Journal of Global Information Management
Volume 30 • Issue 7

Application of Artificial Intelligence Technology Optimized by Deep Learning to Rural Financial Development and Rural Governance

Hongwei Hou, Henan Normal University, China
Kunzhi Tang, The Australian National University, Australia
Xiaoqian Liu, Chengdu Technological University, China
*Yue Zhou, Sichuan Tourism University, China

ABSTRACT

The aim of this article is to promote the development of rural finance and the further informatization of rural banks. Based on DL (deep learning) and artificial intelligence technology, data pre-processing and feature selection are conducted on the customer information of rural banks in a certain region, including the historical deposit and loan, transaction record, and credit information. Four DL models are proposed with a precision of more than 87% by test to improve the simulation effect and explore the application of DL. The BLSTM-CNN (bi-directional long short-term memory-convolutional neural network) model with a precision of 95.8%, which integrates RNN (recurrent neural network) and CNN (convolutional neural network) in parallel, solves the shortcomings of RNN and CNN separately. The research result can provide a more reasonable prediction model for rural banks and ideas for the development of rural informatization and promoting rural governance.

KEYWORDS

Artificial Intelligence Technology, Deep Learning, Rural Finance, Rural Governance

INTRODUCTION

With the widespread popularity of mobile devices, the booming mobile Internet provides enterprises with more access to customer information, thus promoting the rapid development and broad application of big data, data storage, cloud computing, and other technologies (Lee and Oh, 2020). Besides, AI (artificial intelligence) technology has fast grown, benefiting from the rapid growth of data and the enhancement of computing power (Wang et al., 2020a). In this context, the traditional rural financial industry is facing peer competition within the industry and challenges from the Internet financial industry (Ye et al., 2018). For data-driven business industries such as banks, there are urgent requirements for business according to different demands, including obtaining practical value from the data, conducting business more effectively for various customer groups, exploring the potential value of data, and earning more incredible benefits. Bank enterprises combine data analysis, data mining, and machine learning models with actual business for customer group division, customer churn prediction, customer behavior information analysis, and customer credit prediction division.

DOI: 10.4018/JGIM.289220 *Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

However, most of these methods are in the stage of research and preliminary application. In the actual production, the use of data is not sufficient and reasonable. Furthermore, the application of big data technology and AI technology in the banking industry is still in its infancy, and there is still much room for development. Rewilak (2017) used AI technology to study whether financial development was conducive to rural poverty reduction. They found that financial deepening and greater physical access are conducive to reducing people below the poverty line. They used alternative measures of economic instability, and the results also challenged existing findings that economic instability may increase the incidence of poverty (Rewilak, 2017). To lessen the complexity and time consumption of traditional statistical and mathematical planning methods, Metawa et al. (2017) used AI technology to conduct research and analysis on rural finance. They proposed an intelligent credit decision model for rural banks based on a genetic algorithm. This model could provide a framework for optimizing bank objectives when constructing loan portfolios and seek dynamic loan decisions by maximizing bank profits and minimizing bank default probability. This model was more intelligent than the previous methods, which could shorten the screening time of bank's loan by 12% to 50%, and meanwhile increase rural bank profit by 3.9% to 8.1% (Metawa et al., 2017).

It is worth exploring how to extract effective and valuable information from numerous customer data and how to provide accurate and efficient banking services, to retain existing customers and take precautionary measures to maintain customers at risk of loss. These measures are of vital significance for the healthy and stable development of banks (Ran et al., 2020). But in reality, the risk of customer churn is unpredictable and irregular. Learning from massive data of customers is an important method to predict the probability of customer churn. With the help of big data technology and AI technology, more reliable solutions are available for massive data processing and analysis (Lee et al., 2018). Customer churn prediction means that banks use specific means or methods to find out the customers with potential loss risk and take appropriate marketing means or specific retention strategies to reduce the loss of customers more efficiently (Kinjo et al., 2018). Traditional customer churn prediction methods are mainly based on statistics and machine learning. With the popularity of computers, mobile devices, digital cameras, and other electronic devices, data increases geometrically. In addition, the continuous breakthrough of Graphics Processing Unit technology and the significant improvement of computer computing ability make the DL model have an outstanding performance in dealing with massive data and establishing complex and accurate learning models (Kalaiselvi et al., 2017). Besides, the innovation and transformation of the algorithm significantly improve the performance and training speed of DL (Kamilaris et al., 2018). The traditional machine learning model has limited space to enhance the effect when processing large amounts of data, while the DL model has natural advantages in processing mass data (Esteva et al., 2019). It can store and manage massive data of the financial industry with the help of big data technology and process or mine data maturely (Zou et al., 2019).

The primary purpose of this work is to study and analyze the financial development of rural areas by AI technology and promote rural governance. Firstly, data pre-processing and feature selection are performed after investigating the historical deposit and loan, transaction record, and credit information of rural bank customers. Secondly, four DL (deep learning) models are adopted for analysis on the data set to improve the simulation effect and application effect of DL models, including the RNN (Recurrent Neural Network), CNN (Convolutional Neural Network), BLSTM-CNN (Bi-directional Long Short-Term Memory-CNN), and Attn-BLSTM-CNN (Attention-BLSTM-CNN). Then, a comparative analysis is performed on the data analysis effects of the four models to enable rural banks to better deal with financial-related businesses. The innovation lies in the parallel integration BLSTM-CNN model combining RNN and CNN and introducing the attention mechanism into the BLSTM-CNN model, forming the Attn-BLSTM-CNN model. The Attn-BLSTM-CNN model achieves better accuracy and effect than the BLSTM-CNN model. It is hoped that this work can effectively promote the rural financial industry and further help the rural population get rid of poverty.

APPLICATION OF AI IN THE FIELD OF FINANCE

Positive Effect of Ai On The Development Of Rural Finance

The application and influence of AI are increasingly wide from academia to industry (Rahardja and Harahap, 2019). The financial sector is one of the earliest and most comprehensive industries integrated with AI. A series of AI technologies such as data mining, accurate portrait, machine learning, and neural network brings a new round of changes to financial products, service channels, service methods, risk management, credit financing, and investment decisions. The in-depth analysis of the basis, influence mechanism, risk, and development countermeasures of the integration between AI and finance has crucial theoretical value and practical significance for promoting the deep integration and healthy development of AI and rural finance.

The irreversible integrated development of AI and rural finance faces coexisting risks and opportunities. Thus, the rural financial industry must accelerate adjustment and development (Dewi and Darma, 2019; Irshat et al., 2018). In the era of AI, it is essential to transform the traditional rural financial talents into compound talents. Meanwhile, rural financial institutions should invest more in the research and development of intelligent technologies (Lu et al., 2020). Besides, rural financial supervision departments need to strengthen risk prevention and control and further clarify the regulatory responsibility. They can also apply AI technology to supervision methods and means to improve the degree of automation and intelligence of supervision (Li et al., 2017).

Customer Churn Prediction in The Development Of Rural Finance

Customer churn is the phenomenon that customers abandon the service provided by the company. In recent years, the customer churn rate of various industries, including the financial sector, is increasing. In contrast, the cost of obtaining new customers is more than five times the cost of retaining existing customers. Therefore, it is the key for enhancing the core competitiveness of banks to establish a customer churn prediction model for early warning and effective prediction, so that customer managers can take effective measures in advance to prevent customer churn (Milošević et al., 2017).

Customer churn prediction has attracted wide attention from many industries such as banks, telecoms, insurance companies, the gambling industry, and academia. Nowadays, many traditional machine learning models have achieved good simulation results in customer churn prediction, such as logistic regression, support vector machine, decision tree, and random forest model (Berger and Kompan, 2019). Under the background of fast-growing mobile Internet, banks can extract the potential temporal information of many customers by massive customer log data. The traditional machine learning model requires considerable manual design but cannot achieve excellent effects when processing mass data (Li et al., 2018). RNN is a network structure connected into a loop. In this network, the output of the neural unit is not only related to the input of the current moment but also the value of the last moment. This structural feature is appropriate for dealing with timing problems. Figure 1 illustrates the structure of RNN.

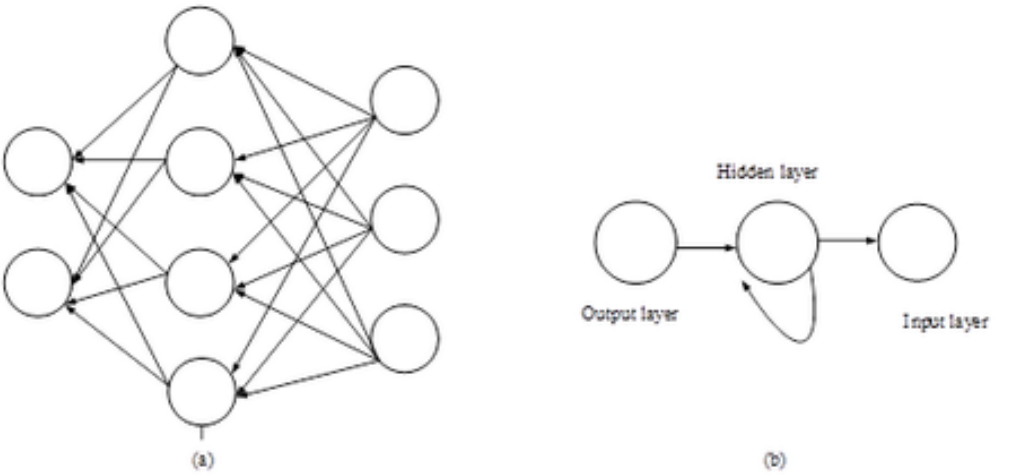
The calculation method of RNNs can be expressed as Equations (1) and (2).

$$O_{\mathrm{i}}=\mathrm{g}(\mathrm{V}\bullet S_{i}) \tag{1}$$

$$S_{i}=f(\mathrm{U}\bullet \mathrm{X}_{\mathrm{i}}+W\bullet S_{i-1}) \tag{2}$$

Among Equations (1) and (2), X represents the input data, U denotes the weight vector between the input layer and the hidden layer, and S refers to the hidden layer data. Besides, V signifies the

Figure 1. The structure of RNNs (a: structural representation; b: structural sketch)



weight vector between the hidden layer and the output layer, *O* means the output data, and *W* stands for the weight of the hidden layer to the hidden layer (Sherstinsky, 2020).

