Article

Credit Risk Prediction Model for Listed Companies Based on CNN-LSTM and Attention Mechanism

Jingyuan Li [1], Caosen Xu [1,*], Bing Feng [1] and Hanyu Zhao [2]

[1] School of Management, Wuhan Institute of Technology, Wuhan 430205, China
[2] Beijing Academy of Artificial Intelligence, Beijing 100084, China
* Correspondence: 12006010002@stu.wit.edu.cn

Abstract: The financial market has been developing rapidly in recent years, and the issue of credit risk concerning listed companies has become increasingly prominent. Therefore, predicting the credit risk of listed companies is an urgent concern for banks, regulators and investors. The commonly used models are the Z-score, Logit (logistic regression model), the kernel-based virtual machine (KVM) and neural network models. However, the results achieved could be more satisfactory. This paper proposes a credit-risk-prediction model for listed companies based on a CNN-LSTM and an attention mechanism, Our approach is based on the benefits of the long short-term memory network (LSTM) model for long-term time-series prediction combined with a convolutional neural network (CNN) model. Furthermore, the advantages of being integrated into a CNN-LSTM model include reducing the complexity of the data, improving the calculation speed and training speed of the model and solving the possible lack of historical data in the long-term sequence prediction of the LSTM model, resulting in prediction accuracy. To reduce problems, we introduced an attention mechanism to assign weights independently and optimize the model. The results show that our model has distinct advantages compared with other CNNs, LSTMs, CNN-LSTMs and other models. The research on the credit-risk prediction of the listing formula has significant meaning.

Keywords: CNN; attentional mechanisms; LSTM; credit risk

1. Introduction

With the continuous development of the financial market, the number of listed companies is increasing, which is an essential thrust of the financial market. Still, opportunities and risks coexist, some listed companies have credit-risk problems, and there are many debt and loan defaults. Therefore, it becomes necessary to predict the credit risk of listed companies and to minimize the impact of credit risk in listed companies to an extent. In the field of financial risk management, the analysis of the credit risk of an enterprise has always been a hot issue [1].

At present, the research on credit-risk prediction for listed companies is insufficient, there is a lack of a default database, and the level of credit-risk prediction needs to be improved. With the economic and financial globalization trend, some advanced international experiences and models for managing credit risk have been gradually developed and widely used [2]. These risk-management experiences play an essential role in improving banks' credit risk profile, enhancing the financial system's competitiveness and ensuring the financial market's stable and prosperous development [3].

The traditional methods of credit-risk prediction for listed companies are the Z-score model and Logit model; the Z-score model is based on multivariate statistical methods, and the Logit model has dichotomous explanatory variables (i.e., default and non-default), uses the coefficients of model variables obtained by estimating the sample and then determines whether the company has a high credit risk by the calculated *p*-value. The KMV model is the most popular model among credit-risk-prediction models.

The default distance of the KMV model can represent the magnitude of credit risk and can be used to approximate the probability of default of a firm using mathematical inference. In addition, the continuous development of artificial intelligence technology and big data has led more scholars to use machine-learning methods in assessing credit risk [4], including GBDT, support vector machines, XGboost, MLP neural networks, LSTM and integrated learning models. The explanatory variables of the model are dichotomous, which may not be able to truly reflect the relationship between default and the risk of listed companies. KMV requires a large amount of historical default data for processing [5]. The neural network can effectively solve the nonlinear problem in predicting credit risk; however, the operation is complicated and requires a large number of samples.

First, our method is based on the importanceof the credit-risk prediction for listed companies to the financial market to stabilize the financial market and make it healthy and orderly. The development of our proposed research is based on the advantages of the LSTM model in long-term time-series prediction as the basis of the model and then combined with the benefits of the CNN network in feature extraction to reduce the amount of computation and parameters of the LSTM model and improve the model performance.

Finally, to solve the problem of missing historical data that may occur in the long-term time-series prediction of an LSTM model, the attention mechanism is introduced, the calculation weight is reasonably allocated through independent learning, the model is optimized, and the prediction accuracy and operation of the CNN-LSTM model are improved. Finally, we propose the CNN-LSTM-AM model to solve the prediction of the credit risk of listed companies [6].

This paper proposes a credit-risk-prediction model for listed companies based on CNN-LSTM and an attention mechanism. The CNN is first applied to convolve the input data to enhance the correlation between the input and output; then, the LSTM network is used to predict the time-series data; the attention mechanism is added to the LSTM output; and finally, the trained network is used to indicate the credit risk of the listed companies. The experimental results show that the proposed method can effectively improve the prediction accuracy.

The contribution points of this paper are as follows.

- Compared with the traditional Z-score and Logit models, the improved CNN-LSTM model used in this paper has a more vital information selection ability and time-series data-learning ability and can make accurate predictions for time-series data. The attention module can automatically judge and learn the importance of different features of credit indicators of listed companies and the derived importance relationship to assign weights, which significantly improves the prediction ability of the LSTM model for long input series and effectively improves the prediction ability for the credit risk of listed companies.
- The model proposed in this paper can effectively solve the nonlinear problem of predicting credit risk, has more applicability than the Z-score, Logit and KMV models and does not require many samples compared with the latest neural network model.
- It can genuinely reflect the relationship between the default and credit risk of listed companies, which makes commercial banks and investors better able to make reasonable and timely responses to the credit-risk problems of companies.

In the rest of this paper, we present recent related work in Section 2. Section 3 offers our proposed methods: an overview, convolutional neural networks, attentional mechanisms and long short-term memory networks. Section 4 presents the experimental part, details and comparative experiments. Section 5 concludes this work.

2. Related Work

2.1. Logit Model and the Z-Score

The logistic model is a statistical method of nonlinear classification—an extension of the ordinary multiple linear regression model. The logistic model uses maximum likelihood parameter estimation and does not require the data sample to be normally distributed [7].

It solves the problem of discontinuous regression of the dependent variable, which is one of the model's highlights, especially when the dependent variable is categorical [8]. As the logistic model uses a logistic probability distribution function, the traditional methods for forecasting the credit risk of listed companies focus on using the historical financial data of listed companies [9]; however, these models can only predict the future from the past, which is a significant drawback of the model.

