Mine Water and the Environment (2024) 43:707–726
https://doi.org/10.1007/s10230-024-01011-2

TECHNICAL ARTICLE

Transfer Learning with Transformer-Based Models for Mine Water Inrush Prediction: A Multivariate Analysis Using Sparse and Imbalanced Monitoring Data

Huichao Yin[1,2] · Gaizhuo Zhang[1] · Qiang Wu[1] · Fangpeng Cui[1] · Bicheng Yan[3] · Shangxian Yin[4] · Mohamad Reza Soltanian[5] · Hung Vo Thanh[6,7,8] · Zhenxue Dai[9,10]

Received: 20 April 2024 / Accepted: 24 October 2024 / Published online: 16 November 2024

Abstract

Predictions of mining-induced water inrush accidents are challenged by data sparseness and imbalances, as very few high-quality datasets can be obtained for successfully modeling data variation. By using the concept of transfer learning, we employed a well recorded borehole group water level dataset as a source dataset to train a selection of Transformer-based multivariate prediction models with state-of-the-art performance including PatchTST, InFormer, and AutoFormer, to capture data variation patterns in a statistically similar target dataset from a site with similar geological and mining conditions and examined the models' accident prediction performance. Additionally, the frequently used MLP-based Nbeats, RNN-based LSTM, and CNN-based TCN were adopted for the same task. In contrast to models trained merely on the target dataset, the Transformer-based models, especially PatchTST, achieved satisfactory zero-shot prediction performances in terms of accuracy, responsiveness, and anomaly detections for the early warning of accidents, proving their promising generalization capabilities for leveraging existing datasets for forecasting future accidents with data obtained in similar geological conditions. This has broad implications for mining accident prediction and groundwater risk assessment using data-driven approaches.

Keywords Underground mining · Multivariate prediction · Borehole group · Transformer models · Anomaly detection

Huichao Yin and Gaizhuo Zhang contributed equally to this work.

✉ Shangxian Yin
yinshx03@126.com

✉ Zhenxue Dai
dzx@jlu.edu.cn

1 College of Geoscience and Surveying Engineering, China University of Mining and Technology, Beijing 100083, China

2 Department of Plant and Environmental Sciences, New Mexico State University, Las Cruces, NM 88003, USA

3 Energy Resources and Petroleum Engineering, King Abdullah University of Science and Technology, 23955-6900 Thuwal, Saudi Arabia

4 College of Safety Engineering, North China Institute of Science and Technology, Langfang 065201, China

5 Departments of Geosciences and Environmental Engineering, University of Cincinnati, Cincinnati, OH 45220, USA

6 Laboratory for Computational Mechanics, Institute for Computational Science and Artificial Intelligence, Van Lang University, Ho Chi Minh City, Vietnam

7 Faculty of Mechanical-Electrical and Computer Engineering, School of Technology, Van Lang University, Ho Chi Minh City, Vietnam

8 MEU Research Unit, Middle East University, Amman, Jordan

9 College of Construction Engineering, Jilin University, Changchun 130026, China

10 Institute of Intelligent Simulation and Early Warning for Subsurface Environment, Jilin University, Changchun 130026, China

Introduction

Water inrush accidents pose severe risks to underground mining activities in major coal producing countries around the world (Chen et al. 2015; Dong et al. 2020; Liu et al. 2017). In China, such issues have become more complicated with increased mining depth (Gu et al. 2020; Wang et al. 2022; Zhang et al. 2023b). Due to the stochastic occurrence and drastic development of such accidents, massive casualties and property losses have been sustained by the mining industry over the years (Singh 2015). In addition, long recovery periods and large investments are required to restore the mining process (Cui et al. 2018; Dash et al. 2016; Zeng et al. 2024) and reduce the environmental effects (Liu et al. 2024; Liu et al. 2022; Yin et al. 2018; Yu et al. 2020; Zeng et al. 2023b).

Drilling exploration is one of the most common ways to obtain direct, reliable, and real-time observations of variations of water properties, including level (Yin et al. 2023a), temperature (Li et al. 2023), and pressure (Li et al. 2021; Xu et al. 2021), through boreholes for monitoring and controlling the risk of mining-induced water inrush (Bao 2019; Hungerford and Ren 2013; Kang et al. 2023). This allows precursory information of potential disasters to be detected and early warnings to be issued (Cao et al. 2022). As a result, massive amounts multisource data are dynamically obtained using high-precision sensors (Miao et al. 2022; Zhang et al. 2023a), allowing data-driven approaches to be used for pattern analysis and risk prediction (Li et al. 2024; Yin et al. 2022; Zhang et al. 2023e).

With the fast development of data science and deep/machine learning algorithms, large amount of efforts has been applied to develop data-driven models to predict water inrush and many other mining-induced hazards (Dong and Zhang 2023). This includes regression and statistical analysis based models (Han et al. 2023; Qiao et al. 2019; Sun et al. 2022; Qu et al. 2023; Wu et al. 2017; Zhang et al. 2022), machine-learning based methods (Chen et al. 2022; Huang et al. 2019; Li et al. 2022; Lin et al 2021; Wei et al. 2022; Yan et al. 2021a; Zhang 2023; Zhang et al. 2023e; Zhao and Wu 2018; Zhao et al. 2018), and various neural network models (Fang 2022; Oreshkin et al. 2021; Prasad et al. 2024; Yang et al. 2018; Yin et al. 2023a, b; Zhou et al. 2023a; Zhu et al. 2022). Employing their capabilities of extracting both long- and short-time dependencies in data variations, most of the methods are able to achieve satisfactory predictive results by effectively capturing seasonal and trend variation patterns in the data.

However, these methods' effectiveness to accurately predict water inrush accidents are affected by imbalances and variations in the obtained data, which are often nonlinear, non-stationary, and contain a high level of noise and randomness (Choi et al. 2023; Cui et al. 2023; Yedla et al. 2020). Their predictive accuracy and timeliness is especially affected in disastrous scenarios when data can vary drastically and unexpectedly under exogenous interferences (Cui et al. 2022). In many cases, even when the overall predictive accuracy is adequate, the response speed of the models during potential disasters are not good enough to provide useful predictions (Zhang et al. 2023d).

Various techniques have been adopted to alleviate this effect and improve their predictive performance, including applying hyperparameter optimization (Ji et al. 2022; Yan et al. 2021a; Yin et al. 2023a, b; Zhou et al. 2023a; Zhu et al. 2022), adopting data-driven anomaly detection methods (Su et al. 2021; Teng et al. 2020; Zhang et al. 2023c), AutoEncoder (AE) models (Asahi et al. 2021; Gong et al. 2022; Nicholaus et al. 2021; Skaf and Horváth 2022; Yokkampon et al. 2020, 2022), using composite prediction models like (Dey et al. 2021; He et al. 2022; Yang et al. 2024; Yokkampon et al. 2022), and the coupling of predictive and anomaly detection models (Yin et al. 2023a, b). Despite some of their success, the issue persists and inhibits their broad implementation (Zhang and Wang 2021; Zhang et al. 2023e), since well-recorded datasets obtained throughout water inrush accidents are still very rare and data sparseness and imbalances often lead to inadequately trained models for predicting accidents at various mine sites (Ahmadzadeh et al. 2021; Lyu et al. 2021; Shen et al. 2023; Zeng et al. 2022).

To overcome this issue, we explored the potential of transfer learning on data-driven approaches of water inrush prediction by pre-training (Hendrycks et al. 2020) a few of the state-of-the-art Transformer-based predictive models including PatchTST (Nie et al. 2023), InFormer (Zhou et al. 2020), and AutoFormer (Wu et al. 2021) using a large multivariate borehole water level dataset fully recorded before and during a major water inrush accident. Taking advantage of the exceptional generalization (Cho et al. 2023; Mu et al. 2023) and adaptation (Shamsabadi et al. 2022; Thomas et al. 2023) capabilities of Transformer-based models, we examined their performance in terms of various metrics by conducting zero-shot predictions (Sahin et al. 2023; Su et al. 2022) of a major water inrush incident on an unseen dataset that was obtained at a mining site with very similar hydrogeological and working conditions.

In addition, we compared the models' performance with and without the pre-training process in terms of both predictive accuracy and timeliness to investigate the imposed effect of transfer learning on the models. Furthermore, a collection of frequently used neural networks multivariate prediction models including NBeats (Oreshkin et al. 2021; Prasad et al. 2024), LSTM (Mahmoodzadeh et al. 2021; Yin et al. 2023a, b), and TCN (Yang et al. 2023) were also thoroughly compared with their performance using identical forms of inputs on the same task. Finally, as a frequently adopted approach of reflecting unexpected interference to data variations, all models were examined for their abilities to detect data anomalies in terms of predictive errors and thereby capture early signs of potential accidents (Sun and Li 2022; Toufigh and Ranjbar 2023; Yan et al. 2021b). The basic information of the models used in this study and their notations is shown in Table 1.

To summarize: (a) Transfer learning was applied to Transformer models, using a large and diverse source water level dataset for pretraining, to perform zero-shot predictions on a target accident dataset. (b) Frequently used neural network prediction models were also compared for the same task. Their performances were evaluated in terms of accuracy, efficiency, responsiveness, and anomaly detection. (c) Models pretrained on the source dataset were compared with those trained merely on the target dataset. The benefits of transfer learning were analyzed in detail, focusing on addressing data sparseness, imbalance, and prediction timeliness in mining water inrush accidents.

The results of this work provide a way to alleviate data sparseness and improve predictive performance of mining-induced water inrush accidents. The findings thus potentially have broad implications for using field-observed high-quality datasets to pretrain Transformer-based models to achieve accurate and timely data predictions in similar mining conditions and thereby improve mining safety and prevent accidents.

Studied Mining Site and Dataset

Overview of the Mine Sites

The two mine sites being studied were the Xingdong and Xin'an coal mines in Hebei Province in China, both located in the Hanxing mining area in the middle of the eastern foot of the Taihang Mountains. These mines are representative of the North China Carboniferous-Permian coalfield and are threatened by high-pressure confined water in the Ordovician limestone that underlies the working face. As shown in Fig. 1, the two mine sites are less than 90 km apart. The two mines share most of the same geological and mining conditions, as shown in supplemental Fig. S-1.

Table 1 Basic information of model used in the study

Full name	Abbreviation	Model type
Patch Time Series Transformer	PatchTST	Transformer-based
Information Fusion Transformer	InFormer	Transformer-based
AutoFormer	AutoFormer	Transformer-based
Neural Basis Expansion Analysis Time Series	Nbeats	MLP-based
Long Short-Term Memory	LSTM	RNN-based
Temporal Convolutional Network	TCN	CNN-based

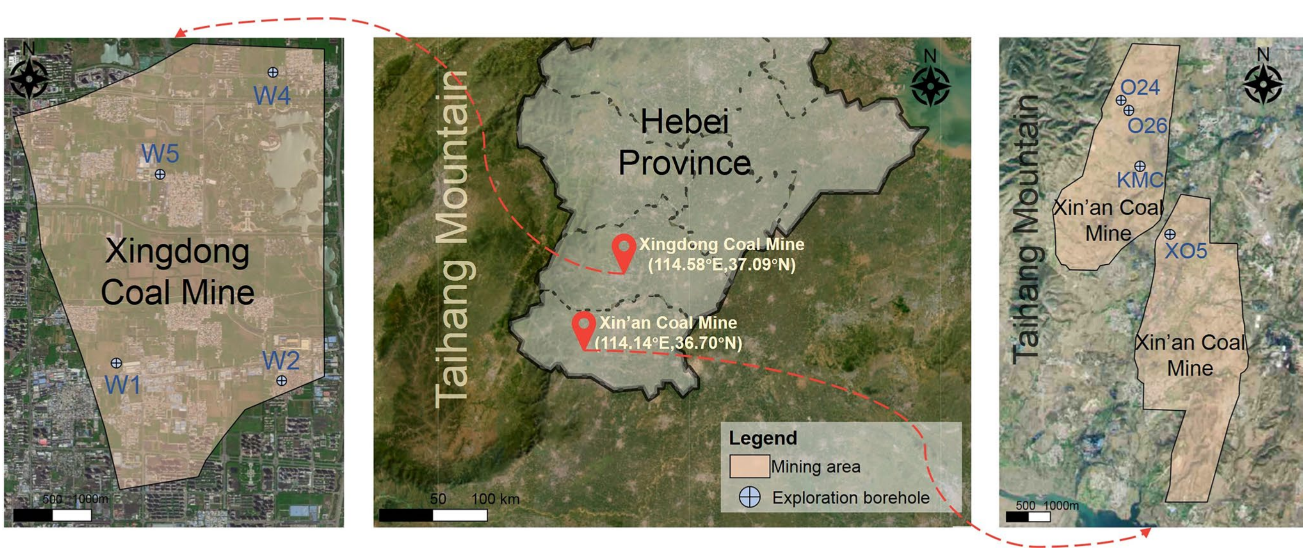

Fig. 1 Locations of the Xingdong and Xin'an coal mines. The borehole groups of the Xindong and Xin'an mines include boreholes W1, W2, W4, and W5, and boreholes XO5, KMC, O24, and O26, respectively

Figure S-1 shows first that the sedimentary condition of coal formation of the two mines are similar: both are located in the Hanxing coalfield (Yin et al. 2021) and share the same structure, which is several thin layers of limestone between the major coal seam and the thick Ordovician limestone aquiclude. Second, the tectonic evolution of the two mines is similar, since both belong to the geological structural belt in the middle section of the eastern foot of the Taihang Mountains. Also, both have similar structural patterns and characteristics, which result from the fault block structure and are mainly influenced by the major ground stress field in the northeast direction. Last, both are longwall coal mines with water pressures on the coal seam floor that exceeds 10 MPa. Therefore, the physical mechanisms of water inrush from the mining floor of the two mines are similar, as the confined water in the Ordovician limestone aquifer below the mine enters the mine after breaking through the aquiclude, which has already caused hundreds of inrush incidents over the years.

