scientific reports

OPEN

Deep insight: an efficient hybrid model for oil well production forecasting using spatio-temporal convolutional networks and Kolmogorov–Arnold networks

Yandong Hu[1], Xiankang Xin[1,2,3] ✉, Gaoming Yu[1,2,3] & Wu Deng[1]

As global energy demand grows, the oil and gas industry faces increasing challenges in optimizing production while achieving sustainability. Accurate oil well production forecasting is essential for effective resource management and operational decision-making. However, traditional mathematical models struggle with the nonlinear and dynamic characteristics of production data, while existing hybrid neural networks often lack sensitivity to operational changes and suffer from overcomplexity due to numerous parameters. This study proposes a novel hybrid model, TCN–KAN, combining temporal convolutional networks (TCN) and Kolmogorov–Arnold networks (KAN), to address these challenges. By integrating feature selection informed by reservoir engineering expertise and Spearman correlation analysis, the model effectively reduces input dimensionality while ensuring physically meaningful feature representation. Experimental results demonstrate the TCN–KAN model's superior ability to capture nonlinear interactions and long-term temporal dependencies, achieving the highest predictive accuracy among tested models. Additionally, a modified whale optimization algorithm (WOA) is employed for hyperparameter tuning, further enhancing the model's robustness. Validation on Volvo oil field data (2008–2016) highlights the model's operational sensitivity and practical value, providing actionable insights for optimizing oilfield management strategies.

Well production prediction plays a critical role in reservoir engineering, significantly influencing financial decisions, development optimization, production operations, and uncertainty management. Accurate short-term forecasts ensure daily operational targets are met while optimizing resource allocation and costs[1]. Robust long-term forecasts guide financial planning, reserve evaluations, and economic assessments of reservoirs, informing budgets and investment strategies[2]. By quantifying forecast uncertainty, engineers can better manage risks and develop flexible plans that adapt to evolving reservoir conditions[3]. Moreover, optimizing well placement and production strategies maximizes reservoir recovery and economic benefits[4].

Decline curve analysis (DCA) methods are widely used in oil and gas forecasting due to their simplicity and computational efficiency, particularly when detailed geological models are unavailable or when analyzing large datasets[5–7]. However, the accuracy of DCA is limited by its reliance on model selection, data quality, and assumptions about driving mechanisms. Different models, such as the Arps and exponential decline models, may fit the same dataset well but produce divergent predictions, leading to inaccuracies if the wrong model is chosen[8]. Additionally, traditional DCA often requires manual intervention for outlier detection and data fitting, introducing subjectivity and limiting standardization[5].

Numerical simulation(NS) offers detailed modeling of reservoir conditions, providing critical insights into flow behavior and informing optimized production strategies[9,10]. NS often achieve high accuracy in production forecasts than DCA, particularly when integrated with machine learning (ML) techniques[11]. However, NS is computationally intensive, requiring significant time and resources for model development, particularly in large-scale reservoirs with complex physical phenomena[12,13]. Additionally, uncertainties in model parameters can

[1]School of Petroleum Engineering, Yangtze University, Wuhan 430100, China. [2]Hubei Key Laboratory of Oil and Gas Drilling and Production Engineering, Yangtze University, Wuhan 430100, China. [3]School of Petroleum Engineering, National Engineering Research Center for Oil and Gas Drilling and Completion Technology, Yangtze University, Wuhan 430100, China. ✉email: xiankang.xin@hotmail.com

reduce forecast reliability, and changes in operating conditions may render previously insignificant parameters critical[14]. These challenges limit the broad applicability and efficiency of NS[15,16].

ML techniques have demonstrated significant potential in production forecasting by processing large datasets and capturing complex nonlinear relationships. Support Vector Regression (SVR), enhanced by metaheuristic optimization for hyperparameter tuning, achieves high precision[17–19]. Random Forest (RF) excels in ranking variable importance and forecasting cumulative production[20], while Gradient Boosting Trees (GBT) effectively model variable interactions and nonlinearity[21]. However, ML models are prone to overfitting and often struggle with generalization across reservoirs with differing geological and operational conditions[20,22].

Deep learning (DL) techniques, particularly Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), have advanced production forecasting by capturing long-term dependencies and temporal patterns[23,24]. Despite these improvements, traditional DL models like RNN, LSTM, and GRU still face limitations in capturing the nonlinear relationships and long-term temporal dependencies present in oil production data. These models also lack the capacity to fully address the dynamic interactions between operational behaviors and reservoir conditions, leading to suboptimal predictive performance in complex scenarios. Recent studies, such as Zhen et al.'s work on attention-enhanced temporal convolutional networks (TCNs), have demonstrated that incorporating attention mechanisms into TCNs can effectively extract temporal features and improve the performance of well production forecasting models[25]. By dynamically focusing on critical time steps and key operational factors, attention mechanisms enhance the model's ability to learn intricate patterns and dependencies in oil production data, thereby addressing some of the limitations of traditional DL models.

Hybrid DL models, such as CNN-LSTM, Attention-CNN-LSTM, and CNN-GRU, have further improved forecasting accuracy by combining CNN's feature extraction capabilities with recurrent networks' temporal modeling strengths[26–28]. Bruyker et al.[29–31] integrate data-driven approaches with physics-based principles, offering significant potential for reservoir management strategies. Recent advancements, such as the causal-based temporal graph convolutional network (CGTCN) developed by Bin et al., demonstrate the value of incorporating causal knowledge into deep learning prediction methods for CCUS-EOR. By identifying dynamic mechanisms and causal pathways in CCUS-EOR systems, CGTCN provides a template for causal discovery and prediction in energy engineering, highlighting the importance of domain knowledge integration in improving interpretability and model robustness[32]. However, a limitation in existing studies remains the lack of effective integration of domain knowledge into the feature selection process, often resulting in reduced interpretability, increased data complexity, and limited model robustness. Moreover, the substantial parameter interdependence in hybrid DL models complicates hyperparameter tuning, making it challenging to achieve robust performance across varied production data distributions. These models also often require extensive re-training to adapt to changing conditions, increasing maintenance costs and limiting their industrial application[33,34].

To address these challenges, metaheuristic algorithms, such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Whale Optimization Algorithm (WOA), have emerged as effective tools for parameter optimization in hybrid models. WOA, in particular, has demonstrated stronger global search capability, faster convergence, and better diversity preservation compared to traditional metaheuristic methods, making it well-suited for optimizing high-dimensional hybrid DL models[35].

In this context, this study develops a novel hybrid neural network model (TCN-KAN) to address these challenges. The key contributions of this research are as follows: (1) Feature Selection Strategy: A feature selection strategy guided by reservoir engineering expertise and Spearman correlation analysis to reduce input dimensionality while retaining critical production behaviors. (2) Optimization Framework: Validation of Enhanced Adaptive Learning WOA's superior global search capability for hyperparameter optimization in hybrid neural networks. (3) Hybrid Model Design: Development of the TCN-KAN model, which effectively captures long-term temporal dependencies and nonlinear relationships in oil production data, demonstrating improved forecasting accuracy and robustness.

Methodology
Proposed model
The structural design of the proposed TCN-KAN neural network, as illustrated in Fig. 1, outlines the key components and data flow of the model. The TCN layer first extracts the spatio-temporal features affecting the well production from the input layer, and then inputs them into the KAN layer. The KAN neurons then perform an approximation and recombination to predict the production rate. The entire model is optimized during the training process using the improved Whale Optimization Algorithm.

TCN-KAN model
Figure 2 illustrates the transition from traditional Multi-Layer Perceptrons (MLP) to Kolmogorov–Arnold networks (KAN). Rooted in the Kolmogorov–Arnold representation theorem, this evolution represents a fundamental shift in neural network design. The theorem states that any multivariable continuous function can be represented as a finite sum of continuous single-variable functions and their compositions. In simple terms, it decomposes complex nonlinear problems into a linear combination of simpler nonlinear functions.

The Kolmogorov–Arnold representation theorem is as follows:

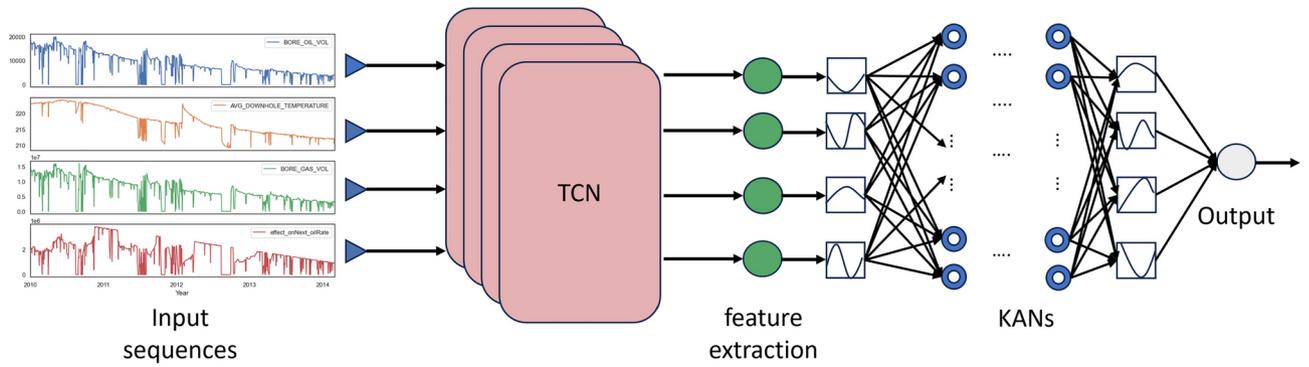$$f\left(x\right) = \sum_{q=1}^{2n+1} \phi_q \left(\sum_{p=1}^{n} \phi_{q,p}\left(x_q\right) \right) \tag{1}$$

Fig. 1. Structure of TCN-KAN.