Domestic scholars have also analyzed and applied the Z-score and Logit models. Qiu Yunlai empirically tested the Z-score model based on data from 46 listed companies in China. The results show that the prediction accuracy of the Z-score model was higher after the introduction of net cash flow [10].

He Zhanxiong and Tang Xiangjin argued that, to make the Z-score model better predict [11] the market value of equity, the Z-score model should be improved to better predict the credit risk of listed companies in China. Then, 20 listed companies were selected for the Z-score model. The results showed that ST stocks had higher credit risk than blue chip stocks.

Beaver created the earliest univariate forecasting model, which he believed could determine a company's financial condition by using only a single variable [12] regarding the financial condition of a company. The Z-score model is a model that distinguishes between insolvent and non-insolvent firms by assigning different weights to several financial ratios that reflect the firm's financial position [13]. The model assigns different weights to some financial ratios, which can reflect the company's financial status, and then calculates the total risk value of each listed company, i.e., the Z value, after weighting [14]. By comparing this value with the critical value, we can determine the degree of the financial crisis of listed companies [15].

2.2. KMV Model

On the stock market using B-S stock option pricing to derive the probability of future defaults of listed companies, the investors' knowledge of the stock's future is included in the stock market, and thus the model is sensitive and forward-looking, and the prediction is more accurate and objective than the Logit model [16]. The KMV model was developed by Kealhofer, McQuown and Vasick, a company specializing in credit risk analysis [17]. It treats the company's equity as a European call option, and when the market value of the company's assets is higher than the debt at maturity, the debt is repaid; if the market value of the company's assets is less than its debt, the company chooses to default [18].

2.3. ANN

An artificial neural network (ANN) is an abstraction of the neuronal network of the human brain from the perspective of information processing to build a simple model according to different connections to form various networks; it has been a hot research topic in the field of artificial intelligence since the 1980 [19]. A neural network is a model of operations consisting of interconnections between neurons—also known as nodes [4]. With continuous in-depth research on artificial neural networks, they have been used to successfully solve practical problems that are difficult for modern computers in many fields, such as intelligent robotics, automatic control, economics, biology and medicine [20]. Artificial neural networks now have a wide range of applications; in financial markets, they can predict the market price and assess the risk.

Market price forecasting: The movement of market commodity prices is a comprehensive analysis of the various factors affecting supply and demand in the market [21]. Traditional methods of statistical economics have difficulty in making reasonable forecasts of market price changes [22]. At the same time, artificial neural networks can handle incomplete relevant data with obscure regularity or vague uncertainty; therefore, artificial neural networks have a massive advantage over traditional methods in making price forecasts [23].

Taking the market price determination mechanism as the starting point, the corresponding neural network model is built based on the following factors: the loan interest rate, per capita disposable income and urbanization level [24]. It is possible to make a reasonable and scientific forecast of the price trend of commodities.

Risk assessment: The application of an artificial neural network for risk prediction can be used to construct a credit-risk model suitable for the actual specific situation according to the particular accurate risk sources, obtain the risk evaluation coefficient [25] and then take reasonable measures to cope with the possible risks through comparative analysis, which is of great significance to the stability of the financial industry [26].

3. Methodology

3.1. Overview of Our Network

The long short-term memory network (LSTM) is an excellent variant of RNN, which can solve the gradient explosion problem in long-term sequence prediction. It is now commonly used in time-series prediction. This article explains the projection of listed companies' credit risk and focuses on their credit indicators. Then, it performs risk prediction so that the LSTM model can be used for prediction.

In addition, to improve the training time and speed of the LSTM model and reduce the number of parameters, we introduce a convolutional neural network (CNN); before the data enters the LSTM model, we use the CNN for feature extraction, select indicators that are more relevant to the company's credit-risk prediction, reduce the complexity of the data, combine the advantages of the LSTM model and the CNN model to form a CNN-LSTM model and, finally, introduce the attention mechanism. The attention mechanism can learn independently, select more relevant feature vectors, optimize the model and improve the prediction accuracy of the CNN-LSTM model—finally, combined into a CNN-LSTM-AM model.

This paper uses a CNN-LSTM-AM model for credit-risk prediction for listed companies. The model structure is shown in Figure 1. One of the features of CNN is that it can extract local key features for feature processing. LSTM can use previous historical information to deal with sequence problems, which has clear advantages over RNN in long sequence prediction. Therefore, a new CNN-LSTM model is formed by combining CNN and LSTM. To improve the accuracy and efficiency of this model, the attention mechanism is further introduced into the model.
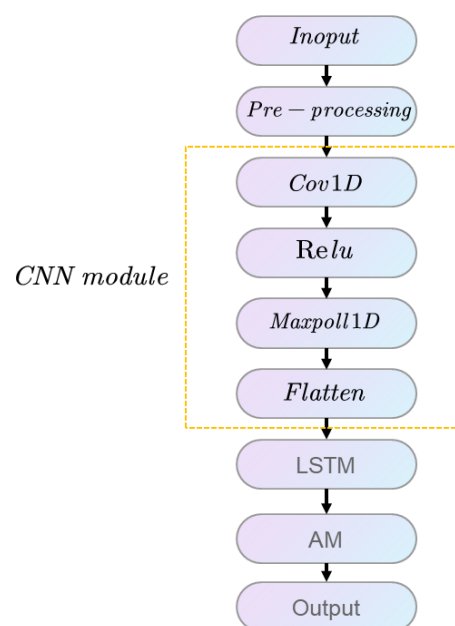

Figure 1. Overall flowchart of CNN-LSTM-AM model.

The attention model (AM) has been widely used in deep learning tasks in recent years [27]. When a large amount of information is input to a neural network, the impacts of different knowledge of the input on the output are different [28], and allocating more computational power to important input information can improve the computational speed and efficiency of the neural network. Thus, we introduce the attention mechanism.

The structure is shown in Figure 2. The computation of the attention mechanism can be divided into two steps. One is calculating the corresponding weight coefficients based on the input information.