One of the major accidents (Accident I) took place in the Xingdong mine at 4 pm on 3 March 2018 and lasted for more than a month (Yin et al. 2023a). During this accident, the water inrush quantity in the mine tunnels drastically increased from 60 m^3/h and peaked at 2649 m^3/h in less than 3 days. Meanwhile, the water level in some of the exploration boreholes dropped rapidly, with a maximum decrease of more than 70 m in less than 10 days. Afterwards, it took more than 3 months of grouting to restore production. Starting from February 2015 and throughout the accident, the water level data of the four exploration boreholes near the accident location, namely W2, W4, W1, and W5, were fully recorded, generating a complete multivariate time series dataset with more than 3 years of valuable water level variations including those immediately prior to the occurrence of the accident. The large sample size, data continuity, and the diversity of data variation features all make the dataset ideal for training our chosen neural network models and identifying water level variations for detecting potential accidents in the future.

Another major accident (Accident II) occurred in the Xin'an mine at 4 pm on 25 October 2021. While it was of smaller scale and lasted a shorter period than the Xingdong mine accident, it also experienced a rapid water inrush quantity increase from 55.5 to 94.5 m^3/h in less than 10 days. Significant drops of water levels were also demonstrated in four of its exploration boreholes (namely XO5, O24, O26, and KMC), with a maximum decrease of almost 1 m in less than 4 days. Although the water level data were also well recorded, the time series dataset are of a shorter time span, starting from October 2020 and ending December 2021. Therefore, our chosen predictive models were tested on the dataset obtained in this case to examine their adaptability and effectiveness in detecting early signs of water inrush accidents.

Despite Accident II being less severe with milder disturbances on water level data variations, these two accidents have several similarities. First, the water source of both were the Ordovician limestone aquifer. The water-conducting channels were both concentrated cylindrical and tubular channels between the coal seam and the thick Ordovician limestone aquiclude. Second, the water inrush quantities were both large, and gradually increased in a stepwise manner to their peak values. This is due to the Ordovician limestone water's gradual rise through the strata and sudden breakage through the thick aquicludes, forming the water inrush. Finally, both of the observed water inrush quantities' variations noticeably lag behind the exploration boreholes' water level variations, allowing early warnings of accidents to be issued by analysis of the water levels' anomalous variations.

Data Description

The water level dataset obtained from the borehole groups in Xingdong (Source Dataset) and Xin'an mine (Target Dataset) are plotted in Fig. 2. As shown in Fig. 2, both datasets demonstrate a variety of water level features. The sampling period of the Target Dataset was 5 min, whereas the Source Dataset was sampled every hour originally, but shortened to 5 min as well using linear interpolation to match the sampling period of the Xin'an mine data and to allow more precise and timely predictions in further analysis. Linear interpolation can adequately restore the water level trend since the variation was mostly steady and stable within the sampling period.

As a result of the similar hydrogeologic conditions shared by the two mines, these two datasets also have statistically much in common. First, the water level data of each borehole in the same group have high monotonic correlations to each other during normal periods without water inrush accidents, exhibiting very similar Spearman correlation coefficients (Wu & Wang 2022) that are both close to 1 (Fig. 3). The Spearman coefficient is employed in this case for its adaptivity in identifying the linear and monotonic non-linear correlations between the continuous variables with no normal distribution of the analyzed data.

Second, by decomposing the data by an additive model (Wahyuningsih et al. 2017), as shown in supplemental Fig. S-2, data in both mining sites were found to have a dominant trend component with negligible seasonal and residual components during normal periods, whereas their residual components both exhibited noticeable spikes and rapid changes during the accident periods.

Third, further specific relationships between the variables and the distribution of each variable in each dataset

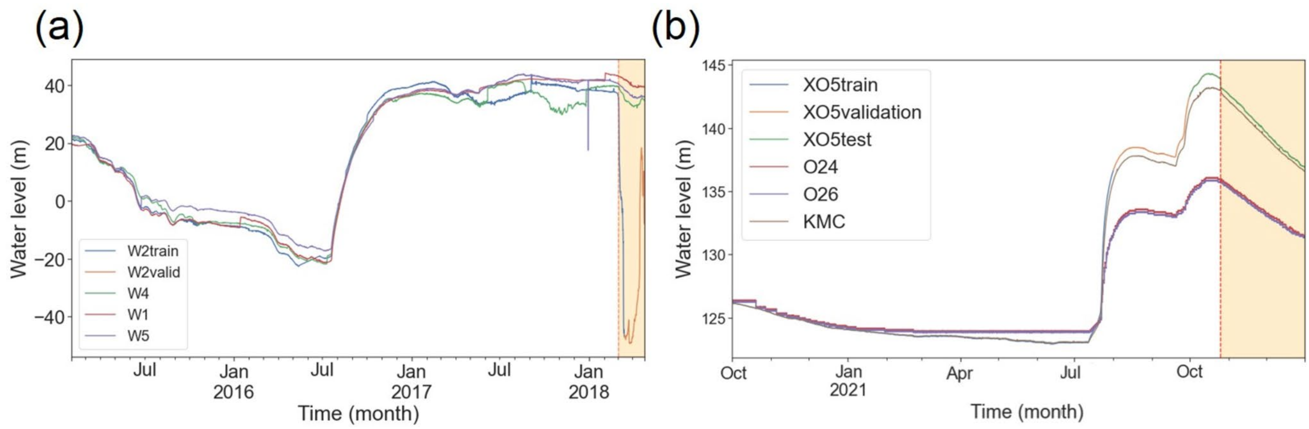

Fig. 2 Borehole water level variations of the **a** Xingdong and **b** Xin'an mine. Yellow shades are the periods of the two water inrush accidents. The red vertical dashed lines represents the time when the two accidents started and the water level of all boreholes, especially W2 and XO5, experienced obvious sharp drops. The heavy precipitation from July to September 2016 and 2021 caused the water level of boreholes in both mines to raise substantially

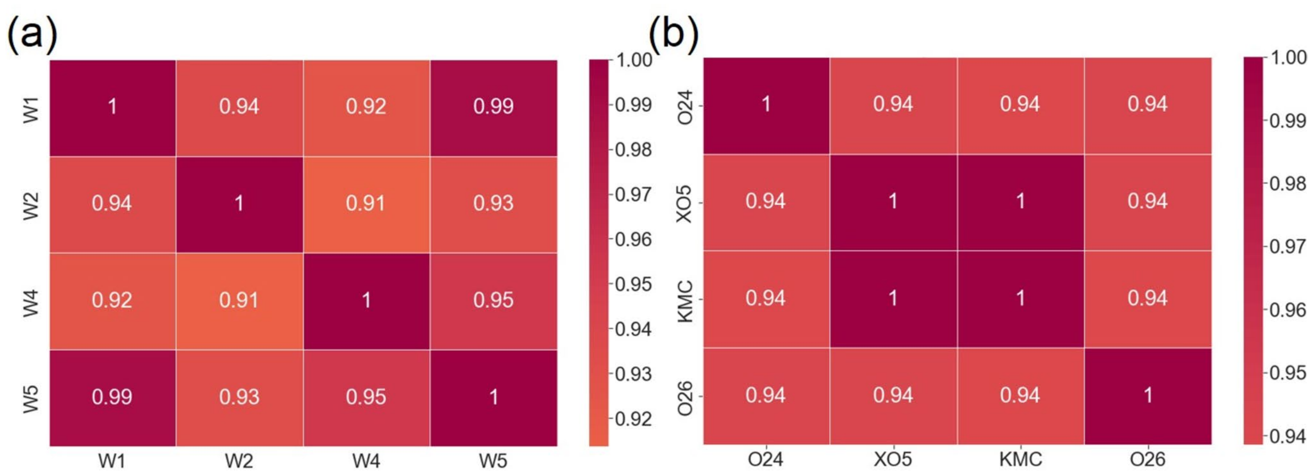

Fig. 3 Spearman correlation heatmap of water level data in borehole group of the **a** Xingdong and **b** Xin'an mines

are demonstrated using the pair plot (Fig. 4), which shows clear linear relationships amongst all of the variables in each group during their normal periods, and that the kernel density estimations of each variable are very unevenly distributed in both groups, indicating high similarity between the two datasets during the periods before the accidents. However, since the accidents took place when the variables in each group reached relatively large value (red scatters), the Source Dataset variables all decreased moderately and retained similar linear relationships, whereas the W2 variable in the Target Dataset drastically decreased and demonstrated distinctively different relationships with other variables, which indicates that the water level of any one of the boreholes may demonstrate abnormal and critical changes in disastrous scenarios. This diverse feature set can help the model learn more robust patterns, potentially leading to better generalization and improved performance for timely and accurate predictions for deep learning models.

In summary, although the Target Dataset is smaller in terms of sample size with a shorter observation period prior to the water inrush accident, both datasets are similar in terms of their statistical and variation features that are due to the common hydrogeological and working conditions of the mines, which suggests the high feasibility of applying transfer learning to alleviate data sparseness and improve predictive performance. In addition, using the data recorded during the major accident of the Source Dataset to pretrain the models could also improve their adaptiveness to future unseen data in disastrous scenarios like the accident of the Target Dataset.

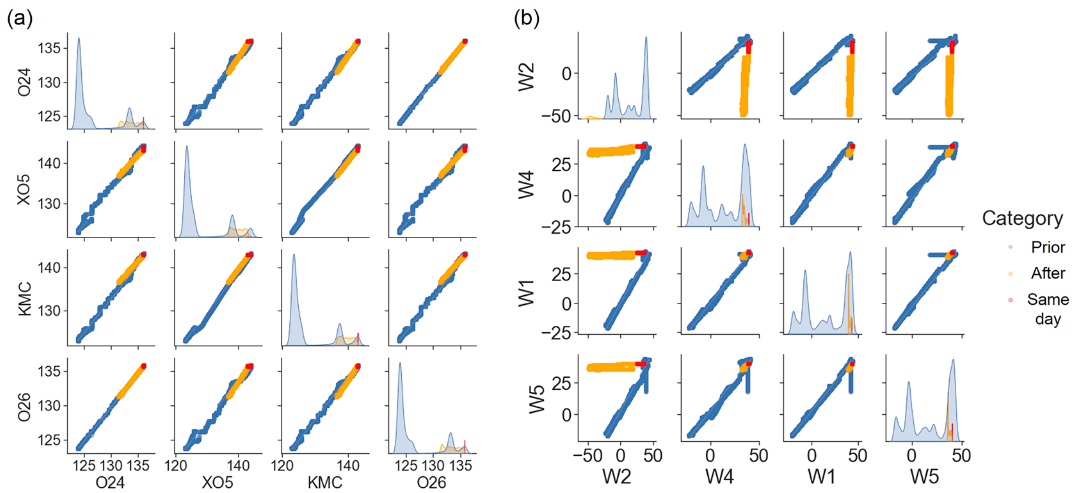

Fig. 4 Pairplot of water level data of the **a** Source Dataset and **b** Target Dataset. Blue, yellow, and red scatters respectively represent data prior to, after, and on the same day of the accident of each dataset. Plots on the diagonal lines are the kernel density estimations of each variable

Methodology

The analysis in this study was conducted following the procedures depicted in Fig. 5. The training and validation sets were first split from the Source Dataset with Accident I being included for pretraining of the models. Comparatively, the Target Dataset was also split into two sections with the first section to train the models, and the second containing Accident II, which was used as the testing set for the models. As comparisons for the Transformer based models on the testing set, we also examined the performance of a few frequently used MLP-, RNN-, and CNN-based prediction models including NBeats, LSTM, and TCN with identical input setups from the two datasets. Utilization of the Source Dataset for model training not only allowed the models to extract data variation features in rare and extreme scenarios like Accident I, but also provided a large amount of continuously recorded data over a relatively long period in a typical hydrogeological