The most significant advantage of RNN is that each hidden layer is connected. The hidden layer contains the current input information, and meanwhile, the information of the time node is transmitted to the hidden layer at the next moment to cross the time node. However, in the actual learning process of the traditional RNN model, with the increase of time nodes and input, it was challenging to learn the relationship in the context, along with long-term dependence, gradient disappearance, and explosion (Kumar et al., 2018). Then, the threshold mechanism was introduced into the traditional RNN to propose the LSTM (Long Short-Term Memory) network, which overcame the disadvantages of RNN. Moreover, the BLSMT (bidirectional long short-term memory) network based on the LSTM network was used for the processing of context information. The BLSMT network has achieved excellent application effects in speech recognition, and it has remarkable performance in the model with many hidden layers. In addition to learning the correlation information on time series, it is also vital to extract the essential characteristics of different layers from high-dimensional features. The outstanding performance of CNN in image recognition and speech is different from RNN, which focuses more on extracting local features and has good representation ability. In 1989, LeCun constructed a neural network similar to modern CNN to solve computer vision problems. Besides, in 1989, LeCun also used random gradient descent to learn, which was retained in the subsequent DL technology, and the word "convolution" was first used in his paper. The standard CNN network structure generally includes a convolution layer, pooling layer, and fully connected layer. The most prominent advantage of RNN is that each hidden layer is connected. The hidden layer contains the current input information, and meanwhile, the information of the time node is transmitted to the hidden layer at the next moment to cross the time node. However, in the actual learning process of the traditional RNN model, with the increase of time nodes and input, it was difficult to learn the relationship in the context, along with long-term dependence, gradient disappearance and explosion. Later, the threshold mechanism was introduced into the traditional RNN to propose the LSTM network, which well overcame the disadvantages of RNN. Then, the BLSMT (bidirectional long short-term memory) network based on LSTM network was used to facilitate the processing of context information. The BLSMT network has been successfully applied in speech recognition with a good result, and it has better performance in the model of more hidden layers (Kim and Cho, 2019; Feng et al., 2021). In addition to learning

the correlation information on time series, it is also important to extract the essential characteristics of different layers from high-dimensional features. The outstanding performance of CNN in image recognition and speech is different from RNN, which focuses more on extracting local features and has good representation ability. In 1989, LeCun constructed a neural network similar to modern CNN to solve computer vision problems. In 1989, LeCun also used random gradient descent to learn, which was retained in the subsequent DL technology, and the word "convolution" was first used in his paper. CNN is also a multi-layer network structure, mainly composed of the convolution layer, pooling layer and fully connected layer. The convolution layer can select the local region features of the upper layer after convolution operation. During image processing, the convolution operation of a photo is essentially image filtering by a convolution kernel. Convolution of images can be written as Equation (3).

$$f(x,y) * w(\mathrm{x},\mathrm{y}) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} w(\mathrm{s},\mathrm{t})\, \mathrm{f}(\mathrm{x}-\mathrm{s},\mathrm{y}-\mathrm{t}) \tag{3}$$

In Equation (3), $f(x, y)$ represents the pixel grey value of a specific coordinate point in the selected image, and $w(x, y)$ denotes the convolution kernel. In addition, a and b refer to the size of the convolution kernel.

RNN and CNN have their specific advantages and disadvantages. In 2019, some researchers proposed the DLCNN (Densely-connected LSTM-CNN) model of serial integration of densely connected LSTM network and CNN (Shahid et al., 2020; Qin et al., 2021; Jian et al., 2021). DLCNN can better balance the local characteristics and timing characteristics. However, in the serial structure, the output results of the LSTM layer will ignore some local information when input to the convolution layer. Besides, the LSTM module is prone to over-fitting when there are numerous hidden layers (Wang et al., 2020b; Yang et al., 2021). The attention mechanism model was first used in machine translation. Attention has become an essential concept in the field of neural networks. Its main idea is to obtain the target information requiring concern by targeting the area that needs to be focused, and suppress other useless information, which can effectively solve the long-distance dependence of neural networks. The behavior data of bank customers is characterized by large amounts, time sequence, and complexity, so the attention mechanism is introduced to improve the performance of the existing RNN model (Ullah et al., 2017; Bie et al., 2021).

Customer Churn Prediction Model Based on DL

1. Overview of the BLSTM (Bi-directional Long Short-Term Memory) network

RNN is a kind of neural network with serial row data as input, recursion in the direction of the sequence, and chain connection of all nodes (Song et al., 2019; Rui et al., 2021; Deng et al., 2021). In theory, the traditional RNN model can learn the information of any time series. However, in a real scenario, with the increase of input and the growth of time series, RNN cannot learn the relationship in the context of the time series, and it may produce a long-term dependence problem. Although the gradient descent algorithm can optimize RNN, gradient disappearance or gradient explosion may occur when the gradient propagates backward along the sequence (Kim et al., 2019; Li et al., 2021). Equation (4) illustrates the forward propagation of RNN.

$$\begin{cases} h_t = \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \\ o_{t+1} = W_{hy} + b_y \\ y_t = soft\max(o_t) \end{cases} \tag{4}$$

The output at h_{t-1} is essential for calculating the output at h_t, and the output at h_{t-2} is required by the output at h_{t-1}. The state at a particular time of RNN depends on all the time states in the past.

The structure of the LSTM network is similar to that of RNN. However, the LSTM network is more refined in the internal processing unit than the traditional RNN. It can better retain valid information or discard invalid details. The interior unit structure consists of three gate control units, including the input gate, forgetting gate, and output gate (Wu et al., 2018; Chen et al., 2020; Chen et al., 2021). The hidden layer of the LSTM network mainly contains three kinds of gate structures, including output gate, forgetting gate, and input gate. These three gates can realize the control of information transmission. In each door, there is also a dot multiplication operation and a sigmoid layer. The output value range of the sigmoid layer is (0, 1), describing the amount of information passed through each part, where "0" means no passing, and "1" means all passing.

1) The input layer information Ct is calculated according to Equation (5), where W_{ij} represents the weight vector between the input data and the hidden layer, W_{jc} refers to the output weight at the last moment, and b_c signifies the offset.

$$C_t = \tanh(\mathrm{w_{ij}x_t} + \mathrm{w_{jc}C_{t-1}} + \mathrm{b_c}) \tag{5}$$

2) The input gate it is calculated according to Equation (6), where W_{xi} denotes the weight between the input gate and the input information, and w_{ih} represents the weight between the input gate and the output at the last moment. In addition, w_{ci} refers to the weight between the input gate and the Cell at the last moment, and b_i means the offset.

$$i_t = \psi(\mathrm{w_{xi}x_i} + \mathrm{w_{hi}h_{i-1}} + \mathrm{w_{ci}c_{t-1}} + \mathrm{b_i}) \tag{6}$$

3) Similarly, the forgetting gate is calculated according to Equation (7).

$$f_t = \psi(\mathrm{w_{xf}x_t} + \mathrm{w_{hf}h_{t-1}} + \mathrm{w_{cf}c_{t-1}} + \mathrm{b}_f) \tag{7}$$

4) The state value in the Cell is calculated according to Equation (8).

$$c_t = f_t \otimes c_{t-1} + i_t \otimes c_t \tag{8}$$

5) The state of the input gate is calculated according to Equation (9).

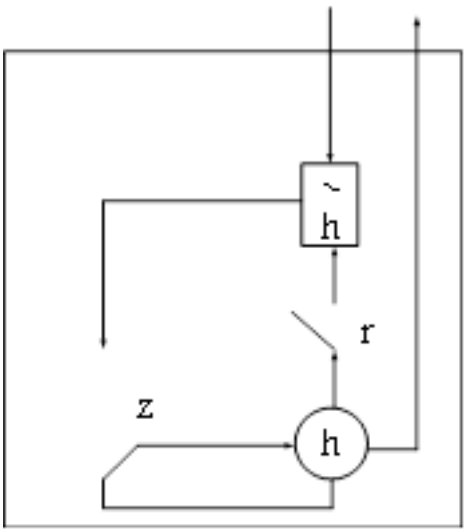
$$O_t = \psi(\mathrm{w_{xo}x_t} + \mathrm{w_{ho}h_{t-1}} + \mathrm{w_{co}c_{t-1}} + \mathrm{b}_o) \tag{9}$$

6) The final input of the LSTM network can be displayed as Equation (10).

$$h_t = o_t \otimes \tanh(c_t) \tag{10}$$

The LSTM network is further simplified to obtain the GRU (Gated Recurrent Unit) neural network. Considering that the LSTM network has a singularly complex "gate" structure, the detection process by the LSTM network is highly complex, consuming lots of time. Therefore, the simplified GRU neural network is responsible for reducing the gates of the LSTM network to two, namely the update gate and the reset gate. The role of the reset gate is mainly to control the degree of discarding the state information in the GRU structure at the last moment, and the update gate is responsible for regulating the degree of dumping the memory unit information at the last moment. Figure 2 reveals the unit structure of GRU.