Model	**Multi-Layer Perceptron (MLP)**	**Kolmogorov-Arnold Network (KAN)**
Theorem	**Universal Approximation Theorem**	**Kolmogorov-Arnold Representation Theorem**
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(\epsilon)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right)$
Model (Shallow)	(a) *fixed* activation functions on *nodes* / *learnable* weights on *edges*	(b) *learnable* activation functions on *edges* / sum operation on *nodes*
Formula (Deep)	$\text{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$	$\text{KAN}(\mathbf{x}) = (\mathbf{\Phi}_3 \circ \mathbf{\Phi}_2 \circ \mathbf{\Phi}_1)(\mathbf{x})$
Model (Deep)	(c) $\text{MLP}(\mathbf{x})$ — $\mathbf{W}_3$, σ_2 *nonlinear, fixed*, $\mathbf{W}_2$, σ_1, $\mathbf{W}_1$ *linear, learnable*, $\mathbf{x}$	(d) $\text{KAN}(\mathbf{x})$ — $\mathbf{\Phi}_3$, $\mathbf{\Phi}_2$ *nonlinear, learnable*, $\mathbf{\Phi}_1$, $\mathbf{x}$

Fig. 2. Comparison of KAN and MLP structures[36].

where ϕ_q and $\phi_{q,p}$ are continuous functions learned during training. This decomposition reduces complex multivariable interactions into manageable one-dimensional functions, forming the backbone of KAN. Nonlinear layers define the complexity of these functions, while linear combination layers determine the number of summation terms.

Unlike traditional neural networks, which use neuron-specific activation functions prone to redundancy and overfitting, KAN employs shared B-Spline activation functions. B-Splines are smooth, flexible piecewise polynomial functions that approximate any continuous function with high precision while maintaining computational efficiency. Their shared mechanism minimizes parameter count, and their piecewise smooth nature ensures efficient function approximation without extensive parameter learning[37,38].

Since the activation function is fitted in layers, KAN, directly addresses the need for global approximation of high dimensional problems. However, high-dimensional inputs can still pose challenges to KAN's performance, as the complexity of the input space may exceed the network's approximation capacity.

The oil production forecasting task has multiple features and both time-series dependencies and complex interactions. With the help of causal and dilation convolution, the coupled TCN can efficiently extract temporal patterns, thus complementing the time-series fitting capability of KAN; while feature engineering incorporating

reservoir engineering experience can greatly reduce the input dimensions, thus guaranteeing the nonlinear approximation capability of KAN.The capability of TCN-KAN is further optimized by meta-heuristic algorithms.

Enhanced adaptive learning whale optimization algorithm
The Whale Optimization Algorithm (WOA) is a heuristic algorithm proposed by Seyedali Mirjalili and Andrew Lewis in 2016, simulating the foraging behavior of humpback whales to optimize objective functions. Figure 3 illustrates a schematic diagram of a humpback whale using a bubble attack to gather prey. WOA is known for its simplicity, minimal parameter settings, and strong optimization performance. However, it also has drawbacks, such as low convergence precision and a tendency to get trapped in local optima[39].

To address these limitations, an Enhanced Adaptive Learning Whale Optimization Algorithm (EALWOA) is introduced. EALWOA aims to improve the search capability and convergence speed of the traditional WOA by incorporating adap- tive parameter adjustment, a learning mechanism, and a hybrid strategy. Specifically, EALWOA includes the following enhancements:

1. Adaptive Parameter Adjustment: This mechanism helps in broad exploration during the initial stages and focuses on fine-tuned exploitation in later stages. It adjusts parameters dynamically to balance exploration and exploitation throughout the optimization process.
2. Learning Mechanism: By introducing a mechanism where individual solutions learn from randomly selected peers, EALWOA increases randomness and enhances global search capabilities. This mechanism diversifies the search paths and helps escape local optima[40].
3. Hybrid Strategy: EALWOA combines random search strategies with precise optimization techniques. Dur- ing the global search phase, it employs random search to explore the solution space extensively. In the local search phase, it uses strategies such as prey encirclement and bubble net attack to refine the search around promising solutions, thus improving the speed and precision of convergence[41].

To better understand the improved algorithm used in this study, the pseudo-code of the EALWOA is provided in Algorithm 1, and Fig. 4 shows the workflow of the whale algorithm to optimize the KAN neural network. The pseudo-code outlines the step-by-step procedure of EALWOA, detailing how it initializes whale positions, updates parameters, and incorporates adaptive learning mechanisms to improve convergence and accuracy.
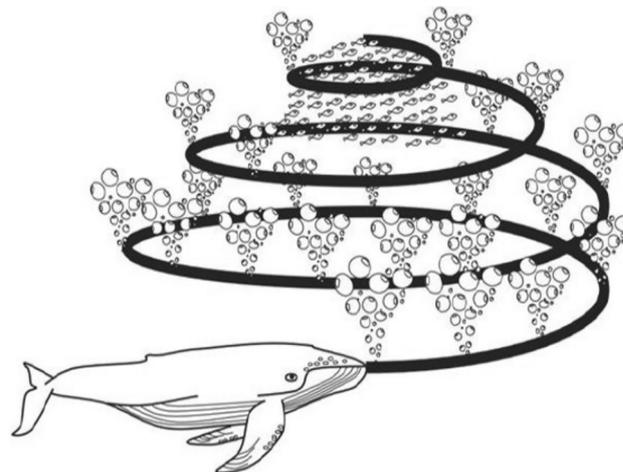

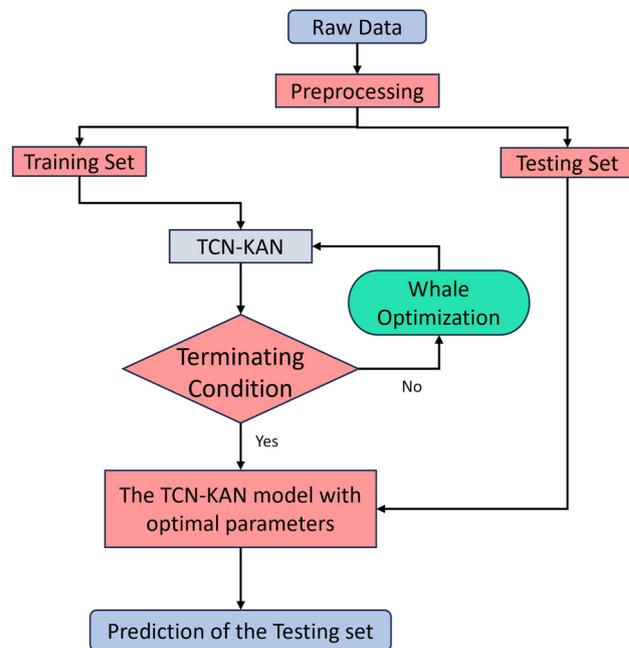**Fig. 3**. A humpback whale using a bubble attack to round up prey[39].

Fig. 4. The workflow of EALWOA optimized hyper-parameters of KAN.

1: **Input**: Objective function f, lower bound **lb**, upper bound **ub**, dimension d, number of whales N, max iterations T_{max},
    optional thresholds ($fitness\_threshold$, $convergence\_threshold$, $patience$).
2: **Initialize**: Randomly generate N whales $\mathbf{X}_i$ ($i = 1,...,N$) within [**lb**, **ub**].
3: Evaluate $f(\mathbf{X}_i)$ and find the best whale $\mathbf{X}^*$ with fitness f^*.
4: **for** $t \leftarrow 1$ to T_{max} **do**
5: Update $a = 2 - 2t/T_{max}$, $a2 = -1 - 1/T_{max}$
6:      **for** $i \leftarrow 1$ to N **do**
7:          Generate random values r_1, r_2, p, l, b.
8:          $A = 2 \cdot a \cdot r_1 - a$, $C = 2 \cdot r_2$.
9:          **if** $p < 0.5$ **then**
10:             **if** $|A| \geq 1$ **then**
11:                 Select a random whale $\mathbf{X}_{rand}$; update $\mathbf{X}_i \leftarrow \mathbf{X}_{rand} - A \cdot |\mathbf{X}_{rand} \cdot C - \mathbf{X}_i|$.
12:             **else**
13:                 Update $\mathbf{X}_i \leftarrow \mathbf{X}^* - A \cdot |\mathbf{X}^* \cdot C - \mathbf{X}_i|$.
14:             **end if**
15:         **else**
16:             // Spiral search
17:             $\mathbf{X}_i \leftarrow |\mathbf{X}^* - \mathbf{X}_i| \exp(b \cdot l) \cos(2\pi l) + \mathbf{X}^*$.
18:         **end if**
19:         // Learning mechanism (10% probability)
20:         **if** rand() < 0.1 **then**
21:             Randomly select $\mathbf{X}_{rnd}$;
22:             $\mathbf{X}_i \leftarrow \mathbf{X}_i + (\mathbf{X}_{rnd} - \mathbf{X}_i)/2$.
23:         **end if**
24:         // Ensure $\mathbf{X}_i$ stays in [**lb**, **ub**] and is integer if needed
25:         $\mathbf{X}_i \leftarrow \text{clip}(\mathbf{X}_i, \mathbf{lb}, \mathbf{ub})$.
26:     **end for**
27:     Evaluate $f(\mathbf{X}_i)$ for all whales.
28:     Update best whale $\mathbf{X}^*$ and f^* if a better fitness is found.
29:     // Check optional stopping criteria (fitness threshold, convergence threshold, patience)
30:     **if** stopping criteria met **then**
31:         **break**
32:     **end if**
33: **end for**
34: **Return** $\mathbf{X}^*, f^*$.

Algorithm 1. Enhanced adaptive learning whale optimization algorithm (EALWOA)

Reference models

The proposed WOA-TCN-KAN network is compared with several classical deep learning networks to evaluate its performance and advantages. A brief description of the deep learning networks used for comparison in this study is provided below.

RNN network

Recurrent Neural Networks (RNNs) are a class of Artificial Neural Networks (ANNs) specifically designed for modeling temporal dependencies in sequential data[42–44]. They have found wide applications in fields such as time series analysis, natural language processing, and medical research[44–46]. Figure 5a illustrates the basic structure of an RNN, where hidden states are connected across time steps to capture sequential patterns.

Its operations can be described by the following equations:

$$h_t = \sigma \left(W_{xh} x_t + W_{hh} h_{t-1} + b_h \right), \tag{2}$$

$$y_t = W_{hy} h_t + b_y. \tag{3}$$

Here, h_t represents the hidden state at time t, updated based on the input x_t and the previous hidden state h_{t-1}. W_{xh} and W_{hh} are weight matrices connecting the input and hidden states, respectively, while b_h is the bias term. The activation function σ, commonly tanh or ReLU, introduces non-linearity. The output y_t is computed using the hidden state and the output weight matrix W_{hy}, along with the bias b_y.