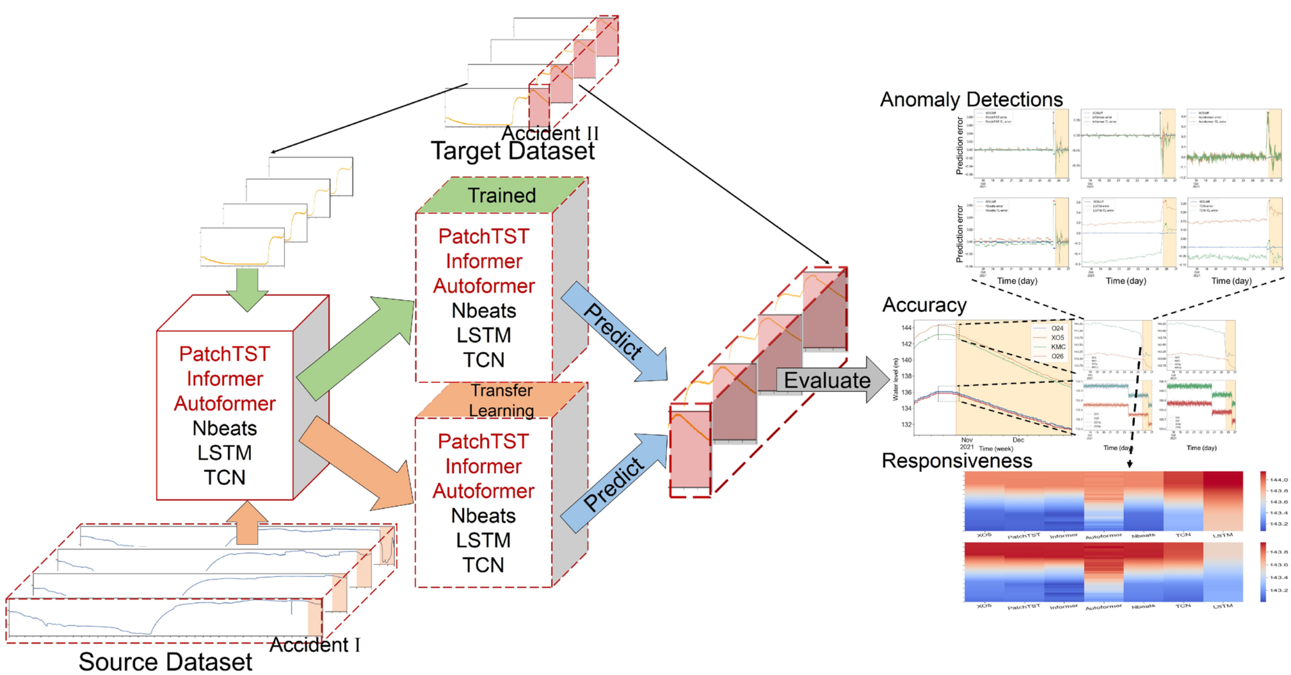

Fig. 5 Schematic diagram of the modeling process

condition of the mining area. This was especially vital to the transfer learning process to enable pretraining of the Transformer prediction models. Meanwhile, the Target Dataset was also comparatively used by itself as in normal cases where training, validation, and testing sets are continuous. Subsequently, the trained models were used to perform predictions on the testing set. The performance of the models was thoroughly evaluated for their accuracy and timeliness for capturing precursory information of the accident. Furthermore, the predictive errors of all of the models were studied to further understand how data anomalies can be used as early signs of potential accidents.

Transfer Learning

Transfer learning is a machine-learning technique where a model trained on one task is adapted for use on a second related task. It leverages the knowledge and representations learned during the training of the source model and applies them to the target task. This approach is particularly useful when the target task has a limited amount of data available, as it can benefit from the broader knowledge and generalizations of the source task (Kimura et al. 2020).

As a powerful technique for adapting machine-learning models on time-series forecasting tasks, transfer learning is able to reduce training time, improve predictive performance, and boost generalizations on the target task dataset (Ye and Dai 2021; Zhou et al. 2023b). The transfer learning process typically includes pre-training a model on a large-source dataset for a specific task and then applying it to a smaller related target dataset for feature extraction, testing, and evaluation (Pal and Kar 2022; Xu and Meng 2020). Although it is usually considered beneficial to fine-tune the models with the target dataset, the zero-shot learning capability of the models, which involves making predictions for unseen data during training, were prioritized in this study over few-shot learning for testing model adaptiveness, since gathering enough target data for adequate model generalization would be impractical in entirely novel disastrous scenarios. It would also require extra training time and computing resources before deployment, making it less effective for early warning of potential disasters.

Transformer-Based Models

The Transformer-based PatchTST, InFormer, and AutoFormer were the primary models used in this study to extract useful pre-accident information from the multivariate borehole group water level data. Transformer models employ a highly parallelization ability, multi-attention mechanism, and state-of-the art performance (Sitapure and Kwon 2023; Zeng et al. 2023a). This effectively and efficiently captures the highly complex long- and short-time data dependencies amongst multiple variables (Zhang et al. 2021) and leverages the valuable temporal patterns exhibited during disastrous scenarios to adapt to various specific prediction domains and applications, enabling the successful adoption of transfer learning for alleviating any potential data sparseness of the target dataset (Yildiz et al. 2022; Zerveas et al. 2021). In addition, Transformers can automatically learn relevant features from the multivariate data (Deihim et al. 2023), reducing the need for extensive manual feature engineering, which saves time and effort in preprocessing the data.

PatchTST

PatchTST is a recently developed Transformer-based model with state-of-the-art performance in multivariate time series prediction tasks (Nie et al. 2023). Because of its specially designed patching and channel independence mechanism, it is able to capture highly complex data variation patterns of multiple variables over long time-spans with an accuracy that is superior to other frequently used linear, DNN, MLP, and Transformer-based models.

PatchTST emphasizes channel independence, as each input to the Transformer backbone contains only the information of one channel or time series (Nie et al. 2023). As shown in Fig. 6, the multivariate time series is decomposed into single sequences, and each sequence is fed into the Transformer backbone. Afterwards, each series is predicted with the results concatenated to obtain the final predicted result, with attention weights applied separately to each channel, allowing the distinct features and patterns of each channel to be better captured by the model. In addition, a unique technique known as patching in the Transformer backbone (supplemental Fig. S-3) improves the model's accuracy and computational efficiency (Nie et al. 2023).

As shown in Fig. S-3, partitioning input sequences into smaller subsequences known as patches allows better modeling of localized correlations in long-range dependencies while eliminating memory constraints. Patching is used to divide each input time series, which allows it to independently analyze each input, providing effective and efficient in modeling complex patterns across the entire sequence.

InFormer

InFormer is also a Transformer-based model recently designed for time series forecasting (Zhou et al. 2020). It offers a combination of global and local attention mechanisms to effectively capture both long and short-term patterns in high-dimensional data (supplemental Fig. S-4; Zhou et al. 2020).

As shown in supplemental Fig. S-4, the InFormer model adopts a multi-scale encoder and decoder structure, which can simultaneously consider information at different time

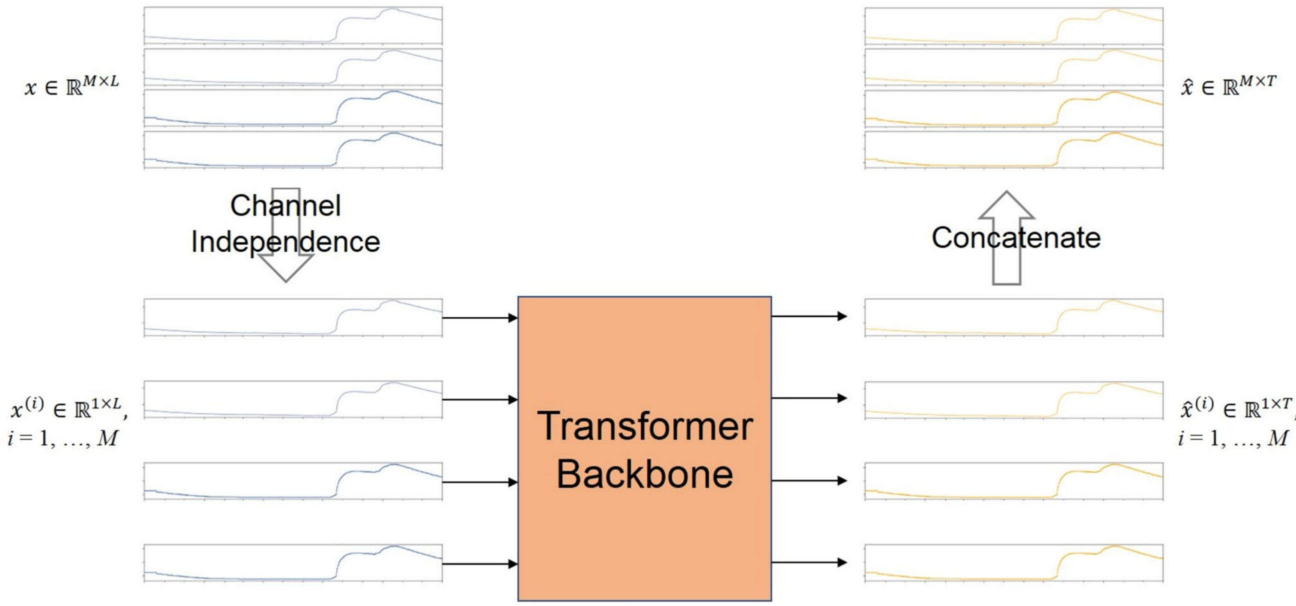

Fig. 6 Channel independence mechanism of PatchTST. x denotes the multivariate series with M variables, look back window size of L, and number of predicted values T; $x^{(i)}$ and $\hat{x}^{(i)}$ respectively denotes the ith input and output series

scales. In addition, an adaptive length attention mechanism called ProbSparse allows it to adjust the attention range based on the length of the sequence. The InFormer model's flexibility, parallel processing capabilities, and multi-scale attention mechanisms have made it a promising choice for multivariate time series forecasting tasks in many cases, where understanding and capturing complex temporal patterns are essential.

AutoFormer

AutoFormer is another recently designed time series prediction model that incorporates traditional time series decomposition and auto-correlation concepts into Transformer (Wu et al. 2021). Compared to ordinary Transformer structures, AutoFormer decomposes the original input data into seasonal and trend-cyclical components to be processed by its Decoder structure. AutoFormer is known for its highly predictive performances on time series with periodic variations.

Results Comparison and Discussion

Pre-training

First, the input and output forms of the prediction models were configured for data preparation. Both the input and output were multivariate, containing the four variables of either the Source Dataset or Target Dataset, since the water levels in the same borehole group are usually analyzed synchronously for geological purposes and it is uncertain which variable would exhibit data anomalies during accidents. A prediction horizon (the time span of the prediction on future data) of at least 5 min of water level data is usually required for any early warning of potential water inrush accidents to be practical. As the water level was being sampled every 5 min, the input sequence length of 12 time-steps (data samples in the past hour) for the prediction of one time-step (data sample in the next 5 min) has proven to produce good results in similar applications (Yin et al. 2023a). However, we extended the prediction horizon to 30 min (six time-steps in this study) to further examine the performance of the models for capturing early signs of potential accidents and allow additional time for responsive decisions to be made. As a result, the input sequence length was proportionally increased to 72 time-steps (data samples in the past 3 h). Additionally, the MinMax scaler of the training set of each dataset was applied throughout the respective dataset to reduce the impact of outliers and improve model performance, since Minmax scaler is better than other scalers such as the standard scaler for handling non-Gaussian distributions for a bounded output range suitable for neural network models.

For pretraining, the Source Dataset was split into a training and validation set at the cutoff of 15 March 2018, which is when Accident I roughly ceased. This allowed the data during the occurrence of Accident I to be included in the training set while preserving the continuity of the time series, which increased data feature diversity and was beneficial for model training. Considering the sample size of the entire dataset was very large, it also allows enough data to be included in the validation set.

Recent advancements in neural network models for multivariate time series prediction include the MLP-based Nbeats, RNN-based LSTM, and CNN-based TCN, all of which have shown effectiveness in various applications. These models were compared with the Transformer models in terms of prediction accuracy, timeliness, and anomaly detection. While suitable for transfer learning, these models often require fine-tuning on target datasets, which is less practical for this study. Nbeats uses a stacking mechanism for efficient parallel data processing, LSTM captures long-term sequential dependencies, and TCN offers high parallelization and speed, making them all appropriate for forecasting multiple related water level time series in this study. Along with these comparative models, the Transformer based PatchTST, InFormer, and AutoFormer models were configured following their earlier setups with minor adjustments based on proven promising performances in generic or similar datasets in existing studies, as shown in Table 2.

For the Transformer-based models, n_head is the number of multi-head's attention, hidden_size is the units of embeddings and encoders, ProbSparse is the number of ProbSparse attention factors, conv_ and linear_ hidden_size is respectively the channels of the convolutional encoder and linear layer. The Adam optimizer and the batch size of 32 were used for all the comparative models. The learning rate of 1×10^{-3} and the loss function of the mean absolute error (MAE) were universally used for all the models. The MAE is defined as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_i - y_i\right| \times 100\% \tag{1}$$

where $\hat{y}_i$ and y_i is respectively the i-th predicted value and actual value, and N is the sample size. The MAE is a preferred loss function in similar cases (Zhang et al. 2023e) over other frequently used ones like mean squared error for its robustness in reflecting prediction errors, and the MAE of each model's output is also subsequently analyzed as the most direct and linear representation of prediction error in this study, which required model evaluation metrics to be less sensitive to few outputs' large errors.