Figure 2. The unit structure of GRU



Equation (12) expatiates the update gate in the unit at time *t* for the *j*th unit in the GRU network.

$$Z_t^j = \sigma(W_z x_t + U_z h_{t-1})^j \tag{11}$$

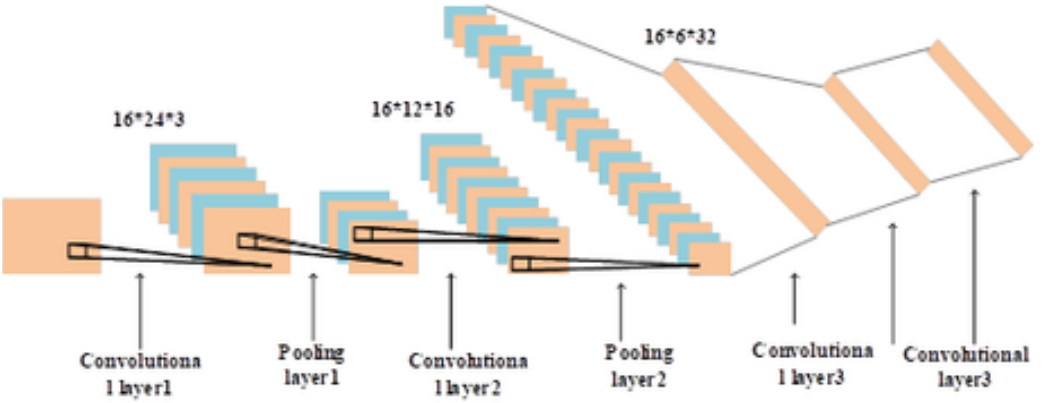
The reset gate at time *t* can be expressed as Equation (13).

$$r_t^j = \sigma(W_r x_t + U_r h_{t-1})^j \tag{12}$$

Then:

$$h_t^j = (1 - z_t^j) h_{t-1}^j + z_t^j \tanh(W_r x_i + U(r_r \odot h_{t-1})^j \tag{13}$$

Figure 3. Structure of the CNN model



In the BLSTM network, the input of the current moment depends on not only the information learned previously but also the information learned later, to take full account of the previous and later temporal information, as shown in the following five equations.

$$h_t = sigmoid(W_1 x_t + W_2 h_{t-1} + b^{(1)}) \tag{14}$$

$$h'_t = sigmoid(W_3 x_t + W_5 h_{t+1} + b_t^{(2)}) \tag{15}$$

$$o_t = \tanh(W_4 h_t + b_t^{(3)}) \tag{16}$$

$$o'_t = \tanh(W_6 h_t + b_t^{(4)}) \tag{17}$$

$$o_t = \frac{(o'_t + o''_t)}{2} \tag{18}$$

In the above five equations, $b_t^{(1)}$, $b_t^{(2)}$, $b_t^{(3)}$ and $b_t^{(4)}$ represent the bias of the hidden unit in the BiLSTM network. o and o_t¢ are output results of the feature vectors of two LSTM units dealing with hidden layer information at the corresponding time. Among the equations, the value of the feature vector o_t is the mean of feature vectors at two moments.

2. Construction of the customer churn prediction model based on DL

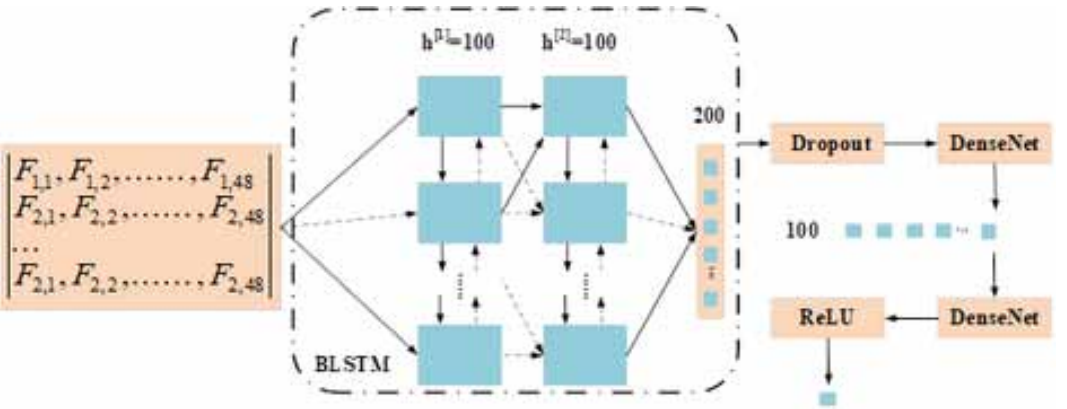
The CNN model in the experiment uses three convolution processes with the final full connection for prediction results. Figure 3 reveals the structure of the CNN model.

Independent identical distribution is an essential assumption in the field of machine learning. As one of the critical achievements of DL in recent years, Batch Normalization keeps the input of each neural network with the same distribution in the training process of DL. Leaky-ReLU (Rectified Linear Unit) is a variant of ReLU, one of the commonly used activation functions in DL, as shown in Equation (19).

$$f(x) = \begin{cases} x & if \quad x > 0 \\ \lambda x & if \quad x \leq 0 \end{cases} \tag{19}$$

The BLSTM model consists of two layers of BLSTM structure and the final complete connection process. Figure 4 displays the structure of the BLSTM model.

Figure 4. Structure of the BLSTM model



Chen et al. (2020) believed that RNN could learn potential temporal information from time-series data but could not capture critical local features (Chen et al., 2020; Zheng et al., 2021). On the contrary, CNN can learn local features but ignores temporal information. The LSTM module in the original DLCNN model is replaced by the BLSTM network to verify the performance of the integrated DLCNN model and the pre-integrated BLSTM model. Figure 5 demonstrates the structure of the DLCNN model before the replacement. Figure 6 shows the structure of the DLCNN model after the replacement.

As shown in Figure 5, the DLCNN model integrates RNN with CNN in a serial structure to eliminate their shortcomings. The BLSTM network substitute for the LSTM layer in the original DLCNN model to prevent DLCNN from being interfered with by other factors in the comparison test with the CNN model and BLSTM model, and to verify its performance after integration. Figure 6 represents the model structure after replacing its LSTM module.

BLSTM-CNN, a parallel integration of BLSTM and CNN, is proposed to alleviate poor simulation results caused by the neglect of some local information when the output results of the BLSTM layer are input to the convolution layer. In the BLSTM-CNN model, Feature_RNN is used as the input of the BLSTM module and the CNN module, to avoid the information dropout and increases the input of feature information. Figure 7 provides the structure of the BLSTM-CNN model.

As shown in Figure 7, the CNN module and BLSTM module are combined in parallel to avoid the feature loss caused by the feature passing through the BLSTM module, which is the most obvious