Despite their simplicity and effectiveness in capturing short-term dependencies, RNNs suffer from significant drawbacks. Specifically, they are prone to vanishing and exploding gradient problems during training, particularly when modeling long sequences[47,48]. These issues hinder the network's ability to learn long-term dependencies,
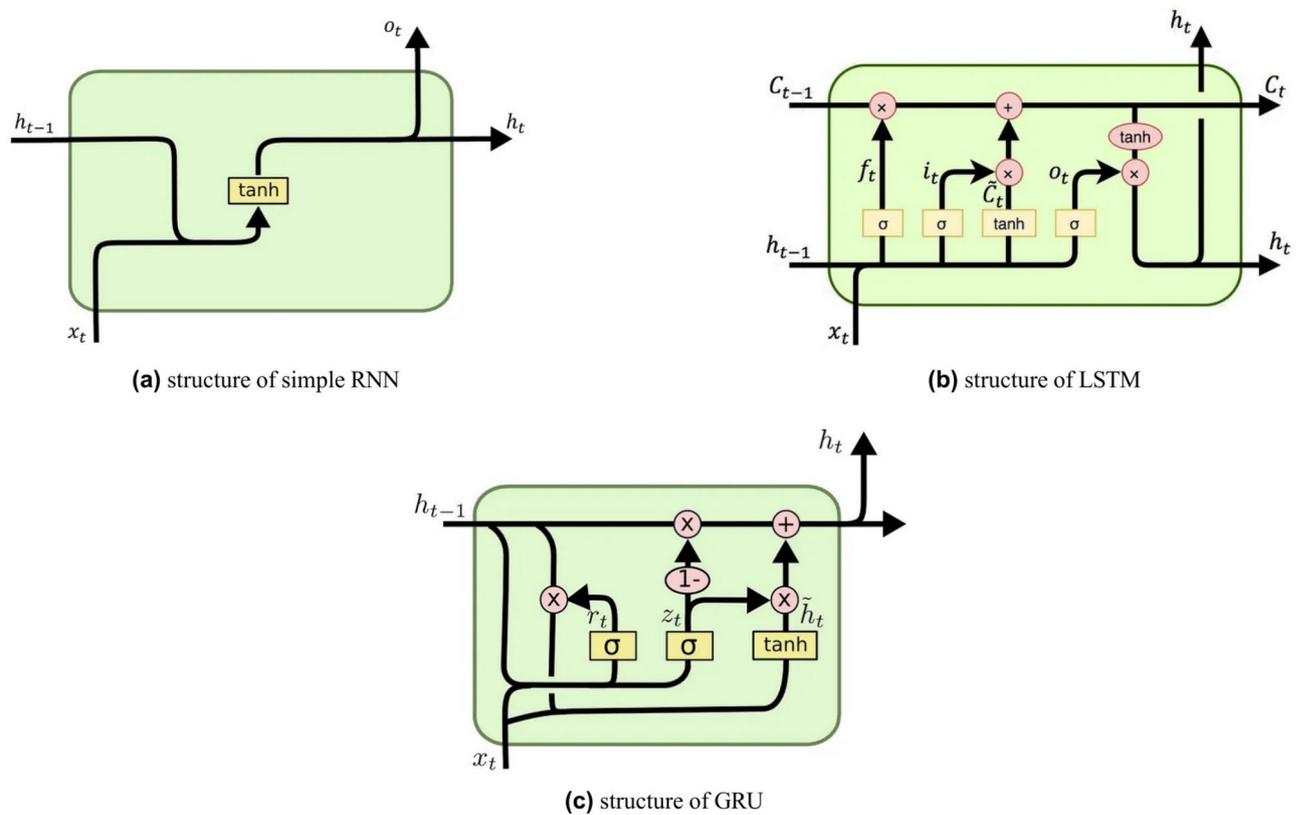
Fig. 5. Comparison of different RNN structures: (**a**) Simple RNN, (**b**) LSTM, (**c**) GRU.

as gradients decay or explode during backpropagation through time[49,50]. To address these limitations, advanced architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks have been developed, offering improved capabilities for managing long-term dependencies and ensuring training stability. In this study, RNN serves as a baseline model for predicting oil production.

LSTM and GRU networks
To address the gradient-related issues in RNNs and capture patterns in long sequences of production data, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks were developed.

The LSTM network introduces three gating mechanisms: the forget gate (f_t), input gate (i_t), and output gate (o_t), which enable effective management of long-term dependencies while mitigating the vanishing gradient problem. The key equations are as follows:

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

$$\tilde{C}_t = tanh \left(W_C \cdot [h_{t-1}, x_t] + b_C \right), C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{5}$$

$$o_t = \sigma \left(W_o \cdot [h_{t-1}, x_t] + b_o \right), h_t = o_t * tanh \left(C_t \right) \tag{6}$$

The GRU network simplifies the LSTM architecture by merging the forget and input gates into a single update gate (z_t) and introducing a reset gate (r_t). This reduces the model's complexity while maintaining competitive performance. The GRU equations are as follows:

$$z_t = \sigma \left(W_z \cdot [h_{t-1}, x_t] + b_z \right), r_t = \sigma \left(W_r \cdot [h_{t-1}, x_t] + b_r \right) \tag{7}$$

$$\tilde{h}_t = \tanh \left(W_h \cdot [r_t * h_{t-1}, x_t] + b_h \right), h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{8}$$

While both LSTM and GRU can capture long-term dependencies in sequential data, GRU offers computational efficiency due to its simpler structure. Figure 5b, c illustrate their respective architectures. GRUs are particularly advantageous in scenarios with limited computational resources or shorter input sequences[51]. This study utilizes LSTM and GRU networks as improved baseline models for RNN networks.

TCN network
Temporal Convolutional Networks (TCNs) are specifically designed for sequence modeling tasks and employ causal convolu- tions and dilated convolutions to capture temporal dependencies efficiently. The structure of
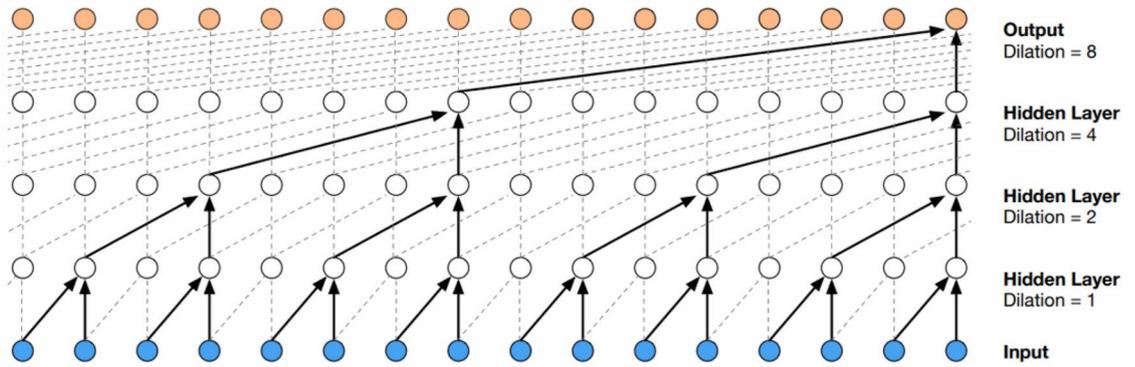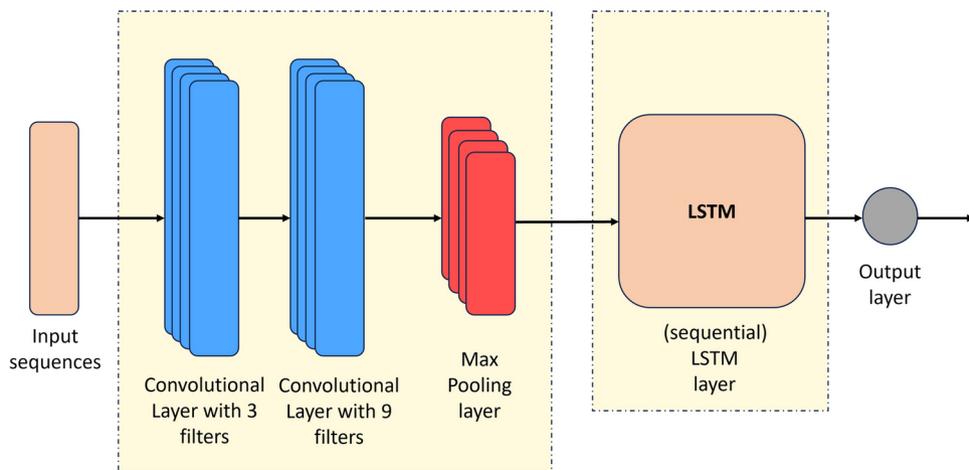
Fig. 6. Structure of TCN.

Fig. 7. Structure of CNN-LSTM.

TCN is illustrated in Fig. 6. In TCN, causal convolutions ensure that the output at time step t depends only on inputs from time t and earlier, maintaining a unidirectional flow of information. This property aligns with the temporal nature of sequential data and prevents information leakage from future time steps[52,53].

Dilated convolutions play a critical role in TCN by enabling the network to capture long-term dependencies with fewer layers. By introducing gaps between input samples during the convolution operation, the effective receptive field grows exponentially with the number of layers, allowing TCN to model long-range dependencies efficiently[54,55]. This makes TCN particularly suitable for tasks like time series forecasting, where capturing both short- and long-term temporal patterns is essential. In this study, TCN serves as an advanced baseline model to capture timing dependencies.

CNN-LSTM network
Convolutional Neural Networks (CNN) are powerful deep learning models designed for extracting spatial features from high-dimensional data. They consist of convolutional layers, pooling layers, and fully connected layers. The convolutional layers apply filters to identify patterns such as edges and textures, while pooling layers reduce spatial dimensions to decrease computational complexity and prevent overfitting. Fully connected layers integrate extracted features to produce the final output, typically for classification or regression tasks[56]. CNNs are widely recognized for their ability to capture spatial hierarchies in data and their computational efficiency through weight sharing and local connectivity[57] (Fig. 7).

The CNN-LSTM model combines the spatial feature extraction capability of Convolutional Neural Networks (CNN) with the temporal sequence modeling strength of Long Short-Term Memory (LSTM) networks. In this hybrid architecture, CNN first extracts spatial features from the input data, which are then fed into the LSTM component to capture temporal dependencies. This design enables CNN-LSTM to effectively model spatiotemporal correlations, making it a classical architecture widely used in tasks involving both spatial and temporal information. In this study, CNN-LSTM is used as a baseline hybrid model for production forecast.

