

# Reservoir Production Prediction Model Based on a Stacked LSTM Network and Transfer Learning

Yukun Dong,\* Yu Zhang, Fubin Liu, and Xiaotong Cheng



Cite This: *ACS Omega* 2021, 6, 34700–34711



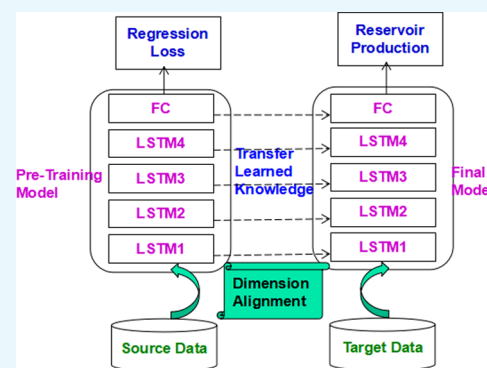
Read Online

ACCESS |

Metrics & More

Article Recommendations

**ABSTRACT:** Gas injection and water injection are common and effective methods to improve oil recovery. To ensure its production effect, it is necessary to simulate the oilfield production process. However, traditional composition simulation runs a large number of calculations and takes a long time. Through the analysis of relevant data, we found that production is affected by many factors and has a strong sequential character. Therefore, this paper proposes a deep learning model for reservoir production prediction based on stacked long short-term memory network (LSTM). It is applied to other well patterns with a short production time and a few samples in the same oilfield block by transfer learning. The model achieves an effective combination with the actual reservoir production process. At the same time, it uses the knowledge learned from the well pattern with sufficient historical data to assist in the establishment of the model of the well pattern with limited data. This can obtain accurate prediction results and save the model training time, thus getting more effective application effects than composition simulation. This paper verifies the effectiveness of the proposed method through the data and multiple different injection combinations of the Tarim oilfield.



## 1. INTRODUCTION

Gas injection and water injection are the most common exploitation methods used in the process of oilfield development, but the cost of injection, water treatment, and liquid extraction is relatively large. Therefore, the selection and reasonable construction of injection and production parameters are particularly important for the development effect of injection. Due to the complexity and unpredictability of the oil production process, it is necessary to adjust the injection strategy and the corresponding working system in time according to the problems encountered in the implementation of the oilfield exploitation process. Due to the difference in the well pattern, injection volume, and the development stage in the oilfield, there are many different working systems. The application of reservoir composition simulation to study the influence of various working systems on the effect of injection is a commonly used method at present.<sup>1–6</sup> But, because of the large amount of calculation in composition simulation, it is impossible to return an optimization result in a short time. The simulation of a large amount of data using composition simulation is relatively imperfect, the time effectiveness cannot meet the business needs, and the feasibility is not high. As a result, the oil production prediction model based on an agent model comes into being. Using agent models to replace composition simulations is an important idea in many fields, and it is also suitable for oilfield production predictions. The agent model in this paper learns rules from the existing data and finally achieves fast prediction and gets good results. This

is of vital importance to scientific decision-making and reasonable guidance of oilfield production.

With the rapid development and wide application of machine learning technology, machine learning algorithms have been used in oil reservoir production prediction. Negash et al.<sup>7</sup> and Wu et al.<sup>8</sup> used simple neural networks to learn the characteristics of historical data from oil wells to achieve simple production prediction. These prediction models can be used to predict production accurately and quickly instead of numerical simulation after training with a large amount of actual data, but the majority of them cannot predict the oil production of new wells, nor can they obtain the dynamic effects of different systems; therefore, they still have certain limitations.

With the development of a recurrent neural network (RNN), a variety of algorithms based on the RNN have begun to be used in reservoir development. Liu et al.<sup>1</sup> combined an LSTM network, a traditional machine learning algorithm, and mean decrease impurity (MDI) feature selection to realize oil production prediction. Tatsipie et al.<sup>9</sup> used the RNN to create a data-driven model capable of

Received: September 15, 2021

Accepted: November 25, 2021

Published: December 7, 2021



generating logging curves. Zhang et al.<sup>10</sup> proposed a method using the LSTM network to establish a prediction model to predict the distribution of reservoir water saturation and oil production. Ki et al.<sup>11</sup> proposed a data-driven method based on the LSTM network to recover the lost pressure data in gas wells. Li et al.<sup>12</sup> used a bidirectional gated recurrent unit (Bi-GRU) and sparrow search algorithm (SSA) to improve the prediction accuracy of oil production. Zhou et al.<sup>13</sup> achieved high-precision fluid identification using a bidirectional long short-term memory (Bi-LSTM) network.

Although the abovementioned neural network models can achieve good performance in the prediction of oil production, the superior performance of these methods largely depends on sufficient historical data to train the model, and not every well pattern can establish the deep learning model because there is insufficient data for training. Traditional machine learning is only a small part of the research, with the above experience of these traditional machine learning methods, deep learning methods are naturally derived, and it solves some problems of machine learning. Therefore, methods should be explored to overcome or alleviate the problems of insufficient historical data and long training duration, and this paper hopes to provide a deep learning model that combines them to promote the understanding and research of production prediction.

In the development process of an oilfield block, there may be different existing situations of injection wells and oil production wells in the oilfield block, and there are different combinations of injection strategies and timings; therefore, it is very important to accurately predict the production under different conditions. If the trained production prediction deep learning model can be applied to other well patterns in the same oilfield block, the time cost can be saved to a great extent. Considering that the data of different well patterns may have differences, therefore, the trained deep learning model cannot be directly applied in other well patterns of the same oilfield block. This paper applies dimension alignment to solve the data differences between them. Aiming at the accurate prediction of oil production with different well patterns in the same oilfield block, this paper proposes a method based on transfer learning to solve the problem when the data does not meet the requirements of model construction. When the data of well patterns is limited, training the model by transfer learning can improve the model performance and save model training time.

To sum up, this paper proposes a deep learning model of oil production in reservoirs based on a stacked LSTM network and applies this model to other well patterns in the same oilfield block through transfer learning. The method proposed in this paper has high prediction precision and good efficiency, and through transfer learning, the model of training time can be saved. For well patterns with insufficient historical data, using transfer learning can improve the prediction accuracy and save time cost.

## 2. ESTABLISHMENT OF THE DEEP LEARNING PREDICTION MODEL

For any reservoir, economic evaluation is necessary at its development stage, and then the optimal plan is selected for investment. In the process of injection development, because of the need for frequent injection and production operations, the remaining oil will inevitably change in a certain form, which makes the production capacity and net present value (NPV) change. In fact, the value of NPV is the most important

factor affecting the development, and the calculation and analysis of the NPV will determine the priorities of different strategies. The calculation formula of NPV can be expressed as

$$NPV = \sum_{k=1}^{N_t} \left[ \frac{1}{(1+b)^k} \left( \sum_{i=1}^{N_p} r_{gp} q_{gp}^{i,k} + \sum_{i=1}^{N_p} r_{op} q_{op}^{i,k} - \sum_{i=1}^{N_p} c_{wi} q_{wi}^{i,k} - \sum_{i=1}^{N_i} c_{gi} q_{gi}^{i,k} - \sum_{i=1}^{N_p} cd_{wp} q_{wp}^{i,k} \right) \right] \quad (1)$$

where  $b$  is the basic discount rate,  $N_t$  is the calculation period of the project,  $k$  represents the  $k$ th period,  $N_p$  and  $N_i$  are the number of production wells and the number of gas injection wells, respectively,  $r_{gp} q_{gp}$  is the gas sales revenue,  $r_{op} q_{op}$  is the oil sales revenue,  $c_{gi} q_{gi}$  is the gas injection cost,  $c_{wi} q_{wi}$  is the water injection cost, and  $cd_{wp} q_{wp}$  is the water treatment cost. The optimization problem is considered as shown in eq 2

$$\max_{\delta} NPV(\delta) = \max_{\delta} [NPV_1(\delta), NPV_2(\delta), \dots, NPV_m(\delta)] \quad (2)$$

where  $\delta$  is the parameter to be optimized. According to the characteristics and geological situation of a certain oilfield block, different wells will be drilled, and different well pattern deployment strategies will be designed, that is, the distribution mode of oil wells, gas wells, and water injection wells. For a certain well pattern, different values of its parameters correspond to different working systems, that is, the working method of the oil well, such as the flowing bottom-hole pressure, and the amount of liquid injection of the oil well, which is the source of the above parameter  $\delta$ . Finally, the well pattern and working system that can maximize the oil fields' income<sup>14–16</sup> and give full play to its reasonable productivity are explored.

