Coalbed Methane Production Forecasting Based on LSTM and Transfer Learning Integration

Wenjie Li[1,2], Fengnian Wang[1,2], Chenglong Qiu[1], Chenchen Jia[1], Jinxin Meng[3], Yuxia Wang[3], Bei Zhu[3]*

1.Xinjiang Yaxin Coalbed Methane Resources Technology Research Co., Ltd., Urumqi 830009, Xinjiang Uygur Autonomous Region, China
2.Key Laboratory of Coalbed Methane Exploration and Development in Xinjiang
3. School of Petroleum Engineering, Yangtze University, Wuhan 430100, Hubei, China
**Corresponding Author:Jinxin Meng*
Email:i2684501257@163.com

Received 26 March 2026; Accepted 12 July 2026

Abstract Traditional Long Short-term memory (LSTM) method faces a lot of challenges during coalbed methane production prediction, such as poor generalization on the small datasets, difficult hyper parameter turning and poor accuracy in prediction. Due to these challenges this research introduces a new method which is the combination of LSTM and Transfer Learning inorder to encounter to these issues. This approach combines the Person Correlation Coefficient together with the Grey relational degree to identify production factors which is more important by using a "Pre-training Hierarchical Fine-Tuning" transfer learning framework. In order to maintain the ability of extracting temporal features, the LSTM layer uses a low learning rate while fully connected layer uses a high learning rate for faster convergence. When we applied this new strategy to the FSL-34 well, it's R^2 value increased from 0.8 to 0.93 and the Mean Absolute Percentage Error (MAPE) decreased from 0.614% to 0.288%, without forgetting in the situation of small datasets around 100, R^2 increased greatly from -2.3 to 0.94 and the MAPE decreased by 37%. These two examples shows clearly the advantages of the new proposed method over Traditional LSTM models. While using Maximum Mean Discrepancy (MMD) domain arrangement helps bridge the gap between separate domains, utilizing the transfer of knowledge from a source domain reduces the data requirement in the target domain. I general, this strategy provides a very effective coalbed methane production prediction in cases with scanty data samples.
Keywords Coalbed methane; Long Short-Term Memory network; Transfer learning; Production prediction; Small-sample prediction

1 Introduction

To predict accurately and correctly production of coalbed methane (CBM) is very important for maximizing recovery rate and improve development plan[1]. Its very important to Characterize in a correct way the complex relationships between different parameters including production parameters, engineering parameters and geological. Production decline analysis methods and Traditional numerical simulations offering a more effective way to address reservoir heterogeneity and production uncertainties due to

Copyright © 2026 The Authors. Published by Emedia/Online Inc. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

raised of Machine Learning in recent years[2]. One of the very important and well known application method in the field for it's ability of identifying long-term dependencies in time series data is Long Short-Term memory denoted as LSTM in short. Due to it's special "Gating Mechanism", the model proved to very good and flexible in predicting CBM Production. Apart from it's greater efficiency and it's important to the field, LSTM models also faces some challenges in real time applications. One of the basic Challenge faces LSTM model lie in it's complex hyperparameter combinations[3] which needs a number of experiment inorder to get ideal configurations. Apart from that challenge, LSTM models incase of small data samples become very sensitive to noise otherwise the performance decrease it's ability. These constraints hinder the training efficiency and generalization performance of LSTM models in real-world applications.[4-5].

The problem of limited data and model customization is solved by using the knowledge from a source domain to improve performance in a target area. In this study, a new prediction method is introduced that combines Long Short-Term Memory (LSTM) networks with transfer learning techniques [6]. Improving accuracy of prediction, simplifying hyper parameter tuning and reducing errors on dealing with small datasets are the three main Challenge discussed by this new research by using a "Pre-training-hierarchical fine-tuning" technique to refine the model building and integrating multi-dimensional feature selection techniques inorder to identify critical variables. In a case study of horizontal wells in the Fuchang District's eight coal seam, the technique shows very good results in providing a promising technical path for predicting coal seam gas production. The Potential of transfer learning in improving prediction models and opens up new possibilities for optimizing gas production forecasts is highlighted by this study.

2 Principles and Construction of LSTM Transfer Learning

2.1 Fundamentals of LSTM Neural Networks

An improved version of traditional RNNs, pointing the problem of vanishing or exploding gradients by utilizing a "Gate mechanism" is known as Long Short-term Memory (LSTM) Networks [7]. This mechanism contain the Forget gate, output gate, input gate and Cell state, which make the network to gain important information over long range. LSTM networks are very excellent at identifying long-term dependencies in data.

Forget Gate: It Uses the sigmoid function, which is expressed in equation number (1) below to determine whether to keep or discard historical data.:

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

Where: x_t is the input information at time step t, σ is the sigmoid activation function, h_{t-1} is the hidden layer state at time step t-1, w_f , b_f are the weights and bias of the forget gate.

Input gate: Filters new information and updates candidate states. The Formulas are as follows:

$$\boldsymbol{i_t}=\sigma(\boldsymbol{w_i}[\boldsymbol{h_{t-1}},\boldsymbol{x_t}]+\boldsymbol{b_i}) \tag{2}$$

$$\boldsymbol{C_t'}=\mathbf{tanh}\,(\boldsymbol{w_C}[\boldsymbol{h_{t-1}},\boldsymbol{x_t}]+\boldsymbol{b_C}) \tag{3}$$

Where: C_t' represents the candidate state after tanh transformation; information in may be updated to the current memory time step. w_i and b_i denote the input gate's weight and bias, respectively. w_C and b_C denote the candidate state's weight and bias, respectively.