Figure 5. Structure of the DLCNN model before the replacement

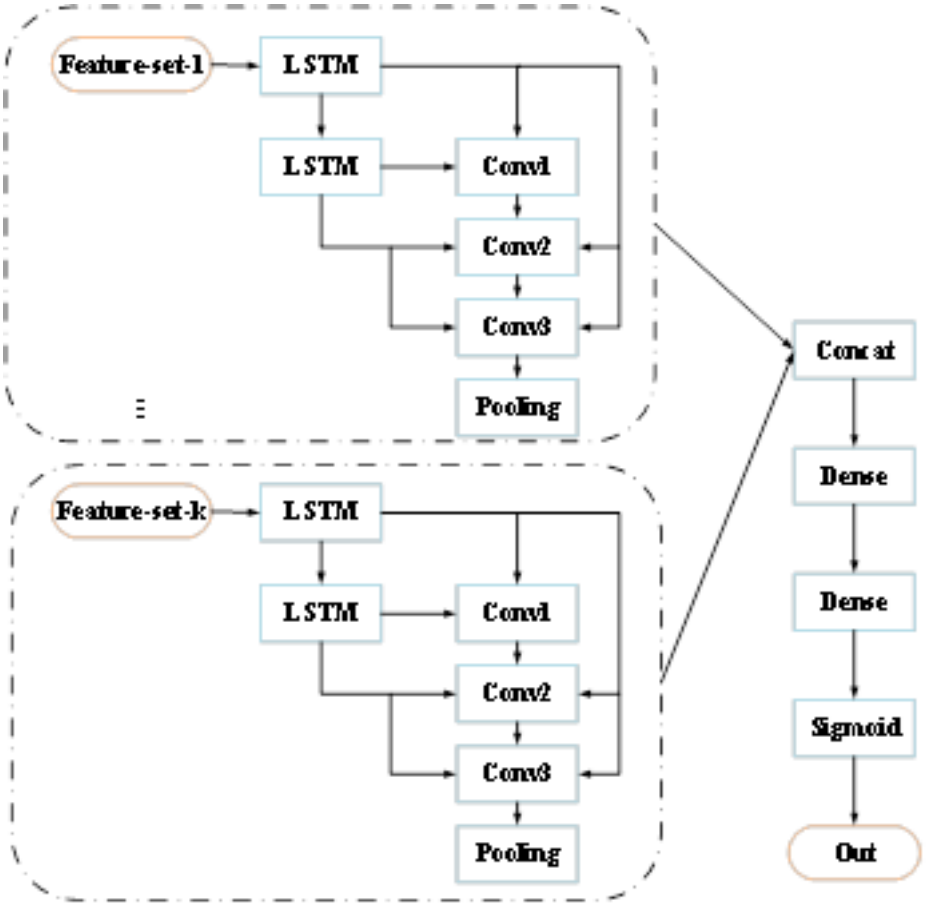
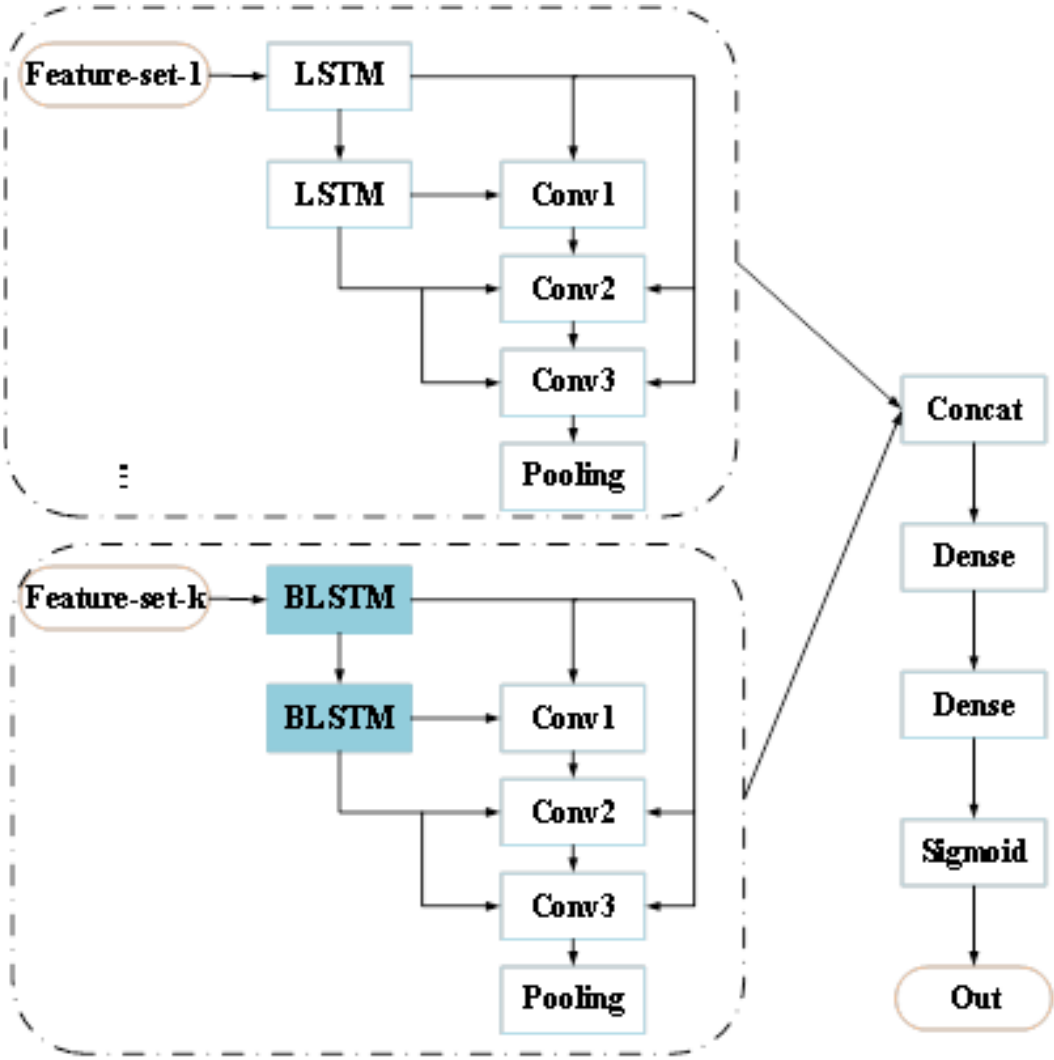


Figure 6. Structure of the DLCNN model after the replacement



difference from the DLCNN model. Besides, Feature_CNN is used as the independent input of the CNN module, further increasing the feature input of the BLSTM-CNN model.

Optimization of the Customer Churn Model Based on DL

Gao et al. (2018) stated that the attention-based BLSTM-CNN model had been widely used in various DL tasks, such as natural language processing, image recognition, and speech recognition. It has been one of the core technologies in DL technology worthy of attention and in-depth understanding (Gao et al., 2018; Wu et al., 2021). It is mainly combined with RNNs to improve network performance. Here, Attn-BLSTM-CNN is proposed based on BLSTM-CNN with the introduction of the attention mechanism to explore whether it can further enhance the model's performance. For comparison, the structures of the three models are shown in Figure 8, Figure 9, and Figure 10, respectively.

Figure 8 represents the structure of the BLSTM-CNN model, and Figure 9 denotes the structure of BLSTM-CNN with the attention mechanism. Figure 10 shows the location and system of the

