Acta Geodyn. Geomater., Vol. 23, No. 3 (223), 417–427, 2026
DOI: 10.13168/AGG.2026.0028

journal homepage: https://www.irsm.cas.cz/acta

ORIGINAL PAPER

RESIDUAL-BASED ARIMA–CNN–LSTM MODELING FOR BOREHOLE STRESS TIME-SERIES FORECASTING

Jun TANG [1)] Ting WANG [1)] *, Guolin GUO [2)], Li LIU[3] and Shuxin NI [4)]

[1)] *Xuzhou College of Industrial Technology, Xuzhou 221000, China*
[2)] *Fujian Research Institute of Water Conservancy and Hydropower, Fuzhou 350000, China*
[3)] *China National Building Waterproof Association, Beijing 100055, China*
[4)] *Nanjing Huizhu Information Technology Research Institute Co., Ltd, Nanjing, 210000, China*

**Corresponding author's e-mail:* zcwang0505@163.com

ARTICLE INFO

Article history:

Received 27 July 2026
Accepted 15 September 2026
Available online 23 September 2026

Keywords:

Borehole stress
ARIMA –CNN–LSTM
Time-series prediction
Minepressure monitoring
Deep learning

ABSTRACT

This research establishes are residual-based ARIMA–CNN–LSTM model for predicting borehole stress in non-stationary, nonlinear, and noise-affected monitoring situations. The ARIMA (2,1,5) initially models the foreseeable linear and autocorrelated part. Subsequently, a CNN–LSTM network captures short-term patterns and long-term dependencies left in the residual. The stored dataset consists of 2,975 data points sampled at intervals of 15 minutes. The final 595 data points result in 594 matched one-step-ahead forecast-observation pairs. In this reserved block, the combined model attains MAE = 12.34 SU and RMSE = 15.00 SU. When compared with 51.72 and 60.26 SU of the independent ARIMA, it leads to reductions of 76.14 % and 75.11 %. In comparison to the stored signal span, NMAE and NRMSE are 0.86 % and 1.05 %. Since the acquisition export fails to maintain a verified channel-to-MPa conversion, observations, forecasts, residuals, MAE, and RMSE are constantly reported in software units (SU). An analysis of post hoc sensitivity encompassing filter quantity, recurrent depth, and learning rate reveals only slight changes in validation error, advocating for an accuracy–complexity interpretation instead of an assertion of universal architectural optimality. The outcomes show that residual-driven statistical-deep learning is able to enhance borehole-stress forecasting, whereas multi-location validation, calibrated collection metadata, and systematic examinations of sampling and data absence are still required.

1. INTRODUCTION

Borehole stress surveillance offers continuous data regarding stress development in coal seams and neighboring rock. It plays a significant role in pinpointing mining-induced stress accumulation, evaluating rockburst hazard, and safeguarding underground engineering safety. In the actual monitoring situation, the measured signal is rarely a simple stationary sequence. It is a situation which the gradual loading and unloading can hold with the autocorrelation, the sudden local changes, the short term drifts and irregular noise. In this case, the dependable forecasting is difficult especially if a model has to express the long lasting background evolution and the fast local disturbance. The data driven investigation has showed the importance of the sequence forecasting and the early-alert analysis about longwall roof stress, mine ground pressure and the rockburst hazard (Qin et al., 2022; Xiao et al., 2022; Guo et al., 2023).

The functions of existing prediction approaches are complementary. However, they also have different limitations if used separately. For example, ARIMA is useful for linear dependency, trend elimination, and autocorrelated framework. However, its linear construction is less fitting for unaddressed nonlinear and transitory conduct. LSTM is capable of retaining information across more extended temporal spans, whereas one-dimensional CNN layers effectively capture short-term temporal patterns (Hochreiter and Schmidhuber, 1997; LeCun et al., 2015). In mining applications, transfer-learning LSTM, dual-layer GRU, and LSTM–Transformer models have improved stress or hydraulic-support pressure prediction (Qin et al., 2022; Li and Lv, 2026; Tao et al., 2026). Recent mining-focused studies have further applied deep learning to strata-pressure forecasting (Gu et al., 2025), LSTM–PatchTST with cross-mine transfer validation (Yu et al., 2025), and CNN–LSTM–attention ensemble strategies for hydraulic-support pressure (Deng and Ding, 2026). Broader geophysical studies also support the usefulness of machine-learning and attention-based sequence models for complex monitoring signals (Bergen et al., 2019; Kong et al., 2019; Kubo et al., 2024; Perol et al., 2018; Mousavi et al., 2020; Zhu and Beroza, 2019). Together, these studies motivate combining explicit statistical structure with nonlinear residual learning.

Related borehole-strain research further illustrates the need to model nonstationary in-borehole signals at multiple temporal scales. Chi et al. (2023) combined signal decomposition with a GRU-LUBE network to forecast borehole strain and identify pre-earthquake anomalies; Yu et al. (2024) used one-dimensional CNN features for four-component borehole-strain analysis; and Li et al. (2024, 2025) applied graph neural networks to temporal and inter-station relationships. Borehole strain and borehole stress are different physical quantities, yet both are continuous in-borehole monitoring sequences affected by environmental interference, local transients, and temporal dependence. Despite this progress, relatively few studies have evaluated a residual-based statistical–deep-learning framework for continuous mine borehole-stress forecasting using controlled comparisons on the same monitoring sequence. This constitutes the principal methodological gap addressed here.

Accordingly, this study develops a residual-based ARIMA–CNN–LSTM model in which ARIMA (2, 1, 5) first represents predictable linear/autocorrelated behavior and CNN–LSTM learns unresolved residual dynamics. The proposed model is compared with standalone ARIMA, LSTM, and CNN–LSTM on a chronologically held-out test block to avoid future-data leakage. The analysis uses MAE and RMSE together with range-normalized errors, computational time, and contextual comparison with related mining studies. Extra sensitivity tests explore the quantity of convolutional filters, the depth of recurrence, and the learning rate. Meanwhile, the discourse connects the statistical decomposition to the recorded geological context and clearly deals with single-site verification, sampling precision, absent data, and engineering-unit calibration. The aim is not to assert a globally optimal network, but to ascertain whether residual-based hybridization offers a replicable accuracy benefit for the accessible borehole-stress sequence.

2. MATERIALS AND METHODS

2.1. *MODEL FRAMEWORK*

ARIMA represents a nonstationary series through autoregressive (AR), differencing (I), and moving-average (MA) components. The first use of differencing is to remove trend and to improve the stationarity; the resulting sequence is then described by a linear combination of previous values and previous errors. We denote this model by ARIMA (p, d, q), with the autoregressive order, the differencing order and the moving-average order. The three components give us an explicit statistical description of the persistent and autocorrelated temporal behaviour before the nonlinear residual learning.

The autoregressive component is expressed as:

$$y_t = c + \sum_{i=1}^{p} \varphi_i \, y_{t-i} + \varepsilon_t \tag{1}$$

where p is the AR order, φ_i is the autoregressive coefficient, and ε_t is the error term.