Cell state update: Updated based on the decision from the forget gate and input gate, as follows:

$$C_t=f_tC_{t\text{-}1}+i_tC_t' \tag{4}$$

Where: C_t and $C_{t\text{-}1}$ represent the memory cell states at time t and t-1, respectively.

Output Gate: Uses the following formula to generate the hidden layer state at the current time step:

$$o_t=\sigma(w_o[h_{t\text{-}1},x_t]+b_o) \tag{5}$$

$$h_t=o_t\tanh C_t \tag{6}$$

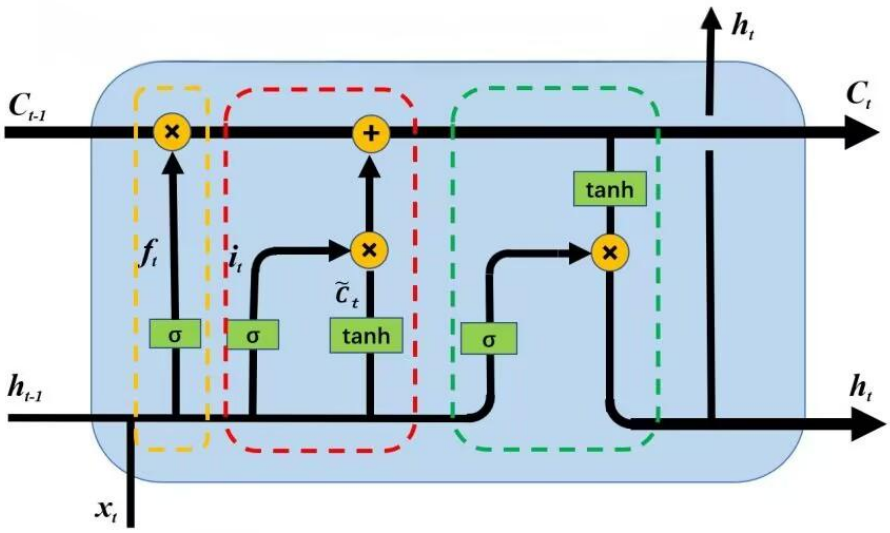


Fig. 1 LSTM Network Unit Structure at Time t

2.2 Transfer Learning Fusion Strategy

Pre-training and fine-tuning are the primary uses of transfer learning, an effective tool in machine learning. This study focuses on the "pre-training - hierarchical fine-tuning" framework in the context of coalbed methane wells[1,13]. The three-step approach, which shows how transfer learning can optimize methane well analysis, consists of data pre-processing, model pre-training, and fine-tuning.

For building a basic or fundamental model during the first training phase, the source domain of FSL-30 dataset is used. The aim was to gain a basic and fundamental understanding of extracting significant temporal patterns associated with the different generation of coalbed methane using effective “LSTM + Fully connected layer” approach.

Parameters of the model are shifted from the source domain towards the targeted ones, this is happening during Fine-Tuning phase of the FSL- 32 wells and FSL-34 wells. To insure peak performance, a Hierarchical learning rate technique is used. The LSTM Layer was given a lower learning rate of 1e-5 inorder to extract time features efficiently while the fully connected layer was given a greater learning rate of 1e-4 so as to quickly adjust to the data distribution of the target domain. Apart from that, Techniques for feature engineering optimization are also used. In order to prevent future information leakage, training samples are enlarged using a sliding window technique, lag features are added to capture temporal autocorrelation, and data is divided depending on time order.