Abbreviation from database	Description
DATEPRD	Date of record
ON_STREAM_HRS	On stream hours
AVG_DOWNHOLE_PRESSURE	Average downhole pressure
AVG_DOWNHOLE_TEMPERATURE	Average downhole temperature
AVG_ANNULUS_PRESS	Average annular pressure
AVG_CHOKE_SIZE_P	Average choke size percentage
AVG_WHP_P	Average wellhead pressure
AVG_WHT_P	Average wellhead temperature
BORE_OIL_VOL	Oil volume from well
BORE_WAT_VOL	Water volume from well
BORE_GAS_VOL	Gas volume from well
BORE_WI_VOL	Water volume injected
FLOW_KIND	Type of flow (production or injection)
WELL_TYPE	Type of well (oil production or water injection)

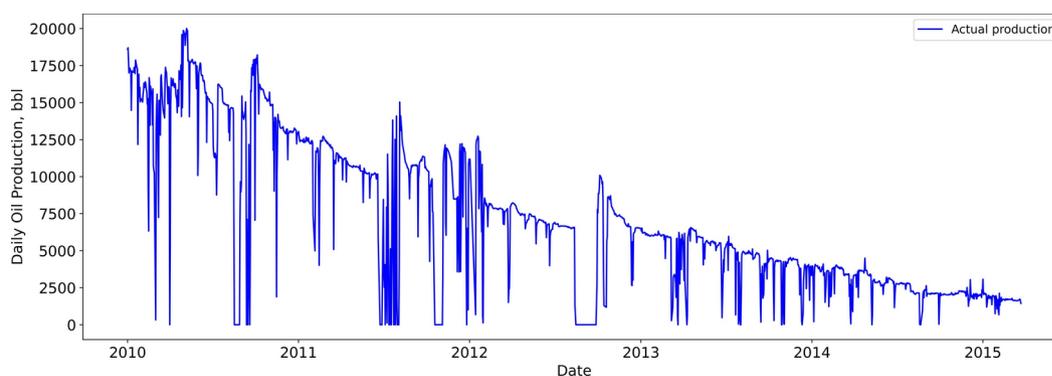**Table 1.** Data provided for each well in the Volve field.

Fig. 8. Oil production of the well NO159–F–14H.

Well production forecasting

Data description

The Volve oil field, operated by Equinor, is located in the North Sea, approximately 200 km west of Stavanger, Norway. Discovered in 1993, the field started production in 2008 and ceased operations in 2016. The Volve field was developed using a floating production storage and offloading (FPSO) unit and included both production wells and injection wells. The field is known for its challenging reservoir conditions, which provided a valuable opportunity to test advanced reservoir management and production optimization techniques. Equinor released a comprehensive dataset from the Volve field in 2018, aimed at fostering research and development in reservoir engineering and data analytics. This dataset includes seismic data, well logs, reservoir simulation models, and actual production data, making it an invaluable resource for academic and industry researchers alike[58].

The public dataset released by Equinor (2018) for research purposes includes a variety of data, such as seismic, logging, and reservoir simulation. The actual production data comprises five production wells and two injection wells: NO15/9-F-1 C, NO15/9-F-11H, NO15/9-F-12H, NO159-F-14H, NO15/9-F-15D, NO15/9-F-4AH and NO15/9-F-5AH. The data for each well are shown in Table 1. To investigate the effect of the injection and extraction relationship on reservoir production, data from three wells—NO159-F-14H, NO15/9-F-4AH, and NO15/9-F-5AH—were selected, resulting in a dataset of 1903 data points for this study. The daily production profile for the NO159-F-14H production well is shown in Fig. 8.

Data preprocessing and feature engineering

Spearman correlation analysis is a feature selection method frequently employed in industrial applications to enhance decision- making processes, improve model performance, and effectively manage data complexity[59–63]. In Fig. 9, we use a feature selection strategy that combines Spearman correlation analysis with reservoir engineering expertise. Spearman correlation was utilized as a preliminary screening tool to evaluate nonlinear relationships between features and the target production rate. Subsequently, reservoir engineering knowledge was employed to design and select features capable of capturing operational behaviors and reservoir dynamics, as well as to eliminate redundant features.
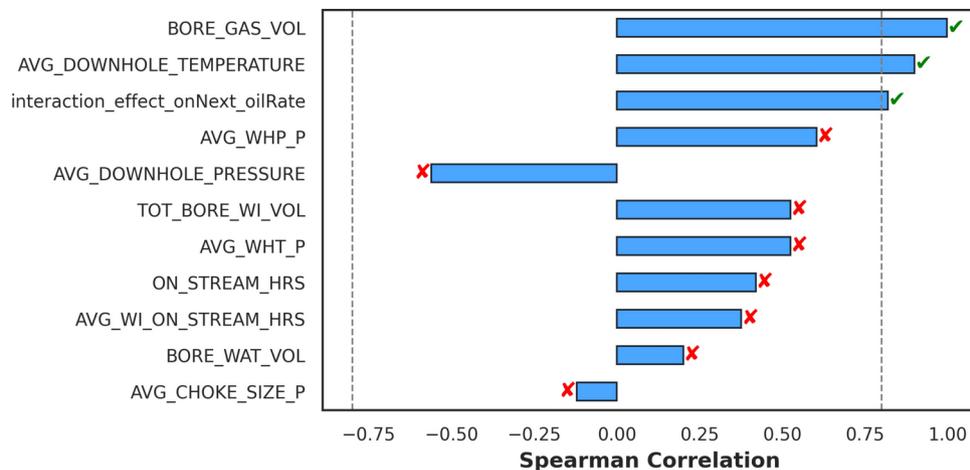
Fig. 9. Feature selection based on Spearman Correlation.

Abbreviation	Conversion formula
TOT_BORE_WI_VOL	F_4_BORE_WI_VOL + F_5_BORE_WI_VOL
AVG_WI_ON_STREAM_HRS	(F_4_ON_STREAM_HRS + F_5_ON_STREAM_HRS)/2
effect_onNext_oilRate	AV G_CHOKE_SIZE_P(-1)×ON_STREAM_HRS(-1)

Table 2. Interaction item design.

By integrating Spearman analysis with domain-specific knowledge, this approach not only minimizes input dimensionality but also ensures the statistical and physical relevance of the selected features, aligning with the objectives of model simplification and robust performance. Compared to post hoc importance ranking methods like Random Forest or Lasso regression, which rely on specific model structures and hyperparameters, the proposed method offers stronger interpretability and robustness, particularly in scenarios requiring a balance between dimensionality reduction and feature relevance.

To determine the relationship between injection and production wells, the data from injection wells were first processed to calculate the total injection volume and average running time for NO15/9-F-4AH and NO15/9-F-5AH. Additionally, a time lag term was introduced to capture the delayed effect of choke size and runtime on production. The details of the interaction item design are shown in Table 2.

In Fig. 9, the feature `effect_onNext_oilRate` represents the interaction of choke size and runtime at the previous time step. This selected interaction term was specifically designed to capture the delayed influence of operational behaviors, such as well shut-in or production parameter adjustments, on subsequent oil production. Notably, the correlation analysis shows a significant increase in correlation for the feature `effect_onNext_oilRate` after combining the throttle size and runtime, effectively capturing the impact of pressure recovery following shut-in operations on production rates. Moreover, the average annular pressure (`AVG_ANNULUS_PRESS`) is influenced by the choke size percentage (`AVG_CHOKE_SIZE_P`), which is already incorporated as part of the interaction term `AVG_ANNULUS_PRESS` in the selected features. To avoid feature redundancy and potential interference between correlated inputs, `AVG_ANNULUS_PRESS` was excluded from the final feature set.

Additionally, the two water injection-related features, `AVG_WI_ON_STREAM_HRS` and `TOT_BORE_WI_VOL`, are in- tended to reflect the impact of water injection in flooding and replenishing reservoir pressure to enhance production. However, their correlation coefficients with the production rate are not significant. A possible explanation is that the pressure recovery induced by water injection does not immediately manifest in daily oil production, and the heterogeneity of the reservoir may lead to the establishment of preferential flow pathways during the flooding process, causing partial pressure dissipation. For these reasons, injection-related variables were excluded from the final model despite their theoretical relevance.

The other three selected features are `AVG_DOWNHOLE_TEMPERATURE`, `BORE_GAS_VOL`, and `BORE_OIL_VOL`. The feature `AVG_DOWNHOLE_TEMPERATURE` reflects fluid properties such as oil viscosity, gas-oil ratio (GOR), and oil–water relative permeability. In field operations, maintaining a controlled and low GOR is a common practice to stabilize reservoir pressure and enhance oil recovery. By incorporating GOR-related inputs, the model indirectly captures the mechanisms through which operational adjustments control production parameters and optimize production. Changes in GOR indicate long-term reservoir pressure trends, while the pressure effects of Enhanced Oil Recovery (EOR) operations are effectively controlled by the interaction term, reducing the need for pressure-related features in the final model. The final selection of input features is shown in Table 3.

Input data	Unit
effect_onNext_oilRate	–
BORE_OIL_VOL	m3 (daily)
BORE_GAS_VOL	m3 (daily)
AVG_DOWNHOLE_TEMPERATURE	°C
Output data	
Oil Production	m3 (daily)

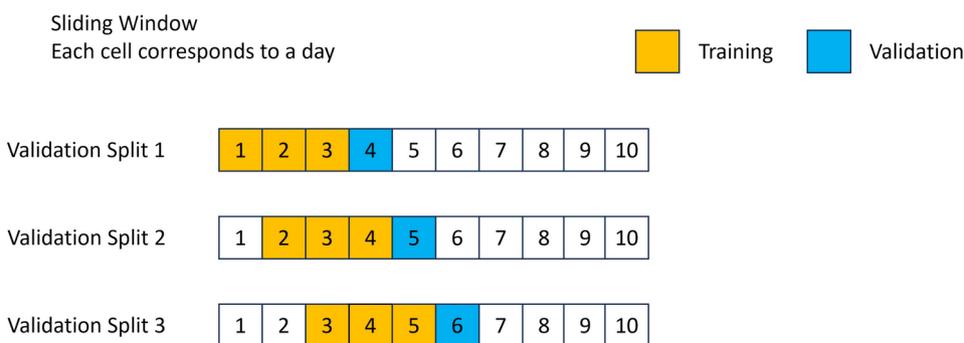**Table 3**. Selected input and output data for data-driven modeling.

Fig. 10. Sliding window mechanism.