Through the comprehensive study of injection development, it can be found that after the gas drive operations, the saturation value of the remaining oil in the stratum will change to a certain extent. With the gradual increase of the development time, the cumulative production value of the oilfield will gradually increase, and it is bound to have a certain law. To study the influence of various factors in the gas injection reservoir on the production, it is necessary to make sure of the production influencing factors. There are many factors that affect oilfield production, such as the injection rate,<sup>2</sup> the number of producing wells,<sup>17</sup> injection–production ratio,<sup>18</sup> etc. This paper further screened the main influencing factors of oilfield injection development, and eventually, the injection volume of injection wells,<sup>19</sup> flowing bottom-hole pressure of production wells,<sup>3</sup> and production time<sup>20</sup> are selected as several indicators that have a direct impact on production changes, that is, the injection and production parameters are expressed as <gas injection rate of gas injection wells, the water injection rate of water injection wells, flowing bottom-hole pressure of production wells, time  $t$ , and the dimensions of gas injection and water injection correspond to the number of gas injection and water injection wells in each well pattern, corresponding to  $q_{gi}$  and  $q_{wi}$  in eq 1, and their values correspond to the injection rate. The injection and production parameter matrix corresponding to water injection and gas injection are

$$x = [T_1, \dots, T_n, Q_{g1}, \dots, Q_{gn}, Q_{w1}, \dots, Q_{wn}, P_1, \dots, P_m] \quad (3)$$

Table 1. Example Table of Injection and Production Parameter Values

working system for 5 gas injection wells					working system for 6 producing wells						time
I1	I2	I3	I4	I5	P1	P2	P3	P4	P5	P6	T
0	0	0	0	0	27 000	11 000	22 000	25 000	18 000	22 000	30
0	0	0	0	0	27 000	11 000	22 000	25 000	18 000	22 000	60
0	0	57 779	0	0	27 000	11 000	22 000	25 000	18 000	22 000	90
0	0	57 779	0	0	27 000	11 000	22 000	25 000	18 000	22 000	120
0	0	57 779	46 438	0	27 000	11 000	22 000	25 000	18 000	22 000	150
0	0	57 779	46 438	0	27 000	11 000	22 000	25 000	18 000	22 000	180
0	0	57 779	46 438	0	27 000	11 000	22 000	25 000	18 000	22 000	210
58 359	0	57 779	46 438	0	27 000	11 000	22 000	25 000	18 000	22 000	240
58 359	0	57 779	46 438	53 934	27 000	11 000	22 000	25 000	18 000	22 000	270
58 359	80 333	57 779	46 438	53 934	27 000	11 000	22 000	25 000	18 000	22 000	300
58 359	80 333	57 779	46 438	53 934	27 000	11 000	22 000	25 000	18 000	22 000	330
58 359	80 333	57 779	46 438	53 934	27 000	11 000	22 000	25 000	18 000	22 000	360

where,  $m$ ,  $n$ , and  $z$  are the number of production, gas injection, and water injection wells, respectively;  $T_i$  and  $Q_{gi}$  are the timing and speed of gas injection, respectively,  $i = 1, 2, \dots, n$ ;  $Q_{wj}$  is the water injection rate,  $j = 1, 2, \dots, z$ ; and  $P_k$  is the flowing bottom-hole pressure of the production well,  $k = 1, 2, \dots, m$ . Moreover,  $Q_{gi}$ ,  $Q_{wj}$ , and  $P_k$  meet the following constraints

$$Q_{gi}^{\min} \leq Q_{gi} \leq Q_{gi}^{\max} \quad (4)$$

$$Q_{wj}^{\min} \leq Q_{wj} \leq Q_{wj}^{\max} \quad (5)$$

$$P_k^{\min} \leq P_k \leq P_k^{\max} \quad (6)$$

Obviously, these injection and production parameters are easy to obtain and quantify.

An injection and production working system have a great influence on the development effect of an oilfield block, and for an oilfield block, the controllable parameters are also mainly injection and production parameters. Therefore, the injection and production parameters, the important factors that affect the development effect of the gas drive oilfield, are analyzed first. The parameters that affect the development of gas drive oil fields mainly include the injection rate and flowing bottom-hole pressure. Table 1 is an example of the injection and production parameters of an oilfield block in the Tarim oilfield, and the corresponding values mentioned above are expressed as  $((Q_{g1}, Q_{g2}, Q_{g3}, Q_{g4}, Q_{g5}), \emptyset, (P_1, P_2, P_3, P_4, P_5, P_6), T)$ .

Table 1 is the working system data established when five gas injection wells and six production wells are used to simulate the development, according to the actual situation of the well pattern of an oilfield block in the Tarim oilfield, on the premise of ensuring compliance with the current production status and taking into account the actual demand. A total of 11 wells are used in the model, among which I1, I2, I3, I4, and I5 are five gas injection wells, and the rest P1, P2, P3, P4, P5, and P6 are six production wells. The values of each parameter are shown in Table 1. Among them, the working system parameters of the gas injection well are partly derived from the gas injection rate and those of the production well is derived from the flowing bottom-hole pressure. For example, the injection volume of I1 is 58 359 m<sup>3</sup>/day, and the well would be opened from the 8th month. The injection volume of I2 is 80 333 m<sup>3</sup>/day. From the 10th month, the well was opened and maintained upto the 12th month, the flowing bottom-hole pressure of P1 is 27 000 kPa, and that of P2 is 11 000 kPa. To generate sufficient and diverse data as well as keep the difference between different

working systems in the process of generating the database, that is, to make the parameter values of each well under each system evenly distributed, Latin hypercube sampling<sup>21</sup> is selected in this paper. Compared with the pure stratified sampling method, its biggest advantage is that any number of samples can be easily produced.

The deep learning model of this paper is to study a rule of oil production change during a certain period, and then use the trained model to predict the production of the new injection and production strategy because the production dynamics slowly changes as the injection development progresses. Therefore, if the development time is short, the oil production will not change obviously, and the knowledge learned by the model will be very limited. A well can be productive for more than 10 years, 20–30 years, or even longer; this paper chooses to use 10 years of development data. It is obtained by recording the cumulative oil production every other month. The data in Table 1 only represents one kind of working system. The Latin hypercube sampling method described above should be carried out to generate data samples to obtain hundreds of working systems, and then the numerical simulation is used to obtain the final samples.

Using composition simulation to simulate the above data and analyze the monthly production of the reservoir is an important task. Similarly, the composition simulation is also of great value to the method presented in this paper. When the well patterns with no historical data are predicted, a small number of samples are generated using composition simulation for transfer. But, the reservoir composition simulation also has its limitations, mainly embodied in the simulation of a working system will need 10 min. Since the time of composition simulation is related to whether the development work can be carried out efficiently and directly affects the final injection and production strategy, it is of great significance to analyze the existing problems, study how to predict oil production efficiently in an oilfield, and form an effective production prediction method.

In this paper, the production prediction deep learning model takes the oil production prediction as the research object, and the influencing factors are the injection and production parameters. The constructed model is based on the stacked LSTM network and transfer learning, and it mainly includes three parts:

- (1) Based on the initial historical data or the data generated by the composition simulation, a reservoir oil production



- prediction model that meets the current accuracy requirements is formed.
- (2) The actual output value of the component numerical simulation is compared with the predicted value obtained by the constructed prediction model, and the accuracy of the latter is calculated.
  - (3) If the accuracy shows that the accuracy of the prediction model is sufficient to reflect the true dynamics of the current reservoir, the model structure and parameters can be transferred to the new well pattern in the same oilfield block, and the new well pattern data can be used to train the model.

Affected by the working system, the investment and recovery factor of the oil reservoir will be different, and the NPV will also be different. To draw the most economic benefit of the development plan, it is necessary to compare the NPV of different development plans. Through the constructed deep learning model, we can achieve rapid production prediction, so as to select the most optimal working system. The prediction model can be used to simulate oil reservoir production, so the model can provide guidance for production.