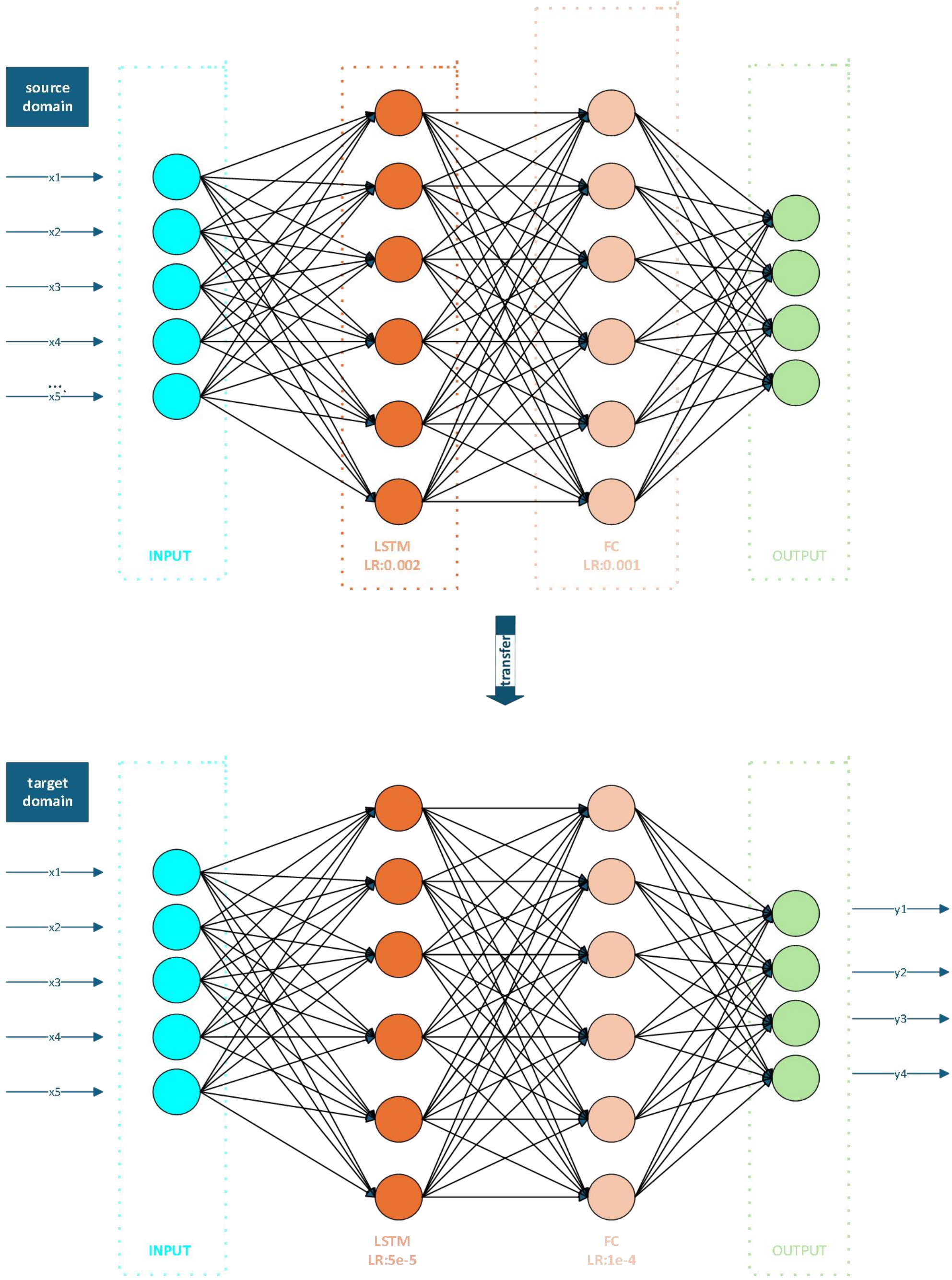


Fig. 2 Schematic Diagram of Transfer Learning Principle

2.3 MMD for Assessing Domain Similarity Distribution

In transfer learning, the similarity between the source domain and the target domain is the core prerequisite for selecting transfer strategies and ensuring transfer effectiveness, especially in energy engineering scenarios such as coalbed methane/oil daily gas production prediction - these scenarios have strong domain dependencies, and data distribution is significantly affected by geological and engineering conditions. Similarity directly determines the feasibility and efficiency of "transfer". The Maximum Mean Discrepancy (MMD) can be used to compare the distribution differences between two datasets using a Reproducible Kernel Hilbert Space (RKHS), which can help determine the similarity between source and target domain data. Its square form is theoretically defined as:

$$MMD^2(P,Q)=E_{x,x'\sim P}(k(x,x'))+E_{y,y'\sim Q}(k(y,y'))-2E_{x\sim P,y\sim Q}(k(x,y))$$

(7)

$E_{x,x'\sim P}(k(x,x'))$: The expected value of the kernel function for two independent samples x and x 'in distribution P, reflecting the correlation between samples within P;$E_{y,y'\sim Q}(k(y,y'))$: reflects the correlation of samples within the distribution Q; $E_{x\sim P,y\sim Q}(k(x,y))$: The expected value of the kernel function between sample x of distribution P and sample y of distribution Q, reflecting the correlation between samples of different distributions

However, in practical applications, it is not possible to directly obtain the true expected distribution, and empirical estimation is required through a limited sample set. Assuming sample set $X=\{x_i\}_{i=1}^{n}$ (from distribution p), ($Y=\{y_j\}_{j=1}^{m}$) (from j=1 to m, from distribution q), the unbiased empirical estimation formula for MMD is:

$$MMD^2=\frac{1}{n(n-1)}\sum_{i\neq j}k(x_i,x_j)+\frac{1}{m(m-1)}\sum_{i\neq j}k(y_i,y_j)-\frac{2}{nm}\sum_{i,j}k(x_i,y_j)$$

(8)

n: Source domain sample size, m: Target domain sample size;$\sum_{i\neq j}k(x_i,x_j)$: The sum of the kernel function values of all non repeating sample pairs in the source domain;$\frac{1}{m(m-1)}\sum_{i\neq j}k(y_i,y_j)$: The sum of the kernel function values of all non repeating sample pairs in the target domain;$\sum_{i,j}k(x_i,y_j)$: The sum of the kernel function values of all cross sample pairs in the source and target domains (i traverses the source domain, j traverses the target domain).

2.4 Principles of MMD Domain Alignment

One of a widely used technique in transfer learning and domain adaptation is called MMD domain alignment. The method involves mapping data from the source/target domain to a Reproducing Kernel Hilbert Space (RKHS) and then minimizing the distance between their distributions to ensure convergence, aiming to achieve efficient distribution arrangement.

In feature mapping, the feature extractors include LSTM hidden layers that seek to extract features of both two ways which is source domain sample and target domain samples.

These features are then transformed to the RKHS with the help of kernel functions, such as Gaussian or linear kernels. The method has the advantage of efficiently capture a nonlinear correlations between characteristics without having to do more complicated computations in higher dimensions.

The value of MMD is calculated by an independent empirical formula. This involves the determination of the mean cross-correlation of the source domain and target domains and the mean autocorrelation in each of them. Due to analytical reasons, a measure of the difference between the distributions can be generated by comparing these values.