To eliminate the effect of data scale on the model, the dataset was normalized to the interval [0,1]. Normalization improves the speed of convergence and reduces training error. The formula for normalization is:

$$X_{i,normalized} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{9}$$

The normalized data was then processed into a sliding window format, transforming the prediction from a time series problem to a supervised learning problem. A schematic of the data processing using the sliding window is provided in Fig. 10. The normalized and reconstructed data were divided into training and test sets in an 8:2 ratio.

Evaluation criteria
In this study, the coefficient of determination (R^2) and the root mean square error (RMSE) were used as indicators to evaluate the prediction of precision with the following expressions:

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{10}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{11}$$

In the above equations, n denotes the number of predicted points, y_i denotes the actual value, and $\hat{y}_i$ denotes the predicted value at the ith point.

Seven deep learning networks were built to predict well production, and the predictive performance of the models was evaluated using predefined metrics. The significance of the two metrics is as follows: R^2 indicates the model's ability to explain the variance of the data. Its value ranges from 0 to 1, with values closer to 1 indicating a better fit of the model. RMSE measures the squared mean of the differences between the predicted and actual values. Due to the squaring operation, RMSE is more sensitive to larger errors, making it effective in penalizing large deviations.

Oil production prediction
The development of the TCN-KAN network involved careful hyperparameter selection and optimization to ensure robust and accurate predictions. The hyperparameter ranges for the model were determined based on prior methodologies and recommendations from relevant literature[64–68], as well as domain expertise. Table 4 summarizes the hyperparameter search space for the TCN-KAN network, while Table 5 details the parameter settings for the three metaheuristic optimization algorithms employed in this study. Fairness across the three algorithms was ensured by using consistent settings for population size, iteration limits, and convergence thresholds.

Hyperparameter	Search space
Batch size	2–8
Window size	2–8
Hidden layers (KAN)	1–4
Number of neurons (KAN)	20–160
Kernel size (TCN)	1–4
Number of kernels (TCN)	3–9

Table 4. Search space for hyperparameters of the TCN-KAN network.

Algorithm	Parameters
GA	Population size (10)
	Generations (50)
	Crossover probability (0.8)
	Mutation probability (0.2)
	Convergence threshold (1×10^{-5})
PSO	Number of particles (10)
	Iterations (50)
	Inertia weight (0.5)
	Cognitive/social coefficients ($c_1 = 1.5$, $c_2 = 1.5$)
	Convergence threshold (1×10^{-5})
WOA	Number of agents (20)
	Iterations (50)
	Search range control (a: $2 \rightarrow 0$)
	Search probability weight (a_2: $-1 \rightarrow 0$)
	Stochastic learning probability (0.1), switch probability (0.5)
	Convergence threshold (1×10^{-5})

Table 5. Parameter settings for different optimization algorithms.

For the TCN-KAN model, the kernel size specifies the receptive field dimensions, determining the size of the local features captured by the network. The number of kernels represents the total number of convolutional filters applied in a layer, which governs the depth of feature extraction. Additionally, 'Hidden layers' and 'Number of neurons' refer to the number of layers and neurons in the KAN component, respectively. These parameters influence KAN's nonlinear approximation capacity, allowing it to model complex relationships in the input data effectively.

The performance evaluation of the TCN-KAN model was conducted using two key metrics, R^2 and RMSE, which provide a comprehensive assessment of predictive accuracy and error magnitude. Additionally, visual analyses, including predicted vs. true production profiles, intersection plots, and residual plots, were employed to investigate the model's behavior and its ability to generalize across various regions of the data distribution. These evaluation methods aim to identify the model's strengths and limitations in capturing the underlying patterns of oilfield production data.

Results and discussion

The study developed 10 neural network models to predict the daily oil production of the F-14 well, including multiple variants of the proposed TCN-KAN framework, alongside traditional and hybrid deep learning models. The objectives of this section are: (1) to validate the effectiveness of the feature engineering approach by comparing the performance of models using all features versus selected features; (2) to evaluate the effectiveness of the proposed Whale Optimization Algorithm (EALWOA) by comparing fitness curves with GA and PSO in optimizing the proposed models; (3) to analyze the proposed model's ability to capture temporal dependencies and nonlinear relationships, as well as its robustness, through production profiles, cross plots, and residual plots. This section presents the experimental results to assess the proposed methods in detail. It provides a comprehensive performance analysis, highlighting the advantages of the WOA-TCN-KAN model. Finally, the section discusses potential applications in reservoir management and explores future research directions.

Table 6 validates the effectiveness of the proposed feature selection method, which integrates Spearman correlation analysis with reservoir engineering expertise to identify key variables relevant to production. Across all models, the performance with selected features is comparable to or slightly lower than that with all features, indicating that the selected features effectively retain critical production information while significantly simplifying the model by reducing dimensionality. For instance, TCN-KAN achieved $R^2 = 0.9407$ and RMSE = 12.45 with selected features, which is very close to its performance with all features ($R^2 = 0.9487$,

Models	R²	RMSE	R²	RMSE	Models	R²	RMSE	R²	RMSE
	Selected features		All features			Selected features		All features	
RNN	0.9084	14.87	0.9150	13.95	KAN	0.9026	15.47	0.9105	14.35
LSTM	0.9177	14.42	0.9240	13.11	TCN-KAN	0.9307	13.55	0.9401	12.74
GRU	0.9115	14.72	0.9205	13.55	GA-TCN-KAN	0.9642	11.66	0.9700	10.95
TCN	0.8958	15.33	0.9052	14.27	PSO-TCN-KAN	0.9711	10.59	0.9765	9.88
CNN-LSTM	0.9140	16.04	0.9195	14.90	WOA-TCN-KAN	0.9815	9.93	0.9850	8.95

Table 6. Performance of models on the test set for different feature sets.

Models	Hidden layers	No. of neurons	kernel size	No. of kernel	Batch size	Window size
TCN-KAN	3	40	3	{9,6,6}	8	4
GA-TCN-KAN	3	83	3	{8,6,8}	4	4
PSO-TCN-KAN	3	101	3	{8,6,9}	5	4
WOA-TCN-KAN	3	117	3	{9,6,9}	5	4

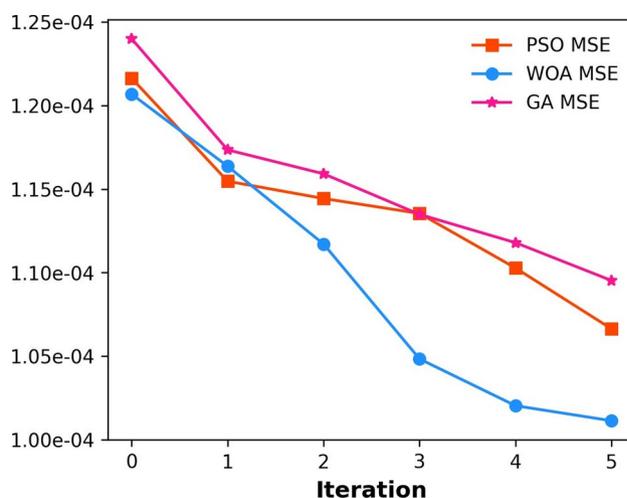**Table 7.** Hyperparameters of TCN-KAN.

Fig. 11. Fitness Curve for GA, PSO, and WOA.

RMSE = 12.21). This demonstrates that the selected features can effectively capture operational sensitivities and the impact of production behaviors on output.

The comparative performance analysis of all developed models, as shown in Table 6, highlights the superiority of the hybrid models in terms of R^2 and RMSE compared to traditional architectures. This suggests their ability to capture the complex nonlinear relationships and temporal dependencies critical(a more detailed discussion will be provided in the production profile analysis section). Among these, the WOA-TCN-KAN model exhibits the best performance, achieving $R^2 = 0.9815$ and RMSE = 9.93 with selected features, and $R^2 = 0.9850$, RMSE = 8.95 with all features. These results reflect the model's strong generalization capability and adaptability to reduced-dimensional features.

Table 6 also highlights the performance improvements achieved by employing optimization algorithms (GA, PSO, and WOA). The hyperparameter identified by each algorithm are summarized in Table 7. Compared to the PSO-TCN-KAN and GA-TCN-KAN models, the WOA-TCN-KAN model demonstrated superior performance, with R^2 improved by 1.04% and RMSE reduced by 6.23% relative to PSO-TCN-KAN, and R^2 improved by 1.76% and RMSE reduced by 14.82% compared to GA-TCN-KAN. The fitness curves shown in Fig. 11 further emphasize WOA's adaptive capabilities. Compared to GA and PSO, WOA exhibits faster convergence and more stable iterative performance, indicating its high search efficiency in high-dimensional search spaces. Notably, WOA achieves a significant reduction in MSE during the early iterations and consistently maintains a lower MSE throughout the optimization process. In contrast, PSO and GA exhibit slower convergence and higher variability in performance across iterations. This demonstrates WOA's superior ability to balance exploration and exploitation, leading to more effective parameter optimization and overall model improvement.

Figure 12 indicates that neither the TCN model (Fig. 12e) nor the KAN model (Fig. 12f) exhibits significant advantages over traditional models in terms of alignment and scatter reduction. However, the integration of TCN and KAN in the TCN-KAN (Fig. 12g) shows a marked improvement over both individual components. The WOA-TCN-KAN model (Fig. 12h) achieves the closest alignment to the diagonal line across all production levels, reflecting its superior accuracy and generalization capabilities.

The residual plots in Fig. 13 compare the prediction errors across all models. Residuals close to the zero line indicate minimal error and unbiased predictions. For traditional DL models (RNN, LSTM, GRU, CNN-LSTM) and baseline models like TCN and KAN (Fig. 13a–f), residuals are more dispersed, particularly in low- and high-production regions, reflecting their limitations in capturing the nonlinear and temporal complexities of production dynamics. Patterns of overestimation and underestimation are noticeable, suggesting challenges in generalizing across varying production levels. In contrast, the WOA-TCN-KAN model (Fig. 13h) demonstrates the most compact and symmetrical residual distribution. This result shows the robustness of WOA-TCN-KAN relative to other models.

The production forecast profiles are shown in Fig. 14. Traditional DL models (RNN, LSTM, GRU, CNN-LSTM) perform well during steady production phases but struggle with early-phase dynamics (e.g., rapid water cut increases) and late-stage complexities introduced by operational adjustments and EOR activities. These limitations highlight their inability to fully capture nonlinear and temporal dependencies in production data.