The second step is calculating the input information's weighted average based on the weight coefficients.

$$a_i = \frac{\exp[s(x_i, q)]}{\sum_{i=1}^{t} \exp[s(x_i, q)]} \tag{1}$$

$$\varepsilon = \sum_{i=1}^{t} a_i \times x_i \tag{2}$$

where q is the query vector and x_i is the input vector; the degree of correlation with q is the basis for the score of the $s(x)$ score function; a_i is the attention distribution; and ε is the weighted average of the input values and the attention distribution.

For ease of reading, Table 1 is a summary table of symbols mentioned in the text.

Table 1. Summary table of the symbols used in the text.

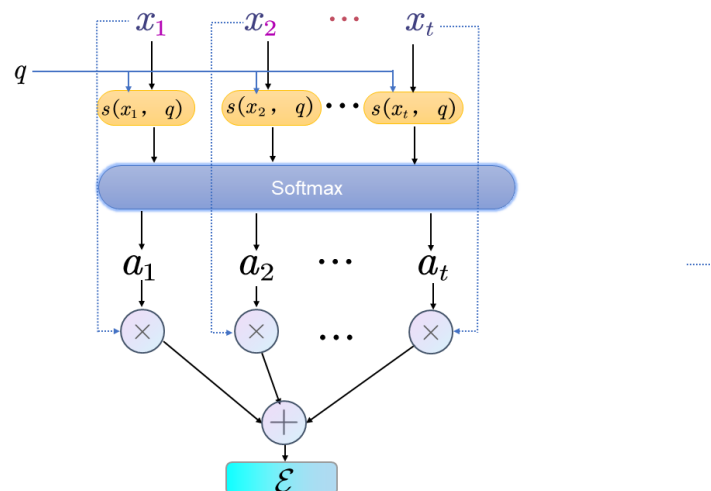| Symbols	Meaning
q	the query vector
X_i	the input vector
a_i	the attention distribution
E	the activation function
M_j	the input feature map
K	the convolution kernel corresponding to the feature
b_j	the bias unit of the layer
$down(x)$	downsampling
σ	the sigmoid function
W	the weight of the neuron
b	the deviation of the neuron

Figure 2. The operation flow chart of the attention mechanism.

3.2. CNN Model

A convolutional neural network (CNN) is a deep neural network that includes convolutional processing. Its unique structure can reduce the risk of overfitting and also reduce the memory occupied by the deep network, which is used to identify the similarity between new features and the original data and can extract features.

CNN mainly consists of convolutions, pooling and full connections (FC). Convolutions and pooling are responsible for feature extraction, while FNN is used for classification recognition [29]. The structure is shown in Figure 3.
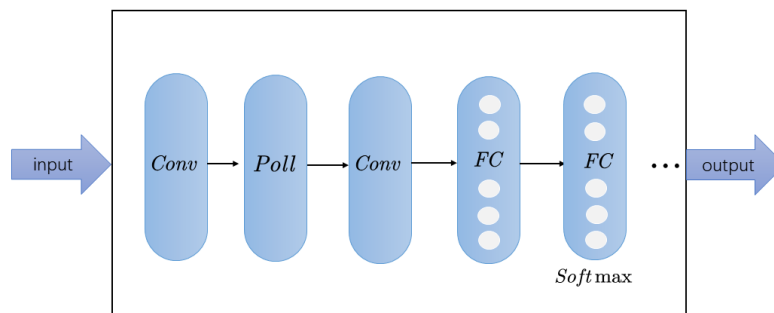

Figure 3. The flow chart of the one-dimensional CNN (convolutional neural network).

In the CNN model, the convolutional layer is the critical layer of the CNN, convolutional kernels can extract the features as feature extractors, and then different weights are set for convolutional computation. Usually, more than one convolutional kernel can be developed to more fully extract the features.

Then, to reduce the complexity and parameters of the model and reduce most of the computation, the pooling operation is used for downsampling—the method usually used is to select the max-pooling layer (max-pooling)—and finally, in the fully-connected layer (FC) is used to weight the sum of the computation results obtained earlier, resulting in the final recognition result. The unique structure of the CNN in processing data reduces the computational effort and makes the network simple and efficient [30].

The convolution layer is calculated as follows.

$$X_j^L = f\left(\sum_{i \in M_j} X_i^{L-1} * K_j^L + b_j^L\right) \tag{3}$$

where X_j^L, X_j^{L-1} is the feature map of the *L*th and *L*-1th layers, respectively; M_j is the input feature map; K is the convolution kernel corresponding to the feature; and b_j^L is the bias unit of the layer.

$$X_j^L = f\left[\beta_j^L \operatorname{down}\left(X_i^{L-1}\right) + b_j^L\right] \tag{4}$$

where $\operatorname{down}(x)$ is downsampling, and β_i^L, b_i^L is the additive and multiplicative bias of the *L*th layer.

3.3. LSTM Model

Long short-term memory (LSTM) is an extended model based on recurrent neural networks (RNN) [31]. Still, the traditional RNN model cannot deal with the long-distance dependence problem well and will produce issues, such as gradient disappearance and gradient explosion. To solve this problem, Hochreiter and Schmidhuber proposed the LSTM model in 1997, which can alleviate these problems to an extent. The gates work together. The forgetting gate reduces the amount of memory for the input information, and the output gate is responsible for updating the memory state and outputs the computational results. The structure is shown in Figure 4.
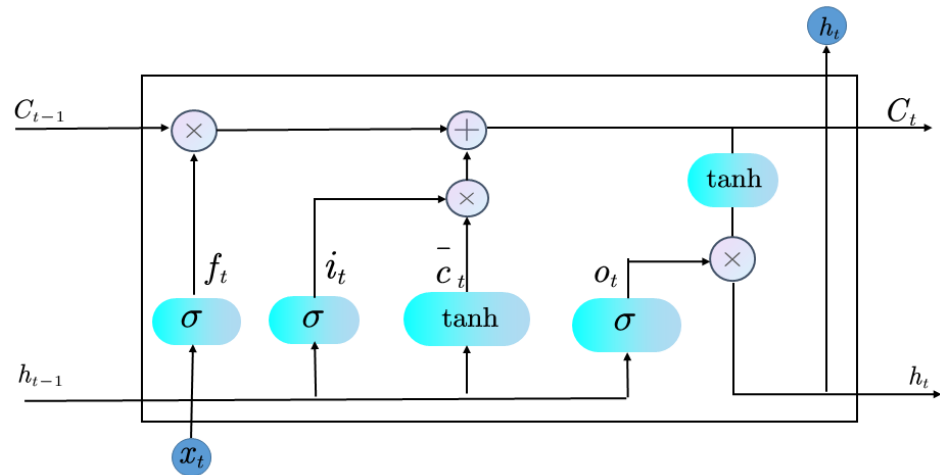
Figure 4. The operation flow chart of a long short-term memory network (LSTM).