MMD calculation: Its an objective empirical formula that measures the difference in distribution by combining source domain/target domain's autocorrelation and cross-correlation between the two domains. In practical model training, the MMD loss and task loss are kept to provide a total loss function that strikes a balance between task performance and domain alignment effectiveness and the formula is as follows:

$$L=L_{task}+\lambda\cdot MMD^2(X_S,X_T) \quad (9)$$

To obtain a balanced relationship between task performance and domain alignment is very important in Machine Learning Models. The core task loss, L_{task} is determined by the Mean Absolute Error (MAE). To control the model focus much on domain alignment at the expense of task performance, a balancing hyper parameter, λ is used to adjust the weight of the MMD loss. X_S represent the source domain and X_T denote the features of the target domains.

3 Experimental Design and Results Analysis

3.1 Experimental Data and Preprocessing for the Study Area

An investigation based on the use of MMD approach was subjected on three horizontal wells in Fukang District 8. The target domain of fine-tuning was Wells FSL-32 and FSL-34 and the source domain of pre-training was well FSL-30. To ensure accuracy and reliability, data were collected on a daily basis and subjected to a thorough planning process.

Malfunction of the equipment and power cuts were also a problem that was occasioned lacking or atypical values in the process of coalbed methane data collection. To make sure that the results are properly analysed and interpreted, Z-score normalisation was used to identify exceptions that would be treated as missing values.

Assuming missing values. The deletion of data might be distroyed since coalbed methane is time-series data, which can result in the time continuity of data being interrupted. Linear interpolation is applied in the case of imputation. The linear interpolation equation is:

$$y=y_0+(y_1-y_0)\cdot\frac{x-x_0}{x_1-x_0} \quad (10)$$

Where x_0, y_0 and x_1, y_1 represent coordinates of two known points.

Kalman filtering is employed to eliminate anomalies caused by random noise deviations, enhancing data reliability.

3.2 Evaluation Metrics and Model Parameters

MAPE, RMSE, MAE and R^2 are the examples of the indicators used to evaluate the performance of the model. A higher accuracy model is reflected in R^2 value closer to 1. These indicators help in comprehensive assessment of the model's effectiveness.

The calculation formulas for the four metrics are as follows:

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \tag{11}$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{12}$$

$$MAPE=\frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \tag{13}$$

$$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{14}$$

Where: y_i represent the predicted coalbed methane production value, and $\hat{y}_i$ is the representation of the actual value.

Setting parameters is an important part of time series analysis. To create sample data, time intervals must be segmented. A sliding window technique is utilized to convert time series data into inputs for the machine learning model[7]. This approach includes adjusting a fixed-size window to obtain repeated training samples. In order to capture autocorrelation and trend information inside the time series, lagged aspects of the target variable are also included. Adam, an adaptive optimisation technique, is used as the optimiser for model training. The initial learning rate (Lr) is set at 0.002, the batch size (batch_size) is 32, the sliding step size (input_size) is 5, and the lag feature step size is 1 through search and debugging. To improve the modelling process, a two-layer LSTM with 128 hidden layer units is used.

3.3 Analysis of Key Factors Affecting Coalbed Methane Production

Coalbed methane production is a complex process influenced by different factors, including geological parameters, engineering measures parameters, and production process parameters[15]. While technical measurements are the outcome of drilling and fracturing operations, geological parameters are established by inherent features of the coal seam[16]. These variables have little effect on the forecast of production dynamics and are fixed for individual wells. The success of coalbed methane generation is largely dependent on the characteristics of the production process after drilling and fracturing. The production process is strongly influenced by variables such bottom-hole pressure, dynamic fluid level, casing pressure, stroke, and stroke rate. These metrics show dynamic trends over time and are gathered as time-series data. To comprehend and maximise coalbed methane production, it is essential to analyse these tendencies. To increase the effectiveness and efficiency of industrial processes, researchers concentrate on finding patterns in these time-series data. In conclusion, optimising coalbed methane production requires the interaction of geological, technical, and production process characteristics. By examining, analyzing and understanding these factors, operators can greatly improve the productivity of coalbed methane wells.

Variance Inflation Factor (VIF) analysis was used in the study to address multi-collinearity. Strong multicollinearity was indicated by a VIF larger than 10, which suggested eliminating one feature. Moderate multicollinearity was indicated by VIF scores between 5 and 10, necessitating additional research. No multicollinearity issues were identified by VIF values less than 5. Table 1 contains detailed results.

Table 1 Variance Inflation Factor (VIF) Analysis

Feature	VIF	Collinearity Level
Cumulative Water Production	999.00	Severe Multicollinearity
Fracturing fluid return rate	999.00	Severe multicollinearity
Daily water production	24.36	Severe multicollinearity
Wellhead Pressure	23.07	Severe multicollinearity
Actual liquid level	17.40	Severe multicollinearity
Torque %/Current A	5.07	Moderate multicollinearity
RPM/Stroke	5.02	Moderate multicollinearity
System Pressure	4.60	No multicollinearity
System Pressure	3.57	No multicollinearity
Pump Efficiency	2.82	No multicollinearity

We inserted strongly associated features using correlation clustering, retain only one representative feature in a group. This improved our model's interpretability and reduced problems of the multicollinearity. Because of this our analysis was more effective and efficient and produced better insights without inflating AIGC rates.

In order to calculate the correlation coefficient matrix, The relation between each pair of original features is assessed. We can identify these correlations by setting a correlation threshold of 0.68, which denotes highly connected characteristics and groupings. In order to create the final feature set, one feature is selected as a representative from each category. This approach guarantees precision and effectiveness in choosing the most pertinent and varied aspects.