The TCN model (Fig. 14e) addresses these issues by extracting temporal dependencies more effectively, as shown by its accurate predictions during the early production phase. However, it cannot model nonlinear interactions, especially those driven by operational adjustments. In contrast, KAN(Fig. 14f) excels at modeling nonlinear relationships, such as the delayed effects of choke size, runtime, and well shut-ins on production.

By integrating TCN's temporal feature extraction capabilities with KAN's nonlinear modeling strengths, the WOA-TCN- KAN model achieves superior accuracy and robustness across the entire production lifecycle
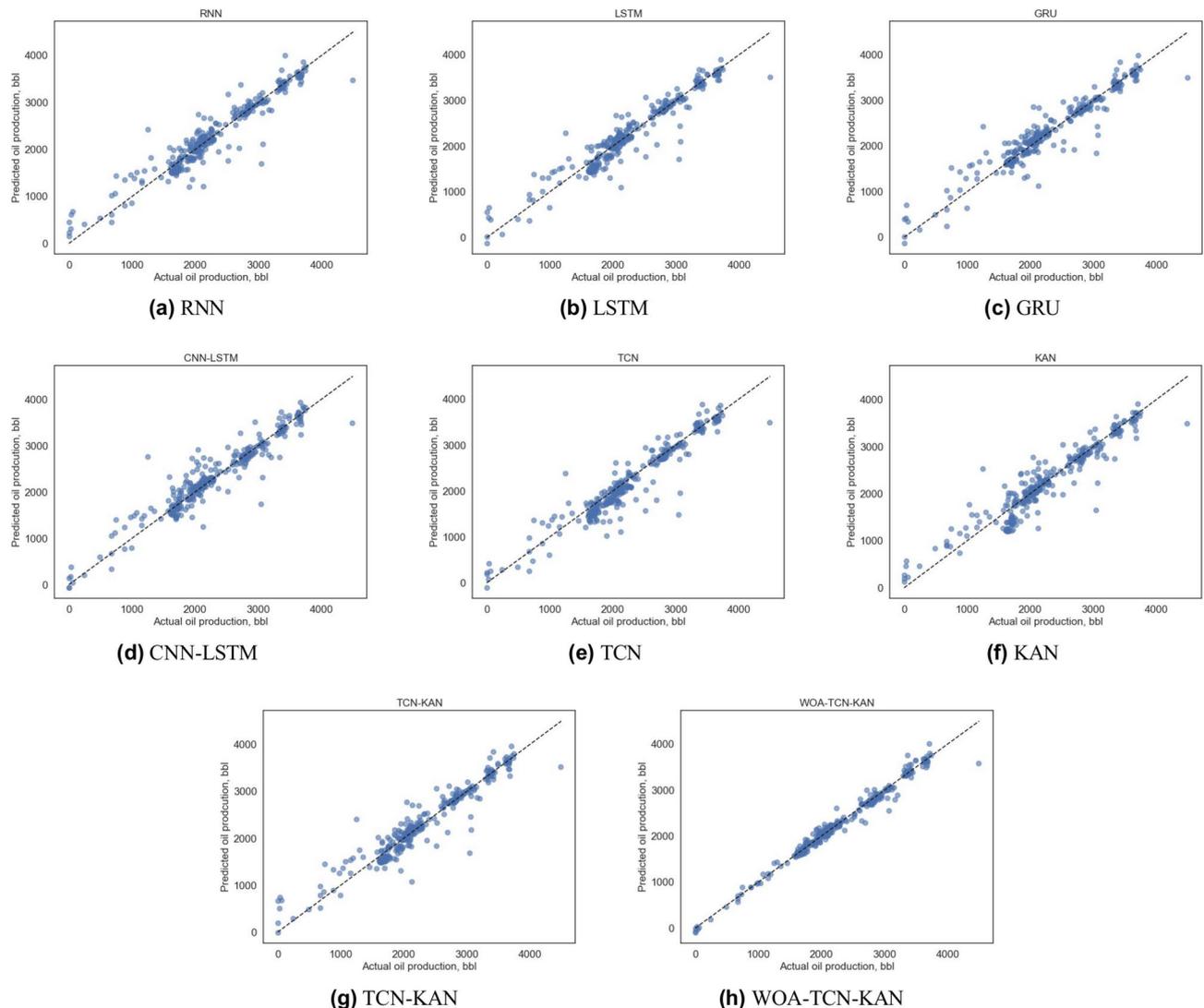

(a) RNN **(b)** LSTM **(c)** GRU

(d) CNN-LSTM **(e)** TCN **(f)** KAN

(g) TCN-KAN **(h)** WOA-TCN-KAN

Fig. 12. Cross plot of the actual and predicted oil production.
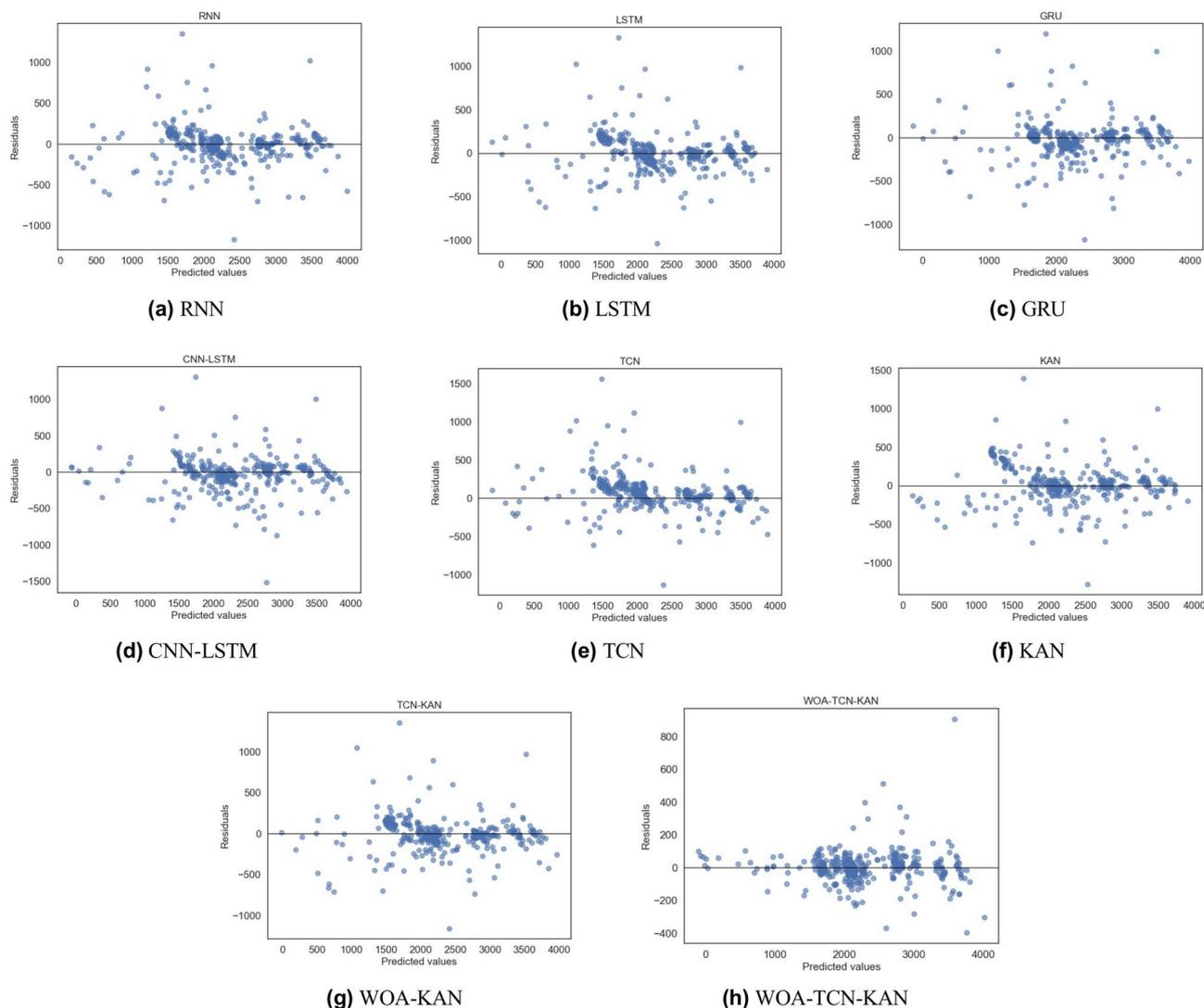
Fig. 13. Residual plots of the models: (**a**) RNN (**b**) LSTM (**c**) GRU (**d**) CNN-LSTM (**e**) TCN (**f**) KAN (**g**) WOA-KAN (**h**) WOA-TCN-KAN.

(Fig. 14h). It captures both localized production adjustments, such as the timing and effects of well shut-ins, and long-term trends, including production declines during late development stages. WOA optimization further enhances parameter tuning, improving the model's generalization and stability, and enabling it to outperform baseline models across all production phases. These results validate the efficacy of combining metaheuristic optimization with hybrid neural networks for production forecasting.

Although the results demonstrate that the WOA-TCN-KAN model achieves high accuracy in forecasting oil well production, the use of data from a single oilfield for model development and validation remains a limit of this study. The applicability of the proposed model to different geological conditions and operational contexts requires further investigation.