In Figure 4, f_t is the forgetting gate; i_t is the input gate; o_t is the output gate; X_t is the input at the current moment; C_{t-1}, h_{t-1} is the output at the previous moment and the cell state; and $C_{t,t}$ is the output and cell status at the current moment. The LSTM uses a unique gate mechanism to control the forgetting gates, input gates, output gates and cell states. The unit state is used to handle the long-term dependencies of the memory units. The following equation can express the relationship between the four.

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \tag{5}$$

$$f_t = \sigma\left(W_f h_{t-1} + U_f x_t + b_f\right) \tag{6}$$

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \tag{7}$$

$$\tilde{C}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c) \tag{8}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{9}$$

$$h_t = o_t \times \tanh(C_t) \tag{10}$$

In this equation, σ is the sigmoid function, W is the weight of the neuron, and b is the deviation of the neuron.

The sigmoid function decides what information is to be removed from the cell state, and this is performed by the sigmoid gate, which is called the forgetting gate [32]. LSTM is a tremendous development for us to use RNN with strong information selection ability and time series data-learning ability, which can greatly improve the accuracy of time series data through the unique gate mechanism [33].

4. Experiment

4.1. Datasets

Our experimental data are mainly from the stock exchange of China, CSMAR, MorningStar database, KMV default database and China Stock Market and Accounting Research Database. CSMAR is a research-oriented and accurate database in the field of economics and finance, which was developed from the needs of academic research and the professional standards of famous international databases, such as CRSP, Compustat, TAQ and

Thomson of the University of Chicago and combined with the actual national conditions of China.

After more than 20 years of continuous improvement and accumulation, the CSMAR database has covered the green economy, stocks, companies, China securities, futures, foreign exchanges, macro finance, industry and other significant areas of economic and financial research-oriented databases, including 200+ databases, 4000+ tables and 60,000+ fields, which can be time, code and other indicators data for the charts in the database; can be exported to Excel with CSV (comma separated values) and other data formats. The stock exchange of China is an essential tool for investment and empirical research.

The MorningStar database is designed to provide investors with professional financial information, fund and stock analysis, ratings and convenient, practical and functional analysis application tools, such as Rating, Investment Style Box and Category Rating [34]. These tools are professional and easy to use and can help investors make informed investment decisions. The archived data is extracted from the Morningstar database, is easily accessible and reproducible and provides a quick way to extract financial data. There are approximately 500,000 investment products with raw data in the Morningstar database. The Morningstar database is now widely used in the literature across multiple disciplines. This paper uses all available data from the database of public companies.

To forecast the credit risk of listed companies, we propose a benchmark for assessing the credit risk of SCF China SMEs through 18 indices [35], which are also the initial independent variables of the CNN-LSTM-AM model. The 18 independent variables are classified into five categories: liquidity, leverage, profitability, activity and non-financial. Among them, the dependent variable is the credit risk status of the listed companies as risky or risk-free. The dependent variable is assigned a value of 1 when the data sample of listed companies shows non-risky but a value of 0 when the data sample release shows a negative signal.

For these 18 independent variables, we define and describe them as shown in Table 2.

Table 2. Credit-risk factors and the classification of listed companies.

Factors	Code	Variable	Categories
Applicant factors	X_1	Current ratio	Liquidity
Applicant factors	X_2	Quick ratio	Liquidity
Applicant factors	X_3	Cash ratio	Liquidity
Applicant factors	X_4	Working capital turnover	Liquidity
Applicant factors	X_5	Return on equity	Leverage
Applicant factors	X_6	Profit margin on sales	Profitability
Applicant factors	X_7	Rate of Return on Total Assets	Leverage
Applicant factors	X_8	Total Assets Growth Rate	Activity
Counter party factors	X_9	Credit rating	Non-finance
Counter party factors	X_{10}	Quick ratio	Liquidity
Counter party factors	x_{11}	Turnover of total capital	Liquidity
Counter party factors	X_{12}	Profit margin on sales	Profitability
ltems' characteristics factors	X_{13}	Price rigidity, liquidation	Non-finance
ltems' characteristics factors	X_{14}	Account receivable collection period	Leverage
ltems' characteristics factors	X_{15}	Accounts receivable turnover ratio	Leverage
Operation condition factors	X_{16}	Industry trends	Non-finance
Operation condition factors	X_{17}	Transaction time and transaction	Non- finance
Operation condition factors	X_{18}	frequency Credit rating of SME	Non-finance

4.2. Experimental Details

First, we used the CNN-LSTM-AM model with the dataset to calculate the results and then compare the speed with the traditional Logistic and KMV models. Then, the processing speed of complex cases is reached. At this point, the models compared are Tree [36] and SVM models [37]. Then, the same results are computed through our model to calculate the degree of accuracy and compare it with other models. We also test the

accuracy of the CNN-LSTM-AM model in different datasets and compare it with other traditional statistical models, KVM models and new neural-network-based models.

We also study the computational effort of the various models, created pictures and then derive accuracy and error charts for the other models to compare the predictive power of our models more visually. Finally, we research the AUC of different models. Furthermore, we test the predictive power of our models in other data sets to thoroughly compare the strengths and weaknesses of our models.

4.3. Experimental Results and Analysis

As shown in Figure 5, we conducted comparison experiments to compare the inference speed of the CNN-LSTM-AM model with the traditional logistic and KMV models, and the results show that our method exhibits faster inference speed in different datasets, and the advantage of the CNN-LSTM-AM model becomes more apparent when the dataset increases.