Differencing eliminates non-stationary trends:

$$\Delta^d y_t = (1 - B)^d y_t \tag{2}$$

where $\Delta^d y_t$ represents the d-th differenced series.

The moving-average part makes use of past prediction mistakes:

$$y_t = \mu + \varepsilon_t + \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j} \tag{3}$$

where q represents the MA order and denotes the moving-average coefficient.

LSTM is a kind of recurrent structure that is designed to keep the information over long time interval and to avoid vanishing-gradient problem (Hochreiter and Schmidhuber, 1997). Input, ignore, and output gates control the way of keeping and refreshing the information. This kind of restricted memory is suitable for stress series where the present value depends on more extended time interval. In this work, we use LSTM both as an independent baseline and inside CNN-LSTM. The gate for forgetting is given by:

$$f_t = \sigma\left(W_f [h_{t-1}, x_t] + b_f\right) \tag{4}$$

In this context, σ represents the sigmoid function, W_f stands for the weight matrix, h_{t-1} denotes the previous hidden state, x_t indicates the current input, while b_f is the bias.

CNN–LSTM integrates one-dimensional convolution for extracting local temporal features together with LSTMs sequence memory ability. Filters work across adjacent time steps to recognize short-term oscillations, sudden alterations, and recurring motifs. The resultant feature maps are then sent to LSTM for long-range dependence modeling (LeCun et al., 2015). As such, CNN highlights local temporal patterns and LSTM combines them over extended intervals. The representation of the mapping is:

$$\hat{y}_t = f_{\mathrm{CNN-LSTM}}(X_t) \tag{5}$$

where X_t represents the input area and $f_{\mathrm{CNN-LSTM}}$ stands for the integrated convolutional–recurrent mapping.

The suggested ARIMA–CNN–LSTM model employs residual learning, as presented in Figure 1. ARIMA is initially applied to the initial time-series to seize the part that can be depicted by a linear stochastic model featuring autocorrelation. The difference between the observation and the ARIMA fit forms the residual sequence. This residual is not assumed to be purely nonlinear; rather, it contains all unresolved structure, including nonlinear variation, transient disturbances, stochastic fluctuations, and measurement-related noise. A one-dimensional CNN extracts local temporal patterns from residual windows, and LSTM then models longer-term dependence among the learned features. This allocation reduces the burden on the neural network because it does not need to relearn the linear/autocorrelated component already represented by ARIMA.

(1) The observed series is decomposed into the ARIMA-fitted component and residual:

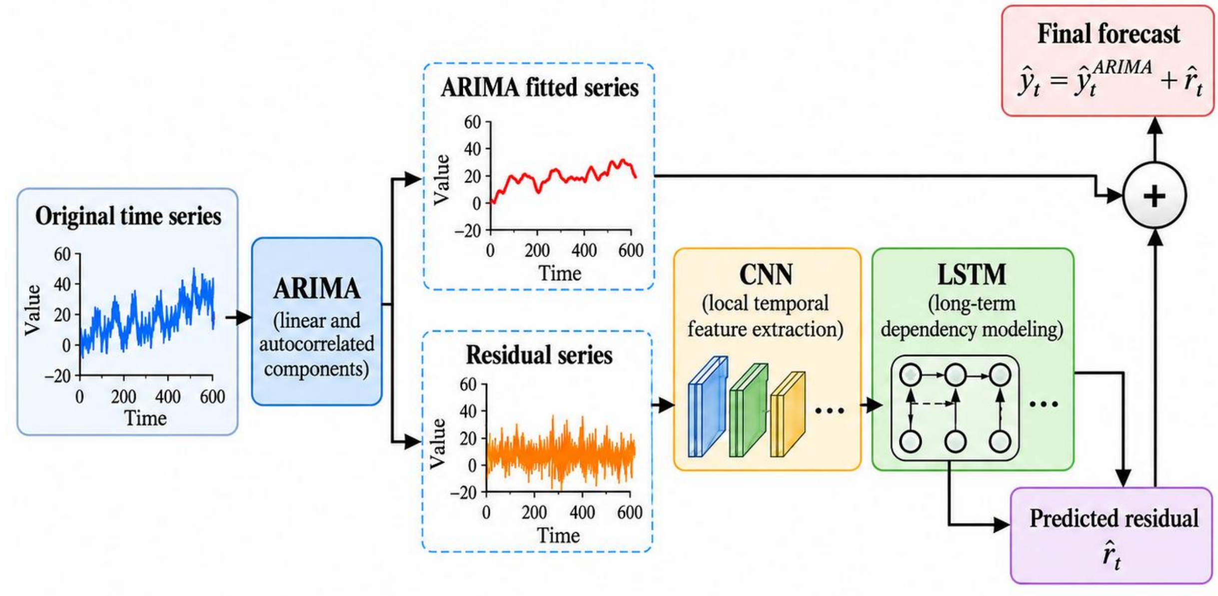

Fig. 1 ARIMA–CNN–LSTM Model.

$$y_t = \hat{L}_t + r_t \quad (6)$$

(2) Convolutional filters transform residual windows into local feature maps:

$$F_t = \mathrm{CNN}(r_{t-k+1:t}) \quad (7)$$

(3) LSTM estimates the residual component from these feature maps:

$$\hat{r}_t = \mathrm{LSTM}(F_t) \quad (8)$$

The final prediction is obtained by adding the ARIMA forecast to the CNN–LSTM residual forecast. The two stages are therefore complementary statistical representations rather than a strict physical separation of stress-generating mechanisms; this interpretation is examined cautiously in Section 3.3.

2.2. DATA ACQUISITION AND PREPROCESSING

Borehole stress was recorded with a YHY-60 infrared memory-type borehole stress meter manufactured by Tai'an Huana Electromechanical Co., Ltd., China (Fig. 2). According to the available technical specifications, the measurement range is 0–60 MPa, the accuracy is ±2.5 % of full scale (Class 2.5), and the resolution is 1 MPa. The instrument is designed for in-situ borehole stress monitoring in coal-mine environments and stores records for later infrared readout. These specifications describe the physical sensor and are reported to document acquisition conditions; they are deliberately separated from the numerical scale retained in the archived modeling file because the original software conversion relationship is unavailable.

Surveillance was carried out at Fengsheng No. 1 Mine in Chongqing, China. The location is within the fold-centered eastern part of the Sichuan Basin. In this area, the structural trends are mostly in the NNE–SSW direction, and joints and fractures have formed. The coal-bearing sequence of the Upper Triassic Xujiahe Formation consists of sandstone, shale, sandy mudstone, carbonaceous mudstone and thin coal seams. This structural and rock-nature heterogeneity offers a reasonable foundation for non-uniform stress transmission and localized mining-caused loading–unloading.