The Transformer-based and comparative models were respectively trained for 100 and 50 epochs, both with an early stop patience of 10 epochs on the validation loss. The implementations were conducted using Python 3.7.9 in a Windows system on an Intel Core i9-13900K @ 3.00 GHz, 128 GB of RAM, and a Nvidia GeForce RTX 4090 for computation.

The variations of the loss function values vs. epochs are shown in Fig. 7. All of the models, especially PatchTST and Nbeats, achieved the convergence of training and validation loss with very low values. Meanwhile, the loss on the Target Dataset of each model generally converged earlier with higher values than corresponding losses on the Source Dataset, demonstrating potential signs of overfitting, which is a result of insufficient data and a lack of diversity of variation features in the Target Dataset.

As for efficiency, PatchTST, InFormer, and AutoFormer respectively had average computing times of 6.12, 36.2, and 138.43 s per epoch. Comparatively, Nbeats, LSTM, and TCN model had average respective computing times of 67, 126, and 141 s per epoch. This comparison highlights the extraordinary efficiency of PatchTST.

Table 2 Summary of model configurations and major parameters

Model	Configuration and major parameters
PatchTST	encoder_layers = 3, n_head = 16, hidden_size = 128, linear_hidden_size = 256, dropout = 0.1, patch_len = 32, stride = 16, activation = ReLU
InFormer	hidden_size = 128, n_head = 4, dropout = 0.05, ProbSparse = 3, Conv_hidden_size = 32, activation = GELU, encoder_layers = 2, decoder_layers = 1
AutoFormer	encoder_layers = 2, decoder_layers = 1, hidden_size = 128, n_head = 4, dropout = 0.05, ProbSparse = 3, conv_hidden_size = 32, activation = GELU
Nbeats	4 generic layers: 512 generic neurons, 30 generic stacks; 4 trend layers: 256 trend neurons, 3 trend stacks; 4 seasonal layers: 2048 seasonal neurons, 3 seasonal stacks
LSTM	1st LSTM layer: 64 neurons, ReLU activation, L2 Regularizers(0.01); 2nd LSTM layer: 32 neurons, ReLU activation, L2 Regularizers(0.01); 1 Dropout layer (0.2); 1 Dense layer: 4 neurons, linear activation
TCN	TCN layer: 64 filters, kernel size = 3, dilations[1, 2, 4, 8], ReLU activation; 1 Dropout layer (0.2); 1 Dense layer: 4 neurons, linear activation

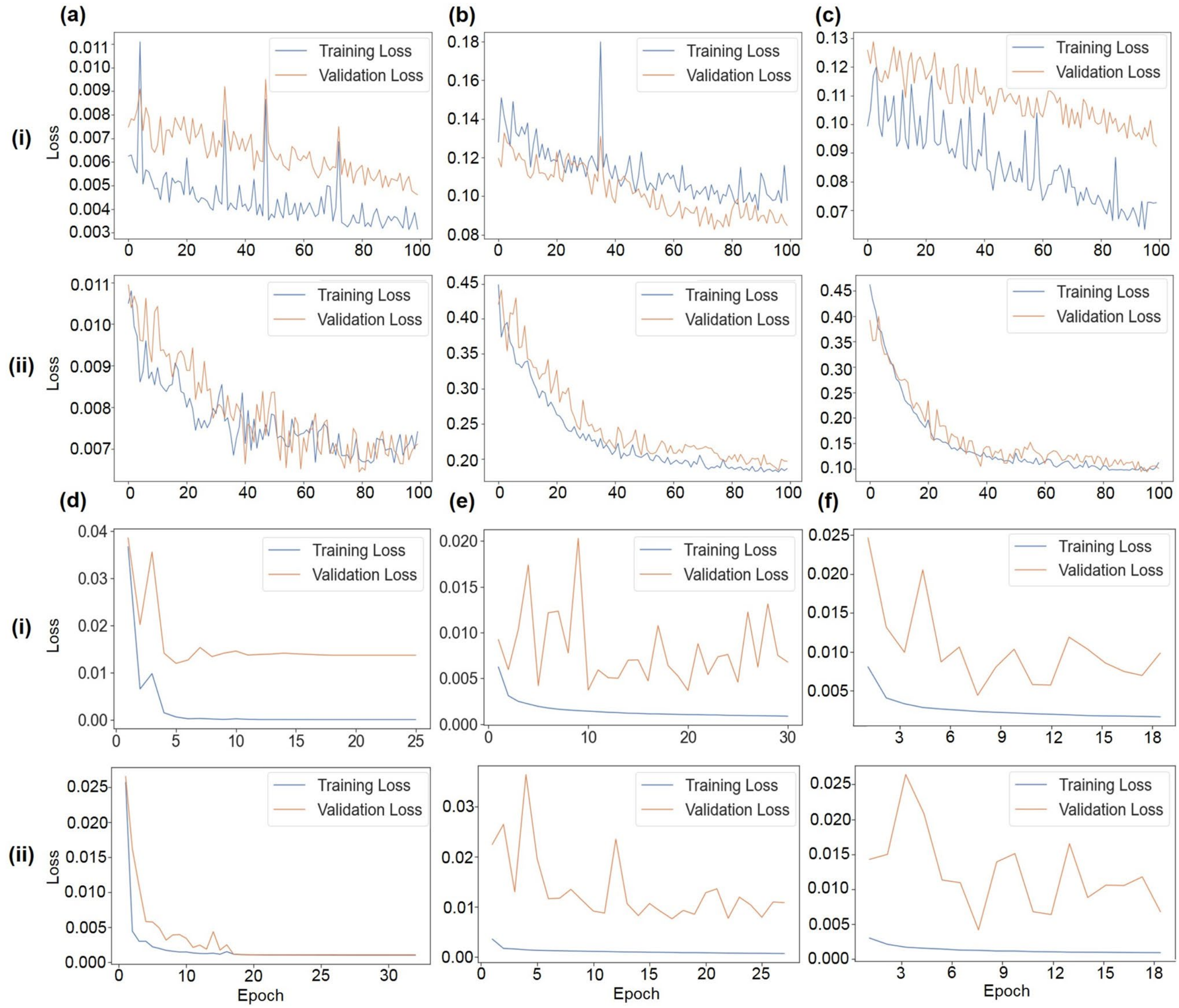

Fig. 7 Loss function variations of **a** PatchTST, **b** InFormer, **c** AutoFormer, **d** Nbeats, **e** LSTM, and **f** TCN when trained with the (i) Source Dataset and (ii) Target Dataset

Prediction Performance

Along with the pretrained models for transfer learning, models with identical configurations that were only trained on the Target Dataset was used to conduct water level predictions during the occurrence of Accident II. The Target Dataset was therefore split into a training, validation, and testing set using the cutoff dates of 1 August and 1 October 2021, allowing data from weeks prior to and during the accident to be included in the testing set, so that the model performances prior to and after the accident could be comparatively studied further. Although this caused the split ratio to be less than usually used values such as 8:2 or 7:3, the size of the validation set was roughly consistent with that of the Source Dataset, and the size of the training set was still acceptable for further comparative studies, considering the limited amount of data in the Target Dataset.

Accuracy

The prediction results of PatchTST on the testing set are displayed in Fig. 8. Both the performance of PatchTST with and without using transfer learning with the Source Dataset was satisfactory, although the model of transfer learning seemingly demonstrates slightly superior adaptiveness to the constant water-level fluctuations of O24 and O26. For comparison, the prediction results of all three Transformer-based models from 17 to 27 October on the testing set are shown in Fig. 9.

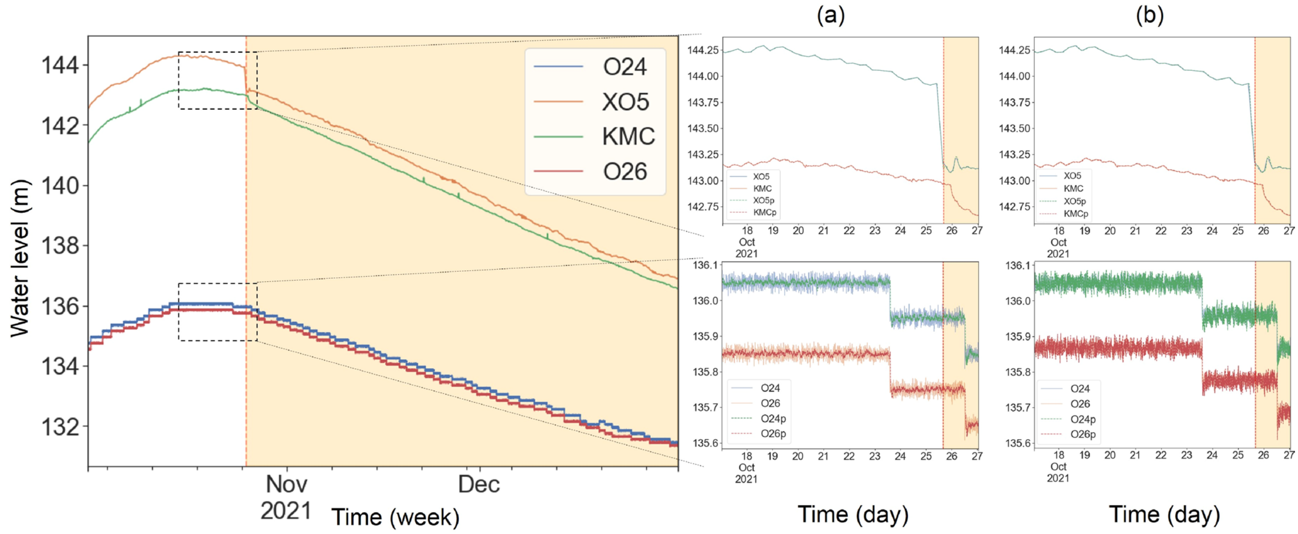

Fig. 8 Predictive results of PatchTST on the testing set of the Target Dataset **a** with and **b** without using transfer learning of the Source Dataset. Both results seem to be satisfactory, as the predicted values of XO5 and KMC water levels almost coincides with their actual values, except for minor lags exhibited moments prior to the accident (vertical dashed line), where XO5 variation abruptly decreased

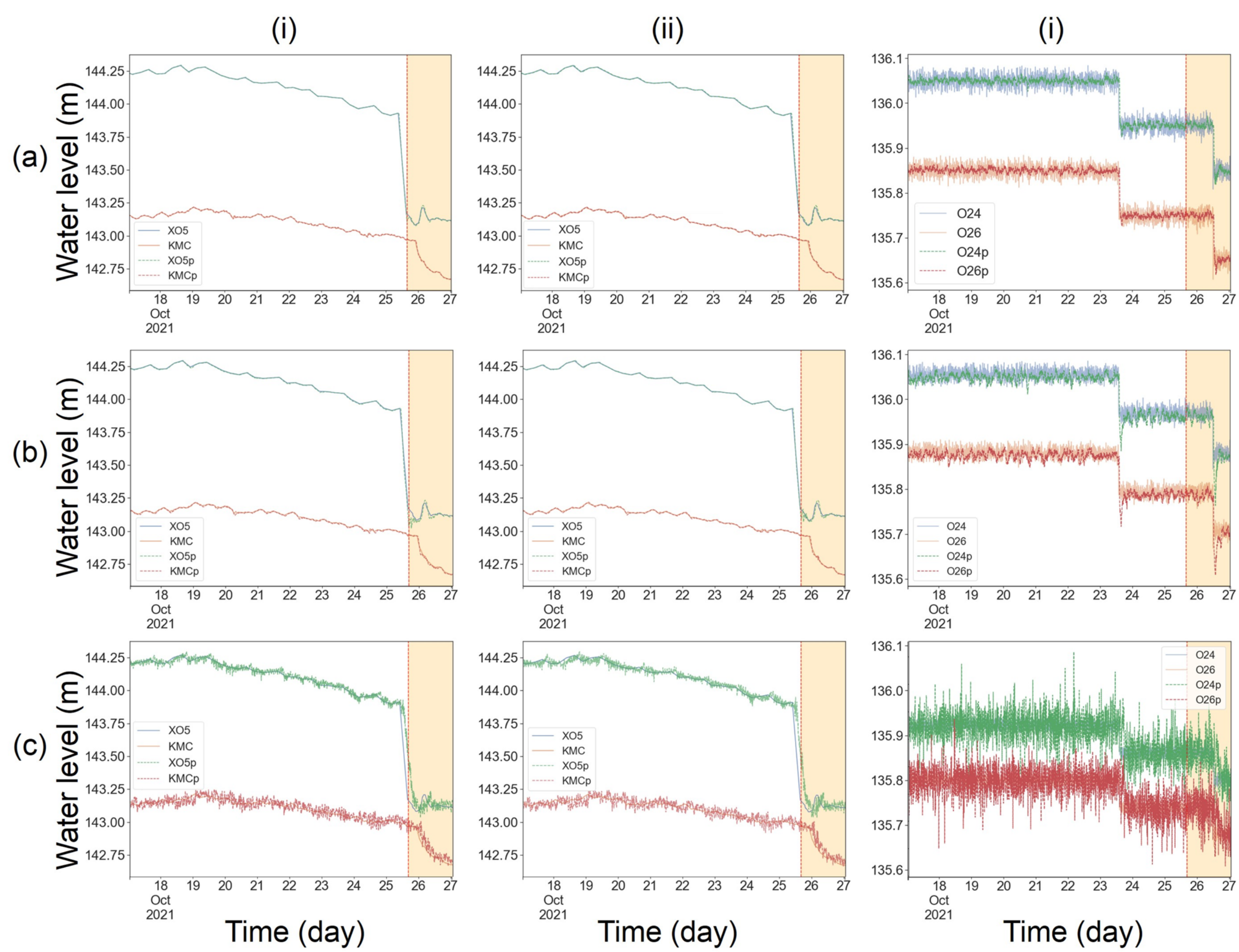

Fig. 9 Predictive results of **a** PatchTST, **b** InFormer, and **c** AutoFormer (i) with and (ii) without using transfer learning from the Source Dataset

Figure 9 shows that the performance of PatchTST and InFormer were promising while that of AutoFormer was less satisfactory as the periodicity feature of the data were over-emphasized. The performance statistics of these models on the entire testing set are shown in Table 3. The transfer learning using the Source Data gave PatchTST an obvious advantage in the prediction task, as it also outperformed the other two models. However, transfer learning had a limited effect of improving InFormer and AutoFormer's performance, resulting in higher metrics for InFormer on only the KMC borehole data.