This is because, when the number of inference cases is small, most of the time overhead is spent on data loading, while as the number of inference cases increases, most of the time overhead is spent on inference. It can be seen that the inference speed of the CNN-LSTM-AM model is significantly due to the other models. In contrast, with the increasing number of inference cases, the equipment appears to heat up, and the time cost rises accordingly. Therefore, our model is more economical than traditional logistic and KMV models.
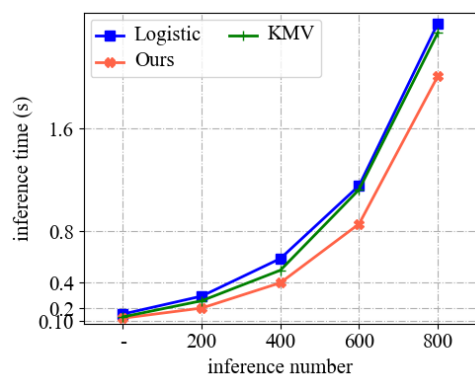

Figure 5. The three models in the case of different reasoning quantities.

In Figure 6, we compare the speed of the different models tested in complex cases. In finance, it is most important to deal with some problematic case situations, and, in corporate credit evaluation, these data are critical. It can be seen from the figure that, in dealing with complex cases, our method has clear advantages compared to the new neural network model, which is an essential inspiration for future credit evaluation models.
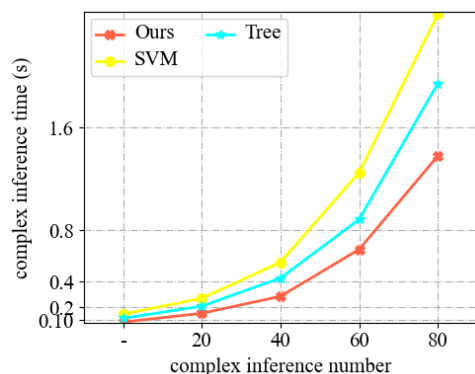

Figure 6. A line chart of the inference speed of three other models in the case of complex and different inference quantities.

We verified the accuracy of different models for predicting the credit risk of listed companies and then presented it graphically to make it more intuitive. In Figure 7, the shades of color are responsible for different degrees with different models' heat maps, and higher values indicate higher correctness. Figure 7 shows that the CNN-LSTM-AM model had higher accuracy than the KMV and SMV models [38].
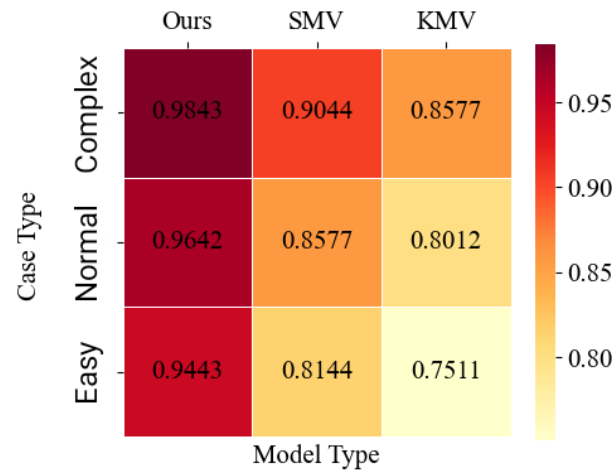

Figure 7. In the same data set, the prediction accuracy of three different models contrasted with data of varying complexity.

In Figure 8, to verify the applicability of our model, we demonstrate the generalization of different models on different datasets.These data sets are the stock exchange of China, CSMAR, MorningStar database and KMV default database. The performance of the CNN-LSTM-AM model in the four data sets is better than other models, showing a wide range of applicability. This makes up for the problem of KVM in dealing with nonlinear data sets and has made significant improvements and breakthroughs.
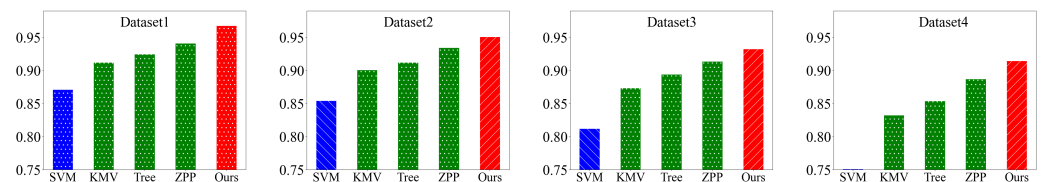

Figure 8. Accuracy in different datasets.

In what follows, the training process Algorithm 1 of this paper is described in detail. Compared with other neural-network models, our training process has removed many weak phase factors in the CNN model. Therefore, compared with Tree, SVM, the training process of the neural network model is more straightforward, and so the iteration speed is faster, and the calculation time and accuracy are improved.

As can be seen from Figure 9, our model has clear advantages in the amount of calculation, mainly because we first perform convolution processing through CNN, which can extract feature vectors well and simplify indicators. Then, we introduce the attention mechanism; the attention mechanism can match the weights more reasonably and effectively so that the deep learning network can focus more energy on more efficient operations [39].

Algorithm 1: Algorithmic representation of the training process in this paper.

Training set $D = \{(x(n), y(n))\}N_N = 1$, validation set V, learning rate φ, regularization coefficient λ, number of network layers L, number of neurons $M_1, I < I < L$.
Random initialization W,b

while *The error rate of the neurat network model on the vatidation set vs. no longer decreases* **do**
    Randomly reorder the samples in the training set D
    for $n = 1...N$ **do**
        Select samples from the training set $(x(n),y(n))$ Feed-forward calculation of net inputs (l) and activations $a(l)$ for each layer until the last layer Backpropagation calculates the error $\delta(l)$ of each layer Calculate the derivative of each layer parameter Update
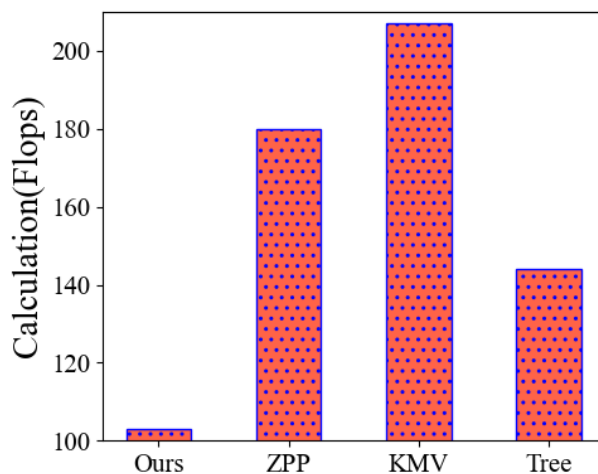   **end**
end

Figure 9. Comparison of the computational size of different models.