The archived dataset for modeling consists of 2,975 observations arranged chronologically, which were recorded at intervals of 15 minutes from 00:15 on 1 March 2011 until 23:45 on 31 March 2011, equivalent to around 31 days of monitoring. The worksheet includes two fields: date-time and the numerical monitoring channel sa. The sa values vary from -169111.75 up to -167677.60, resulting in a span of 1,434.15. These values cannot be directly understood as MPa since the archived files do not maintain a verified acquisition-software conversion from sa to the YHY-60 engineering scale, and the numerical scope is not adaptable for the direct utilization of the instruments 0–60 MPa specification. The scale for modeling is thus denoted as software unit (SU). Measurements, forecasts, deviations, MAE, and RMSE are regularly presented in SU, preventing an unsubstantiated physical transformation. The named sampling frequency is 1/900 Hz. The precise depth of the borehole, the azimuth of the sensor, the installation geometry, and the coordinates of the working face cannot be retrieved from the archived modeling file and are regarded as limitations of the metadata.

The documented sequence shows a trend, significant autocorrelation, irregular oscillations, local deviation, and sporadic sudden alterations. These characteristics might show a mix of mining disruption, local stress realignment, geological non-uniformity, equipment oscillation, and environmental or measurement impacts. Their mutual existence spurs both the pre-treatment and the strategy of residual-based modeling.

Fig. 2 YHY-60 Infrared Memory-Type Borehole Stress Meter Used for Data Acquisition.

Data tidying consisted of numerical outlier filtering and missing-value handling. Outliers beyond three standard deviations from the mean were recognized by applying the 3σ criterion, and absent samples were linearly interpolated to uphold temporal continuity. The standard serves as a practical quality-control guideline instead of proof indicating that every significant deviation is incorrect, as actual mining-triggered stress transients may also lead to large magnitudes when there are no independent equipment-status or operation records available. Linear interpolation is similarly appropriate mainly for short isolated gaps; longer gaps can smooth local variability and distort temporal dependence. These assumptions are retained as explicit limitations and motivate the missing-data sensitivity work proposed for future validation.

A chronological development/test partition was used to prevent future-data leakage. The first 80 % of the sequence, comprising 2,380 observations, formed the model-development block; the remaining 595 observations were reserved as the raw held-out block. No random shuffling was performed across the temporal boundary. Because one-step-ahead prediction requires alignment with the immediately preceding observation, the 595 held-out observations produce 594 valid forecast–observation pairs. Accordingly, N = 594 is used for the test curves and MAE/RMSE calculations in Figures 5–8 and Table 1. Hyperparameter comparisons were conducted only within the non-test portion of the data, so the held-out test block did not redefine the primary architecture or overwrite the archived final prediction outputs.

2.3. MODEL CONFIGURATION AND EVALUATION

Model configuration was kept consistent with the archived result-producing workflow. ARIMA orders were selected from stationarity assessment, ACF/PACF diagnostics, and the Akaike Information Criterion, giving ARIMA (2, 1, 5), with $p = 2$, $d = 1$, and $q = 5$. The standalone LSTM used 32 neurons per layer, dropout = 0.2, Adam with an initial learning rate of 0.005, 100 iterations, and batch size 32. The standalone CNN–LSTM used two one-dimensional convolutional layers with 32 filters, kernel length 2, stride 2, max pooling, and two LSTM layers with 64 and 32 units; Adam was initialized at 0.001. For the primary hybrid model, the ARIMA residual was passed to four convolutional filters and one LSTM layer with 25 units, using dropout = 0.2 and Adam initialized at 0.005. This concise archived setup produced the main prediction results. Different filter quantities, recurrent depths, and learning rates are only explored in the sensitivity analysis within Section 3.2.

The accuracy of predictions was assessed by employing mean absolute error (MAE) and root mean square error (RMSE). MAE offers the mean magnitude of the error, while RMSE places more emphasis on larger discrepancies; smaller values signify superior predictive capabilities. The two measurements are defined in equations (9) as well as (10).

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{9}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{10}$$

In this context, $y_i$and $\hat{y}_i$represent observed and predicted values respectively. Additionally, N stands for the quantity of aligned forecast-observation pairs. For the main held-out assessment, N equals 594. The absolute-value delimiters in Equation (9) and the radical in Equation (10) are upright mathematical operators. Because the archived acquisition conversion is unavailable, MAE and RMSE are reported in SU. To provide a scale-independent indication of error magnitude, range-normalized errors are calculated from the archived span $R = \max(y) - \min(y) = 1{,}434.15$ SU, using NMAE = MAE/R × 100 % and NRMSE = RMSE/R × 100 %. The YHY-60 physical specifications remain acquisition metadata and are not used to relabel SU-domain errors as MPa.

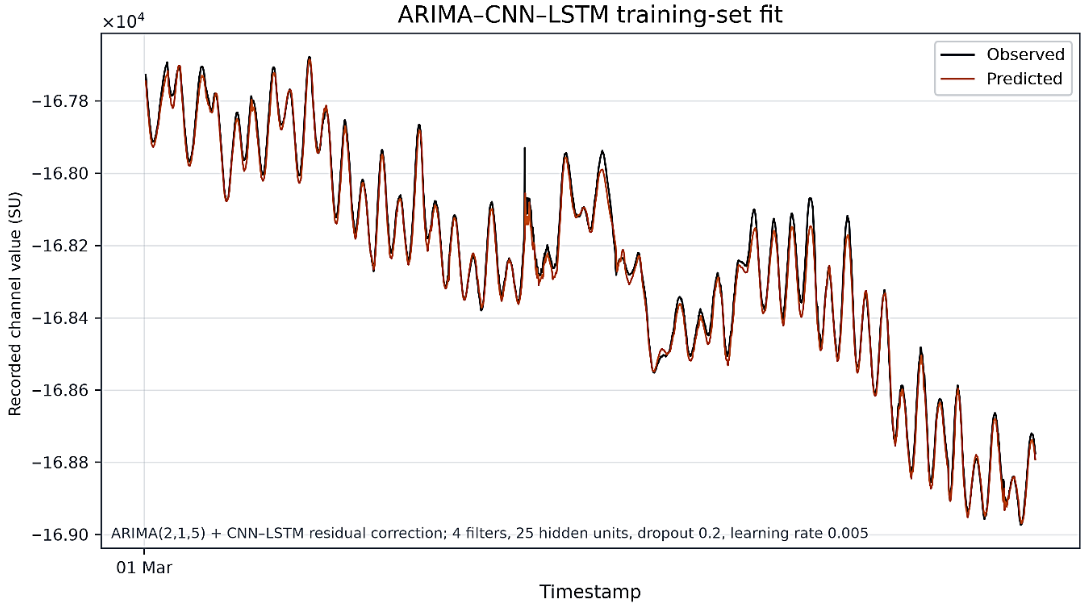

Fig. 3 ARIMA–CNN–LSTM training-set fit (observed and predicted values, SU).