### 3. DIMENSION ALIGNMENT OF HETEROGENEOUS SAMPLES BASED ON THE DOMAIN KNOWLEDGE

The dimensions of injection and production parameters of the samples in the actual new well pattern may not be strictly consistent with those of the training samples. To solve the problem of inconsistency, dimension alignment is introduced, and then the stacked LSTM network is used to predict the aligned data to complete the transfer learning. The alignment idea is as follows: the source well pattern sample has  $G_s$  gas injection wells and  $W_s$  water injection wells; a new well pattern sample has  $G_t$  gas injection wells and  $W_t$  water injection wells. If the same prediction network model is used, the above samples need to be aligned as  $G_r$  gas injection wells and  $W_r$  water injection wells. The injection volume of the expanded injection well is zero. With the dimension alignment, the prediction network can adapt to the mapping of different dimension parameters, so as to prepare for transfer learning. The formal description of the process is expressed by eqs 7 and 8

$$G_r = \max\{G_s, G_t\} \quad (7)$$

$$W_r = \max\{W_s, W_t\} \quad (8)$$

among them, the source well pattern and the target well pattern are located in the same oilfield block of a certain oilfield. This paper verifies through experiment, and it shows that this method has no effect on the prediction accuracy.

For example, the source well pattern that has two gas injection wells and three water injection wells and the target well pattern that has one gas injection well and six water injection wells are the research objects. As mentioned above, the source structure is expressed as  $\langle (Q_{g1}, Q_{g2}), (Q_{w1}, Q_{w2}, Q_{w3}), \emptyset, T \rangle$ , and the target structure is expressed as  $\langle (Q_{g1}), (Q_{w1}, Q_{w2}, Q_{w3}, Q_{w4}, Q_{w5}, Q_{w6}), \emptyset, T \rangle$ . The number of gas injection wells and water injection wells of the two structures is obviously different. The values of the gas injection dimension are 2 and 1, respectively, and the values of the water injection dimension are 3 and 6, respectively. The data of the two well patterns need to be manually aligned to perform dimensional alignment based on domain knowledge and unify the dimensions of the data. The consequences of using the

above formula to align the data of the two structures are that they are both aligned as the data of two gas injection and six water injection wells, i.e.,  $\langle (Q_{g1}, Q_{g2}), (Q_{w1}, Q_{w2}, Q_{w3}, Q_{w4}, Q_{w5}, Q_{w6}), \emptyset, T \rangle$ . For the source well pattern, the injection volume of the expanded water injection wells IW4, IW5, and IW6 is zero, and for the target well pattern, the injection volume of the expanded gas injection well IG2 is also zero.

### 4. DEEP LEARNING MODEL BASED ON THE STACKED LSTM NETWORK

Through a comprehensive analysis of the characteristics of the injection and production data in this paper, we found that it is difficult to characterize the effects of various parameters on the oil production in the reservoir with traditional models. In addition, the data is time dependent, and it is necessary to analyze the effects of last month's production on the next month's production. In the model training stage, the traditional neural network cannot use the information obtained in the previous time step in the current step, which is the main disadvantage of the traditional neural network. However, the RNN tries to pass the information in one step of the network to the next step by using recursions, so that the timing information can be retained, so as to solve this problem. Therefore, this model has advantages in processing time-series data. However, the traditional RNN can only remember short-term information, and it does not have the ability to remember information over a long distance. The LSTM network has solved the problem of gradient disappearance and explosion of the RNN, it can deep mine the information contained in the historical data, at the same time, take into account the scheduling of the data, which meets the requirements of the constructed model. In addition, simple neural networks cannot transfer from one well pattern to another without sufficient data. Thus, the models based on the LSTM network have advantages. To avoid the limitation of the traditional neural network, this paper chooses to construct a deep learning model based on the stacked LSTM network.

There are three types of gates in the LSTM unit: the forget gate ( $f_t$ ), the input gate ( $i_t$ ), and the output gate ( $o_t$ ).<sup>22</sup> The LSTM unit uses the input  $X_t$  at the current time, the hidden state  $H_{t-1}$ , and the cell state  $C_{t-1}$  at the previous time as the input of the LSTM unit at the current time,<sup>23</sup> and the output  $H_t$  and  $C_t$  are used as the input at the next time. The specific details of the LSTM unit at time  $t$  are shown in Figure 1. The specific calculation of the LSTM unit is as follows

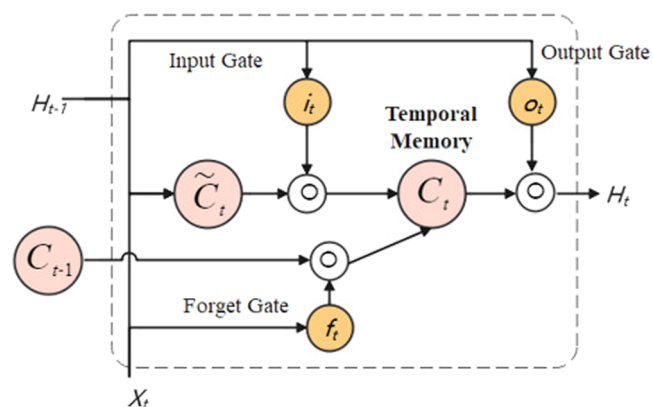


Figure 1. Schematic diagram of the LSTM unit structure.

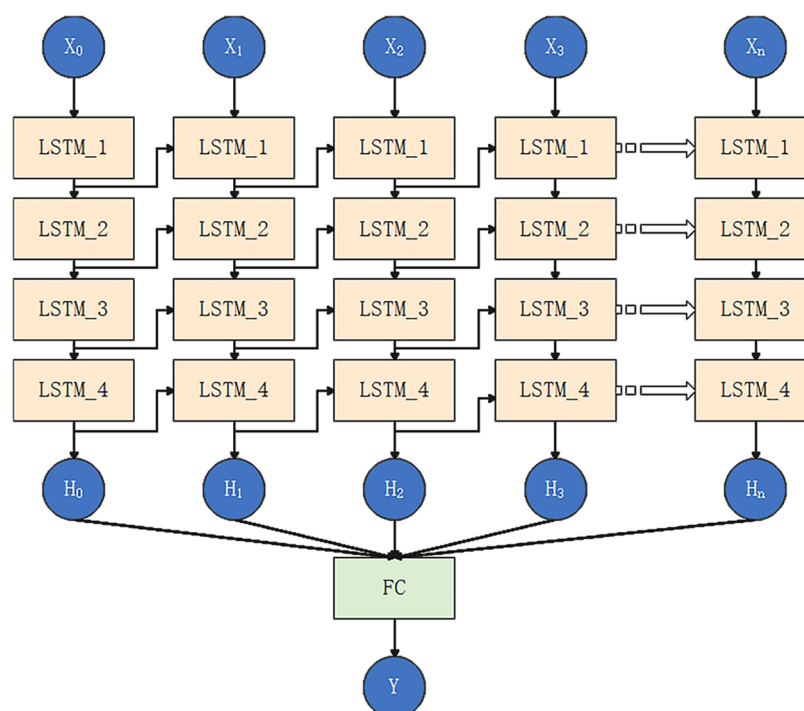


Figure 2. Stacked LSTM network model.

$$\tilde{C}_t = \tan h(W_c[H_{t-1}, X_t] + b_c) \quad (9)$$

$$i_t = \sigma(W_i[H_{t-1}, X_t] + b_i) \quad (10)$$

$$f_t = \sigma(W_f[H_{t-1}, X_t] + b_f) \quad (11)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C}_t \quad (12)$$

$$o_t = \sigma(W_o[H_{t-1}, X_t] + b_o) \quad (13)$$

$$H_t = o_t \circ \tan h(C_t) \quad (14)$$

Among them,  $\tilde{C}_t$  represents the candidate vector;  $\sigma$  represents the Sigmoid activation function, and the output is between 0 and 1;  $\circ$  represents the Hadamard product;  $W_b$ ,  $W_i$ , and  $W_o$  represent the gate weight matrix;  $b_b$ ,  $b_i$ ,  $b_o$ , and  $b_c$  represent the bias; and  $i_t$ ,  $f_t$ , and  $o_t$  control the information flow across the memory state  $C_t$ .