In Figure 10, to verify the prediction accuracy of different models in the CSMAR dataset, we divided the CSMAR dataset into five groups for experiments. As seen in the figure, our method is always more accurate than other models in the five grouping experiments, and all of them have shown high predictive ability. It effectively predicts the credit risk of listed companies in the CSMAR data set and has a more significant advantage than other models.
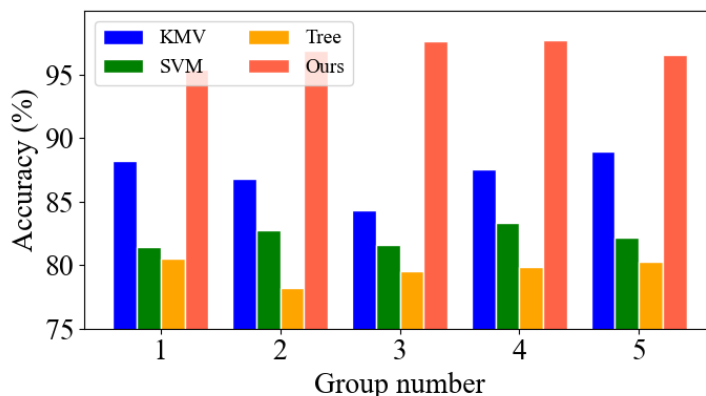

Figure 10. Comparison of prediction accuracy of different models.

In Figure 11, we conduct experiments on the CSMAR dataset, comparing our model with the ZPP model for AUC. As the number of training continues to increase, the precision of the model accuracy also increases, and our model's prediction accuracy and increase are always higher than that of the ZPP model [40]. Compared with the ZPP model, the learning ability of the CNN-LSTM-AM model is more vital than that of the ZPP model.
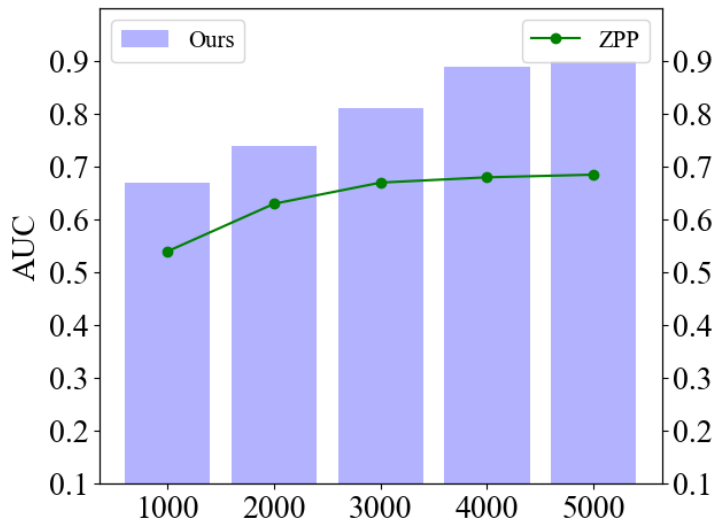

Figure 11. Comparison of the AUC of different models.

Errors in the forecasting model are essential to validating our public company risk forecasting model. In Table 3, we selected the data of the CSMAR and MorningStar databases to verify the prediction errors of different models. It can be seen from the table that the prediction errors of our model in the two data sets are smaller than those of other models.

Table 3. Error comparison of different data sets and models.

Model	e_1	e_2
Ours	2.38	3.34
Kmv	2.58	3.65
Svm	2.81	2.94
Tress	2.41	3.5

In Table 4, we first simplify the operation process by CNN convolutional processing and use the CNN model to find the six indicators with the most significant correlation with the credit risk of listed companies. Then, through the attention mechanism model, we perform deep learning and assign the weight ratio by ourselves to improve the machine-learning efficiency, which makes the LSTM operation process more accessible, and operations target more critical influencing factors, thus, improving the operation efficiency. Finally, by predicting these six indicators, we compare the prediction results of different models for the feature parameters. The results show that our model has a significant advantage in the prediction of these six indicators.

Table 4. Prediction accuracy of different metrics.

Index	Ours	SVM	KMV
Current ratio	0.5653	05023	0.4613
Quick ratio	0.4904	0.4756	0.4653
Cash ratio	0.4545	0.5864	0.5656
Credit rating	0.8623	0.8523	0.8321
Quick ratio	09864	0.9654	0.9451
Industry trends	0.8746	0.8586	0.8321

As shown in Figure 12, we compared the parameters of AM [41], CNN-LSTM, LSTM and our model. The results show that the parameter cost of our model is significantly smaller than other models, which benefits from the CNN network and attention mechanism on LSTM model optimization.
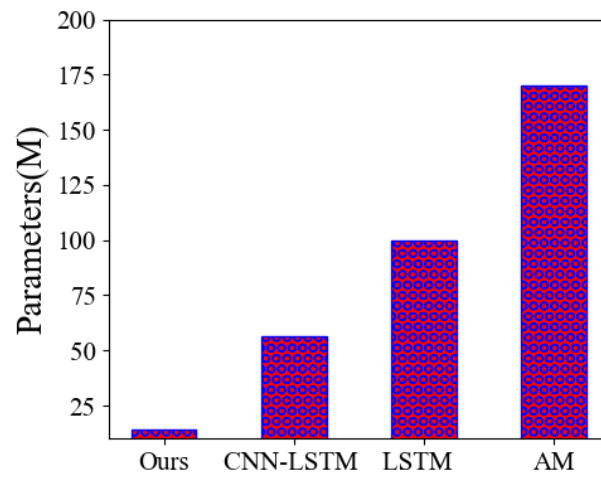

Figure 12. Comparison of the parameter quantities of different models.