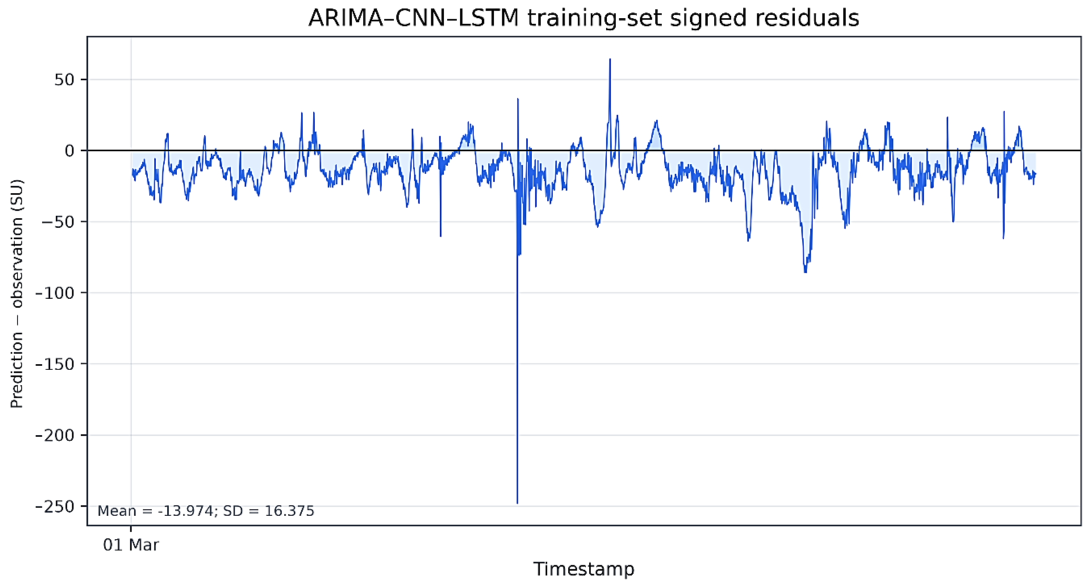

Fig. 4 Training-set signed residuals (prediction − observation, SU).

3. RESULTS AND ANALYSIS

3.1. PREDICTION AND COMPARATIVE PERFORMANCE

ARIMA, LSTM, CNN–LSTM, and ARIMA–CNN–LSTM were evaluated on the same chronological monitoring sequence. Figures 3–6 first show the hybrid model in detail. On the training block, the predicted curve follows the long-term evolution and repeated oscillatory structure of the observed signal (Fig. 3), while the largest signed residuals are concentrated around abrupt changes and the residual mean is slightly negative (Fig. 4). On the held-out block, the hybrid model continues to reproduce the principal peaks, troughs, and overall level changes (Fig. 5). Fiuatuinuationure 6 shows that several test intervals are dominated by negative residuals, together with a smaller number of positive peaks around rapid changes. Residuals are defined as prediction − observation; positive values therefore indicate overprediction and negative values indicate underprediction.

As Figure 3 shows, the forecasted training curve follows very well the observed borehole-stress sequence for most of the duration of the monitoring. The model reproduces both the slow variation of background and the recurring short term oscillations. The only observable differences are mostly around the abrupt changes and the isolated different ones.

As can be seen in Figure 4, most of the training residuals are small when compared to the entire signal span. Nevertheless, the reported negative mean (-13.974SU) is with a slight overall inclination to underprediction. A restricted quantity of larger positive and negative deviations take place around sudden stress alterations, thus, it can be understood that the rapid local variations are harder to reproduce that the constant background pattern.

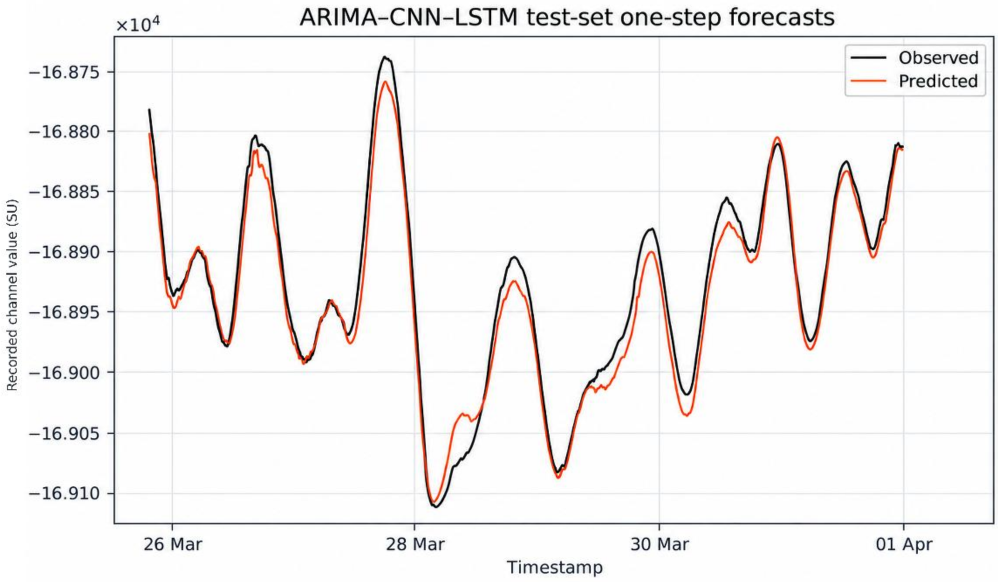

Fig. 5 ARIMA–CNN–LSTM test-set forecasts (N = 594; observed and predicted values, SU).

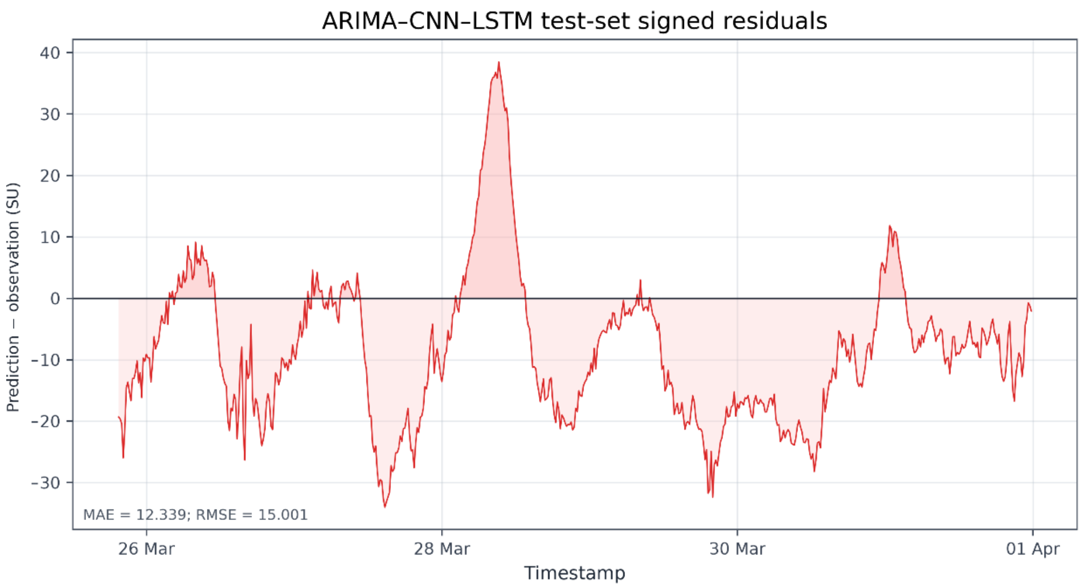

Fig. 6 Test-set signed residuals (N = 594; prediction − observation, SU).