Highly associated features were found in correlation analysis in table 2, and Figure 4 shows how cluster analysis divided them into three different categories. These results highlight the connections between variables and contribute to a better comprehension of the data patterns found in the research.

Table 2 Pairs of Highly Correlated Features

Feature 1	Feature 2	Correlation Coefficient
Bottom Hole Pressure	Torque %/Current A	-0.855279
Bottom Hole Pressure	Actual liquid level	0.961906
Bottom Hole Pressure	Cumulative Water Production	-0.883171

Feature 1	Feature 2	Correlation Coefficient
Bottom Hole pressure	Fracturing fluid recovery rate	-0.885690
Casing pressure	System Pressure	0.749062
Torque %/Current A	Actual Liquid Level	-0.832141
Torque %/Current A	Cumulative Water Production	0.760135
Torque %/Current A	Fracturing Fluid Return Rate	0.762208
Actual liquid level	Cumulative Water Production	-0.867059
Cumulative Water Production	Fracturing fluid return rate	0.999957
Actual liquid level	Fracturing fluid return rate	-0.869312
Daily Water Production	Rotational Speed/Stroke Rate	0.701321
Pump Efficiency	Cumulative Water Production	0.683532

Fig. 3 Correlation-Based Clustering Results

By using a dual correlation verification approach, We examined the complex relationships between daily water production, cumulative water production, casing pressure and daily gas production. Pearson's correlation coefficient for linear relationships and Grey correlation for nonlinear relationships were both used in the investigation. The findings provided very clear insights into the complex relationships between the variables by revealing clear nonlinear correlation patterns.

Table 3 Nonlinear Correlation Matrix

Feature	Grey Correlation Degree	Pearson Correlation Coefficient	Nonlinearity Strength	Relationship Type
Well Pressure	**0.728**	**0.145**	**0.583**	**Moderate Nonlinear Relationship**
Daily Water Yield	**0.560**	**0.120**	**0.440**	**Weakly nonlinear relationship**
Cumulative water production	**0.647**	**0.859**	**-0.212**	**Strong linear relationship**

Table 3 shows the moderate nonlinearity relationship for well pressure which implies not to be much important. Daily water production also not suitable for retention due to its low linear correlation and weak nonlinear association. On the other hand, future model projections will use cumulative water production as an input characteristic because it shows a significant linear relationship.

3.4 Model Performance Validation

This study sought to assess the effectiveness of a long short-term memory network (LSTM) transfer learning fusion model in predicting the performance of coalbed methane wells in the Fukucang area [3, 12]. Using data from three horizontal wells which are FSL-30, FSL-32, and FSL-34 the study compared the capability of the conventional sample prediction model and few sample prediction. Research aim to illustrate the benefits of the LSTM transfer learning fusion model over conventional LSTM models by carrying out experiments, evaluating goodness of fit using R^2 values, and analyzing a variety of error indicators, including mean absolute error, root mean square error, and mean absolute percentage error[17,18,19]. The results indicated reduced in a data dependence, increased in prediction accuracy, and simplifying hyper parameters which highlighting the promise of this model in predicting coalbed methane well performance.

3.4.1 Conventional Sample Prediction Results

Before moving the learnt parameters to the FSL-32 well and FSL-34 wells for additional fine-tuning, the conventional method of sample-based prediction need a pre-training the model on data from the FSL-30 well. This procedure enables the model to capture temporal information from the original source well while adapting to the distinct data patterns of the target wells. This approach enables the model to effectively and correctly generate accurate prediction based on several reliable datasets.

(1) Source Domain Pre-training Results (FSL-30 Well)

Daily samples from the FSL-30 well dataset are subjected to a number of preprocessing procedures, such as data smoothing, missing value imputation and outlier elimination. The 80% of samples dataset is divided into a training set and a 20% of samples test set at an 8 by 2 ratio in accordance with time-series data analysis standards. This division makes sure that the model training process is not interfered by forward information leakage. The FSL-30 well dataset is now ready for efficient analysis and model building.

A fully connected layer structure and a two-layer LSTM were used in the pre-training stage. Adam optimizer with a batch size of 32, 0.002 initial learning rate, a sliding window stride of 5 for the input size, an LSTM hidden layer size of 128, a lag feature stride of 1 and the mean squared error (MSE) as the loss function was chosen after hyper parameters were adjusted using grid search.

The pre-training results are shown in Figure 4, which shows an R^2 value of 0.94 on the test set and a Model Absolute Percentage Error (MAPE) of 3.4feature remarkable point is that the residuals cluster close to zero and exhibit no discernible skew and also the projected values closely match the actual values over time. This implies that the time relationship between coalbed methane production and cumulative water production was successfully captured by the pre-trained model. It can therefore be used as a reliable benchmark for additional improvement in the intended field. As a result, this pre-trained model has the potential to improve accuracy and efficacy in future applications.