A comparison of the prediction metrics of the comparative models on borehole XO5 and KMC water level data are shown in Table 4. Only Nbeats model's performance was on par with that of the Transformer based models, indicating that LSTM and TCN had less promising ability for modeling intercorrelated multiple variates with complex variation patterns. Nevertheless, transfer learning still played a vital role in improving TCN's performance, whereas it respectively had an almost negligible and a detrimental effect on Nbeats and LSTM. LSTM's ability to capture long-term dependencies in the source dataset may lead to overfitting for the task in the Target Dataset, which explains its less satisfying performance.

Table 3 Performance statistics of Transformer-based models

		XO5				KMC			
		MAE ($\times 10^{-3}$)	MSE ($\times 10^{-4}$)	RMSE ($\times 10^{-3}$)	R^2 ($\times 10^{-1}$)	MAE ($\times 10^{-3}$)	MSE ($\times 10^{-4}$)	RMSE ($\times 10^{-3}$)	R^2 ($\times 10^{-1}$)
PatchTST	TL	**2.8177**	**0.4260**	**6.5265**	**9.9999**	**2.6367**	**0.5646**	**7.5140**	**9.9999**
	W/O TL	3.6620	0.4992	7.0655	9.9999	3.3798	0.7089	8.4198	9.9998
InFormer	TL	6.4469	1.9377	13.920	9.9997	**4.7325**	**1.1124**	**10.547**	**9.9998**
	W/O TL	**4.2134**	**0.6934**	**8.3271**	**9.9999**	27.150	11.599	34.058	9.9974
AutoFormer	TL	**26.786**	**13.684**	**36.992**	**9.9976**	**26.656**	**10.802**	**32.867**	**9.9976**
	W/O TL	27.988	14.940	38.652	9.9974	27.150	11.599	34.058	9.9974
		O24				O26			
		MAE ($\times 10^{-2}$)	MSE ($\times 10^{-4}$)	RMSE ($\times 10^{-2}$)	R^2 ($\times 10^{-1}$)	MAE ($\times 10^{-2}$)	MSE ($\times 10^{-4}$)	RMSE ($\times 10^{-2}$)	R^2 ($\times 10^{-1}$)
PatchTST	TL	1.8645	6.4693	2.5435	9.9971	1.8219	6.2157	2.4931	9.9972
	W/O TL	**1.3114**	**4.0931**	**2.0231**	**9.9982**	**1.2745**	**3.8602**	**1.9647**	**9.9983**
InFormer	TL	1.8991	8.1114	2.8481	9.9964	1.8624	7.6720	2.7698	9.9966
	W/O TL	**1.5098**	**5.2990**	**2.3020**	**9.9977**	**1.4697**	**4.9813**	**2.2319**	**9.9978**
AutoFormer	TL	4.3131	32.854	5.7318	9.9854	4.2699	32.076	5.6636	9.9856
	W/O TL	**3.8997**	**27.032**	**5.1992**	**9.9880**	**3.8227**	**25.861**	**5.0853**	**9.9884**

TL stands for transfer learning. Underlined values are of the best performance out of all three models. Values in bold font represents superior performance of the same model with and without transfer learning

Table 4 Performance statistics of the compared models

		XO5				KMC			
		MAE ($\times 10^{-3}$)	MSE ($\times 10^{-4}$)	RMSE ($\times 10^{-3}$)	R^2 ($\times 10^{-1}$)	MAE ($\times 10^{-3}$)	MSE ($\times 10^{-4}$)	RMSE ($\times 10^{-3}$)	R^2 ($\times 10^{-1}$)
Nbeats	TL	3.8799	0.4830	6.9501	9.9999	**2.9364**	**0.42236**	**6.4989**	**9.9999**
	W/O TL	**3.5059**	**0.4808**	**6.9338**	**9.9999**	3.1898	0.44099	6.6407	9.9999
LSTM	TL	416.85	3015.1	549.10	9.4671	305.33	981.77	313.33	9.7833
	W/O TL	**282.57**	**913.07**	**302.17**	**9.8386**	**163.43**	**306.20**	**174.98**	**9.9324**
TCN	TL	**39.894**	**18.790**	**43.347**	**9.9967**	**53.577**	**52.691**	**72.589**	**9.9884**
	W/O TL	166.29	281.18	167.68	9.9503	95.775	98.587	99.291	9.9782

TL stands for transfer learning. Underlined values are of the best performance out of all three models. Values in bold font represents superior performance of the same model with and without transfer learning

Responsiveness

The delay of predicted values to actual values of borehole water levels in response to water inrush accidents is a major issue that affects decisions and judgement on early warnings of the accidents (Yin et al. 2023a). To address this issue, we first examined how the predicted values were distributed vs. the actual values during the days before and after the accidents (17–27 October). As shown in Fig. 10, although predicted values during the accident are distributed close to the diagonal line, positive errors of the predicted values occurred during the abrupt decrease of the water levels, indicating noticeable lagging of the predicted value variations to that of the actual value, despite the use of transfer learning. The variation feature maps of predicted and actual values demonstrate this effect (Fig. 11). Major shifts in the time dimension of the predicted values to the actual values are demonstrated by all three models. However, the transfer learning effectively alleviated this effect, indicating that training the model on the Source Dataset with the existing accident data improved the models' responsiveness to sudden water level decreases. In addition, the adaptiveness of the models to extreme data variation scenarios was also enhanced as transfer learning provided more diversified data variation features to be learned, which better served the models' generalization capabilities.

To further compare the Transformer-based models' temporal performance with other models, a comprehensive heatmap of predicted values were plotted using all of the studied models. Figure 12 shows that prediction delays of Nbeats and TCN were slightly more promising than PatchTST and InFormer when the models were merely trained on the Target Dataset. However, the predicted value variation using PatchTST was the closest to that of the actual values when transfer learning was used, indicating Transformer-based models could potentially benefit more from transfer learning in the studied prediction task.

Anomalies Exhibiting Features

Although delays of the variation of predicted value to actual values can be alleviated to a certain extent by applying transfer learning on Transformer-based models like PatchTST, it is unavoidable for any model to respond with delays when unexpected drastic data variation takes place. As a result, the detection of anomalous data variations prior to water inrush accidents are often relied on for early warnings. In recent studies, anomalous variations of borehole water level prediction errors using deep learning models was proven effective for reflecting unexpected data interferences as precursory signs of the accidents (Yin et al. 2023a). Nonetheless, such anomalies must be distinctively detectable prior to the accidents to be effective early warnings. Therefore, we observed the predicted errors of the studied models, which are given by $Error = \hat{y}_i - y_i$, along with the water level changing rate days before and during the accident.

Figure 13 shows that all of the prediction error curves exhibit clear spikes or level shifts following the sudden decrease of water level change rate, which allows distinctive detections of anomalies. In addition, the utilization of transfer learning has comparatively amplified such anomalous variations as corresponding spikes in the curves are generally of larger values. The prediction error peak times of each model in Fig. 13 are shown in Table 5, which shows that PatchTST's and InFormer's performance are the most promising, demonstrating their ability to reflect distinctive data anomalies almost 6 h prior to the actual occurrence of the accident. The utilization of transfer learning was beneficial in such performances for these two models, whereas its effects on the other models were either detrimental or negligible.

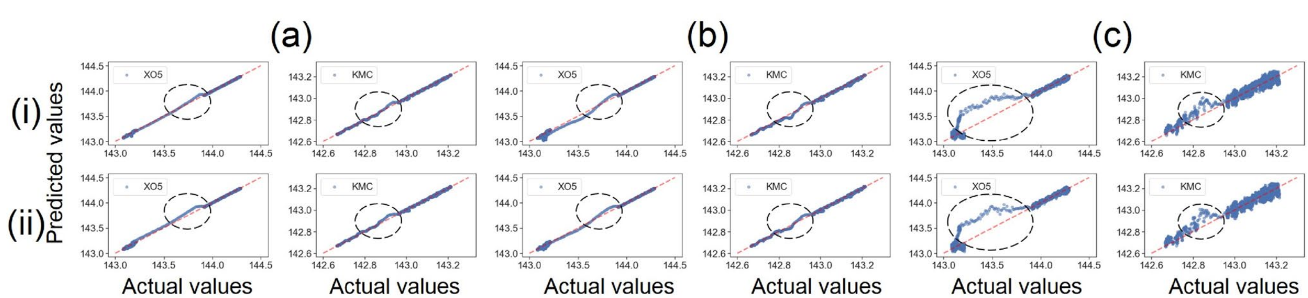

Fig. 10 Forecast evaluation plot of XO5 and KMC borehole water level days before and during the accident using **a** PatchTST, **b** InFormer, and **c** AutoFormer (i) with and (ii) without using transfer learning. All values distribute along the diagonal line in general, indicating satisfactory performances. However, as XO5 and KMC starts to abruptly decrease during the accident, from slightly less than 144.0 and 143.0 m, respectively, the predicted values exhibit noticeable discrepancies to its positive direction (dashed line circles), indicating variation delays to the actual values

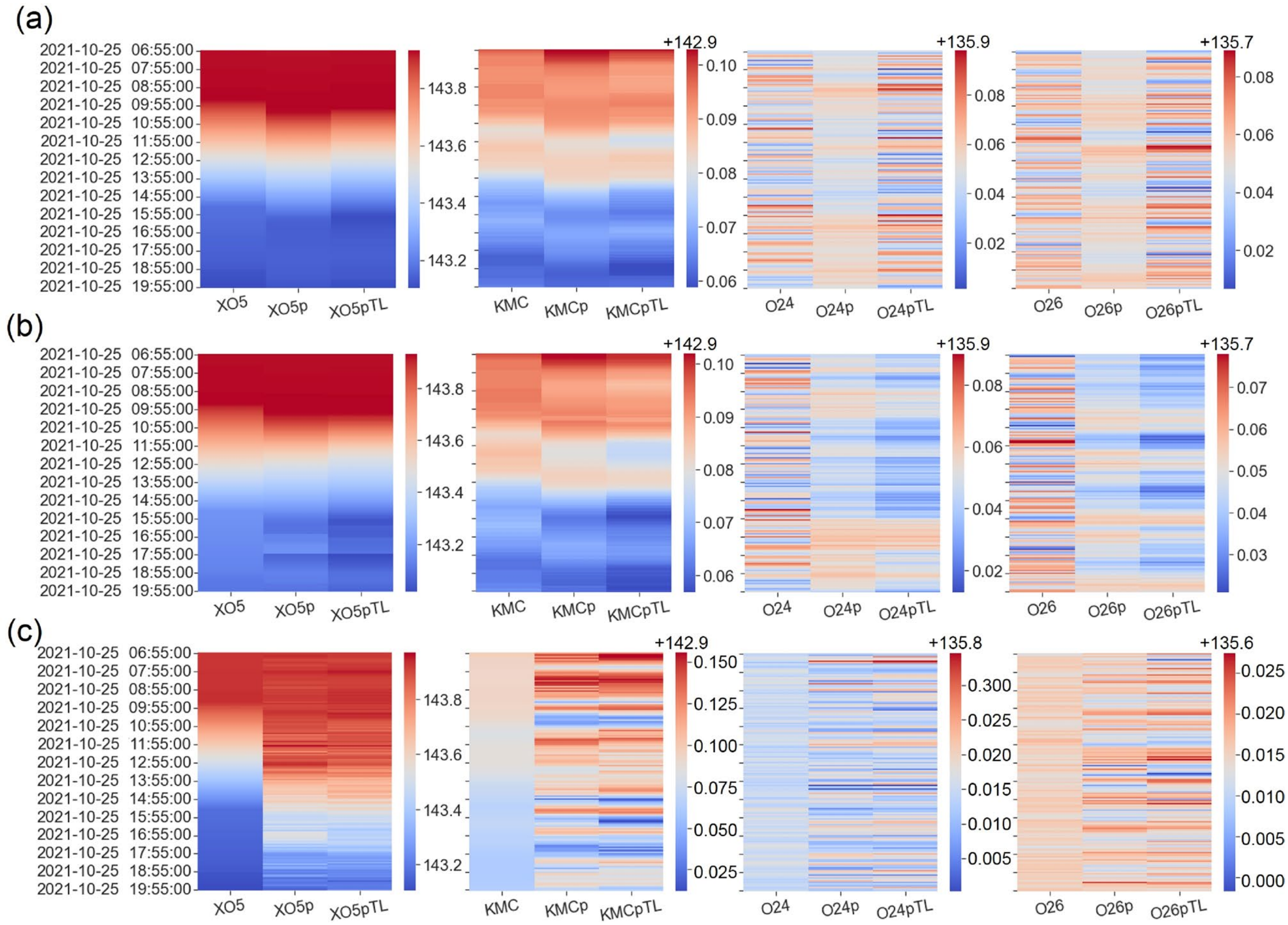

Fig. 11 Variation feature map of the actual and predicted values using: **a** PatchTST, **b** InFormer, and **c** AutoFormer during the hours prior to and during the occurrence of the accident. Strips of each map are, from left to right, the temporal variation of the actual value, the predicted value without, and with using transfer learning. The delaying effect on the predicted values of models with transfer learning are obviously less prominent