Figure 7. Structure of the BLSTM-CNN model

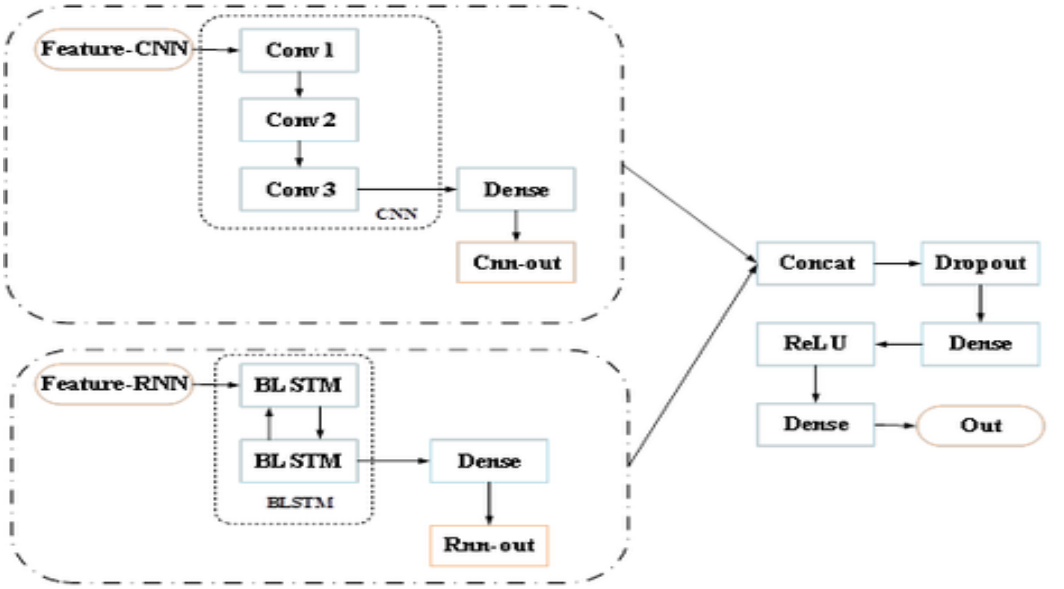


Figure 8. Structure of the BLSTM-CNN model

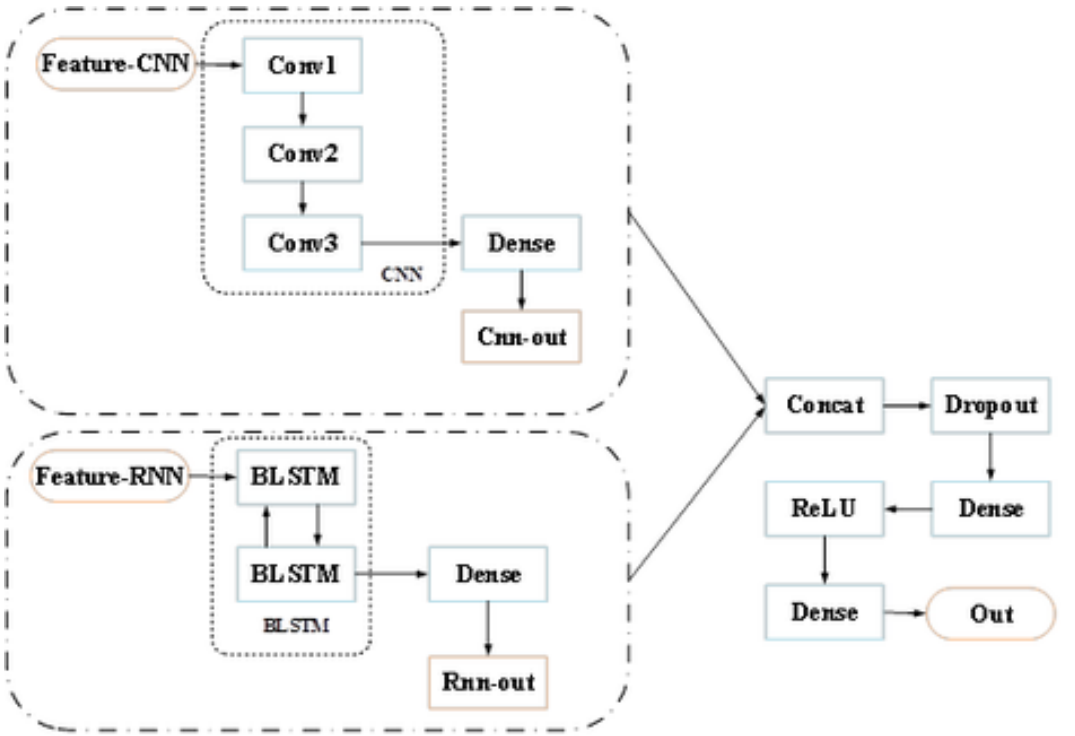


Figure 9. Structure of the Attn-BLSTM-CNN model

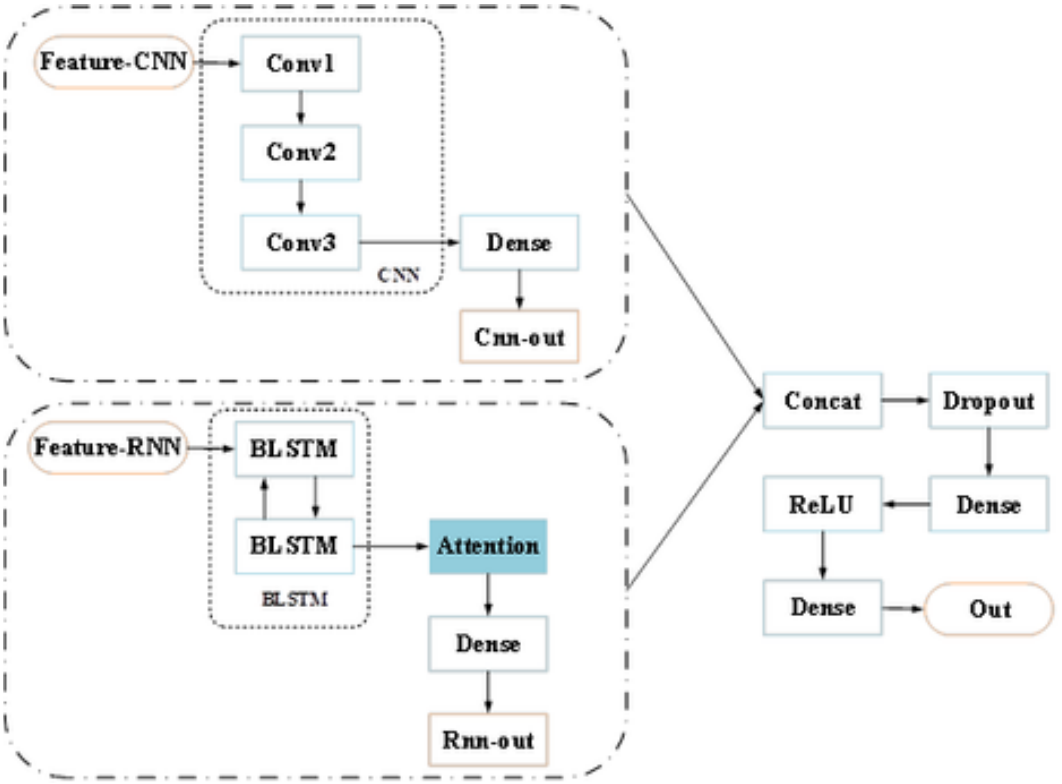
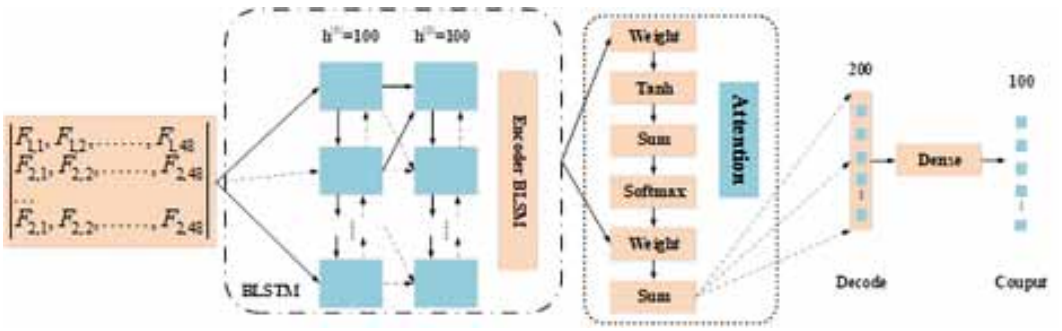


Figure 10. Structure of the AttnBLSTM model

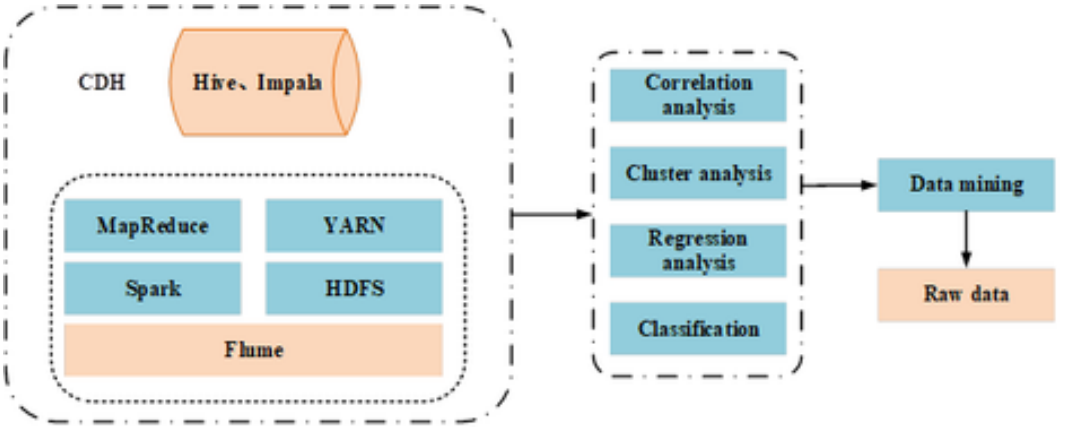


attention layer in the Attn-BLSTM-CNN model, and the Attention layer and the BLSTM layer are combined as the AttnBLSTM module. Here, the Attn-BLSTM-CNN model and BLSTM-CNN model are compared to verify the performance of the BLSTM-CNN model with the attention mechanism.

Data of the Customer Churn Model Based On DL

The original data comes from customer logs, historical deposits and loans, and credit investigation a rural bank in a certain area. The acquisition process of the original data is illustrated in Figure 11.

Figure 11. Acquisition process of the original data



The log data and basic information data of customers are mainly stored in the Oracle database of the bank's big data platform named CDH (Cloudera's Distribution including Apache Hadoop), which are mapped to the data warehouse Impala through relationships. After correlation analysis, cluster analysis, regression analysis, and classification, data mining is carried out to obtain the raw data table. According to the model, the raw data requires data pre-processing, feature construction, and data equalization to form actual feature data as the input of the model.

In the binary classification problem, according to the combination of the actual class and the predicted class by learners, the samples are divided into the TP (true positive) case, FP (false positive) case, TN (true negative) case, and FN (false negative) case. Let TP, FP, TN, and FN denote the corresponding sample number, respectively, and then TP + FP + TN + FN = the total sample number, forming a "confusion matrix". The accuracy, precision, recall rate, and F1 value are calculated according to the confusion matrix to evaluate the performance of different classifiers in customer churn prediction. The accuracy is the ratio of the sum of the correctly predicted positive cases and the correctly predicted negative cases to the total cases, as shown in Equation (20).

$$accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{20}$$

The precision is the proportion of the correctly predicted positive cases to the positive cases, as presented in Equation (21).

$$precision = \frac{TP}{TP + FP} \tag{21}$$

The recall rate refers to how many positive examples in the sample are correctly predicted, that is, the predicted coverage of positive examples, as displayed in Equation (22).

$$recall = \frac{TP}{TP + FN} \tag{22}$$

The F1 value is usually used as an index to evaluate the performance of the classifier. The closer the F1 value is to 1, the better combination precision and recall rate of the model has. F1 value can be calculated according to Equation (23).