Theoretically, the temporal feature extraction capability of TCN and the nonlinear modeling strength of KAN suggest that the proposed approach could be generalized to reservoirs with diverse characteristics and fluid properties. By leveraging these strengths, the model can provide accurate production forecasts and support decision-making in reservoir management. Potential applications include, but are not limited to:

- *Injection rate optimization*: The model can quantify the short- and long-term impacts of EOR techniques, such as water and gas injection, on production rates, providing a basis for optimizing injection strategies.
- *Production decline prediction*: By capturing trends in oil production over time, the TCN-KAN model can help identify potential production decline points, enabling operators to adjust development strategies proactively.
- *Dynamic production adjustments*: Based on model predictions, operators can implement more precise production adjustments, such as optimizing shut-in timing and duration or adjusting choke sizes, to balance short-term production targets with long-term recovery goals.
- *Production anomaly detection*: By monitoring residual changes between predicted and actual values, the model can identify potential production anomalies, such as equipment failures or water breakthrough.
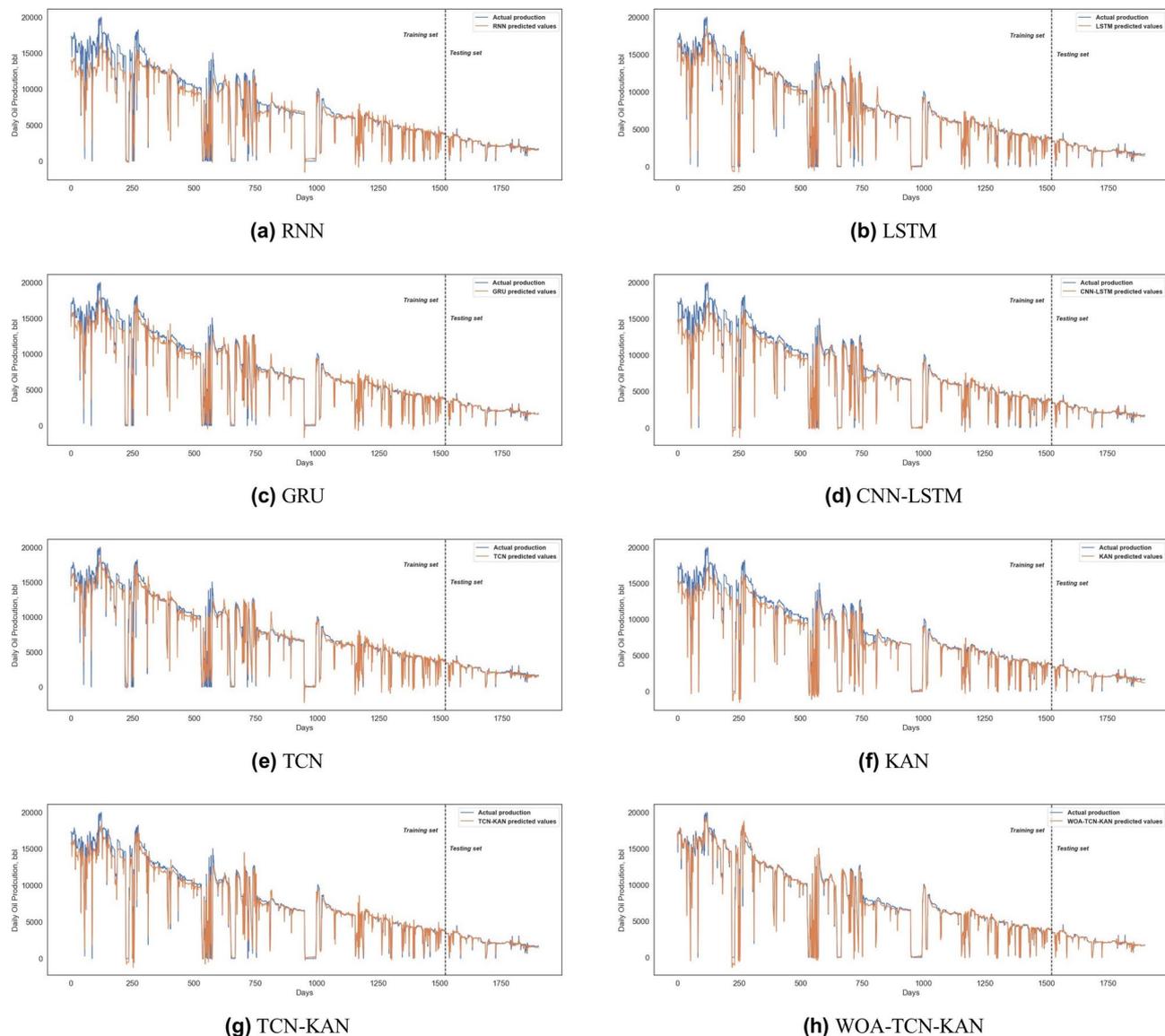
Fig. 14. Comparison of actual and predicted daily oil production using various models.

These applications highlight the model's predictive capabilities and its potential practical value in complex oilfield development scenarios. Future research will shift focus from improving predictive performance on datasets to quantifying the economic impact of various EOR techniques. This will involve using real-world datasets from multiple oilfields or synthetic data generated through numerical reservoir simulations. Combined with a domain knowledge-guided feature engineering and hybrid modeling framework, the objective is to provide reliable insights into reservoir management and serve as a tool for optimizing production strategies and economic outcomes.

Conclusion

This study proposed a novel hybrid model, TCN-KAN, integrating Temporal Convolutional Networks (TCN) and Kolmogorov–Arnold networks (KAN) to address the challenges of oil well production forecasting. By combining TCN's temporal feature extraction capabilities with KAN's nonlinear approximation strengths, the TCN-KAN model effectively captured spatio- temporal dynamics and nonlinear interactions in production data, achieving superior predictive accuracy ($R^2 = 0.9815$, RMSE = 9.93) with fewer input features.

Feature selection guided by reservoir engineering expertise and Spearman correlation analysis ensured that the selected features retained essential production information while reducing dimensionality. The inclusion of a modified Whale Optimiza- tion Algorithm (WOA) enhanced hyperparameter tuning, further improving model robustness. Comparative experiments and ablation studies validated the effectiveness of the proposed architecture and its individual components.

Beyond predictive accuracy, the TCN-KAN model provides interpretability by separating predictions properties in time and nonlinear interactions to provide actionable insights into operating parameters and

reservoir characteristics. This makes it a practical tool for optimizing reservoir management and decision making.

While promising, this study's scope is limited to a single oilfield, and future work should validate the model's generalizability to diverse geological and operational conditions. Further exploration of the economic implications of the model could also strengthen its value in oilfield development strategies.

In summary, the TCN-KAN model demonstrates a powerful and interpretable approach to oil production forecasting, balancing accuracy, computational efficiency, and practical applicability.

Data availability

References

1. Ourir, A., Oukmal, J., Rondeleux, B., Agharzayeva, Z. & Barrault, P. Hybrid data driven approach for reservoir production forecast. In *Abu Dhabi International Petroleum Exhibition and Conference*, D022S169R002 (SPE, 2021).
2. Lia, O., Omre, H., Tjelmeland, H., Holden, L. & Egeland, T. Uncertainties in reservoir production forecasts. *AAPG Bull.* **81**, 775–802 (1997).
3. Busby, D. Deep-dca a new approach for well hydrocarbon production forecasting. In *ECMOR XVII*, vol. 2020, 1–10 (European Association of Geoscientists & Engineers, 2020).
4. Alsaeedi, A. A. et al. Short-term production forecasting to evaluate production target deliverability using end to end integration platform from reservoir to surface network. In *Abu Dhabi International Petroleum Exhibition and Conference*, D012S116R094 (SPE, 2020).
5. Elahi, S. H. A novel workflow for oil production forecasting using ensemble-based decline curve analysis. https://doi.org/10.2118/195916-ms (2019).
6. Miao, Y., Zhao, C. & Zhou, G. New rate-decline forecast approach for low-permeability gas reservoirs with hydraulic fracturing treatments. *J. Petrol. Sci. Eng.* **190**, 107112. https://doi.org/10.1016/j.petrol.2020.107112 (2020).
7. Tan, L., Zuo, L. & Wang, B. Methods of decline curve analysis for shale gas reservoirs. *Energies* **11**, 552 (2018).
8. Tadjer, A., Hong, A. & Bratvold, R. B. Machine learning based decline curve analysis for short-term oil production forecast. *Energy Explor. Exploit.* **39**, 1747–1769 (2021).
9. Sikanyika, E. W. et al. Numerical simulation of the oil production performance of different well patterns with water injection. *Energies* https://doi.org/10.3390/en16010091 (2022).
10. Chen, X., Rao, X., Xu, Y. & Liu, Y. An effective numerical simulation method for steam injection assisted in situ recovery of oil shale. *Energies* https://doi.org/10.3390/en15030776 (2022).
11. He, D. et al. Oil production rate forecasting by sa-lstm model in tight reservoirs. *Lithosphere* https://doi.org/10.2113/2024/lithosphere_2023_197 (2024).
12. Negash, B. & Yaw, A. D. Artificial neural network based production forecasting for a hydrocarbon reservoir under water injection. *Petrol. Explor. Dev.* **47**, 383–392. https://doi.org/10.1016/s1876-3804(20)60055-6 (2020).
13. Amirian, E., Fedutenko, E., Yang, C., Chen, Z. & Nghiem, L. Artificial neural network modeling and forecasting of oil reservoir performance. *Appl. Data Manag. Anal. Case Stud. Soc. Networks Beyond* **8**, 43–67 (2018).
14. Vanegas, G., Nejedlik, J., Neff, P. & Clemens, T. Conditioning model ensembles to various observed data (field and regional level) by applying machine-learning-augmented workflows to a mature field with 70 years of production history. *SPE Reserv. Eval. Eng.* https://doi.org/10.2118/205188-pa (2021).
15. Yu, W. & Sepehrnoori, K. Numerical model for shale gas and tight oil simulation. *Shale Gas Tight Oil Reserv. Simul.* **10**, 11–70 (2018).
16. Tuczynski, T. & Stopa, J. Uncertainty quantification in reservoir simulation using modern data assimilation algorithm. *Energies* **16**, 1153 (2023).
17. Ng, C. S. W., Ghahfarokhi, A. J. & Amar, M. N. Well production forecast in volve field: Application of rigorous machine learning techniques and metaheuristic algorithm. *J. Petrol. Sci. Eng.* **208**, 109468 (2022).
18. Jun, Y., He, S., Kai, Z. & Yang, J. Daily production intelligent prediction model of single well in shb oilfield based on support vector machine. In *2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA)*, 724–729 (IEEE, 2022).
19. Gao, Y. & Liu, Y. Prediction of oilfield systems based on support vector regression. *J. Daqing Petrol. Inst.* **31**, 94 (2007).
20. Schuetter, J., Mishra, S., Zhong, M. & LaFollette, R. A data-analytics tutorial: Building predictive models for oil production in an unconventional shale reservoir. *SPE J.* **23**, 1075–1089 (2018).
21. LaFollette, R. F., Izadi, G. & Zhong, M. Application of multivariate statistical modeling and geographic information systems pattern-recognition analysis to production results in the eagle ford formation of south texas. In *SPE Hydraulic Fracturing Technology Conference and Exhibition*, SPE–168628 (SPE, 2014).
22. Voneiff, G. et al. A well performance model based on multivariate analysis of completion and production data from horizontal wells in the montney formation in british columbia. In *SPE Canada Unconventional Resources Conference*, SPE–167154 (SPE, 2013).
23. Mahzari, P., Emambakhsh, M., Temizel, C. & Jones, A. P. Oil production forecasting using deep learning for shale oil wells under variable gas-oil and water-oil ratios. *Petrol. Sci. Technol.* **40**, 445–468. https://doi.org/10.1080/10916466.2021.2001526 (2021).
24. Vikara, D. M. & Khanna, V. Application of a deep learning network for joint prediction of associated fluid production in unconventional hydrocarbon development. *Processes* https://doi.org/10.3390/pr10040740 (2022).
25. Zhen, Y., Fang, J., Zhao, X., Ge, J. & Xiao, Y. Temporal convolution network based on attention mechanism for well production prediction. *J. Petrol. Sci. Eng.* **218**, 111043 (2022).
26. Abdullayeva, F. & Imamverdiyev, Y. Development of oil production forecasting method based on deep learning. *Stat. Optim. Inf. Comput.* https://doi.org/10.19139/soic-2310-5070-651 (2019).
27. Pan, S., Wang, J. & Zhou, W. Prediction on production of oil well with attention-cnn-lstm. *J. Phys. Conf. Ser.* https://doi.org/10.1088/1742-6596/2030/1/012038 (2021).
28. Haq, K. P. R. A. & Harigovindan, V. P. Water quality prediction for smart aquaculture using hybrid deep learning models. *IEEE Access* **10**, 60078–60098. https://doi.org/10.1109/ACCESS.2022.3180482 (2022).
29. De Bruyker, D. et al. Improving recovery in the yates field using dynamic feedback loop based on physics-informed artificial intelligence. In *SPE Improved Oil Recovery Conference?* D021S021R003 (SPE, 2020).
30. Xue, L. et al. A data-driven shale gas production forecasting method based on the multi-objective random forest regression. *J. Petrol. Sci. Eng.* **196**, 107801 (2021).