Method Limitations

Although promising results were obtained for the studied Transformer models, especially PatchTST, regarding the predictive accuracy, timeliness, and anomaly-detecting abilities for reflecting precursory information of water inrush accidents on a dataset unseen by the models, there are some limitations to this method. First, the success of transfer learning relies heavily on the similarity between the source and target dataset. The two borehole water level datasets being used in this study are statistically similar and have very close geological and engineering origins.

Despite giving a broader range of data collection for accident prediction, the method should also be used in further studies with careful consideration to the similarity between the source and target dataset along with their physical origins for its best performance. Therefore, we recommend that future studies prioritize collecting source datasets in nearby sites or at least in similar geological conditions to the target region. Second, the source dataset employed in this study was about three times the sample size of the target dataset and contained a diversified data variation feature that included variations during different seasons and a major accident. These all helped pretrain the models, whereas in practice, the quality of the source dataset can vary; major quality inconsistencies between the source and target dataset can lead to negative transfer, which is detrimental to model performance. Therefore, it could be beneficial to dynamically and periodically fine-tune or even retrain the models using more sophisticated techniques such as gradual unfreezing of neural network layers with newly observed target datasets under good monitoring conditions and computing resources, as this would further ensure that the latest data variation features are captured and learned for model adaptation and accurate prediction of future data.

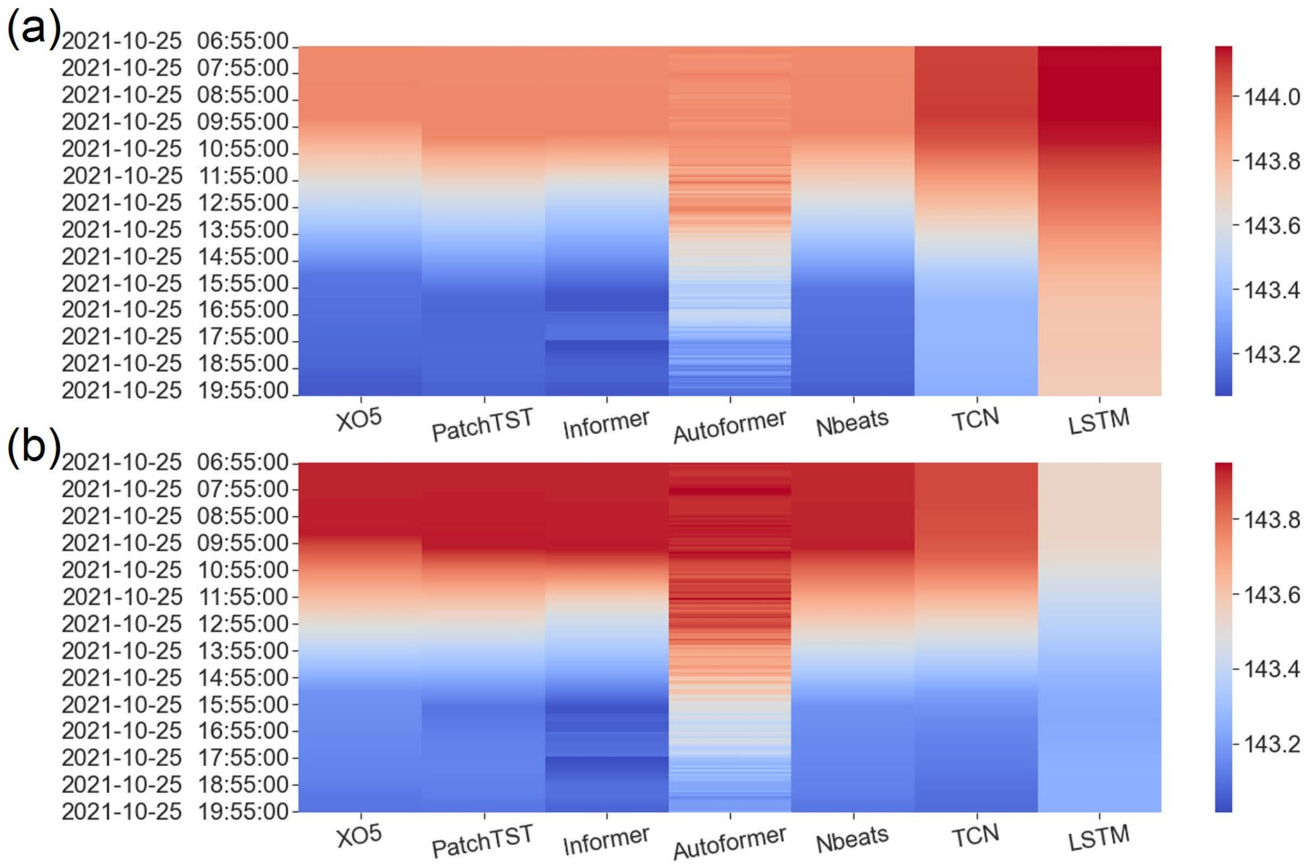

Fig. 12 Comparative variation feature map of predicted values using the studied models: **a** without transfer learning, and **b** with transfer learning. The 'XO5' column stands for the actual value. Nbeats and PatchTST respectively obtained the best results in both **a** and **b**

Conclusion

Along with several frequently used neural networks model including Nbeats, LSTM, and TCN, a selection of the latest Transformer models suitable for multivariate time series predictions were adopted in this study by employing the principle of transfer learning to explore their potential of using an existing field-observed borehole water level dataset with a large sample size and diversified variation features for predicting water inrush accidents in an unseen target dataset.

Compared to the other studied models, PatchTST produced the most promising results in terms of predictive accuracy, timeliness, and abilities to reflect data anomalies with prediction errors, and thereby improve mining safety and water inrush accident prevention. In addition, transfer learning boosted the studied Transformer models' performance in general, whereas its effect on enhancing other comparative models' performance were negligible or even detrimental. The innovative contributions of the study are summarized as follows:

- Transfer learning applied with Transformer models pretrained on a source dataset with abundant data features were shown to help alleviate data sparseness and imbalances in an unseen target dataset, thereby improving predictions of water inrush accidents.
- Model performance was evaluated in terms of multiple metrics to understand their predictive accuracy, timeliness, and ability to reflect anomalous data variations.
- Comparison of several representative neural networks models demonstrated the superior generalization and adaptive abilities of the Transformer models, especially PatchTST, after transfer learning was applied to boost their performance.

To promote the use of the proposed method for mine water inrush accident predictions, further validation of the method's effectiveness and reliability is recommended. Additional source datasets of various qualities and similarities to the target dataset could be experimented with for pretraining the models and to test their potential in cases where the target dataset is inadequate for the direct application of data-driven predictive models. Furthermore, the

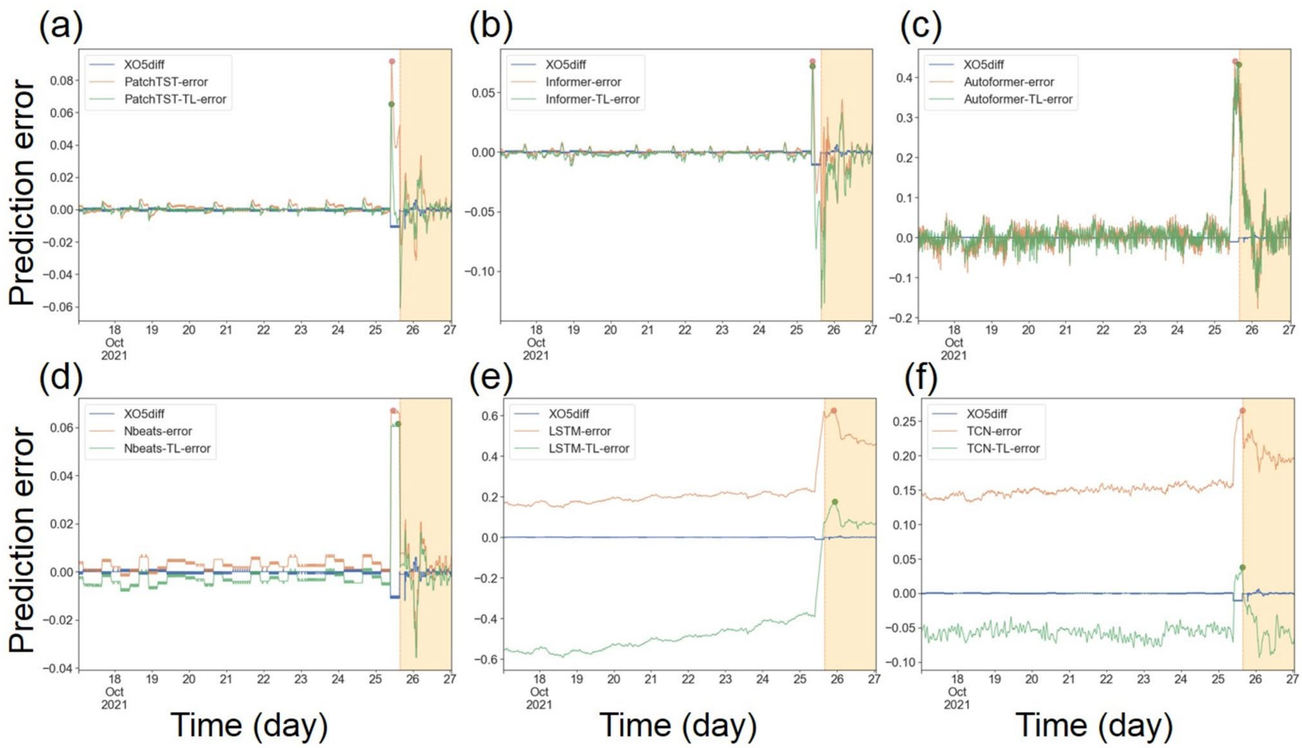

Fig. 13 Prediction errors of **a** PatchTST, **b** InFormer, **c** AutoFormer, **d** Nbeats, **e** LSTM, and **f** TCN models on the testing set from 17 to 27 October. The red vertical dashed line and yellow shade respectively stands for the starting time and period of accident occurrence. The blue curve represents the water level changing rate of borehole XO5. The green and orange curves respectively stand for prediction errors of models with and without transfer learning. The orange and green scatters are respectively the peak value transfer learning

Table 5 Prediction error peak times of each model prior to the accident

	Maximum value time	
	W/O TL	TL
PatchTST	2021-10-25 10:25:00	**2021-10-25 10:05:00**
InFormer	2021-10-25 10:20:00	**2021-10-25 10:10:00**
AutoFormer	**2021-10-25 13:00:00**	2021-10-25 15:30:00
Nbeats	**2021-10-25 11:15:00**	2021-10-25 14:20:00
LSTM	**2021-10-25 21:45:00**	2021-10-25 22:30:00
TCN	2021-10-25 15:30:00	2021-10-25 15:30:00

The bold texts represent better performance of the same model with and without transfer learning. Underlined text represent the best performance in each column

proposed method has the potential to be implemented in related fields such as hydrological modeling, geotechnical, and seismic data analysis, where multivariate time series prediction tasks also face challenges from data sparseness and imbalance. To obtain satisfactory results, data quality, and consistency, careful adaptation of the pre-trained models to the targeted tasks, and adequate consideration of computational resources is necessary for accurate and timely predictions of various geohazards.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10230-024-01011-2.

