时间步注意力机制，以GRU输出的最后一个时间步作为全连接层的输入，以分析时间步注意力对最终预测结果的贡献。利用围护结构侧向位移CX47监测点进行消融实验验证，消融实验模型参数设置与原模型相同，实验结果见表2，可以看出：①DE-SE-Net模型包含所有关键模块，评价指标均为最佳，验证了其在深基坑形变预测任务中的有效性和合理性；②TCM-GRU-SE模型比DE-SE-Net模型在RMSE、MAE和MAPE上分别增加了10.00%、17.86%和23.60%，表明深度位置编码能有效捕捉不同深度间的形变关系，在一定程度上提高模型的预测精度；③DE-TCN-GRU-T模型比DE-SE-Net模型在RMSE、MAE和MAPE分别增加了22.50%、21.43%和7.87%，说明特征强化模块对模型性能的提升较显著，特别是RMSE和MAE；④DE-TCN-GRU-F模型比DE-SE-Net模型在RMSE、MAE和MAPE上分别增加了7.50%、14.29%和6.74%，说明增加时间步注意力提升了模型性能，特别是在对关键时间步的关注上。

表2　消融实验模型评价指标对比

模型	RMSE/mm	MAE/mm	MAPE/%
TCN-GRU-SE	0.44	0.33	2.20
DE-TCN-GRU-T	0.49	0.34	1.92
DE-TCN-GRU-F	0.43	0.32	1.90
DE-SE-Net	0.40	0.28	1.78

5　结　论

以武汉市马湖地铁站深基坑工程实测数据为例，结合围护结构侧向位移的时空特性，本文提出了一种基坑形变预测模型DE-SE-Net。该模型通过引入深度位置编码，有效克服了离散特征难以被神经网络有效利用的问题；同时借助特征和时间步注意力机制，提升了模型对关键特征和关键时间步的敏感性。实验结果表明，该方法显著提高了预测精度，有望应用于类似工程中以提高围护结构形变预测精度，保障施工安全，提升工程管理的科学性和前瞻性。

参考文献

[1] Guo C, Yang X, Guo C, et al. Analysis of Deformation of Deep and Large Foundation Pit Support Structure and Impact on Neighbouring Buildings in Complex Environments[J]. Buildings, 2025, 15(9):1435

[2] Julong D. Introduction to Grey System Theory[J]. The Journal of Grey System, 1989, 1(1):1-24

[3] 苟胜荣, 张文学, 白立东. 基于GM(1,1)模型的地铁基坑变形预测研究[J]. 粉煤灰综合利用, 2024, 38(2):70-74

[4] 万志辉, 刘红艳, 步艳洁. 基于灰色理论的深基坑围护结构变形预测研究[J]. 辽宁工业大学学报(自然科学版), 2014, 34(3):187-189

[5] 王强, 刘松玉, 童立元, 等. 灰色理论在深基坑支挡结构变形预测中应用[J]. 岩土工程学报, 2010, 32(增刊2):69-72

[6] 张磊, 兰孝奇, 向荣荣, 等. 灰色时序组合模型在基坑水平位移预测中的应用[J]. 勘察科学技术, 2016(4):47-50

[7] Cui D, Zhu C, Li Q, et al. Research on Deformation Prediction of Foundation Pit Based on PSO-GM-BP Model[J]. Advances in Civil Engineering, 2021(1):929

[8] Cao J, Ding W Y, Zhao D S, et al. Time Series Forecast of Foundation Pit Deformation Based on BP Neural Network[J]. Applied Mechanics and Materials, 2014, 556:5979-5983

[9] 蒙国往, 刘家梁, 黄劲松, 等. 基于BP人工神经网络的深基坑围护结构水平位移预测研究[J]. 都市快轨交通, 2022, 35(3):80-88

[10] 石荔. 基于粒子群优化BP神经网络的深基坑支护结构变形预测[J]. 地质学刊, 2024, 48(3):327-335

[11] Feng T, Wang C, Zhang J, et al. Prediction of Stratum Deformation During the Excavation of a Foundation Pit in Composite Formation Based on the Artificial Bee Colonyback-propagation Model[J]. Engineering Optimization, 2022, 54(7):1217-1235

[12] Zhang Y X, Li X W. Predicted Water Yield of Open-pit Metal Mines Based on a Bi-RNN and GMS Coupling Model[J]. Coal Geology and Exploration, 2024, 52(12):15

[13] Shiri F M, Perumal T, Mustapha N, et al. A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU[J]. arXiv Preprint arXiv:2305.17473

[14] Li H, Zhao Z, Du X. Research and Application of Deformation Prediction Model for Deep Foundation Pit Based on LSTM[J]. Wireless Communications and Mobile Computing, 2022(1):999

[15] 洪宇超, 钱建固, 叶源新, 等. 基于时空关联特征的CNN-LSTM模型在基坑工程变形预测中的应用[J]. 岩土工程学报, 2021, 43(增刊2):108-111