A multilayer LSTM network can remove some of the constraints of a single-layer LSTM network, which can be viewed as the output of the upper layer and as the input of the next layer. This kind of network uses stacking to stack the layers of the LSTM network.<sup>24</sup> A hierarchical feature representation of the input data can be created through this structure, which can subsequently be used for prediction.

The stacked LSTM network belongs to deep learning,<sup>25</sup> which is a representative optimization to improve the efficiency of training and obtain higher accuracy by adding the depth of the network. In many fields, such as human motion recognition,<sup>26</sup> fault diagnosis,<sup>27</sup> prediction research,<sup>28</sup> the detection algorithm,<sup>29</sup> and so on, stacked networks are more effective than single-layer networks.

The success of a deep neural network (DNN) in predicting classification problems is usually attributed to the depth of the network. Given that the LSTM network operates on sequence data, this means that increasing the layers of the network increases the time to extract the features of the input signal. At

present, the stacked LSTM network is a good model for sequence prediction problems. A stacked LSTM network architecture can be defined as an LSTM network model consisting of multiple LSTM layers. The LSTM layer above provides a sequential output, rather than a single value being the output to the LSTM layer below. Figure 2 shows the model of the stacked LSTM network in this paper.

To better extract features, the LSTM units shown in Figure 1 are stacked to build a stacked LSTM network with stronger representation ability. The stacked LSTM network in this paper is composed of four LSTM layers. Specifically, the LSTM layer is used to extract the features first, and then the output feature vector is sent to the FC layer for prediction, so as to obtain the predicted value of the oil production.

## 5. TRANSFER LEARNING BASED ON THE PRETRAINING MODEL

**5.1. Oil Production Prediction Algorithm Based on Transfer Learning.** In the same oilfield block, the development strategy can be optimized for a more economical and efficient development by adjusting the well pattern. In the transfer learning theory, the learning domain with enough labeled samples is called the source domain,<sup>30</sup> and the target domain usually has limited data. Given a specific data set  $D_s$  of the source well pattern and data set  $D_t$  of the target well pattern, suppose that the labeled data sets are  $D_s = \{X_s, Y_s\}$  and  $D_t = \{X_t, Y_t\}$ . Among them,  $X$  represents the injection and production parameters, and  $Y$  represents the corresponding oil production. Therefore, the problem of transfer learning production prediction is to learn an accurate regression to predict the production of the corresponding oilfield represented by the data in the target domain. The algorithm is based on the constructed stacked LSTM network model, uses the well pattern with abundant available information as the source domain, and that of a small number of samples as the target domain. The prediction of transfer learning is carried

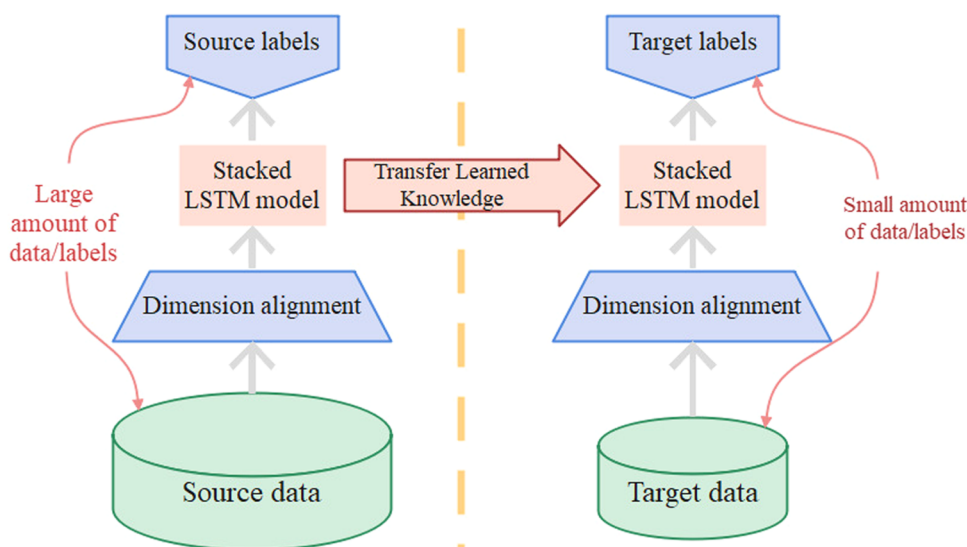


Figure 3. Flow chart of production prediction based on transfer learning.

out by the method of network parameter transfer. Compared with the previous prediction methods, this method only needs a small amount of labeled target domain data to achieve good prediction performance. The corresponding tasks are expressed in eqs 15 and 16

$$Y_s = f_s \{X_s, \delta_s\} \quad (15)$$

$$Y_t = f_t \{X_t, \delta_t\} \quad (16)$$

Among them,  $f_s$  represents the function mapping from  $X_s$  to the predicted value of the source well pattern  $Y_s$ ,  $f_t$  represents the function mapping from  $X_t$  to the predicted value of the target well pattern  $Y_t$ , and  $\delta_s$  and  $\delta_t$  are the model parameters of the source well pattern and the target well pattern, respectively. Transfer learning is to find the relevant features in the source well pattern data  $\{X_s, Y_s\}$  and obtain the mapping  $f_s$ . Then, after the mapping  $f_s$  is transferred to the target well pattern, the mapping  $f_t$  of the target task is learned from the target well pattern data  $\{X_t, Y_t\}$ . Specifically, it can be expressed in eqs 17 and 18

$$\delta'_s = \arg \min \sum_{i=1}^{n_s} L(x_i, y_i, \delta_s) \quad (17)$$

$$\delta'_t = \arg \min \sum_{i=1}^{n_t} L(x_i, y_i, \delta_t) \quad (18)$$

where the subscript  $i$  represents the  $i$ th sample, and  $n_s$  and  $n_t$  are the sample numbers of source well pattern data and target well pattern data, respectively.  $L$  is the loss function, and the model training process is to minimize the loss function.  $x_i$  represents a training sample, and  $y_i$  refers to the label of the sample. The optimal model parameter  $\delta'_s$  can be obtained by fully training the source well pattern model.  $\delta_s$  and  $\delta_t$  are the model parameters of the source well pattern and the target well pattern, respectively. The purpose of transfer learning is to use the optimal model parameter  $\delta'_s$  when training the model in the target task and obtain the optimal model parameter  $\delta'_t$  after training in the target well pattern.

**5.2. Stacked LSTM Network Training Combined with Transfer Learning.** The stacked LSTM network uses a large number of sample data to train the parameters of the model.

The trained model can be used for the sample prediction but it must meet two requirements. First, the source data set must be labeled sample data.<sup>31</sup> Second, the source data set and the target data set must have the same or similar distribution.<sup>32</sup> Therefore, it takes a lot of computing resources to simulate the sample library. In addition, the number of parameters that need to be trained for the stacked LSTM network is very large, and the training process consumes a lot of time and computing resources. Training these parameters requires a large number of sample data, which is very time-consuming in practical application. If the samples are insufficient, it is easy to lead to overfitting problems and the generalization of the trained model is poor.

In view of the above problems, this paper adopts the parameter-based transfer learning strategy to improve the training method of the stacked LSTM network, so as to reduce the requirement of the amount of sample data. The stacked LSTM network training method combined with transfer learning refers to retraining a deep learning model again with the data of the current well pattern, rather than training from the beginning, so as to obtain the model suitable for the current well pattern. Due to the feature extraction part of the stacked LSTM network in different well patterns being interlinked, therefore when using transfer learning, the structure and parameters of the network layer can be retained, and the weights of the model can be initialized and set as a trainable state, and then use the data of the current well pattern to retrain the parameters. Using transfer learning, only a small amount of data is needed to train a model suitable for the current well pattern, which effectively solves the problem of a lack of enough labeled data and the need for a large number of computing resources to train a new model.