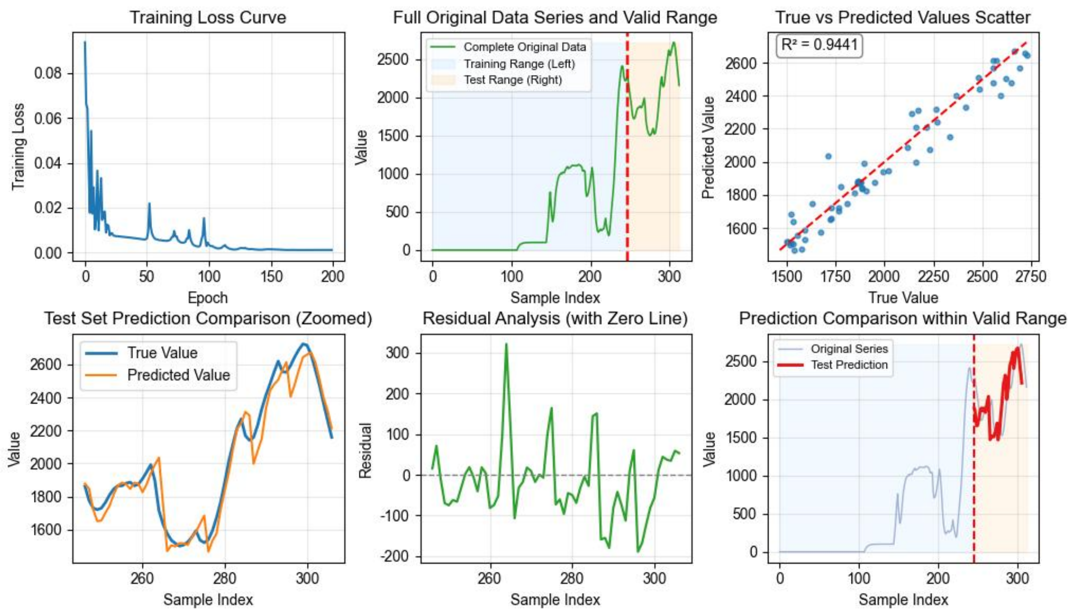
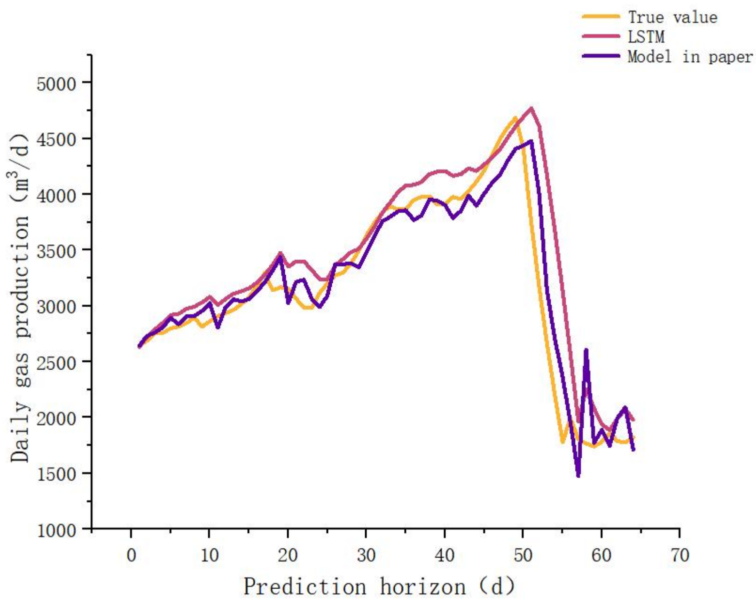


Fig. 4 Pre-training Visualization Results

(2) Target Domain Fine-Tuning Results (FSL-32, FSL-34)

A hierarchical learning rate technique used to maximize the performance of the model during the Fine-tuning process in specific domains. The model was able to maintain the temporal feature extraction skills learnt during pre-training while avoiding the core parameters from being overloaded by the small data samples in the target domain by employing a lower learning rate of 1×10-5 in the long-term and short-term memory layers. The fully connected layer, on the other hand, was given a higher learning rate of 1×10-4 in order to speed up change in parameters and allow the model to quickly adjust to the local data distribution of the FSL-32 and FSL-34 wells. The most important thing is that, the "Cumulative water production" feature from the source domain was directly used, while the input features of the target domain remained unchanged. Additionally, a double-layer learning rate adjustment made the model transfer procedure more simple while maintaining the same structure of the LSTM network. Both well FSL-32 and FSL-34 well saw remarkable performance gains with the improved model. For the FSL-32 well, R^2 rose from 0.70 to 0.90 equivalent to 77% increase, while the test of set MAPE dropped from 10.40% to 6.1%. The model's predicted accuracy was further improved by the notable decreases in the MAE and RMSE values. More Significant improvements were also noted for the FSL-34 well test set. R^2 improved from 0.80 to 0.93 equivalent to 86% percent of increased value, while the MAPE dropped from 0.614% to 0.288%. Additionally, the MAE and RMSE values significantly dropped, confirming the improved model's efficiency precisely forecasting the FSL-34 production of the well.



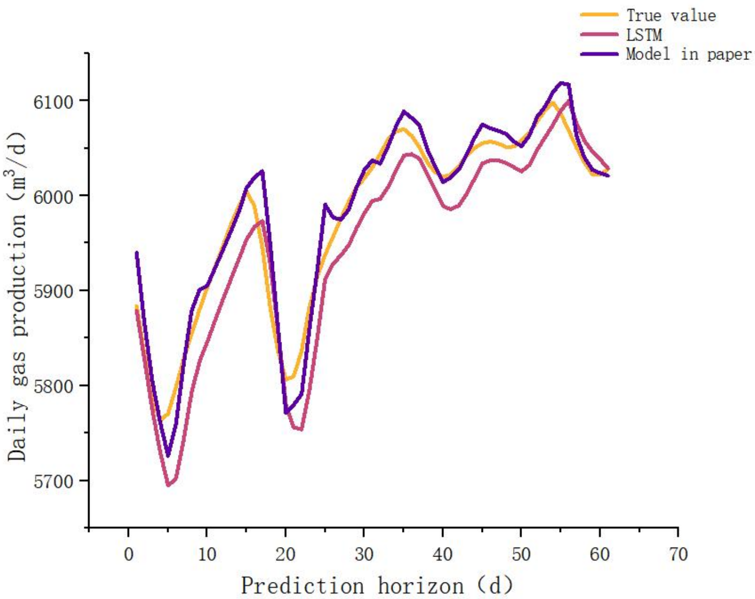


Fig. 5 Comparison of Daily Gas Production Predictions and Actual Values from Different Models for FSL-32 and FSL-34