Figure 5 shows that the hybrid model maintains the main peaks, valleys, and level variations across the 594 reserved forecast pairs. The forecasted curve stays near the observations during the entire test interval, even though minor lag or amplitude disparities emerge near several swiftly changing sections.

Figure 6 indicates that the residuals of the test stay moderate in size throughout most of the held-out duration. Even though negative residuals prevail in several time intervals, it implies a local inclination towards under-predicting. Several positive residual peaks also appear near segments that change rapidly. Therefore, the residual pattern indicates local directional bias instead of a uniformly unbiased distribution throughout the whole test period.

Figure 7 presents a direct comparison of the four models within the same sequence. ARIMA shows the most distinct deviation from the observations in high-variance periods. LSTM enhances the overall temporal tendency, and CNN–LSTM further lessens local discrepancies. The ARIMA–CNN–LSTM curve stays most closely overall near the observed test series.

Figure 8 verifies the numerical superiority of the hybrid model. ARIMA–CNN–LSTM attains the lowest MAE (12.34SU) and RMSE (15.00SU) when contrasted with 51.72/60.26SU applied to ARIMA, 38.15 out of 59.29SU applied to LSTM, and 17.47/17.88SU applied to CNN–LSTM. The reductions compared to ARIMA are 76.14 % and 75.11%, respectively.

The convention of signed residuals in Figures 4 and 6 is in line with the definitions of non-negative MAE and RMSE. Flipping the residual sign can transform the understanding of over-prediction in contrast to under-prediction, yet it wonts change MAE or RMSE. Table 1 thus presents solely magnitude-

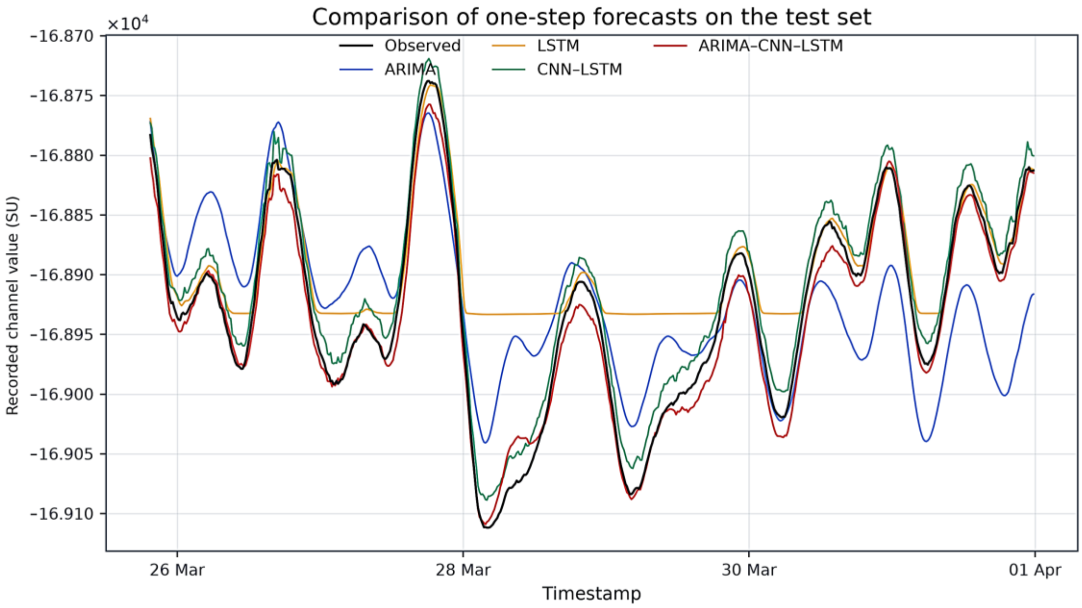

Fig. 7 Test-set forecasts of four models (N = 594; observed vs. predicted values, SU).

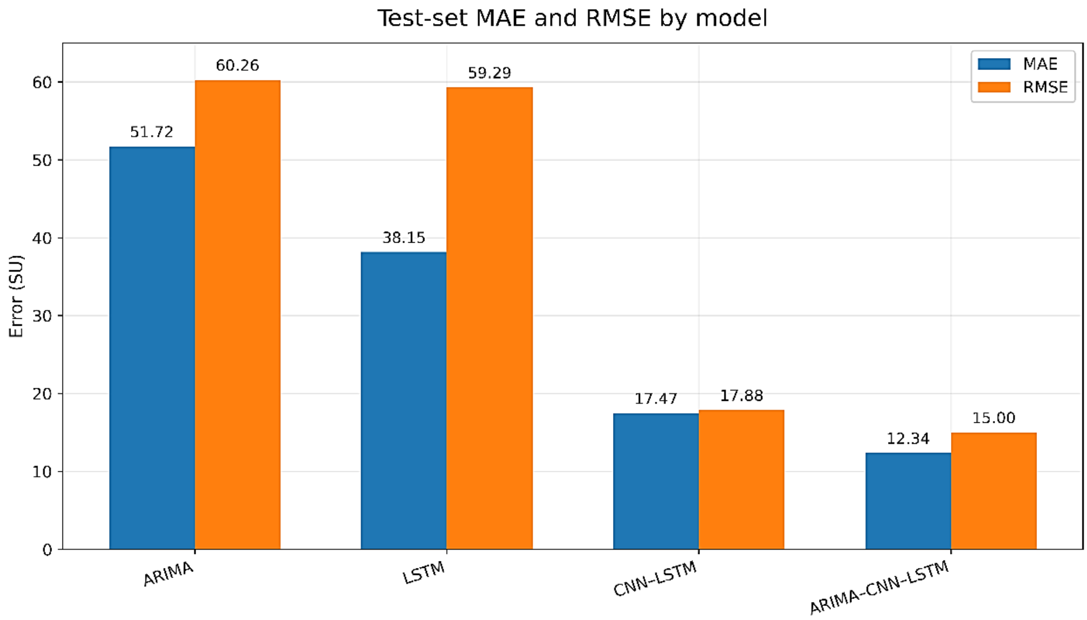

Fig. 8 Test-set MAE and RMSE of four models (N = 594; SU).

based metrics and their decreases, steering clear of the prior ambiguity caused by blending signed extrema and absolute-error statistics. All reported primary data are computed using the same 594 forecast pairs and the identical SU scale.