$$F1 = 2 \times \frac{precision \cdot recall}{precision + recall} \tag{23}$$

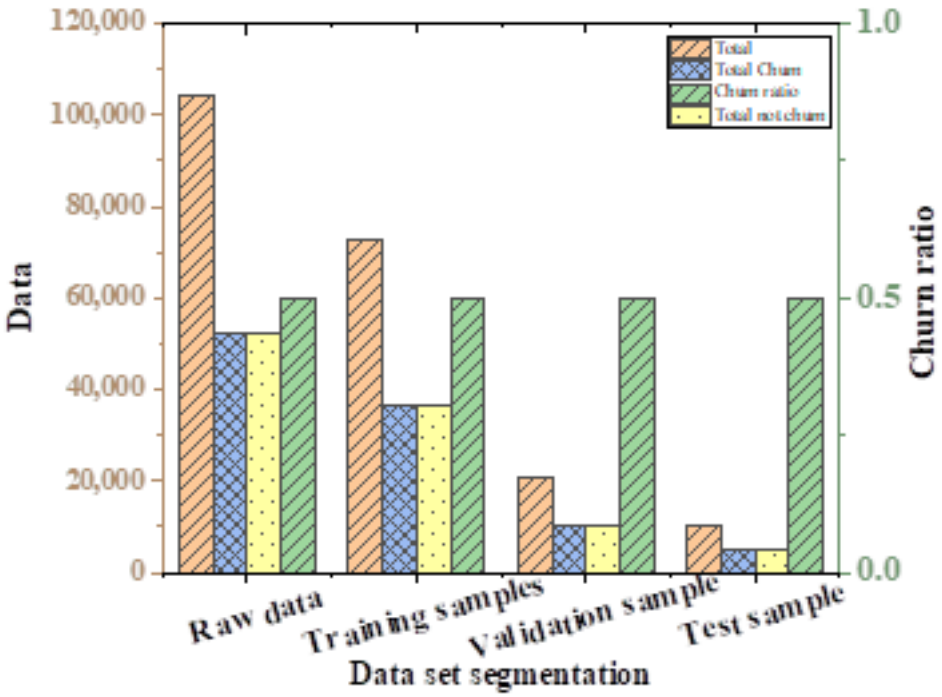
ROC (receiver operating characteristic) curve and AUC (area under curve) value are used in this experiment to evaluate the performance of each model. FPR (false positive rate) represents the abscissa of the ROC curve, and TPR (true positive rate) denotes the ordinate of the ROC curve. In the customer churn prediction model, the ROC curve can reflect the discriminant ability of the model to the churn customer, that is, the sensitivity of the model to the churned customer. The AUC value is the area obtained by the integral of the ROC curve. The closer the AUC value is to 1, the better the performance of the model is.

RESULTS ANALYSIS OF THE CUSTOMER CHURN PREDICTION MODEL FOR THE BANK

Comparison Results of Each Model

The segmentation of data is shown in Figure 12.

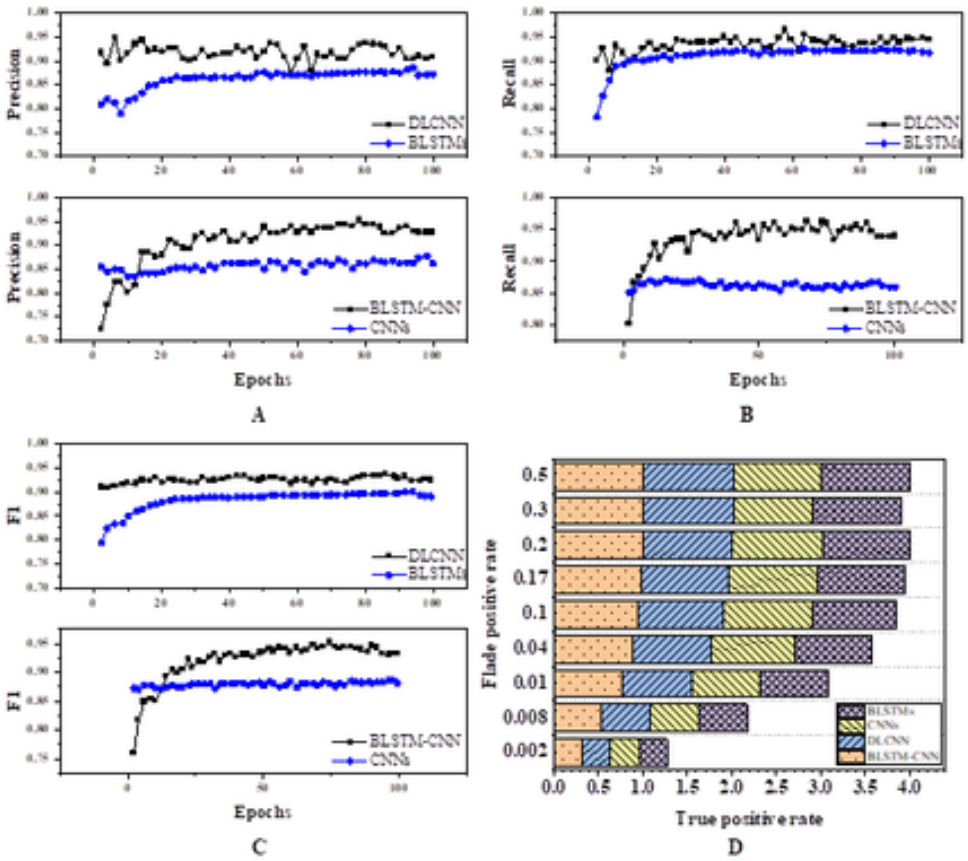
Figure 12. Segmentation of data



From Figure 12, training samples account for 70% of the total sample, validation samples occupy 20%, and test samples are 10% of the whole samples. Training samples are used for model training. Validation samples are used to validate whether the model is overfitting and verify the generalization ability of the model in the training process. Test samples are used to demonstrate the precision of model training.

Figure 13 presents the test results.

Figure 13. Evaluation results of each index (A: precision; B: recall rate; C: FI; D: ROC curve and AUC value)



As shown in Figure 13A, after 100 iterations, the best precision of the LSTM-CNN model, DLCNN model, BLSTM model, and CNN model is 95.8.9%, 94.9%, 88.6%, and 87.9%, respectively. In the iterative process, the BLSTMCNN model and the BLSTM model have better stability than the DLCNN model and the CNN model, which all tend to be stable. Figure 13B shows that the best recall rate of the BLSTM-CNN model, DLCNN model, BLSTM model, and CNN model is 97.1%, 96.25%, 92.31%, and 87.41%, respectively. Besides, the F1 value of the four models in Figure 13C is 0.953, 0.9473, 0.9021, and 0.8828, respectively. The AUC values of the BLSTM-CNN model, DLCNN model, CNN model, and BLSTM model in Figure 13D are 10.9631, 0.9425, 0.9024, and 0.8875, respectively.

The results of precision, recall rate, F1 value and AUC value of each model with the best F1 value are shown in Figure 14.

Figure 14. Results of each dimension under the optimal F1 value (a: comparison between the CNN and DLCNN; B: comparison between the BLSTM and BLSTM-CNN)

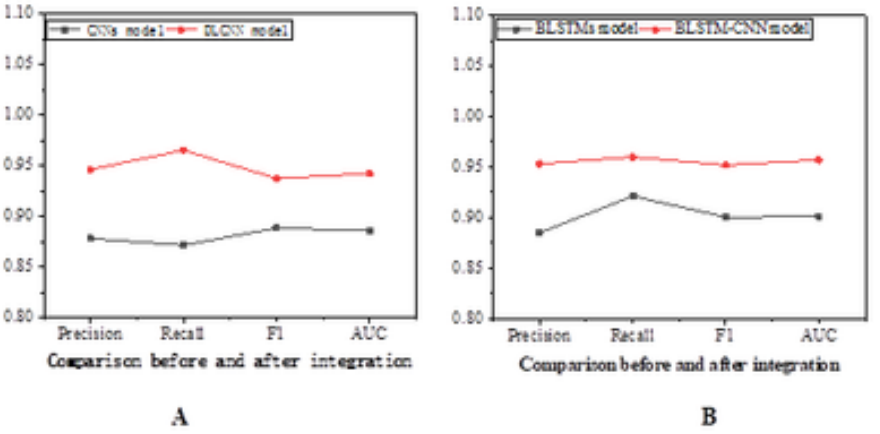


Figure 14 indicates that the four DL models are better than the CART classification tree and the random forest model. In the prediction process of customer churn in rural banks, the DL models show better simulation results, among which the BLSTM-CNN model is the best. The prediction precision of the BLSTM-CNN model is 95.8%, the recall rate is 96.54%, the F1 value is 0.9512, and the AUC value is 0.9575. Moreover, the DLCNN model and the BLSTM-CNN model, which integrate CNN with the RNN, achieve better simulation results than the unintegrated model. This indicates that the integrated model solves the defects of CNN models and RNN models. The indexes of the BLSTM model are slightly better than those of the CNN model, further demonstrating that the RNN model has better advantages than the CNN model when predicting the loss of rural customers with temporal characteristics of rural banks. The F1 value and the AUC value of the BLSTM-CNN model are 1.50% and 1.51% higher than those of the DLCNN model, respectively. It is proved that the BLSTM-CNN model can competently deal with the prediction problem under partial data loss.

Comparison Between the BLSTM-CNN Model and Attn-BLSTM-CNN Model

The experiment evaluates the precision, recall rate, F1 value, ROC curve, and AUC value of the two models, respectively, by the test set and with 100 iterations. The experimental results are shown in Figure 15.

As shown in Figure 15A, the optimal precision of the Attn-BLSTM-CNN model and BLSTM-CNN model is 96.01% and 95.8%, respectively. In the iterative process, the precision of the two models both shows a rising tendency. After reaching the optimum value, the precision tends to be stable and fluctuates around the best precision. According to Figure 15B, the best recall rate obtained by the two models are 98.1% and 95.9%, respectively. In Figure 15C, the best F1values of the two models are 0.958 and 0.952, respectively. Figure 15D shows the ROC curves of the two models corresponding to the iterative results with the highest F1 value of each model, with AUC values of 0.9721 and 0.9582, respectively.