Table 4 Performance Comparison of Different Models for Predicting Daily Gas Production

Well No.	Model	MAE	RMSE	MAPE	R^2
FSL-32	Model in this paper	167.05	250.25	6.1%	0.90
	LSTM	280.73	432.17	10.40%	0.70
FSL-34	Model in this paper	17.07	24.45	0.288%	0.93
	LSTM	36.54	42.19	0.614%	0.80

The traditional LSTM prediction curve shown in Figure 5, shows the yield's overall trend but is inaccurate when it comes to numerical fitting. However, the integrated model's prediction had little inaccuracy and closely matched the actual numbers, especially during periods of stable production. This highlights how well the hierarchical fine-tuning approach adjusts to the particular domain data.

3.4.2 Small-Sample Prediction

We reduced the FSL-34 dataset to about 100 data points while keeping the gas production data from the first 100 days in order to assess the prediction performance of the suggested model. We were able to evaluate the accuracy of the model using this lower sample prediction dataset. We used the MMD domain alignment technique described in Section 2.4 to handle the possible decrease in distribution similarity brought on by data deletion. The prediction result can be seen in Figure 6 below, demonstrating the effectiveness of the model in analyzing gas production trends.

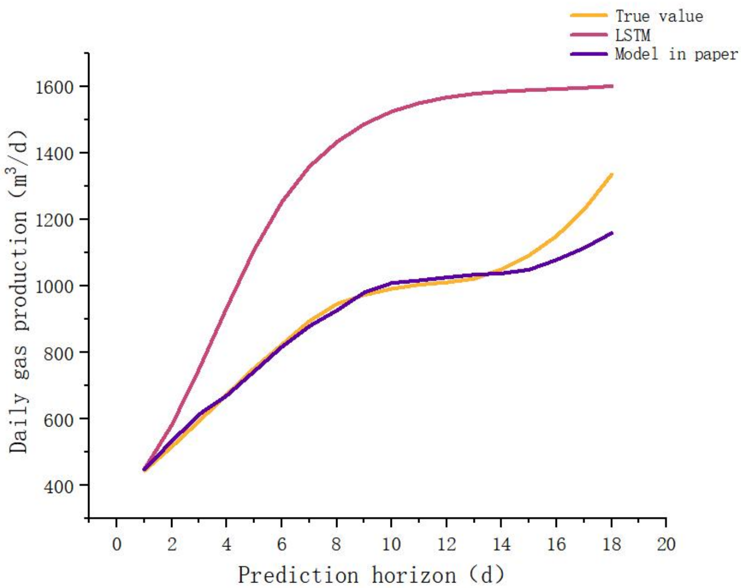


Fig. 6 Small-Sample Prediction Results

Table 6 Performance Comparison of Different Models for Predicting Daily Gas Production

Model	MAE	RMSE	MAPE	R^2
Model in this paper	**32.17**	**54.67**	**3.03%**	**0.94**
LSTM	**390.73**	**425.30**	**40.93%**	**-2.3**

The actual production curve and the predicted curve from the LSTM model differ significantly, as shown in the graph in Figure 6. However, there are very few differences between the prediction curve of an alternate model and the actual production trend. By showing a more realistic depiction of the true production curve, this alternative model performs better than the LSTM model, providing a more accurate reflection of reality.

The suggested model has a reduced prediction error and a higher R^2 value when compared to the LSTM model in Table 6. The suggested model significantly improves from -2.3 to 0.94 with the same expected days. In comparison to the LSTM model, the mean relative error is lowered by about 37%, and the MAE and RMSE errors also show significant reductions. These results indicate the wonderful performance of the proposed model in obtaining accuracy.

Predictive accuracy has significantly improved when LSTM is mixed into transfer learning models. The application of the MMD domain alignment technique helps to reduce inter-domain variations, which facilitates the transfer learning process. This is particularly helpful for datasets that exhibit notable distribution differences. Through fine-tuning the target domain using this model, adjustments to the learning rates of the LSTM layer and the fully connected layer are all that is required[9, 10, 11]. This successfully resolves the problem of excessive parameter adjustment that is frequently linked to LSTM models. A viable approach to maximizing prediction accuracy across many datasets is provided by the LSTM-integrated transfer learning model.

4. Conclusion

1. The Pearson correlation coefficient and grey correlation analysis were used to obtain the required foundation data for the model prediction (cumulative water production).
2. The difficulty of managing several hyper parameters in LSTM models is addressed by varying learning rates across layers. Reducing domain differences, reducing the need for target domain data, and enabling accurate predictions with small sample numbers are all made possible by using source domain knowledge transfer and MMD domain alignment. This method improves LSTM models' efficacy and efficiency across a range of prediction tasks..
3. Significant gains in performance metrics show that the LSTM transfer learning fusion model performs better than the basic LSTM model. For instance, the average relative error of FSL-32 dropped from 10.4% to 6.1%, and the R^2 value increased from 0.7 to 0.9. Similarly, the error rate of FSL-34 decreased from 0.614% to 0.288%, while the R^2 value climbed from 0.8 to 0.93. Furthermore, the R^2 value for small sample prediction saw a great increase from -2.3 to 0.94. Overall, the model achieved an impressive reduction of approximately 37% in the average relative error.