In this paper, transfer learning and the LSTM network are combined to build a deep learning model for the oilfield prediction of the same type. The purpose of neural network training is to minimize the loss function  $L$  through back-propagation and gradient descent. In this paper, mean-square-error (MSE) is used as the loss function of the LSTM network

$$L = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 \quad (19)$$

where  $L$  is the network loss,  $y$  is the expected output of the network,  $f(x)$  is the actual output of the network, and  $n$  is the number of samples.  $L$  gets closer and closer to the minimum by updating the weights and bias parameters in the direction of gradient descent.

Each layer of the stacked LSTM network has its own weight and bias parameters. The weight and bias parameters are obtained by training the source well pattern model, and then the determined parameter  $\delta'_s$  is directly transferred to the target well pattern by transfer learning. In this paper, the trained parameters of the source well pattern model are used as the initial values of the parameters of the target well pattern model. The LSTM layer in the stacked LSTM network is used as a feature extractor to extract the features of the input time-series data, while the information each layer of the neural network extracted is not the same, but the information is universal, so the parameters of each layer are transferred. The target well pattern data is used for training for all layer parameters so that the accuracy and time cost of the prediction results of the target well pattern are optimal.

The entire process of applying transfer learning is shown in Figure 3, and the specific process is as follows:

- (1) First, align the dimension of the data of the source well pattern, construct the model of the source well pattern, train and debug it, find the best prediction model of the source well pattern, and save the model and parameters of the source well pattern for the target well pattern;
- (2) Second, load the saved source well pattern model, transfer the weight parameters of the source well pattern model as the initial value of the parameters of the target well pattern, transfer all layer parameters, align the dimension of the data of the target well pattern and use the data of the target well pattern to train all layers of the network, and get the prediction model of the target well pattern.

## 6. RESULTS AND DISCUSSION

The reservoir data used in this paper is from the Tarim oilfield, and the Donghe sandstone member in Tarim Basin is taken as the research object. The Donghe sandstone member is a typical marine clastic reservoir rich in oil and gas in China.<sup>33</sup> Limited by the “three highs” of high temperature, high salinity, and high pressure, appropriate injection and production systems need to be formulated during the development. After tracking and analyzing the oil and gas production in the process of gas injection and oil displacement, it is found that the oil and gas production of gas injection and oil displacement technology is directly related to the injection volume of gas injection wells, water injection wells, and the flowing bottom-hole pressure of production wells. Therefore, the sample database constructed in this paper mainly includes the working systems of different wells, including thousands of different working systems. To ensure the effectiveness of the work, Latin hypercube sampling is used to design an injection and production parameter suitable for marine clastic reservoirs.

Taking an oilfield in Tarim Basin as an example, there are altogether 37 oil wells in the oilfield, including 21 production wells and 16 injection wells. According to the actual exploitation needs, the four water injection wells, INJ1, INJ2, INJ3, and INJ4, can be adjusted. For example, we can choose 1, 2, 3, or 4 wells for gas injection, and the rest for water injection, so there are  $C_4^1 + C_4^2 + C_4^3 + C_4^4 = 15$  different

injection and production well patterns, as shown in Table 2. If the model training is carried out from the beginning for each

**Table 2. Corresponding Table of the Gas Injection Well and Well Pattern**

well pattern	gas injection well number
1	INJ1
2	INJ2
3	INJ3
4	INJ4
5	INJ1, INJ2
6	INJ1, INJ3
7	INJ1, INJ4
8	INJ2, INJ3
9	INJ2, INJ4
10	INJ3, INJ4
11	INJ1, INJ2, INJ3
12	INJ1, INJ2, INJ4
13	INJ1, INJ3, INJ4
14	INJ2, INJ3, INJ4
15	INJ1, INJ2, INJ3, INJ4

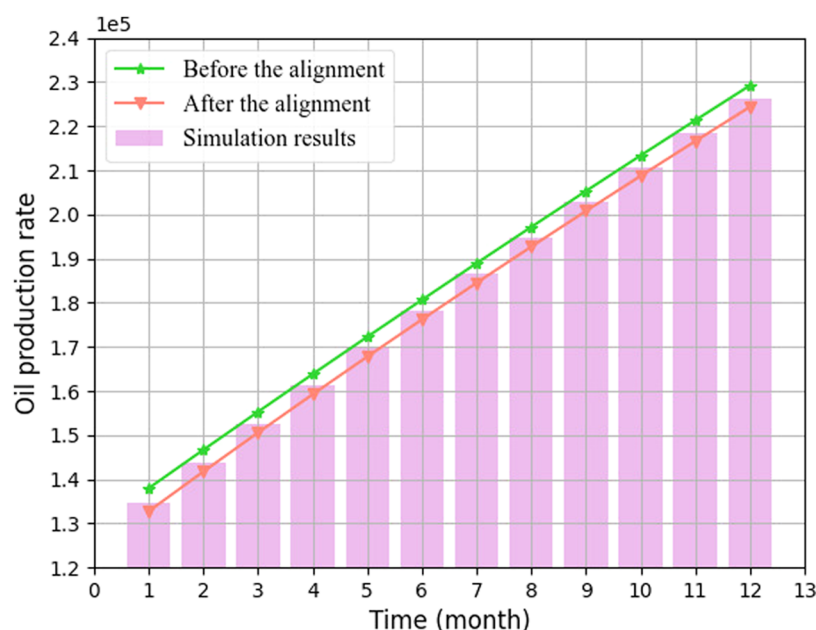
well pattern, a lot of time cost will be lost, and the model trained in a variety of well patterns can also have very good effects in any kind of well pattern. It extracts general information from various well patterns with a large amount of available data and then transfers it to a certain well pattern with different data. The prediction can be made by training the model with a small number of samples.

Using modern reservoir numerical simulation technology to run the sample library, the oil production and gas production of the reservoir in different well patterns and development periods are predicted, and the above data is recorded and organized to obtain a rich data set with 300 systems for each well pattern. Since the recorded data is daily data, data cleaning is required. The final data provided to the model consisted of a monthly record of 10 years of data for different working systems, that is, gas injection volume, water injection volume, flowing bottom-hole pressure, time, and cumulative oil production of the reservoir. The time interval is 1 month, where the first 80% is used as a training set and the last 20% is used as a test set.

Taking into account the complexity of sample generation, taking 50 working systems as an example, the final sample contains a total of 6000 lines of data, so it is estimated that hundreds of systems need to write nearly tens of thousands of lines of data. Such a huge workload requires high-quality completion in a short time. The traditional line-by-line manual design method cannot meet the requirements of the project, and in the repetitive and complicated design, it is prone to make mistakes and errors. Therefore, the python high-level programming language is used to design the program and generate the user interface to generate the injection and production parameters. The user inputs the number of gas injection wells, the number of production wells, the number of production months, and the number of working systems, and the program automatically designs and writes the data. Instead of a manual design, it can greatly reduce the difficulty and workload of the sample generation.

**6.1. Results before and after the Dimension Expansion of a Certain Well Pattern.** Considering the influence of data preprocessing, the accuracy of yield





**Figure 4.** Production fitting before and after the dimension alignment.

prediction cannot be affected in the dimension alignment stage, so its influence should be considered before selecting the data alignment scheme. At the same time, because the alignment methods of different well patterns are also different, considering that evaluating the suitability of the method as accurately as possible before the transfer, this paper selects the injection and production data provided by the well pattern 5 to extend its dimensions. Based on a dimension alignment strategy, and considering the real environment, the two gas injection wells INJ3 and INJ4 and one water injection well were extended, and their impacts were studied as follows: Figure 4 compares the predicted oil production before and after the alignment with the corresponding simulated data; see Table 3 for details. Before and after the expansion of the data dimension, the production prediction accuracy rates of the well pattern are 95.3 and 94.7%, respectively, which are both within the acceptable range. After the alignment, the accuracy

decrease is minimal, and it does not affect the final prediction results.

## 6.2. Source Well Pattern Model and Prediction Results.

The establishment process of the prediction model is as follows. The first part is data preprocessing, considering the well location distribution and incorporating well location coordinates into the features. Performing the correlation analysis on the injection and production data and production of each well, we then obtain the importance of the features of all injection and production wells and use the injection and production data of different injection and production wells multiplied by the corresponding weight as the current input values, and then the data is normalized and reshaped. Consequently, the data set is transformed into the input form of the LSTM network model (batch size, time step, input size). In the second part, the prediction model of the source well pattern is generated and debugged, and the structure of the model and the trained weights are saved to facilitate the subsequent transfer learning in the target well pattern.