The results of precision, recall rate, F1 value, and AUC value of the Attn-BLSTM-CNN model and BLSTM-CNN model with the best F1 value are shown in Figure 16.

From the data statistics shown in Figure 16, the Attn-BLSTM-CNN model performs better in all aspects of data than the BLSTM-CNN model. The precision of the Attn-BLSTM-CNN model reaches 96.01%, the recall rate attains 98.01%, the F1 value is 0.9621, and the AUC value is 0.9647.

Figure 15. Evaluation results of each index (A: precision; B: recall rate; C: FI; D: ROC curve and AUC value)

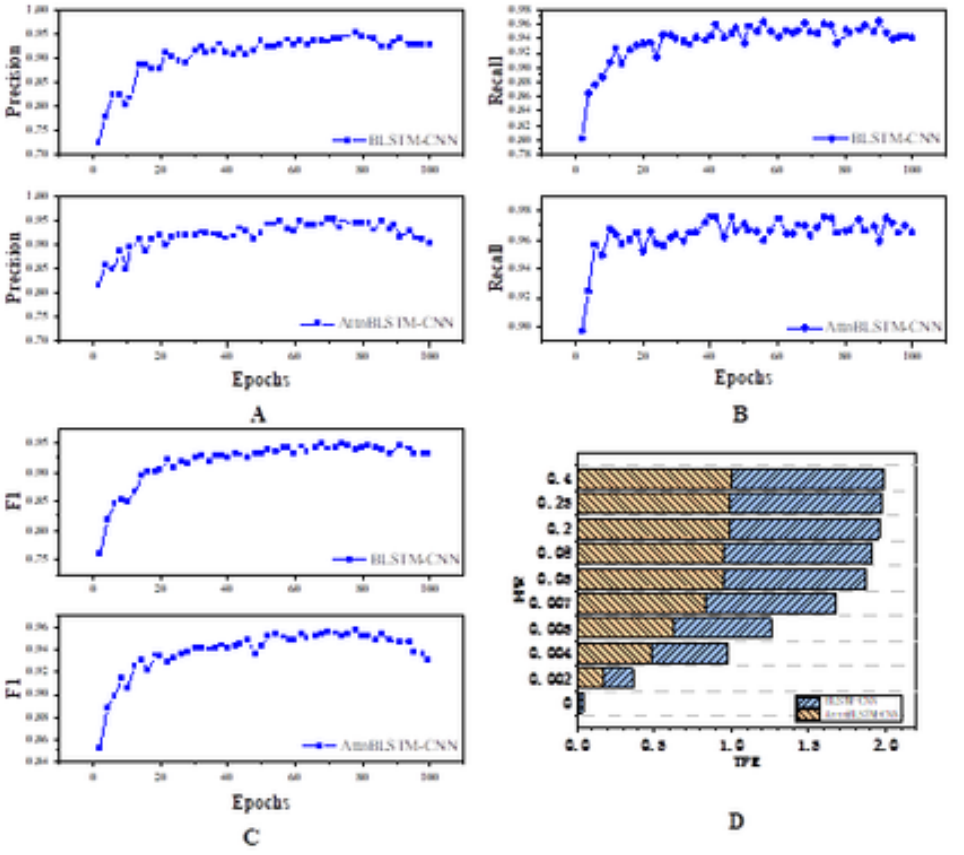
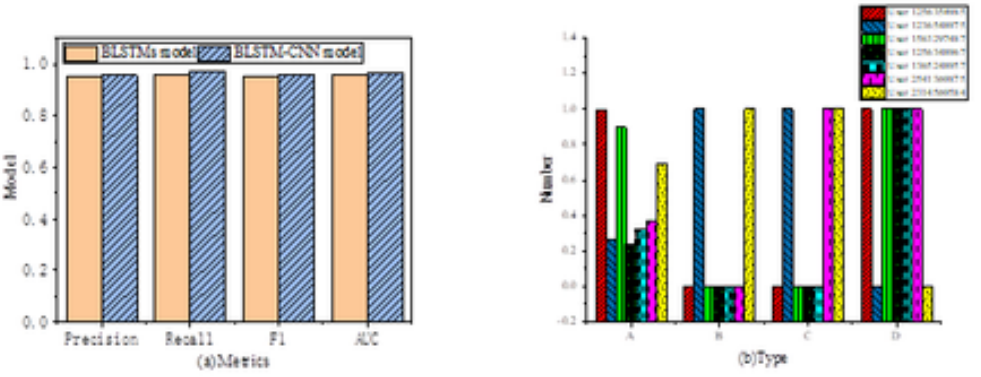


Figure 16. Statistical results of F1 value of two models (a: F1 value; b: sum of test results of some customers; A: prediction results of a customer; B: the prediction results, where 1 means that the customer is lost, and 0 means that the customer is not lost; C: whether customers are lost; D: whether the customer is correctly predicted, where 1 indicates that the prediction is correct, and 0 indicates that the prediction is incorrect.)



It demonstrates that the introduction of Attention into BLSTM can further improve the performance of the model. At the same time, the model's training on some customer data shows high prediction precision of the model, and there are no more prediction errors.

IMPACT OF THE DEVELOPMENT OF FINANCE ON RURAL GOVERNANCE

Issues relating to agriculture, rural regions and farmers have significant impact on political stability, economic growth, and social harmony of China. At present, China has entered a new era of rural governance and development. Lack of talents and funds in rural areas is the status quo of the rural regions and a major issue to be solved in rural revitalization strategy. The current overall scale of the rural financial market is incompetent to support the implementation of the rural revitalization strategy, with the backward development of information technology and prominent redundancy or absence of financial supervision. The reform of rural financial institutions is inseparable from science and technology, mainly from the following aspects to optimize the service of rural financial institutions.

Firstly, it is necessary to increase investment in rural financial science and technology. Policy support and capital investment are the most crucial factors affecting the development in rural areas. Under the leadership of local governments, attentions should be paid to rural financial development policies and capital and the promotion of infrastructure construction such as rural network construction and base station construction. Moreover, publicity and promotion should be increased for various new products and services, focusing more on the cultivation and placement of scientific and technological talents. Secondly, blockchain technology can be used to promote the popularization of inclusive finance in rural areas. The characteristics of open and distributed development make blockchain a significant advantage in rural financial growth. Still, the broad application of blockchain technology in rural financial development needs the joint efforts of the government and rural financial institutions. Finally, it is urgent to strengthen the pastoral financial supervision and risk prevention level. The synchronous development of regulatory technology and compliance technology is the fundamental guarantee of the healthy development of rural finance with less risk. It is essential to strengthen data standardization, comprehensive analysis, and deep utilization of regulatory technology. The supervision and risk prevention of the rural financial system should also be controlled within a certain range. In other words, the environmental standards of rural should be formulated as the guarantee of risk prevention and control, such as customer standards, product standards, and service standards.

CONCLUSION

The historical deposit and loan, transaction record, and credit information of rural bank customers in a certain area are studied to improve the development level of rural finance and help rural banks to further informatization. Based on DL and AI technology, four DL models with high accuracy of prediction results are proposed to improve the simulation effect and explore the application of the DL model in the samples. The BLSTM-CNN model, which integrates RNN and CNN in parallel, effectively solves the problems when RNNs or CNNs operate separately. Meanwhile, it solves the problem that partial local information is ignored when the output of the LSTM layer in the DLCNN model is input to the convolution layer. The precision of the BLSTM-CNN model reaches 95.8%. The Attn-BLSTM-CNN model performs better than the BLSTM-CNN in all aspects. The precision of the Attn-BLSTM-CNN model is 96.01%, the recall rate reaches 98.01%, the F1 value is 0.9621, and the AUC value is 0.9647. It shows that the introduction of the attention mechanism into BLSTM can further improve the performance of the model. Through the comparative analysis of four neural network models, the loss rate prediction of rural banks for rural customers improves, which significantly helps the financial development of rural banks. However, there are some shortcomings in this experiment. As for the model, there are insufficient layers in the BLSTM module of Attn-BLSTM-CNN to accurately predict the customer churn on a mass of rural customer data. Therefore,

in future research, according to the actual needs of the rural banking business, the number of layers can be appropriately increased to ensure the precision of the model. The work reported here provides some references for the informatization development of rural banks and facilitates the steady growth of rural finance, which is feasible for rural governance.

FUNDINGS

This work was supported by Henan Philosophy and Social Science Planning Project (2019CJJ081). This work was also supported by Ph. D foundation program of Chengdu Technological University, the impact of County Administrated by Province reform on counties'economic performance and people's livelihood (No. 2019RC025).

ACKNOWLEDGMENT

Yue Zhou is the corresponding author, email: vvzyzy@163.com.