Acknowledgements

Project Number: 2024B01017-3; Project Name: Evaluation of Deep Coalbed Methane Resources in Key Basins of Xinjiang and Development of Favorable Area Selection and Development Technology - Subproject Three: Research on the Recoverability Evaluation and Engineering Technology System of Deep Coalbed Methane

References

[1] Li Yuan, Guo Dali, Kang Yunwei. Dynamic Prediction of Coalbed Methane Production Integrating Attention Mechanisms [J]. Science, Technology and Engineering, 2023, 23(02): 550-557.

[2] Han Yong. Adaptability Analysis and Optimization Design of Fracturing Technology in Coalbed Methane Fracturing and Production [J]. Geological Equipment, 2025, 26(04): 41-45.

[3] Yu Yong, Dai Wujie, Hu Jun. Application of Intelligent Optimization Machine Learning Methods Based on Cross-Validation in Nozzle Profile Optimization [J]. Journal of Beijing Institute of Technology, 2025, 45(08).
[4] Shang Tingting, Fu Qiming, Cui Zhiming, et al. Research on Building Energy Consumption Prediction Method Based on Sliding Window Transfer LSTM [J]. Journal of Suzhou University of Science and Technology (Natural Science Edition), 2025, 42(03): 58-69.
[5] Huang Qianguang, You Yuwen, Li Jintao. Fault Identification of Variable Air Volume Air Conditioning Systems Based on Transfer Learning Using MCNN-BiLSTM [J]. Journal of Tianjin University of Urban Construction, 2025, 31(04): 285-294.
[6] Zhang Xianmin, Wang Xuebo, Deng Ze, et al. Review of Research Progress on Enhanced Recovery Technologies for Coalbed Methane [J/OL]. Drilling and Production Technology, 2025,(05):169-177.
[7] Yu Dechao, Han Qi, Zhou Junhua, et al. Research on Dynamic Capacity Expansion Risk Prediction Method for High-Voltage Transmission Lines Based on Time-Series Data Correlation Analysis [J]. Intelligent Buildings and Smart Cities, 2025, (09): 26-28.
[8] Kang Lifang, Wang Zhenguo, Zhai Jiayu, et al. Exploration Practice and Insights of Deep Coalbed Methane in Shenfu Area, Eastern Margin of Ordos Basin [J]. Petrochemical Technology, 2025, 32(06): 285-287.
[9] Cang Qing, He Hao, Deng Fan, et al. Research on Changes in Hydraulic Structures Along the Yangtze River Based on Data Analysis and Grey Correlation Degree [J]. Water Resources Science and Technology & Economy, 2025, 31(09): 10-15.
[10] Chen Zhe, Zhou Jinhui, Jin Donghui, et al. Load Forecasting Method for Integrated Energy Systems Based on TimeVAE and Transfer Learning [J]. Zhejiang Electric Power, 2025, 44(09): 58-69.
[11] Yu Hongchi, Cui Mimi. Progress in Multi-Source Sensor Transfer Learning for Heavy Metal Tracing Monitoring in Rivers [J]. Experiment and Analysis, 2025, 3(03): 71-75.
[12] Zhao Mengxue, Shi Xiaohu, Bian Jing, et al. A Transfer Learning-Based Method for Predicting Drilling Overflow[J]. Journal of Inner Mongolia University (Natural Science Edition), 2025, 40(05):42-50.
[13] Zhang E, Liu Y, Lin G, et al. A transfer learning method of collaborating random walk and adaptive instance normalization for inscription image denoising [J]. Engineering Applications of Artificial Intelligence, 2025, 162(PB): 112458-112458.
[14] Klaib FA, Saif A, Alhosanie NT, et al. Automated assessment and detection of third molar and inferior alveolar nerve relations using UNet and transfer learning models[J]. Scientific Reports, 2025, 15(1):34529-34529.
[15] Prabhas S K, Basem A, Lakshmi L, et al. A Deep learning framework for brain tumor detection using CNNs and transfer learning on MRI scans[J]. Systems and Soft Computing, 2025, 7200389-200389.
[16] Li Juhua, Chen Chen, Xiao Jialin, et al. Prediction of Production for Multi-stage Fractured Shale Gas Wells Based on Random Forest Algorithm [J]. Journal of Yangtze University (Natural Science Edition), 2020, 39(6): 168-174.
[17] Ji Lei, Li Juhua, Xiao Jialin, et al. Application of Random Forest Algorithm in the Multi-stage Fracturing Well Reformation of Shale Gas Field [J]. Petroleum Geology & Oilfield Development in Daqing, 2020, 17(4): 34-38
[18] Li Jiaojie, Qin Shunli, Wang Jie, et al. Application of Random Forest Algorithm in Jimsuasal Shale Oil Reservoir [J]. Journal of Yangtze University (Natural Science Edition), 2023, 20(2): 69-76

[19]Liang Chenggang, Li Jihua*, Chen Yiwai, et al. Evaluation of Reservoir Production Capacity of Shale Oil Reservoirs Based on Naive Bayes Algorithm [J]. Journal of Shenzhen University (Science & Technology), 2023, 40(1): 66-73