The simulation is implemented using Python 3.7 in the Ubuntu system, using the Keras deep learning modeling environment in Python, which supports existing common structures, such as RNN, fully connected neural network, etc. The LSTM, dense, and dropout layers can be imported from Keras, and these layers can be put into the sequence model to build the required model in any reasonable way. In this paper, four LSTM layers and one dense layer are selected. The number of hidden units in the four layers of the LSTM network is (50,50,50,50), and Adam and MSE are selected as the optimizer and loss function, respectively.

The abovementioned source well pattern prediction model is used to train the data. ModelCheckpoint is used to store the optimal model, and callbacks are used to control the training model. Figure 5 shows the fitting diagram after 10 years of production, and there is little difference between the predicted and simulated production. When the production cycle reaches 7 years, the prediction error is the largest, and after 7 years, the prediction error decreases. In contrast, the prediction accuracy

**Table 3.** Comparison of the Accuracy of Prediction Results before and after the Dimension Expansion of a Certain Well Pattern

predict time	simulated value	before the alignment		after the alignment	
		predicted value	accuracy (%)	predicted value	accuracy (%)
Month_1	134 759	137 972	97.62	132 759	97.16
Month_2	143 738	146 668	97.96	141 738	97.02
Month_3	152 560	155 284	98.21	150 560	96.96
Month_4	161 239	163 820	98.40	159 239	96.95
Month_5	169 784	172 276	98.53	167 784	96.99
Month_6	178 202	180 651	98.63	176 202	97.05
Month_7	186 505	188 946	98.69	184 505	97.14
Month_8	194 690	197 159	98.73	192 690	97.25
Month_9	202 761	205 292	98.75	200 761	97.37
Month_10	210 723	213 344	98.76	208 723	97.50
Month_11	218 585	221 314	98.75	216 585	97.63
Month_12	226 351	229 205	98.74	224 351	97.77
average			98.48		97.23



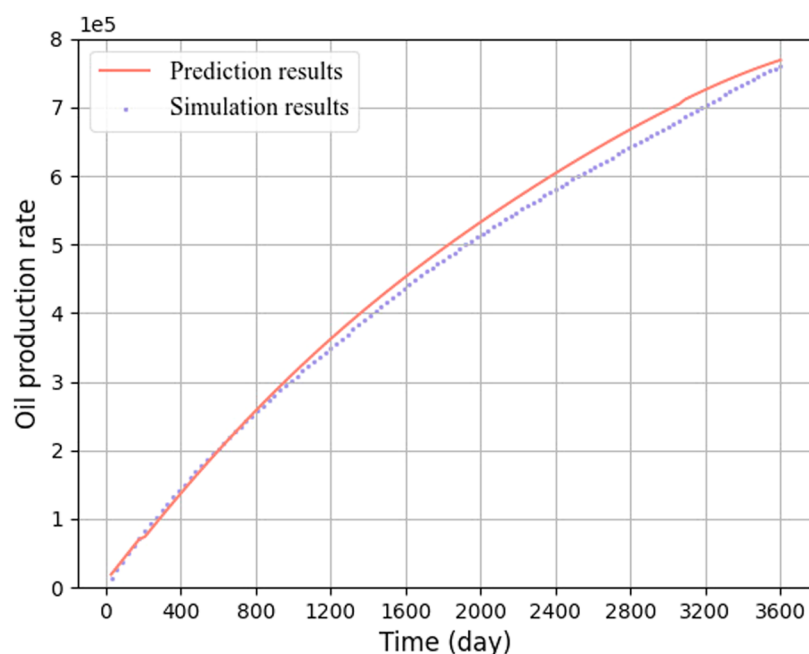


Figure 5. Oil production predicted in 10 years.

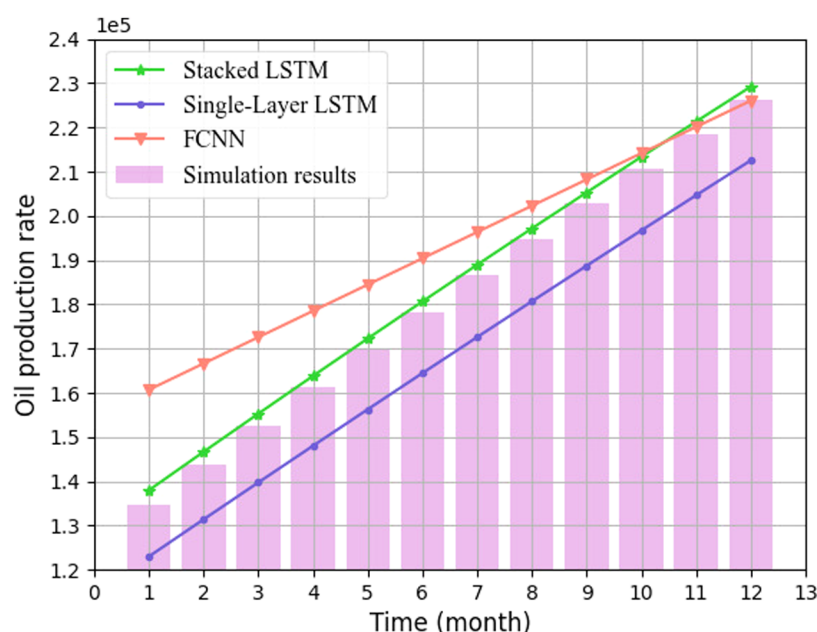


Figure 6. Fitting of oil production under three conditions.

in the early stage is much higher than that in the middle and later stages.

We analyzed and compared three models: a stacked LSTM model, single-layer LSTM model without stacking, and FCNN model. As shown in Figure 6, it can be seen that even if it is predicted for 10 years, the prediction results of the stacked LSTM network prediction model are close to the actual ones. To make the experiment scientific and accurate, compare it with a single-layer LSTM network and fully connected neural network (FCNN), transform the LSTM layer to the dense layer, and retrain the network after adjusting the structure. It can be seen from Table 4 that the prediction results of the stacked LSTM network model are more accurate than those of the single-layer LSTM network and FCNN. The results on the

test set show that comparing and analyzing all data, the accuracy of the stacked LSTM network model is 6.2 and 4.7% higher than those of the two models, respectively.

**6.3. Prediction Effect of Different Well Patterns When Using Transfer Learning.** To verify the transfer learning ability of the designed predictive model, the conditions of different source well patterns and the same target well pattern are considered. Among them, the source well pattern uses several different data such as 1, 14, and 15 in Table 2, and the target well pattern uses the data of the well pattern 8 for model training. The data of the target well pattern and the source well pattern are processed by the dimension alignment. Transfer learning is using the pretraining model to train the data of different working systems of the target well pattern 8, 2 gas

Table 4. Comparison of Accuracy of Prediction Results of Three Models

predict time	simulated value	stacked LSTM		single-layer LSTM		FCNN	
		predicted	accuracy (%)	predicted	accuracy (%)	predicted	accuracy (%)
Month_1	134 759	137 972	97.62	123 000	91.27	160 689	80.76
Month_2	143 738	146 668	97.96	131 390	91.41	166 630	84.07
Month_3	152 560	155 284	98.21	139 731	91.59	172 571	86.88
Month_4	161 239	163 820	98.40	148 022	91.80	178 512	89.29
Month_5	169 784	172 276	98.53	156 263	92.04	184 453	91.36
Month_6	178 202	180 651	98.63	164 454	92.29	190 394	93.16
Month_7	186 505	188 946	98.69	172 594	92.54	196 335	94.73
Month_8	194 690	197 159	98.73	180 683	92.81	202 276	96.10
Month_9	202 761	205 292	98.75	188 722	93.08	208 218	97.31
Month_10	210 723	213 344	98.76	196 709	93.35	214 159	98.37
Month_11	218 585	221 314	98.75	204 646	93.62	220 100	99.31
Month_12	226 351	229 205	98.74	212 531	93.89	226 041	99.86
average			98.48		92.47		92.60

injection, and 14 water injection, and the following results are obtained.