The difference between sensor engineering values and archived modeling values is also crucial for interpretation. The YHY-60 specification spans from 0 to 60 MPa. In contrast, the archived sa series covers a significantly broader negative numerical scope, and the conversion formula has not been retained. Directly designating these numerical forecasts or inaccuracies as MPa would thus lack support. Applying SU steadily maintains the validity of model comparison within the dataset. In contrast, NMAE and NRMSE offer a relative magnitude unassociated with a linear unit transformation. The absent conversion persists as an acquisition-metadata constraint instead of a modeling supposition.

Table 2 shows an accuracy–efficiency trade-off. ARIMA requires 10 min for training and 0.15s per prediction, where as ARIMA–CNN–LSTM requires 380 min and 0.52s per sample. Although the hybrid method is more expensive during model development, inference remains much shorter than the 15-min sampling interval, allowing offline training with online or near-real-time use. Efficiency could be improved through fewer filters or recurrent units, pruning/quantization, mixed precision, and GPU/parallel acceleration. Reported times are hardware-dependent and are interpreted only within the present environment.

Table 1 Final test-set prediction errors and relative reductions among different models (N = 594; software units, SU).

Model	Test MAE	Test RMSE	MAE reduction vs. ARIMA	RMSE reduction vs. ARIMA
ARIMA	51.72	60.26	—	—
LSTM	38.15	59.29	26.24 %	1.61 %
CNN-LSTM	17.47	17.88	66.22 %	70.33 %
ARIMA-CNN-LSTM	12.34	15.00	76.14 %	75.11 %

Table 2 Comparison of computational efficiency among different models.

Model	Training Time (min)	Prediction Time (s/sample)
ARIMA	10	0.15
LSTM	320	0.45
CNN-LSTM	350	0.48
ARIMA-CNN-LSTM	380	0.52

Table 3 Contextual comparison with related borehole-stress and mining stress/pressure time-series prediction studies.

Study	Monitoring target / setting	Model	MAE	RMSE	Reported unit
Qin et al. (2022)	Longwall mining-induced roof stress	TL-LSTM	2.67	5.55	MPa
Li and Lv (2026)	Hydraulic-support mine pressure time series	Dual-layer GRU	0.42	0.73	Study-specific pressure scale
Tao et al. (2026)	Hydraulic-support pressure time series	LSTM–Transformer	0.3034	1.1222	MPa
Present study	Borehole stress monitoring time series	ARIMA–CNN–LSTM	12.34	15.00	Software units (SU); archived acquisition-channel scale

Table 3 places the results in the context of related mining stress and pressure forecasting. Qin et al. (2022) reported MAE = 2.67 MPa and RMSE = 5.55 MPa for transfer-learning-assisted LSTM roof-stress prediction. Li and Lv (2026) reported 0.42 and 0.73 on their study-specific hydraulic-support pressure scale. For Tao et al. (2026), the MAE = 0.3034 MPa and RMSE = 1.1222 MPa values listed in Table 3 correspond specifically to the full-model entry in the authors' ablation comparison and are included only as a contextual reference. These numerical values should not be ranked directly against the present 12.34/15.00 SU because the physical targets, sensors, ranges, preprocessing procedures, and reporting units differ. Consequently, the strongest evidence for model advantage remains the controlled within-dataset comparison of four methods on the same test sequence.

3.2. *HYPERPARAMETER SENSITIVITY AND ABLATION ANALYSIS*

A post hoc one-factor-at-a-time sensitivity analysis was used to examine whether the primary network configuration was unusually sensitive to key hyperparameters. The reference residual network contained four convolutional filters, one LSTM layer, and initial learning rate 0.005. Candidate models changed only one factor at a time: filter number, recurrent depth, or initial learning rate. Each candidate was repeated with three random seeds, and validation MSE within the development data was used for comparison. The independent primary test block was not used to select a replacement architecture. Figure 9 therefore evaluates robustness and the accuracy–complexity trade-off rather than redefining the archived final model.

Figure 9 shows that the lowest mean validation MSE occurs with 16 filters (0.579744), followed

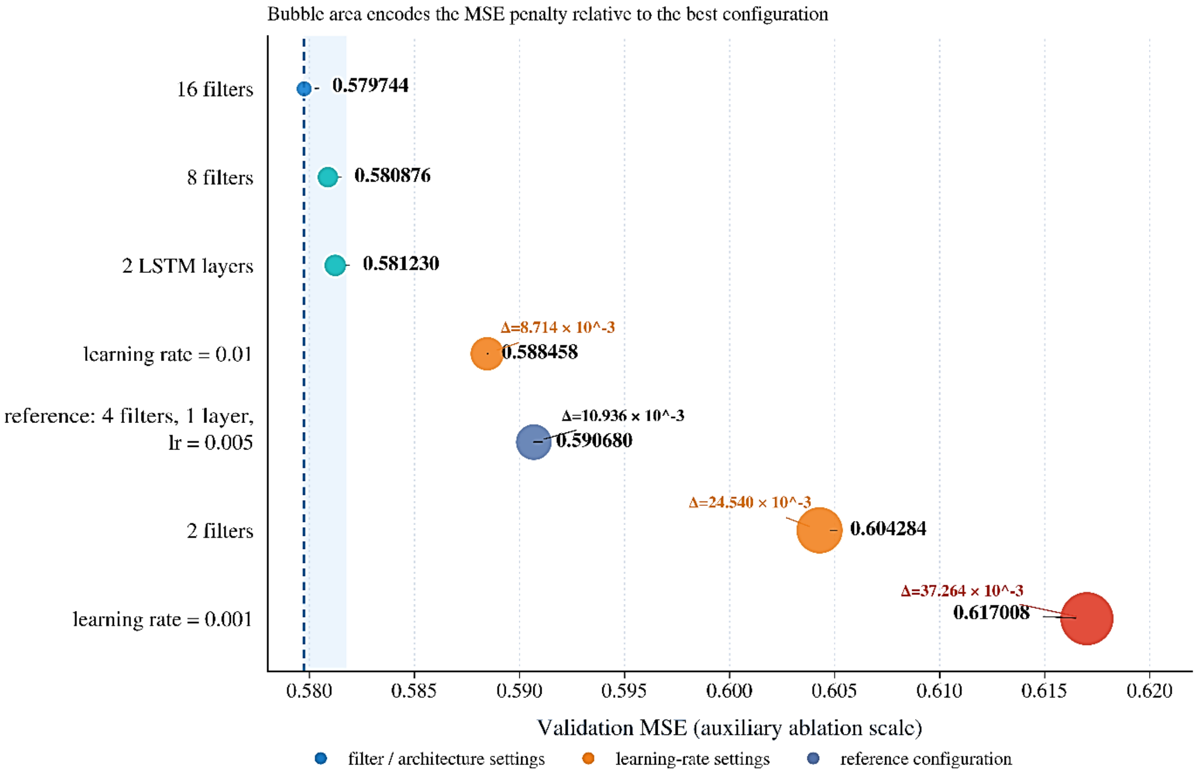

Fig. 9 Post hoc one-factor-at-a-time hyperparameter sensitivity analysis based on validation MSE.

closely by 8 filters (0.580876) and two LSTM layers (0.581230). A learning rate of 0.01 gives 0.588458, the compact reference gives 0.590680, 2 filters give 0.604284, and learning rate 0.001 gives 0.617008. The 16-filter configuration improves validation MSE by only about 1.85 % relative to the four-filter reference. This decrease is minor when compared to the rise in network breadth and computational requirement. Thus, the outcome backs keeping the concise archived setup for the main experiment while recognizing that it is not the sole validation optimum.