Acknowledgements This work was jointly supported by the National Natural Science Foundation of China (no. 42302294, no. 51974126, no. 51774136) and the Program for Innovative Research Team in University sponsored by Ministry of Education of China (no. IRT-17R37).

Data availability Data will be made available on request.

Declarations

Conflict of Interests 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.

References

Ahmadzadeh A, Aydin B, Georgoulis MK, Kempton DJ, Mahajan SS, Angryk RA (2021) How to train your flare prediction model: revisiting robust sampling of rare events. Astrophys J Suppl Ser 254(2):23. https://doi.org/10.3847/1538-4365/abec88

Asahi S, Karadogan C, Tamura S, Hayamizu S, Liewald M (2021) Process data based estimation of tool wear on punching machines using TCN-Autoencoder from raw time-series information. In: Proc IOP conf series-materials science and engineering. https://doi.org/10.1088/1757-899X/1157/1/012078

Bao Y (2019) Study on preventive measures for serious water inrush accident in underground tunnel engineering—the case of water inrush in a tunnel project. In: Proc. 6th annual international conf on material science and environmental engineering. https://doi.org/10.1088/1757-899X/472/1/012073

Cao Z, Gu Q, Huang Z, Fu J (2022) Risk assessment of fault water inrush during deep mining. Int J Min Sci Technol 32(2):423–434. https://doi.org/10.1016/j.ijmst.2022.01.005

Chen Y, Liu R, Barrett D, Gao L, Zhou MW, Renzullo L, Emelyanova I (2015) A spatial assessment framework for evaluating flood risk under extreme climates. Sci Total Environ 538:512–523. https://doi.org/10.1016/j.scitotenv.2015.08.094

Chen Y, Tang L, Zhu S (2022) Comprehensive study on identification of water inrush sources from deep mining roadway. Environ Sci Pollut Res 29(13):19608–19623. https://doi.org/10.1007/s11356-021-16703-3

Cho Y, Yun J, Kwon J, Kim Y (2023) Domain-adaptive vision transformers for generalizing across visual domains. IEEE Access 11:115644–115653. https://doi.org/10.1109/access.2023.3324545

Choi W, Pyun S, Cheon DS (2023) Automatic microseismic signal classification for mining safety monitoring using the WaveNet classifier. Geophys Prospect 72(2):315–332. https://doi.org/10.1111/1365-2478.13398

Cui F, Wu Q, Shuai Z, Wu N, Yuan J (2018) Damage characteristics and mechanism of a strong water inrush disaster at the Wangjialing Coal Mine, Shanxi Province, China. Geofluids 2018:3253641. https://doi.org/10.1155/2018/3253641

Cui Y, Qiu Y, Sun L, Shu X, Lu Z (2022) Quantitative short-term precipitation model using multimodal data fusion based on a cross-attention mechanism. Remote Sens 14(22):5839. https://doi.org/10.3390/rs14225839

Cui W, Ding J, Meng G, Lv Z, Feng Y, Wang A, Wan X (2023) Fault diagnosis of rolling bearings in primary mine fans under sample imbalance conditions. Entropy 25(8):1233. https://doi.org/10.3390/e25081233

Dash AK, Bhattacharjee RM, Paul PS (2016) Lessons learnt from Indian inundation disasters: an analysis of case studies. Int J Disaster Risk Reduct 20:93–102. https://doi.org/10.1016/j.ijdrr.2016.10.013

Deihim A, Alonso E, Apostolopoulou D (2023) STTRE: a spatio-temporal transformer with relative embeddings for multivariate time series forecasting. Neural Netw 168:549–559. https://doi.org/10.1016/j.neunet.2023.09.039

Dey P, Chaulya SK, Kumar S (2021) Hybrid CNN-LSTM and IoT-based coal mine hazards monitoring and prediction system. Process Saf Environ Prot 152:249–263. https://doi.org/10.1016/j.psep.2021.06.005

Dong DL, Zhang JL (2023) Discrimination methods of mine inrush water source. Water 15(18):3237. https://doi.org/10.3390/w15183237

Dong S, Zheng L, Tang S, Shi P (2020) A scientometric analysis of trends in coal mine water inrush prevention and control for the period 2000–2019. Mine Water Environ 39(1):3–12. https://doi.org/10.1007/s10230-020-00661-2

Fang B (2022) Method for quickly identifying mine water inrush using convolutional neural network in coal mine safety mining. Wirel Pers Commun 127(2):945–962. https://doi.org/10.1007/s11277-021-08452-w

Gong X, Liao S, Hu F, Hu X, Liu C (2022) Autoencoder-based anomaly detection for time series data in complex systems. In: Proc IEEE Asia Pacific conf on circuits and systems (APCCAS), So Univ Sci & Technol, Shenzhen, China

Gu Q, Huang Z, Li S, Zeng W, Wu Y, Zhao K (2020) An approach for water-inrush risk assessment of deep coal seam mining: a case study in Xinlongzhuang coal mine. Environ Sci Pollut Res 27(34):43163–43176. https://doi.org/10.1007/s11356-020-10225-0

Han YB, Wang QQ, Li WP, Yang Z, Gu TY, Wang ZK (2023) Predicting the height of the water-conducting fractured zone in fully mechanized top coal caving longwall mining of very thick Jurassic coal seams in western China based on the NNBR model. Mine Water Environ 42(1):121–133. https://doi.org/10.1007/s10230-023-00918-6

He Y, Yan H, Yang W, Yao S, Zhang L, Chen Y, Liu T (2022) Time-series analysis and prediction of surface deformation in the Jinchuan mining area, Gansu province, by using InSAR and CNN-PhLSTM network. IEEE J Sel Top Appl Earth Obs Remote Sens 15:6732–6751. https://doi.org/10.1109/jstars.2022.3198728

Hendrycks D, Liu X, Wallace E, Dziedzic A, Krishnan R, Song D (2020) Pretrained transformers improve out-of-distribution robustness. In: Proc. 58th annual meeting of the assoc for computational linguistics, electr network

Huang P, Yang Z, Wang X, Ding F (2019) Research on Piper-PCA-Bayes-LOOCV discrimination model of water inrush source in mines. Arab J Geosci 12(11):334. https://doi.org/10.1007/s12517-019-4500-3

Hungerford F, Ren T (2013) Drilling for inrush protection and water management. In: Proc. 3rd international workshop on mine hazards prevention and control. CSIRO Earth Sci & Resource Engn, Brisbane

Ji Y, Dong DL, Mei AS, Wei ZL (2022) Study on key technology of identification of mine water inrush source by PSO-LightGBM. Water Supply 22(10):7416–7429. https://doi.org/10.2166/ws.2022.323

Kang M, Hua D, Guo X (2023) Review on the influence of complex stratum on the drilling trajectory of the drilling robot. Appl Sci Basel 13(4):2532. https://doi.org/10.3390/app13042532

Kimura N, Yoshinaga I, Sekijima K, Azechi I, Baba D (2020) Convolutional neural network coupled with a transfer-learning approach for time-series flood predictions. Water 12(1):96. https://doi.org/10.3390/w12010096

Li X, Dong S, Liu K (2021) Prevention and control of water inrushes from subseam karstic Ordovician limestone during coal mining above ultra-thin aquitards. Mine Water Environ 40(2):345–356. https://doi.org/10.1007/s10230-021-00765-3

Li X, Dong DL, Liu K, Zhao Y, Li MM (2022) Identification of mine mixed water inrush source based on genetic algorithm and XGBoost algorithm: a case study of Huangyuchuan mine. Water 14(14):2150. https://doi.org/10.3390/w14142150

Li J, Xu Q, Hu Y, Chen X (2023) Evaluation of control effect of confined water hazard in Taiyuan formation coal seam mining in Huanghebei coalfield. Water 15(11):1973. https://doi.org/10.3390/w15111973

Li W, Wang L, Ye Z, Liu Y, Wang Y (2024) A dynamic combination algorithm based scenario construction theory for mine water-inrush accident multi-objective optimization. Expert Syst Appl 238:121871. https://doi.org/10.1016/j.eswa.2023.121871

Lin G, Jiang D, Dong DL, Fu JY, Li X (2021) A multilevel recognition model of water inrush sources: a case study of the Zhaogezhuang mining area. Mine Water Environ 40(3):773–782. https://doi.org/10.1007/s10230-021-00793-z

Liu B, Malekian R, Xu JP (2017) Groundwater mixing process identification in deep mines based on hydrogeochemical property analysis. Appl Sci Basel 7(1):42. https://doi.org/10.3390/app7010042

Liu Y, Zhu J, Liu Q, Yuan A, He S, Bai Y (2022) Mechanism analysis of delayed water inrush from karst collapse column during roadway excavation based on seepage transition theory: a case

study in PanEr coal mine. Energies 15(14):4987. https://doi.org/10.3390/en15144987
Liu X, Zeng Y, Wu Q, Meng S, Liang J, Hou Z (2024) Ecological-based mining: a coal–water–thermal collaborative paradigm in ecologically fragile areas in western China. Engineering. https://doi.org/10.1016/j.eng.2024.01.019
Lyu S, Liu P, Liu L, Ma S, Wang T (2021) An improved dynamic process neural network prediction model identification method. Microprocess Microsyst 80:103573. https://doi.org/10.1016/j.micpro.2020.103573
Mahmoodzadeh A, Mohammadi M, Noori KMG, Khishe M, Ibrahim HH, Ali HFH, Abdulhamid SN (2021) Presenting the best prediction model of water inflow into drill and blast tunnels among several machine learning techniques. Autom Constr 127:103719. https://doi.org/10.1016/j.autcon.2021.103719
Miao W, Xu Y, Guo Y, Zhang E, Zhuo Y, Huang L, Ma Z, Liang S (2022) The hydrogeological characteristics of thick alluvium with high water level and the influence on Zhaogu mining area, Henan Province, China. Geofluids 2022:9447145. https://doi.org/10.1155/2022/9447145
Mu H, Xia W, Che W (2023) Improving domain generalization for sound classification with sparse frequency-regularized transformer. In: Proc. IEEE international conf on multimedia and expo, Brisbane
Nicholaus IT, Park JR, Jung K, Lee JS, Kang DK (2021) Anomaly detection of water level using deep autoencoder. Sensors 21(19):6679. https://doi.org/10.3390/s21196679
Nie Y, Nam Sinthong P, Kalagnanam J (2023) A time series is worth 64 words: long-term forecasting with transformers. arXiv. https://doi.org/10.48550/arXiv.2211.14730
Oreshkin BN, Dudek G, Pelka P, Turkina E (2021) N-BEATS neural network for mid-term electricity load forecasting. Appl Energy 293:116918. https://doi.org/10.1016/j.apenergy.2021.116918
Pal SS, Kar S (2022) Fuzzy transfer learning in time series forecasting for stock market prices. Soft Comput 26(14):6941–6952. https://doi.org/10.1007/s00500-021-06648-7
Prasad SS, Deo RC, Downs NJ, Casillas-Perez D, Salcedo-Sanz S, Parisi AV (2024) Very short-term solar ultraviolet-A radiation forecasting system with cloud cover images and a Bayesian optimized interpretable artificial intelligence model. Expert Syst Appl 236:121273. https://doi.org/10.1016/j.eswa.2023.121273
Qiao W, Li W, Zhang X, Niu Y, Chen Y, Wang Y, Xing T (2019) Prediction of floor water disasters based on fractal analysis of geologic structure and vulnerability index method for deep coal mining in the Yanzhou mining area. Geomat Nat Hazards Risk 10(1):1306–1326. https://doi.org/10.1080/19475705.2019.1574911
Qu X, Han J, Shi L, Bilal A, Qiu M, Gao W (2023) An extended ITL-VIKOR model using triangular fuzzy numbers for applications to water-richness evaluation. Expert Syst Appl 222:119793. https://doi.org/10.1016/j.eswa.2023.119793
Sahin U, Kucukkaya IE, Ozcelik O, Toraman C (2023) Zero and few-shot hate speech detection in social media messages related to earthquake disaster. In: Proc. 31st IEEE conf on signal processing and communications applications
Shamsabadi EA, Xu C, Dias-da-Costa D (2022) Robust crack detection in masonry structures with transformers. Measurement 200:111590. https://doi.org/10.1016/j.measurement.2022.111590
Shen X, Dai Q, Ullah W (2023) An active learning-based incremental deep-broad learning algorithm for unbalanced time series prediction. Inf Sci 642:119103. https://doi.org/10.1016/j.ins.2023.119103
Singh KKK (2015) MineVue radar for delineation of coal barrier thickness in underground coal mines: case studies. J Geol Soc India 85(2):247–253. https://doi.org/10.1007/s12594-015-0211-x
Sitapure N, Kwon JS-I (2023) CrystalGPT: enhancing system-to-system transferability in crystallization prediction and control using time-series-transformers. Comput Chem Eng 177:108339. https://doi.org/10.1016/j.compchemeng.2023.108339
Skaf W, Horváth T (2022) Denoising architecture for unsupervised anomaly detection in time-series. In: Proc. 26th European conf on advances in databases and information systems. Politecnico Torino, Turin
Su MX, Liu YM, Li HY, Li CC, Wang P, Ma XY (2021) Water inrush detection in limestone pit rock walls using borehole-to-surface ERT. Mine Water Environ 40(4):1061–1072. https://doi.org/10.1007/s10230-021-00809-8
Su A, Wang X, Wang L, Zhang C, Wu Y, Wu X, Zhao Q, Duan H (2022) Reproducing the invention of a named reaction: zero-shot prediction of unseen chemical reactions. Phys Chem Chem Phys 24(17):10280–10291. https://doi.org/10.1039/d1cp05878a
Sun Z, Li J (2022) Anomaly detection for CPS via memory-augmented reconstruction and time series prediction. In: Proc. 19th IEEE international conf on mobile ad hoc and smart systems (MASS), Denver, CO
Sun Z, Bao W, Li M (2022) Comprehensive water inrush risk assessment method for coal seam roof. Sustainability 14(17):10475. https://doi.org/10.3390/su141710475
Teng JW, Li SY, Jia MK, Lian J, Liu HL, Liu GD, Wang W, Schäpe V, Feng L, Yao XS, Wang K, Yan YF, Zhang WP (2020) Research and application of in-seam seismic survey technology for disaster-causing potential geology anomalous body in coal seam. Acta Geol Sin Engl 94(1):10–26. https://doi.org/10.1111/1755-6724.14372
Thomas G, Sahay R, Jahan CS, Manjrekar M, Popp D, Savakis A (2023) On the importance of attention and augmentations for hypothesis transfer in domain adaptation and generalization. Sensors 23(20):8409. https://doi.org/10.3390/s23208409
Toufigh V, Ranjbar I (2023) Unsupervised deep learning framework for ultrasonic-based distributed damage detection in concrete: integration of a deep auto-encoder and Isolation Forest for anomaly detection. Struct Health Monit Int J. https://doi.org/10.1177/14759217231183143
Wahyuningsih S, Goejantoro R, Rizki NA (2017) Forecasting hotspots in East Kutai, Kutai Kartanegara, and West Kutai as early warning information. In: Proc. 1st international conf on tropical studies and its application
Wang D, Sui W, Ranville JF (2022) Hazard identification and risk assessment of groundwater inrush from a coal mine: a review. Bull Eng Geol Environ 81(10):421. https://doi.org/10.1007/s10064-022-02925-3
Wei ZL, Dong DL, Ji Y, Ding J, Yu LJ (2022) Source discrimination of mine water inrush using multiple combinations of an improved support vector machine model. Mine Water Environ 41(4):1106–1117. https://doi.org/10.1007/s10230-022-00884-5
Wu Z, Wang C (2022) Limiting spectral distribution of large dimensional Spearman's rank correlation matrices. J Multivar Anal 191:105011. https://doi.org/10.1016/j.jmva.2022.105011
Wu Q, Guo X, Shen J, Xu S, Liu S, Zeng Y (2017) Risk assessment of water inrush from aquifers underlying the Gushuyuan coal mine, China. Mine Water Environ 36(1):96–103. https://doi.org/10.1007/s10230-016-0410-8
Wu H, Xu J, Wang J, Long M (2021) Autoformer: decomposition transformers with auto-correlation for long-term series forecasting. arXiv. https://doi.org/10.48550/arXiv.2106.13008
Xu X, Meng Z (2020) A hybrid transfer learning model for short-term electric load forecasting. Electr Eng 102(3):1371–1381. https://doi.org/10.1007/s00202-020-00930-x
Xu Z, Sun Y, Gao S, Chen H, Yao M, Li X (2021) Comprehensive exploration, safety evaluation and grouting of karst collapse columns in the Yangjian coalmine of the Shanxi province,