To study the influence of the pretraining model and the number of samples on the prediction accuracy, different values are used for analysis, Table 5 compares the prediction accuracy

Table 5. Comparison of the Prediction Results between the Different Number of Samples and Different Pretraining Models

source of pretraining samples	number	training time (s)	number of working systems			
			5	10	15	20
1 well pattern (15)	100	240	79.7%	80.3%	80.6%	81.3%
	200	476	84.4%	85.4%	86.2%	87.0%
	300	711	88.2%	88.8%	89.2%	89.5%
2 well patterns (1 and 14)	100	240	79.7%	80.1%	80.4%	81.1%
	200	477	88.8%	90.0%	90.3%	90.7%
	300	709	89.1%	90.1%	90.5%	91.0%
2 well patterns (1 and 5)	100	238	80.0%	80.5%	80.9%	81.3%
	200	554	88.3%	88.6%	89.3%	89.7%
	300	709	89.0%	89.3%	89.9%	90.9%
4 well patterns (3, 5, 12, and 14)	100	246	82.4%	82.8%	83.3%	83.6%
	200	476	82.7%	83.2%	83.9%	84.3%
	300	714	88.5%	89.5%	90.4%	91.0%
	400	951	90.1%	90.4%	90.7%	91.6%

rate of oil production using 5, 10, 15, and 20 working systems and different pretraining models during transfer. The numbers in brackets correspond to the well patterns in Table 2, representing the sources of the pretraining samples, and 100, 200, and 300 represent the number of working systems of the training source models. It can be seen in Table 5 that the prediction results obtained when using 400 kinds of systems and four kinds of well patterns data is the best. As shown in Figure 7a–d, the results obtained when using four kinds of well patterns with 200 kinds of systems, data is not as good as when using two kinds of structures with 200 kinds of systems. The reason is that the data of the four well patterns is averaged, and there are only 50 systems for each type, the number decreases and the types increase, and the regularity becomes worse, so the prediction precision is lower. Compared with the 100 systems using two types of samples (1 and 14), the accuracy rate of the 5, 10, 15, and 20 systems in the 200 systems using

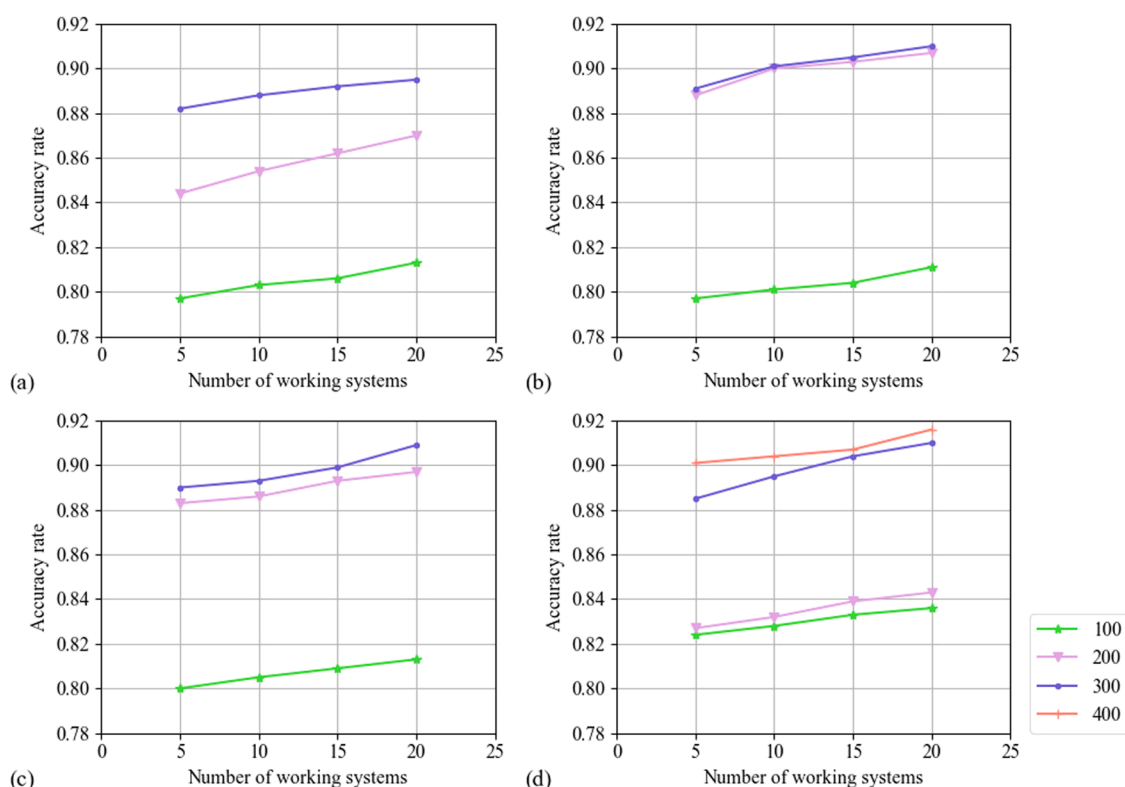
four samples has increased by 3.0, 3.1, 3.5, and 3.2%, respectively. Similarly, the accuracy rate of the 400 systems with four kinds of structures was higher than that of the 200 systems with two kinds of structures. The results show that when the number of samples for each well pattern is the same, the more abundant the source type, the higher the accuracy of the prediction results after transfer.

Table 6 is the comparison of the results of prediction when the source model is trained using 300 working systems from two kinds of samples, and the target well pattern is selected with different well patterns. It is the same as the well pattern 8, several different cases are compared. From Table 6, it can be intuitively seen that the prediction effect of using transfer learning is better when there are more training samples. Table 6 shows the comparison of the prediction results. It can be seen from the prediction results in Table 6 that the accuracy of the prediction results in the target well pattern 2 are 89.2, 89.6, 90.1, and 90.7%, respectively, for the 5, 10, 15, and 20 systems. Compared with other target well patterns, the biggest difference in the four cases is only 0.3% when 10 systems are used. The prediction results of transfer learning are not much different for different target well patterns, indicating that the generalization ability of the used model is good, and it is suitable for the prediction of different well patterns.

It can be seen from the comparison of the above results that the deep learning model based on the stacked LSTM network proposed in this paper has better prediction performance. For the construction of other well pattern prediction models, the prediction precision and the model of training time are comprehensively compared, and the proposed transfer learning saves the model training time and improves the prediction precision, and also receives the recognition from the fields.

## 7. CONCLUSIONS

In this paper, a prediction model based on the stacked LSTM network is proposed to improve the predictive effect of oil production. The actual experimental results show that this method has good practicability and accuracy for oil well production prediction and can be used for oil production prediction of oil wells in oil fields. At the same time, this deep learning model is applied to other well patterns through transfer learning to solve the prediction problem of well patterns with a small number of samples, which can provide a reference for the optimization of the working system, and play an auxiliary role in decision making. Therefore, this algorithm



**Figure 7.** Influence of the source of pretraining samples on the prediction accuracy (a) 15; (b) 1 and 14; (c) 1 and 5; and (d) 3, 5, 12, and 14.

**Table 6.** Comparison of Prediction Results of Transferring to Different Well Patterns

target well pattern	number of working systems			
	5	10	15	20
2	89.2%	89.6%	90.1%	90.7%
3	89.4%	89.8%	90.2%	90.6%
5	89.0%	89.5%	90.2%	90.6%
12	89.4%	89.9%	90.3%	90.7%
15	89.2%	89.6%	90.3%	90.8%

has a good application prospect and certain research value. The conclusions of this paper can be mainly summarized as follows:

- (1) A deep learning framework for the production prediction is proposed. Taking into account the sequential characteristics of the data, a model is constructed through the stacked LSTM network. Compared with composition simulation software, the deep learning model can achieve fast prediction without a significant decrease in accuracy;
- (2) Through transfer learning, the trained deep learning model is applied to other well patterns. Considering that the training of the neural network needs a large amount of data and the available data of the new well patterns are less, the trained parameters of the network of the source well pattern model is chosen to be used as the initial value of the target well pattern, and the pretraining model is used to train the network with the data of the target well pattern. The result of the case analysis shows that this method saves the training time of the model and achieves fast prediction.