31. Bao, A., Gildin, E., Huang, J. & Coutinho, E. J. Data-driven end-to-end production prediction of oil reservoirs by enkf-enhanced recurrent neural networks. In *SPE Latin America and Caribbean Petroleum Engineering Conference*, D021S004R001 (SPE, 2020).

32. Shen, B. et al. Interpretable causal-based temporal graph convolutional network framework in complex spatio-temporal systems for ccus-eor. *Energy* **309**, 133129 (2024).

33. Lwakatare, L. E., Raj, A., Crnkovic, I., Bosch, J. & Olsson, H. H. Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions. *Inf. Softw. Technol.* **127**, 106368. https://doi.org/10.1016/j.infsof.2020.106368 (2020).

34. Kumar, V., Prakash, M., & Thamburaj, S. Deep learning-based predictive maintenance for industrial iot applications. In *2024 International Conference on Inventive Computation Technologies (ICICT)* 1197–1202. https://doi.org/10.1109/ICICT60155.2024.10544377 (2024).

35. He, Z. & Yen, G. Many-objective evolutionary algorithms based on coordinated selection strategy. *IEEE Trans. Evol. Comput.* **21**, 220–233. https://doi.org/10.1109/TEVC.2016.2598687 (2017).

36. Liu, Z. et al. Kan: Kolmogorov–Arnold networks. arXiv preprint arXiv:2404.19756 (2024).

37. Polar, A. & Poluektov, M. A deep machine learning algorithm for construction of the Kolmogorov–Arnold representation. *Eng. Appl. Artif. Intell.* **99**, 104137. https://doi.org/10.1016/j.engappai.2020.104137 (2020).

38. Lin, J.-N. & Unbehauen, R. On the realization of a Kolmogorov network. *Neural Comput.* **5**, 18–20. https://doi.org/10.1162/neco.1993.5.1.18 (1993).

39. Mirjalili, S. & Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **95**, 51–67. https://doi.org/10.1016/j.advengsoft.2016.01.008 (2016).

40. Yan, Z., Zhang, J., Zeng, J. & Tang, J. Nature-inspired approach: An enhanced whale optimization algorithm for global optimization. *Math. Comput. Simul.* **185**, 17–46. https://doi.org/10.1016/j.matcom.2020.12.008 (2021).

41. Liu, L. & Zhang, R. Multistrategy improved whale optimization algorithm and its application. *Comput. Intell. Neurosci.* https://doi.org/10.1155/2022/3418269 (2022).

42. Alakeely, A. & Horne, R. N. Simulating the behavior of reservoirs with convolutional and recurrent neural networks. *SPE Reserv. Eval. Eng.* **23**, 0992–1005 (2020).

43. Alom, M. Z. et al. A state-of-the-art survey on deep learning theory and architectures. *Electronics* **8**, 292 (2019).

44. Mao, S. & Sejdic, E. A review of recurrent neural network-based methods in computational physiology. *IEEE Trans. Neural Networks Learn. Syst.* **34**, 6983–7003. https://doi.org/10.1109/TNNLS.2022.3145365 (2022).

45. Thakkar, U. & Chaoui, H. Remaining useful life prediction of an aircraft turbofan engine using deep layer recurrent neural networks. *Actuators* https://doi.org/10.3390/act11030067 (2022).

46. Karita, S., Kubo, Y., Bacchiani, M. A. U. & Jones, L. A comparative study on neural architectures and training methods for japanese speech recognition. arXiv preprint arXiv:2106.05111 (2021).

47. Ceni, A. Random orthogonal additive filters: a solution to the vanishing/exploding gradient of deep neural networks. ArXiv abs/2210.01245. https://doi.org/10.48550/arXiv.2210.01245 (2022).

48. Monfared, Z., Mikhaeil, J. M. & Durstewitz, D. How to train rnns on chaotic data? ArXiv abs/2110.07238 (2021).

49. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735 (1997).

50. Li, Y., Sun, R.-H. & Horne, R. Deep learning for well data history analysis. https://doi.org/10.2118/196011-ms (2019).

51. Chung, J., Çaglar Gülçehre, Cho, K. & Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. ArXiv abs/1412.3555 (2014).

52. He, Y. & Zhao, J. Temporal convolutional networks for anomaly detection in time series. *J. Phys. Conf. Ser.* https://doi.org/10.1088/1742-6596/1213/4/042050 (2019).

53. Ibrahim, E. A. et al. Dilate-invariant temporal convolutional network for real-time edge applications. *IEEE Trans. Circuits Syst. I Regul. Pap.* **69**, 1210–1220. https://doi.org/10.1109/tcsi.2021.3124219 (2022).

54. Zhu, R., Liao, W. & Wang, Y. Short-term prediction for wind power based on temporal convolutional network. *Energy Rep.* **6**, 424–429. https://doi.org/10.1016/j.egyr.2020.11.219 (2020).

55. Cheng, L., Khalitov, R., Yu, T. & Yang, Z. Classification of long sequential data using circular dilated convolutional neural networks. *Neurocomputing* **518**, 50–59. https://doi.org/10.1016/j.neucom.2022.10.054 (2022).

56. Li, Z., Liu, F., Yang, W., Peng, S. & Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Trans. Neural Networks Learn. Syst.* **33**, 6999–7019. https://doi.org/10.1109/TNNLS.2021.3084827 (2020).

57. Kiranyaz, S. et al. 1d convolutional neural networks and applications: A survey. ArXiv abs/1905.03554. https://doi.org/10.1016/j.ymssp.2020.107398 (2019).

58. Kwon, S., Ji, M., Park, G., Min, B. & Jeong, H. Analysis of data disclosure and reservoir model of the volve oilfield in the north sea. *J. Korean Soc. Miner. Energy Resour. Eng.* **8**, 9. https://doi.org/10.32390/ksmer.2021.58.4.353 (2021).

59. Ghorabaee, M. K. Developing an mcdm method for robot selection with interval type-2 fuzzy sets. *Robot. Comput. Manuf.* **37**, 221–232. https://doi.org/10.1016/J.RCIM.2015.04.007 (2016).

60. Zhu, L. Selection of multi-level deep features via spearman rank correlation for synthetic aperture radar target recognition using decision fusion. *IEEE Access* **8**, 133914–133927. https://doi.org/10.1109/ACCESS.2020.3010969 (2020).

61. Wang, Y. & Zhao, Y. Three-stage feature selection approach for deep learning-based rul prediction methods. *Qual. Reliab. Eng. Int.* **39**, 1223–1247. https://doi.org/10.1002/qre.3288 (2023).

62. Jiarpakdee, J., Tantithamthavorn, C. & Treude, C. The impact of automated feature selection techniques on the interpretation of defect models. *Empir. Softw. Eng.* **25**, 3590–3638. https://doi.org/10.1007/s10664-020-09848-1 (2020).

63. Meng, C., Jiang, X., Wang, J. & Wei, X. The complex network model for industrial data based on spearman correlation coefficient. In *2019 International Conference on Internet Things (iThings) IEEE Green Computer Communication (GreenCom) IEEE Cyber, Physics and Society Computing (CPSCom) IEEE Smart Data (SmartData)* 28–33. https://doi.org/10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00028 (2019).

64. Rodriguez, A. X. & Salazar, D. A. Methodology for the prediction of fluid production in the waterflooding process based on multivariate long–short term memory neural networks. *J. Petrol. Sci. Eng.* **208**, 109715 (2022).

65. Chen, G., Tian, H., Xiao, T., Xu, T. & Lei, H. Time series forecasting of oil production in enhanced oil recovery system based on a novel cnn-gru neural network. *Geoenergy Sci. Eng.* **233**, 212528 (2024).

66. Kocoglu, Y. et al. Improving the accuracy of short-term multiphase production forecasts in unconventional tight oil reservoirs using contextual bi-directional long short-term memory. *Geoenergy Sci. Eng.* **235**, 212688 (2024).

67. Liu, W., Liu, W. D. & Gu, J. Forecasting oil production using ensemble empirical model decomposition based long short-term memory neural network. *J. Petrol. Sci. Eng.* **189**, 107013 (2020).

68. Nikitin, N. O., Revin, I., Hvatov, A., Vychuzhanin, P. & Kalyuzhnaya, A. V. Hybrid and automated machine learning approaches for oil fields development: The case study of volve field, north sea. *Comput. Geosci.* **161**, 105061 (2022).

Author contributions

Hu Yandong: Conceptualization, Methodology, Software, Data curation, Validation, Visualization, Investigation, Writing – original draft, Visualization, Investigation. Xin Xiankang: Writing – review & editing, Supervision. Yu Gaoming:Writing – review & editing, Supervision, Deng Wu:Validation.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to X.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.