Figure 10 presents the auxiliary held-out reruns. The 16-filter configuration yields MAE = 1.624672 ± 0.019978 and RMSE = 2.155219 ± 0.027576 across three seeds. The 2-filter configuration gives slightly smaller mean held-out errors, MAE = 1.589103 ± 0.004892 and RMSE = 2.107076 ± 0.005128, while the two-layer configuration gives RMSE = 2.138430 ± 0.047049. Since the withheld ranking fails to replicate the validation ranking and multiple uncertainty intervals intersect, the experiments do not pinpoint a singularly optimal architecture. Additionally, the stored ablation results employ an auxiliary numerical scale for which a confirmed one-to-one correspondence to the main SU prediction scale is not available. Thus, the absolute MAE/RMSE values of them are solely utilized for comparison within the ablation study and are not numerically contrasted with those in Table 1.

3.3. PHYSICAL INTERPRETATION, APPLICABILITY, AND LIMITATIONS

The geological framework of the Fengsheng No. 1 Mine offers a physical backdrop for the statistical breakdown. Mechanically distinct sandstone, shale or mudstone, carbonaceous mudstone, and coal strata are present in a fold-dominated and fractured rock body. This kind of heterogeneity may lead to non-uniform stress transfer in the process of mining, particularly in the vicinity of lithological interfaces and jointed areas. The comparatively long-lasting and autocorrelated part caught by ARIMA might thus contain gradually changing background loading and step-by-step mining-caused stress redistribution. The residual may include quicker and more erratic responses related to the advancement of the working-face, local loading and unloading, fracture or joint activation, lithological-interface impacts, environmental drift, and measurement noise. This interpretation is intentionally cautious: the two modeled components cannot be mapped one-to-one to specific geological mechanisms because synchronized working-face, fracture-activity, temperature, and equipment-status records are unavailable.

The principal limitation is that validation is based on one YHY-60 device at one mine. Although the regional geological and lithological context is documented, the archived modeling files do not retain exact borehole depth, sensor azimuth, installation geometry, or working-face coordinates. The present

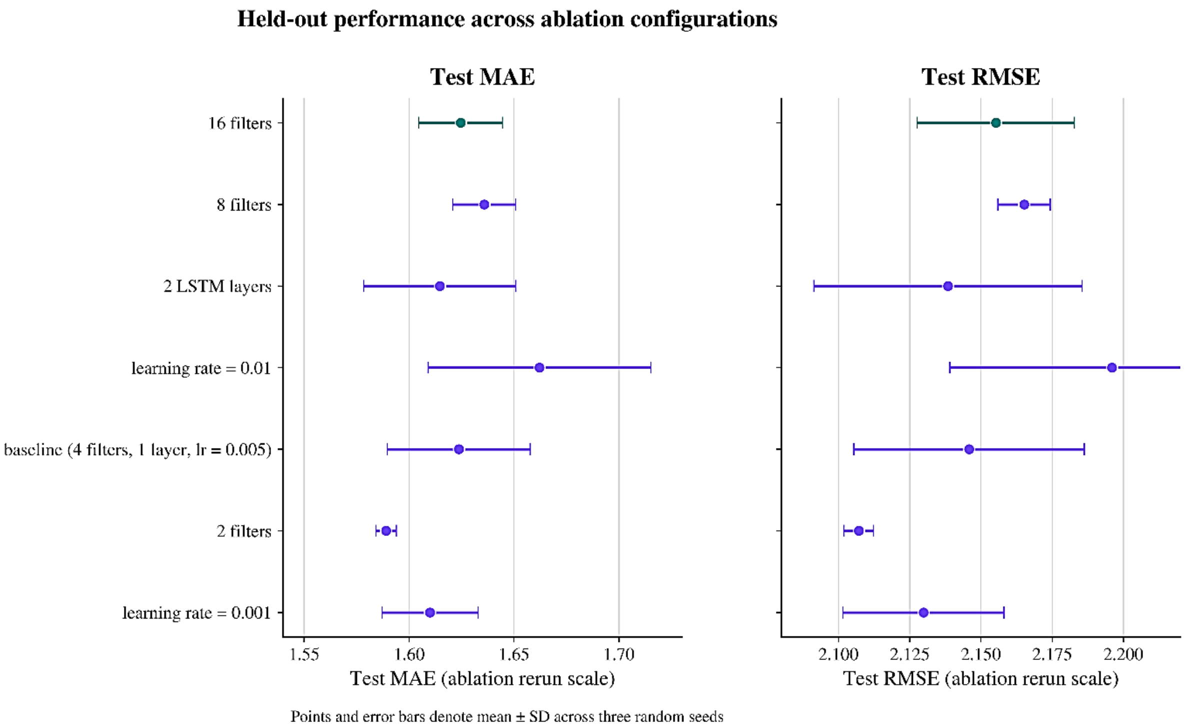

Fig. 10 Held-out MAE and RMSE across ablation configurations on the auxiliary rerun scale (mean ± SD, three random seeds).

results therefore establish a site-specific proof of concept rather than universal transferability. Application to other mines, minerals, lithologies, borehole layouts, or sensors should use new validation data and may require site-specific normalization, transfer learning, domain adaptation, or limited local fine-tuning. Multi-site joint validation is the most important next step for evaluating whether the residual-learning advantage persists under different geological and monitoring conditions.

Sampling precision and data comprehensiveness are extra constraints. The present interval stands at 15 minutes. Lower sample frequencies can hinder the capture of brief-duration patterns that the convolutional filters are designed to detect. On the other hand, significantly higher sample rates might raise redundancy and computational expense. Linear interpolation keeps the continuity for short gaps, but it can produce artificial smoothness when successive observations are missing for long periods. In further work, it has to be compared with the present series with those under the control of 30-and 60-minute down-sampling. It also should assess missing data blocks of increasing in duration. Recurrent models which are aware of the missingness, the masked model and the probabilistic imputing are also contestants for long gaps. Such assessments would give more precise the practical operative scope of the model.