## AUTHOR INFORMATION

### Corresponding Author

Yukun Dong – College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China; Email: [dongyk@upc.edu.cn](mailto:dongyk@upc.edu.cn)

### Authors

Yu Zhang – College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China; [orcid.org/0000-0001-9467-8733](https://orcid.org/0000-0001-9467-8733)

Fubin Liu – College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China

Xiaotong Cheng – College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China

Complete contact information is available at: <https://pubs.acs.org/10.1021/acsomega.1c05132>

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This research was funded by the Fundamental Research Funds for the Central Universities (20CX05016A) and the Major Scientific and Technological Projects of CNPC under Grant ZD2019-183-007.

## REFERENCES

- (1) Liu, W.; Liu, W.; Gu, J. Oil production prediction based on a machine learning method. *Oil Drill. Prod. Technol.* **2020**, *42*, 70–75.
- (2) Gu, J.; Ren, Y.; Wang, Y. Prediction methods of remaining oil plane distribution based on machine learning. *J. China Univ. Pet., Ed. Nat. Sci.* **2020**, *44*, 39–46.

- (3) Wang, L.; Zhang, L.; Lai, F. Multi-objective optimization design of reservoir injection-production parameters based on substitution model. *Sci. Technol. Eng.* **2019**, *19*, 178–185.
- (4) Tang, H.; Juan, H.; Rong, Y. Study on water drive law and characteristics of remaining oil distribution of typical fault-karst in fault-karst reservoirs, Tahe Oilfield. *Pet. Geol. Recovery Effic.* **2018**, *25*, 95–100.
- (5) Wang, R.; Yuan, S.; Wang, J. Mechanism of the injection-production coupling technique for complex fault-block oil reservoirs. *Pet. Geol. Oilfield Dev.* **2018**, *10*, 38–42.
- (6) Fang, N.; Zhang, Z.; Cheng, M. Injection-production parameter optimization study of cyclic water injection at different development levels of fractures. *Spec. Oil Gas Reservoirs* **2019**, *26*, 131–135.
- (7) Negash, B. M.; et al. Artificial neural network based production forecasting for a hydrocarbon reservoir under water injection. *Pet. Explor. Dev.* **2020**, *47*, 383–392.
- (8) Wu, J.; Li, Z.; Sun, Y. Neural network-based prediction of remaining oil distribution and optimization of injection-production parameters. *Pet. Geol. Recovery Effic.* **2020**.
- (9) Tatsipie, N. R. K.; Sheng, J. J. Generating Pseudo Well Logs for a Part of the Upper Bakken using Recurrent Neural Networks. *J. Pet. Sci. Eng.* **2020**, *200*, No. 108253.
- (10) Zhang, Q.; Wei, C.; Wang, Y.; et al. Potential for Prediction of Water Saturation Distribution in Reservoirs Utilizing Machine Learning Methods. *Energies* **2019**, *12*, No. 3597.
- (11) Ki, S.; Jang, I.; Cha, B. Restoration of Missing Pressures in a Gas Well Using Recurrent Neural Networks with Long Short-Term Memory Cells. *Energies* **2020**, *13*, No. 4696.
- (12) Li, X.; Ma, X.; Xiao, F.; et al. Time-series production forecasting method based on the integration of Bidirectional Gated Recurrent Unit (Bi-GRU) network and Sparrow Search Algorithm (SSA). *J. Pet. Sci. Eng.* **2021**, *208*, No. 109309.
- (13) Zhou, X.; Zhang, Z.; Zhu, L. A new method for high-precision fluid identification in bidirectional long short-term memory network. *J. China Univ. Pet.* **2021**, *45*, 69–76.
- (14) Liu, Z.; Reynolds, A. Gradient-Enhanced Support Vector Regression for Robust Life-Cycle Production Optimization with Nonlinear-State Constraints. *SPE J.* **2020**, *26*, 1590–1613.
- (15) Liu, Z.; Reynolds, A. A sequential-quadratic-programming-filter algorithm with a modified stochastic gradient for robust life-cycle optimization problems with nonlinear state constraints. *SPE J.* **2020**, *25*, 1938–1963.
- (16) Liu, Z.; Reynolds, A. Robust Multiobjective Nonlinear Constrained Optimization with Ensemble Stochastic Gradient Sequential Quadratic Programming-Filter Algorithm. *SPE J.* **2021**, *26*, 1964–1979.
- (17) Gu, J.; Zhou, X.; Wang, S. A prediction method for oilfield production based on Lasso algorithm. *Sci. Technol. Eng.* **2020**, *20*, 10759–10763.
- (18) Ren, Y.; Gu, J.; C, W. Oilfield production prediction model based on improved fruit fly algorithm and long-short term memory neural network. *Sci. Technol. Eng.* **2020**, *20*, 7245–7251.
- (19) Feng, Q.; Li, Y.; Wang, S. Predicting gas migration development using deep convolutional generative adversarial network. *J. China Univ. Pet.* **2020**, *44*, 20–27.
- (20) Gu, J.; Sui, G.; Li, Z. Oil well production forecasting method based on ARIMA-Kalman filter data mining model. *J. Shenzhen Univ., Sci. Eng.* **2018**, *35*, 575–581.
- (21) Andrés-Martínez, O.; Flores-Tlacuahuac, A. Optimal Molecular Design of Low-Temperature Organic Fluids under Uncertain Conditions. *Ind. Eng. Chem. Res.* **2018**, *57*, 5058–5069.
- (22) Müller, A. T.; Hiss, J.; Schneider, G. Recurrent Neural Network Model for Constructive Peptide Design. *J. Chem. Inf. Model.* **2018**, *58*, 472–479.
- (23) Zou, M.; Zhu, S.; Gu, J.; et al. Heating and Lighting Load Disaggregation Using Frequency Components and Convolutional Bidirectional Long Short-Term Memory Method. *Energies* **2021**, *14*, No. 4831.
- (24) Lateko, A.; Yang, H.; Huang, C.; et al. Stacking Ensemble Method with the RNN Meta-Learner for Short-Term PV Power Forecasting. *Energies* **2021**, *14*, No. 4733.
- (25) Khatibi, T.; Karampour, N. Predicting the number of hospital admissions due to mental disorders from air pollutants and weather condition descriptors using stacked ensemble of Deep Convolutional models and LSTM models (SEDCMLM). *J. Cleaner Prod.* **2021**, *280*, No. 124410.
- (26) Wang, M.; Zhang, Y.; Cui, G. Human motion recognition exploiting radar with stacked recurrent neural network. *Digit. Signal Process.* **2019**, *87*, 125–131.
- (27) Fang, N.; Zhang, Z.; Cheng, M. Injection-production parameter optimization study of cyclic water injection at different development levels of fractures. *Spec. Oil Gas Reservoirs* **2019**, *26*, 131–135.
- (28) Rabby, M. F.; Tu, Y.; Hossen, M. Stacked LSTM based deep recurrent neural network with kalman smoothing for blood glucose prediction. *BMC Med. Inf. Decis. Making* **2021**, *21*, 101.
- (29) Dandil, E.; Karaca, S. Detection of pseudo brain tumors via stacked LSTM neural networks using MR spectroscopy signals. *Biocybern. Biomed. Eng.* **2020**, *41*, 173–195.
- (30) Cheng, H.; Liu, Y.; Huang, D.; et al. Adaptive Transfer Learning of Cross-Spatiotemporal Canonical Correlation Analysis for Plant-Wide Process Monitoring. *Ind. Eng. Chem. Res.* **2020**, *59*, 21602–21614.
- (31) Cai, C.; Wang, S.; Xu, Y.; et al. Transfer Learning for Drug Discovery. *J. Med. Chem.* **2020**, *63*, 8683–8694.
- (32) Bukhsh, Z. A.; Jansen, N.; Saeed, A. Damage detection using in-domain and cross-domain transfer learning. *Neural Comput. Appl.* **2021**, *10*, 16921–16936.
- (33) Sun, P.; et al. Siliciclastic depositional model within a subsurface coastal parasequence based on barriers and baffles – An application to the Donghe Sandstone Member in the Hudson Oilfield, Tarim Basin, NW China. *J. Pet. Sci. Eng.* **2020**, *195*, No. 107947.