As shown in Figure 13, we compared the computational load of different models to test our model's performance further. The results show the reasonable distribution of weight and computational load in the feature processing and attention mechanism of the CNN network of our model, compared with the other CNN [42], LSTM [43] and CNN-LSTM [44].
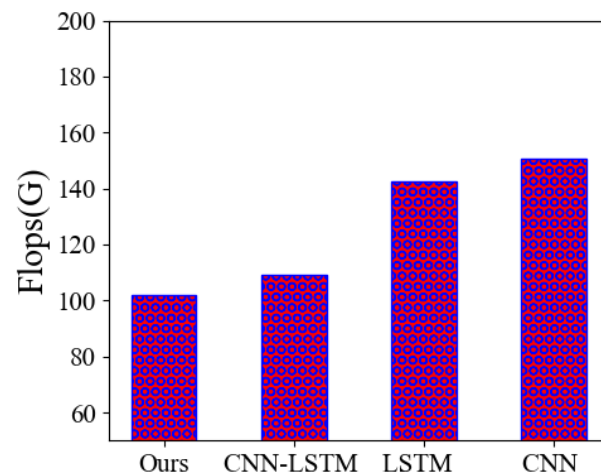

Figure 13. Comparing the amount of computation of different models.

In Figure 14, we compare the training time of the CNN, CNN-LSTM, LSTM and our model, which is also one of the essential indicators to measure the model's performance. The results show that, in different amounts of data processed, the training time of our model time, is better than for the other models.
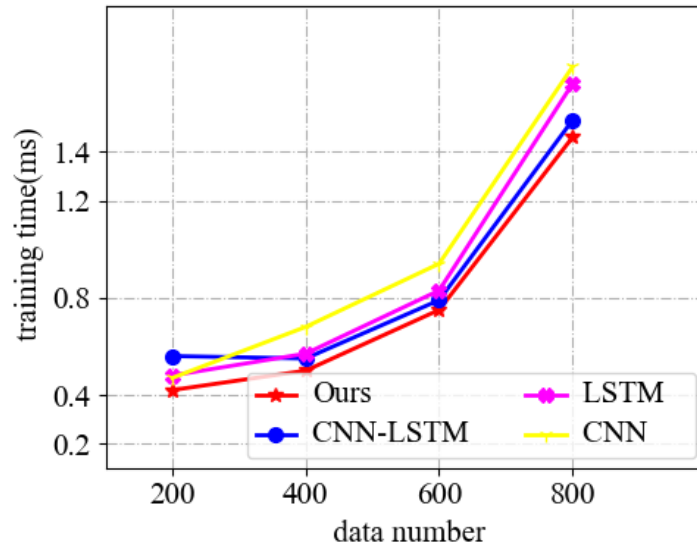

Figure 14. Comparison of the training times of different models.

In Figure 15, to further test the performance of our model, we compared the inference time of different models in complex data. The results show that our model has clear advantages in the performance of more complex data sets [45].
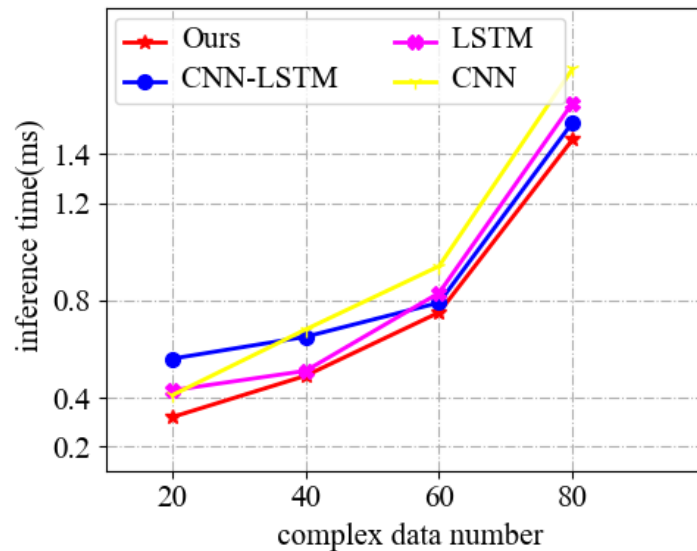

Figure 15. Inference time comparison of different models.

In Figure 16, to verify the generalization of our model, we compared the prediction accuracy of different models in four different data sets. These are the stock exchange of China, CSMAR, MorningStar database and KMV default database. The results show that, whether in linear or nonlinear data sets, in the prediction of our model, the accuracy is higher than in other models, showing good generalization [46].
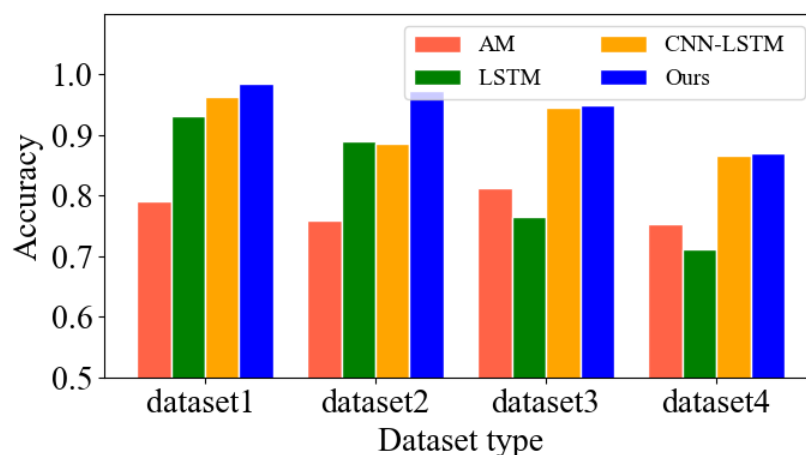
Figure 16. Prediction accuracy of different models in different datasets.

Table 5 compares the accuracy, computation and parameter size of the models mentioned in the paper with our model. The table shows that our model has significant advantages in these aspects.