China. Carbonates Evaporites 36(1):16. https://doi.org/10.1007/s13146-021-00675-z

Yan PC, Shang SH, Zhang CY, Yin NN, Zhang XF, Yang GK, Zhang Z, Sun QS (2021a) Research on the processing of coal mine water source data by optimizing BP neural network algorithm with sparrow search algorithm. IEEE Access 9:108718–108730. https://doi.org/10.1109/access.2021.3102020

Yan Q, Lu Z, Wang P, Ding X, Cheng F, Zhang Y (2021b) A new method for anomaly detection and diagnosis of ocean observation system based on deep learning. In: Proc. 40th Chinese control conf, Shanghai

Yang Y, Yue J, Li J, Yang Z (2018) Mine water inrush sources online discrimination model using fluorescence spectrum and CNN. IEEE Access 6:47828–47835. https://doi.org/10.1109/access.2018.2866506

Yang S, Lian H, Xu B, Thanh HV, Chen W, Yin H, Dai Z (2023) Application of robust deep learning models to predict mine water inflow: implication for groundwater environment management. Sci Total Environ 871:162056. https://doi.org/10.1016/j.scitotenv.2023.162056

Yang S, Lian H, Soltanian MR, Xu B, Liu W, Thanh HV, Li Y, Yin H, Dai Z (2024) Hybrid approach for early warning of mine water: energy density based identification of water-conducting channels combined with water inflow prediction by SA-LSTM. IEEE Trans Geosci Remote Sens 62:1–12. https://doi.org/10.1109/TGRS.2024.3384990

Ye R, Dai Q (2021) Implementing transfer learning across different datasets for time series forecasting. Pattern Recognit 109:107617. https://doi.org/10.1016/j.patcog.2020.107617

Yedla A, Kakhki FD, Jannesari A (2020) Predictive modeling for occupational safety outcomes and days away from work analysis in mining operations. Int J Environ Res Public Health 17(19):7054. https://doi.org/10.3390/ijerph17197054

Yildiz AY, Koc E, Koc A (2022) Multivariate time series imputation with transformers. IEEE Signal Process Lett 29:2517–2521. https://doi.org/10.1109/lsp.2022.3224880

Yin H, Zhou W, LaMoreaux JW (2018) Water inrush conceptual site models for coal mines of China. Environ Earth Sci 77(22):746. https://doi.org/10.1007/s12665-018-7920-6

Yin H, Xu B, Yin S, Tian W, Yao H, Meng H (2021) Prevention of water inrushes in deep coal mining over the Ordovician aquifer: a case study in the Wutongzhuang coal mine of China. Geofluids 2021:5208670. https://doi.org/10.1155/2021/5208670

Yin H, Xu G, Zhang Y, Zhai P, Li X, Guo Q, Wei Z (2022) Risk assessment of water inrush of a coal seam floor based on the combined empowerment method. Water 14(10):1607. https://doi.org/10.3390/w14101607

Yin H, Wu Q, Yin S, Dong S, Dai Z, Soltanian MR (2023a) Predicting mine water inrush accidents based on water level anomalies of borehole groups using long short-term memory and isolation forest. J Hydrol 616:128813. https://doi.org/10.1016/j.jhydrol.2022.128813

Yin H, Zhang G, Wu Q, Yin S, Soltanian MR, Thanh HV, Dai Z (2023b) A deep learning-based data-driven approach for predicting mining water inrush from coal seam floor using microseismic monitoring data. IEEE Trans Geosci Remote Sens 61:1–15. https://doi.org/10.1109/tgrs.2023.3300012

Yokkampon U, Chumkamon S, Mowshowitz A, Fujisawa R, Hayashi E (2020) Improved variational autoencoder anomaly detection in time series data. In: Proc. IEEE international conf on systems, man, and cybernetics, pp 82–87. https://doi.org/10.1109/SMC42975.2020.9283010

Yokkampon U, Mowshowitz A, Chumkamon S, Hayashi E (2022) Robust unsupervised anomaly detection with variational autoencoder in multivariate time series data. IEEE Access 10:57835–57849. https://doi.org/10.1109/access.2022.3178592

Yu S, Xu J, Zhu W, Wang S, Liu W (2020) Development of a combined mining technique to protect the underground workspace above confined aquifer from water inrush disaster. Bull Eng Geol Environ 79(7):3649–3666. https://doi.org/10.1007/s10064-020-01803-0

Zeng C, Huang J, Wang H, Xie J, Huang S (2022) Rail break prediction and cause analysis using imbalanced in-service train data. IEEE Trans Instrum Meas 71:3527114. https://doi.org/10.1109/tim.2022.3214494

Zeng P, Hu G, Zhou X, Li S, Liu P (2023a) Seformer: a long sequence time-series forecasting model based on binary position encoding and information transfer regularization. Appl Intell 53(12):15747–15771. https://doi.org/10.1007/s10489-022-04263-z

Zeng Y, Meng S, Wu Q, Mei A, Bu W (2023b) Ecological water security impact of large coal base development and its protection. J Hydrol 619:129319. https://doi.org/10.1016/j.jhydrol.2023.129319

Zeng Y, Mei A, Wu Q, Meng S, Zhao D, Hua Z (2024) Double verification and quantitative traceability: a solution for mixed mine water sources. J Hydrol 630:130725. https://doi.org/10.1016/j.jhydrol.2024.130725

Zerveas G, Jayaraman S, Patel D, Bhamidipaty A, Eickhoff C (2021) A transformer-based framework for multivariate time series representation learning. In: Proc. 27th ACM SIGKDD international conf on knowledge discovery and data mining. https://doi.org/10.1145/3447548.3467401

Zhang FD (2023) Dynamic monitoring and an early warning model of a floor water disaster: a case study. Mine Water Environ 42(1):158–169. https://doi.org/10.1007/s10230-023-00925-7

Zhang Q, Wang Z (2021) Spatial prediction of loose aquifer water abundance mapping based on a hybrid statistical learning approach. Earth Sci Inform 14(3):1349–1365. https://doi.org/10.1007/s12145-021-00640-3

Zhang H, Xia Y, Yan T, Liu G (2021) Unsupervised anomaly detection in multivariate time series through transformer-based variational autoencoder. In: Proc. 33rd Chinese control and decision conf. https://doi.org/10.1109/CCDC52312.2021.9601669

Zhang WR, Liu TX, Duan LM, Zhou SH, Long-Sun S, Shi ZM, Qu S, Bian MM, Yu DG, Singh VP (2022) Forecasting groundwater level of karst aquifer in a large mining area using partial mutual information and NARX hybrid model. Environ Res 213:113747. https://doi.org/10.1016/j.envres.2022.113747

Zhang D, Qian L, Meng M, Cui K (2023a) Research on the development regularity of Ordovician limestone karst in Liangzhuang coalfield. Carbonate Evaporite 38(3):53. https://doi.org/10.1007/s13146-023-00876-8

Zhang J, Wu J, Yang T, Yang S, He Y, Gao S (2023b) Analysis of water inrush disaster mechanism of inter-layer rocks between close coal seams under the influence of mining. Appl Sci Basel 13(15):9043. https://doi.org/10.3390/app13159043

Zhang J, Yang C, Liu SD, Li CY (2023c) Detection of geological anomalies in coal mining working faces using a scattered-wave imaging method. J Pet Explor Prod Technol. https://doi.org/10.1007/s13202-023-01619-0

Zhang W, Cui G, Wang Y, Zheng C, Zhu Q (2023d) A human comfort prediction method for indoor personnel based on time-series analysis. Build Simul 16(7):1187–1201. https://doi.org/10.1007/s12273-023-1010-8

Zhang Y, Tang S, Shi K, Tong X (2023e) An evaluation of the mine water inrush based on the deep learning of ISMOTE. Nat Hazards 117(2):1475–1491. https://doi.org/10.1007/s11069-023-05912-3

Zhao DK, Wu Q (2018) An approach to predict the height of fractured water-conducting zone of coal roof strata using random forest regression. Sci Rep 8:10986. https://doi.org/10.1038/s41598-018-29418-2

Zhao DK, Wu Q, Cui FP, Xu H, Zeng YF, Cao YF, Du YZ (2018) Using random forest for the risk assessment of coal-floor water inrush in Panjiayao coal mine, northern China. Hydrogeol J 26(7):2327–2340. https://doi.org/10.1007/s10040-018-1767-5

Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W (2020) Informer: beyond efficient transformer for long sequence time-series forecasting. Arxiv PDF. https://doi.org/10.48550/arXiv.2012.07436v3

Zhou J, Shen XJ, Qiu YG, Shi XZ, Du K (2023a) Microseismic location in hardrock metal mines by machine learning models based on hyperparameter optimization using Bayesian optimizer. Rock Mech Rock Eng. https://doi.org/10.1007/s00603-023-03483-0

Zhou X, Zhai N, Li S, Shi H (2023b) Time series prediction method of industrial process with limited data based on transfer learning. IEEE Trans Ind Inform 19(5):6872–6882. https://doi.org/10.1109/tii.2022.3191980

Zhu ZJ, Sun C, Gao XC, Liang Z (2022) A prediction method for floor water inrush based on chaotic fruit fly optimization algorithm-generalized regression neural network. Geofluids 2022:9430526. https://doi.org/10.1155/2022/9430526