4. CONCLUSION

This research generated and evaluated the residual-centered ARIMA-CNN-LSTM model for the prediction of borehole stress of Fengsheng No. 1 Mine. ARIMA (2, 1, 5) models the foreseen linear and autocorrelated element, and CNN-LSTM learns the unconsidered residual structure. Using a chronologically held-back test, the within-dataset baseline comparisons, the computational assessment and some more supplementary sensitivity trials, the research investigated the prediction precision, the robustness, the interpretability and the practicability.

In the set of 594 aligned single-step-ahead test pairs the ARIMA-CNN-LSTM model has a MAE of 12.34 SU and a RMSE of 15.00 SU. With respect to the standalone ARIMA model, the corresponding errors is reduced of 76.14 % and 75.11 %. The values of NMAE and NRMSE are 0.86 % and 1.05 % respectively of the archived signal range. Alterations in the hyperparameters results in only slight differences during the validation: the top-tested 16-filter version reduced the validation MSE by about 1.85 % with respect to the compact four-filter reference. Even if the hybrid model needs the longest training duration (380 min), inferences take only 0.52s per specimen, that back up the offline training and then the operational forecasting.

The ARIMA and CNN—LSTM components should be interpreted as complementary statistical expressions rather than as completely separable physical processes. The different coal-containing strata, folds, joints, and fissures of the research location do provide a plausible setting for the co-occurrence of a continuous development of stress and of more rapid local variations. Nevertheless, single remaining events could not be assigned to a single cause on the basis of the available data. The proof that we have is for one and only one monitoring location and for an archived SU scale which does not have a validated MPa conversion. Further work should keep the calibrated acquisition metadata, check the model on different sites and sensor systems, check sampling-frequency and missing-data sensitivity, see transfer/domain adaptation and cut down training cost by using lightweight architectures and fast computation.

ACKNOWLEDGEMENTS

The research was financially supported by the “Blue Project” of Jiangsu Universities in 2023 and Xuzhou 2023 Social Undertakings Project (Project No. KC23319).

REFERENCES

Bergen, K.J., Johnson, P.A., de Hoop, M.V. and Beroza, G..: 2019, Machine Learning for Data-Driven Discovery in Solid Earth Geoscience. Science, 363, 6433, eaau0323. DOI: 10.1126/science.aau0323

Chi, C., Li, C., Han, Y., Yu, Z., Li, X. and Zhang, D.: 2023, Pre-earthquake anomaly extraction from borehole strain data based on machine learning. Sci. Rep., 13, 20095. DOI: 10.1038/s41598-023-47387-z

Deng, C. and Ding, Z.: 2026, Prediction of hydraulic support pressure in a coal longwall face by integrating multi-head attention and temporal feature learning. Sci. Rep., 16, 20772. DOI: 10.1038/s41598-026-51773-8

Gu, X., See, K.W. and Zhou, X.: 2025, Deep Learning for Time Series Prediction of Strata Pressure in Coal Mining. IEEE Access, 13, 124068–124085. DOI: 10.1109/ACCESS.2025.3589493

Guo, Y., Gu, S., Du, R. and Shen, J.: 2023, Multi-parameter comprehensive early warning of coal pillar rockburst risk based on DNN. Front. Earth Sci., 11, 1201946. DOI: 10.3389/feart.2023.1201946

Hochreiter, S. and Schmidhuber, J.: 1997, Long Short-Term Memory. Neural Comput., 9, 8, 1735–1780. DOI: 10.1162/neco.1997.9.8.1735

Kong, Q., Trugman, D.T., Ross, Z.E., Bianco, M.J., Meade, B.J. and Gerstoft, P.: 2019, Machine Learning in Seismology: Turning Data into Insights. Seismol. Res. Lett., 90, 1, 3–14. DOI: 10.1785/0220180259

Kubo, H., Naoi, M. and Kano, M.: 2024, Recent Advances in Earthquake Seismology Using Machine Learning. Earth Planets Space, 76, 36. DOI: 10.1186/s40623-024-01982-0

LeCun, Y., Bengio, Y. and Hinton, G.: 2015, Deep Learning. Nature, 521, 436–444. DOI: 10.1038/nature14539

Li, C., Duan, Y., Han, Y., Yu, Z., Chi, C. and Zhang, D.: 2024, Extraction of pre-earthquake anomalies from borehole strain data using Graph WaveNet: a case study of the 2013 Lushan earthquake in China. Solid Earth, 15,7, 877–893. DOI: 10.5194/se-15-877-2024

Li, C., Qin, C., Zhang, J., Duan, Y. and, Chi, C.: 2025, Analysis of borehole strain anomalies before the 2017 Jiuzhaigou Ms 7.0 earthquake based on a graph neural network. Nat. Hazards Earth Syst. Sci., 25, 1, 231–245. DOI: 10.5194/nhess-25-231-2025

Li, Q. and Lv, S.: 2026, Time Series Prediction of Mine Pressure Using a Dual-Layer GRU with Multi-Feature Enhancement. Appl. Sci., 16, 13, 6564. DOI: 10.3390/app16136564

Mousavi, S.M., Ellsworth, W.L., Zhu, W., Chuang, L.Y. and Beroza, G.C.: 2020, Earthquake Transformer—An Attentive Deep-Learning Model for Simultaneous Earthquake Detection and Phase Picking. Nat. Commun., 11, 3952. DOI: 10.1038/s41467-020-17591-w

Perol, T., Gharbi, M. and Denolle, M.: 2018, Convolutional Neural Network for Earthquake Detection and Location. Sci. Adv., 4, 2, e1700578. DOI: 10.1126/sciadv.1700578

Qin, C., Zhao, W., Zhong, K. and Chen, W.: 2022, Prediction of longwall mining-induced stress in roof rock using LSTM neural network and transfer learning method. Energy Sci. Eng., 10, 2, 458–471. DOI: 10.1002/ese3.1037

Tao, R., Lei, X., Wan, L., Wang, Y. and Xu, N.: 2026, LSTM–Transformer-Based Mine Pressure Prediction Using Hydraulic-Support Monitoring Data. Sensors, 26, 14, 4423. DOI: 10.3390/s26144423

Xiao, Y., Deng, H., Xie, Z. and Lu, H.: 2022, Mine Ground Pressure Monitoring and Early Warning Based on Deep Learning Data Analysis. Mob. Inf. Syst., 2022, 1, 6255119. DOI: 10.1155/2022/6255119

Yu, Q., Sun, C., Yang, Y. and Yang, P.: 2025, Hydraulic support pressure prediction via deep learning with multilevel temporal feature integration. Sci. Rep., 15, 45680. DOI: 10.1038/s41598-025-13089-x

Yu, Z., Li, H., Jing, X., Chi, C. and Zheng, H.: 2024, Borehole strain data based seismicity prediction analysis using a neural network. Acta Seismol. Sin., 46, 2, 327–339. DOI: 10.11939/jass.20230122

Zhu, W. and Beroza, G.C.: 2019, PhaseNet: A Deep-Neural-Network-Based Seismic Arrival-Time Picking Method. Geophys. J. Int., 216, 1, 261–273. DOI: 10.1093/gji/ggy423