REFERENCES

Berger, P., & Kompan, M. (2019). User modeling for churn prediction in E-commerce. *IEEE Intelligent Systems*, *34*(2), 44–52. doi:10.1109/MIS.2019.2895788

Bie, Y., Ji, J., Wang, X., & Qu, X. (2021). Optimization of electric bus scheduling considering stochastic volatilities in trip travel time and energy consumption. *Computer-Aided Civil and Infrastructure Engineering*, mice.12684. Advance online publication. doi:10.1111/mice.12684

Chen, D., Gao, H., & Ma, Y. (2021). Human capital-driven acquisition: Evidence from the Inevitable Disclosure Doctrine. *Management Science*, *67*(8), 4643–4664. doi:10.1287/mnsc.2020.3707

Chen, Y., Hu, S., Mao, H., Deng, W., & Gao, X. (2020). Application of the best evacuation model of deep learning in the design of public structures. *Image and Vision Computing*, *102*, 103975. doi:10.1016/j.imavis.2020.103975

Deng, R., Li, M., & Linghu, S. (2021). Sensitivity analysis of steam injection parameters of steam injection thermal recovery technology. *Fresenius Environmental Bulletin*, *30*(05), 5385–5394.

Dewi, M. V. K., & Darma, G. S. (2019). The Role of Marketing & Competitive Intelligence In Industrial Revolution 4.0. *Jurnal Manajemen Bisnis*, *16*(1), 1–12. doi:10.38043/jmb.v16i1.2014

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, *25*(1), 24–29. doi:10.1038/s41591-018-0316-z PMID:30617335

Feng, J., Liu, Z., & Feng, L. (2021). Identifying opportunities for sustainable business models in manufacturing: Application of patent analysis and generative topographic mapping. *Sustainable Production and Consumption*, *27*, 509–522. doi:10.1016/j.spc.2021.01.021

Gao, X., Zhao, S., & Yibo, S. (2018, August). An analysis of the current status and countermeasures of bike-sharing in the background of Internet. In *2018 International Conference on Virtual Reality and Intelligent Systems (ICVRIS)* (pp. 469-472). IEEE. doi:10.1109/ICVRIS.2018.00121

Irshat, K., Petr, R., & Irina, R. (2018, October). The selecting of artificial intelligence technology for control of mobile robots. In *2018 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)* (pp. 1-4). IEEE.

Kalaiselvi, T., Sriramakrishnan, P., & Somasundaram, K. (2017). Survey of using GPU CUDA programming model in medical image analysis. *Informatics in Medicine Unlocked*, *9*, 133–144. doi:10.1016/j.imu.2017.08.001

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90. doi:10.1016/j.compag.2018.02.016

Kim, J., Moon, J., Hwang, E., & Kang, P. (2019). Recurrent inception convolution neural network for multi short-term load forecasting. *Energy and Buildings, 194*, 328-341. 10.1016/j.enbuild.2019.04.034

Kim, T. Y., & Cho, S. B. (2019). Predicting residential energy consumption using CNN-LSTM neural networks. *Energy*, *182*, 72–81. doi:10.1016/j.energy.2019.05.230

Kinjo, A. R., Bekker, G. J., Wako, H., Endo, S., Tsuchiya, Y., Sato, H., Nishi, H., Kinoshita, K., Suzuki, H., Kawabata, T., Yokochi, M., Iwata, T., Kobayashi, N., Fujiwara, T., Kurisu, G., & Nakamura, H. (2018). New tools and functions in data-out activities at Protein Data Bank Japan (PDBj). *Protein Science*, *27*(1), 95–102. doi:10.1002/pro.3273 PMID:28815765

Kumar, J., Goomer, R., & Singh, A. K. (2018). Long short term memory recurrent neural network (LSTM-RNN) based workload forecasting model for cloud datacenters. *Procedia Computer Science*, *125*, 676–682. doi:10.1016/j.procs.2017.12.087

Lee, H. J., & Oh, H. (2020). A Study on the deduction and diffusion of promising artificial intelligence technology for sustainable industrial development. *Sustainability*, *12*(14), 5609. doi:10.3390/su12145609

Lee, J., Davari, H., Singh, J., & Pandhare, V. (2018). Industrial Artificial Intelligence for industry 4.0-based manufacturing systems. *Manufacturing Letters*, *18*, 20–23. doi:10.1016/j.mfglet.2018.09.002

Li, G., Hou, Y., & Wu, A. (2017). Fourth Industrial Revolution: Technological drivers, impacts and coping methods. *Chinese Geographical Science*, *27*(4), 626–637. doi:10.1007/s11769-017-0890-x

Li, S., Li, W., Cook, C., Zhu, C., & Gao, Y. (2018). Independently recurrent neural network (indrnn): Building a longer and deeper RNN. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5457-5466). doi:10.1109/CVPR.2018.00572

Li, X., Li, Z., Jia, T., Yan, P., Wang, D., & Liu, G. (2021). The sense of community revisited in Hankow, China: Combining the impacts of perceptual factors and built environment attributes. *Cities (London, England)*, *111*, 103108. doi:10.1016/j.cities.2021.103108

Lu, Y., Hu, X., & Su, Y. (2020). Framework of industrial networking sensing system based on edge computing and artificial intelligence. *Journal of Intelligent & Fuzzy Systems*, *38*(1), 283–291. doi:10.3233/JIFS-179403

Metawa, N., Hassan, M. K., & Elhoseny, M. (2017). Genetic algorithm based model for optimizing bank lending decisions. *Expert Systems with Applications*, *80*, 75–82. doi:10.1016/j.eswa.2017.03.021

Milošević, M., Živić, N., & Andjelković, I. (2017). Early churn prediction with personalized targeting in mobile social games. *Expert Systems with Applications*, *83*, 326–332. doi:10.1016/j.eswa.2017.04.056

Qin, C., Jin, Y., Tao, J., Xiao, D., Yu, H., Liu, C., Shi, G., Lei, J., Liu, C., & Liu, C. (2021). DTCNNMI: A deep twin convolutional neural networks with multi-domain inputs for strongly noisy diesel engine misfire detection. *Measurement*, *180*, 109548. doi:10.1016/j.measurement.2021.109548

Rahardja, U., & Harahap, E. P. (2019, July). Implementation of Information Planning and Strategies Industrial Technology 4.0 to Improve Business Intelligence Performance on Official Site APTISI. *Journal of Physics: Conference Series*, *1179*(1), 012111.

Ran, M., Chen, L., & Li, W. (2020). Financial deepening, spatial spillover, and urban–rural income disparity: Evidence from China. *Sustainability*, *12*(4), 1450. doi:10.3390/su12041450

Rewilak, J. (2017). The role of financial development in poverty reduction. *Review of Development Finance, 7*(2), 169-176. 10.1016/j.rdf.2017.10.001

Shahid, F., Zameer, A., & Muneeb, M. (2020). Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons, and Fractals*, *140*, 110212. doi:10.1016/j.chaos.2020.110212 PMID:32839642

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D. Nonlinear Phenomena*, *404*, 132306. doi:10.1016/j.physd.2019.132306

Song, X., Yang, F., Wang, D., & Tsui, K. L. (2019). Combined CNN-LSTM network for state-of-charge estimation of lithium-ion batteries. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 88894–88902. doi:10.1109/ACCESS.2019.2926517

Ullah, A., Ahmad, J., Muhammad, K., Sajjad, M., & Baik, S. W. (2017). Action recognition in video sequences using deep bi-directional LSTM with CNN features. *IEEE Access: Practical Innovations, Open Solutions*, *6*, 1155–1166. doi:10.1109/ACCESS.2017.2778011

Wang, F., Xuan, Z., Zhen, Z., Li, K., Wang, T., & Shi, M. (2020a). A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. *Energy Conversion and Management*, *212*, 112766. doi:10.1016/j.enconman.2020.112766

Wang, X., Yang, L. T., Song, L., Wang, H., Ren, L., & Deen, M. J. (2020b). A tensor-based multiattributes visual feature recognition method for industrial intelligence. *IEEE Transactions on Industrial Informatics*, *17*(3), 2231–2241. doi:10.1109/TII.2020.2999901

Wu, B., Fang, H., Jacoby, G., Li, G., & Wu, Z. (2021). Environmental regulations and innovation for sustainability? Moderating effect of political connections. *Emerging Markets Review*, *100835*, 100835. Advance online publication. doi:10.1016/j.ememar.2021.100835

Wu, Y., Yuan, M., Dong, S., Lin, L., & Liu, Y. (2018). Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing, 275*, 167-179. 10.1016/j.neucom.2017.05.063

Yang, Y., Liu, Y., Lv, X., Ai, J., & Li, Y. (2021). Anthropomorphism and customers' willingness to use artificial intelligence service agents. *Journal of Hospitality Marketing & Management*, 1–23. doi:10.1080/19368623.2021.1926037

Ye, Z., Qian, L., & Na, Z. (2018). Financial development, urbanization, and urban-rural income disparity: Evidence based on Chinese provincial data. *Modern Economy*, *9*(1), 31–60. doi:10.4236/me.2018.91003

Zheng, Y., Han, W., & Yang, R. (2021). Does government behaviour or enterprise investment improve regional innovation performance?-evidence from China. *International Journal of Technology Management*, *85*(2-4), 274–296. doi:10.1504/IJTM.2021.115266

Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., & Telenti, A. (2019). A primer on deep learning in genomics. *Nature Genetics*, *51*(1), 12–18. doi:10.1038/s41588-018-0295-5 PMID:30478442

Kunzhi Tang was born in Qinhuangdao, Hebei. P.R. China, in 1997. He received the bachelor's degree from China University of Petroleum, Beijing, P.R. China. Now, he studies at The Australian National University, Au. His research interest includes data science, machine learning and artificial intelligence.

Yue Zhou is the corresponding author, email: vvzyzy@163.com.