Table 5. A comparison of different models.

Model	Accuracy	Flops	Parameters (M)
Logistic [4]	0.7762	212	27.03
Tree [35]	0.914	140	140.47
KMV [38]	0.8577	205	11.69
ZPP [40]	0.8454	180	15.79
AM [41]	0.790	125.77	169.99
CNN [42]	0.897	150.66	177.17
LSTM [43]	0.931	142.43	99.86
CNN-LSTM [44]	0.964	109	56.44
SMV [47]	0.9044	113.4	122.86
Ours	**0.9843**	**102**	**14.14**

5. Conclusions and Discussion

First, a CNN model was used for feature processing to obtain more useful independent variables among the 18 independent variables, which improved the operation speed and accuracy of the model. Then, the output results entered the LSTM model, which can perform long-term sequence prediction, learn independently without a large number of samples and positively affect the credit-risk prediction for listed companies. Finally, the attention mechanism was introduced in the CNN-LSTM model, and several listed companies were locked through independent weight ratios. Several variables were most closely related to credit risk to enhance the authenticity of the forecast.

However, our method currently has certain limitations. First, to improve the accuracy of prediction, we used the CNN-LSTM-AM model, which is more complicated and has more processes than the KMV and Logit models; thus, our model can deal with the same situation at a slower speed. In the future, we can simplify the model and make the calculation speed faster. Secondly, we used data from four datasets. Collecting these data requires a large workload; however, the verification results are promising. After the verification is completed, more data can be used to predict the credit risk of listed companies.

This paper adopted a CNN-LSTM model based on the attention mechanism, which has improved the accuracy of credit-risk prediction for listed companies compared with traditional models and is a good application of deep machine learning in the financial neighborhood. The CNN-LSTM-AM model proposed in this paper has a wide range of credit-risk prediction, applies to both linear and nonlinear data sets and does not require a large number of historical default records as the branching point, which makes it easier to apply the scenario.

Its prediction accuracy has dramatically improved compared to traditional and new machine-learning models. The error is small, the application range is wide, the operation speed is fast, the AUC value is high, and it can better handle more complex data sets, which can predict the default risk of listed companies well and effectively in advance. The comparative study of credit-risk-prediction models of listed companies has important practical significance, which is reflected in the following points. We maintain a relatively stable investment environment for the Chinese stock market by focusing on the operations of these listed companies by the regulators.

The comparative study of different credit-risk-prediction models can make the regulators more timely and effective in finding problems that may exist in the operating conditions of these listed companies in time to maintain the excellent and healthy development of the stock market. Our work can provide investors with relatively objective investment advice, which can reduce the influence of artificial factors, such as expert opinions, so that investors can avoid misunderstandings and obtain fairly accurate investment advice.

Since most of the listed companies' sources of liabilities are related to commercial banks, whether the listed companies can repay on time and have high credit risk is also directly related to the interests of the commercial banks themselves, and thus the commercial banks themselves attach great importance to the credit risk of listed companies. Therefore, the comparative study of credit-risk-prediction models by merchants also has great significance for commercial banks.

The study can help commercial banks find the appropriate credit-risk-prediction model to reduce the economic loss of commercial banks to a certain extent. The comparative analysis of credit-risk-prediction models of listed companies can help predict the risk of default or non-default of listed companies in advance to reduce the economic loss of commercial banks to a certain extent, which is of great significance to the stability and prosperity of the financial market.

Author Contributions: Conceptualization, J.L. and C.X.; methodology, B.F. and H.Z.; software, J.L., C.X. and B.F.; validation, C.X., B.F. and H.Z.; formal analysis, C.X., B.F. and H.Z.; investigation, C.X., B.F. and H.Z.; data curation, C.X., B.F. and H.Z.; writing—original draft preparation, J.L., C.X. and B.F.; writing—review and editing, J.L. and B.F.; visualization, J.L., C.X. and B.F.; supervision, J.L.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

References

1. Rostami, M.; Berahmand, K.; Nasiri, E.; Forouzandeh, S. Review of swarm intelligence-based feature selection methods. *Eng. Appl. Artif. Intell.* **2021**, *100*, 104210. [CrossRef]
2. Ning, X.; Xu, S.; Nan, F.; Zeng, Q.; Wang, C.; Cai, W.; Li, W.; Jiang, Y. Face editing based on facial recognition features. *IEEE Trans. Cogn. Dev. Syst.* **2022**. Available online: https://ieeexplore.ieee.org/document/9795907 (accessed on 22 March 2023). [CrossRef]
3. Yoosefdoost, I.; Basirifard, M.; Álvarez-García, J. Reservoir Operation Management with New Multi-Objective (MOEPO) and Metaheuristic (EPO) Algorithms. *Water* **2022**, *14*, 2329. [CrossRef]
4. Zhu, Y.; Xie, C.; Wang, G.J.; Yan, X.G. Predicting China's SME credit risk in supply chain finance based on machine learning methods. *Entropy* **2016**, *18*, 195. [CrossRef]
5. Ning, X.; Tian, W.; Yu, Z.; Li, W.; Bai, X.; Wang, Y. HCFNN: High-order coverage function neural network for image classification. *Pattern Recognit.* **2022**, *131*, 108873. [CrossRef]

6. Cai, W.; Ning, X.; Zhou, G.; Bai, X.; Jiang, Y.; Li, W.; Qian, P. A Novel Hyperspectral Image Classification Model Using Bole Convolution with Three-Directions Attention Mechanism: Small sample and Unbalanced Learning. *IEEE Trans. Geosci. Remote. Sens.* **2022**, *61*, 5500917. [CrossRef]

7. Manab, N.A.; Theng, N.Y.; Md-Rus, R. The determinants of credit risk in Malaysia. *Procedia-Soc. Behav. Sci.* **2015**, *172*, 301–308. [CrossRef]

8. Ning, X.; Tian, W.; He, F.; Bai, X.; Sun, L.; Li, W. Hyper-sausage coverage function neuron model and learning algorithm for image classification. *Pattern Recognit.* **2023**, *136*, 109216. [CrossRef]

9. Chen, Z.; Huang, J.; Ahn, H.; Ning, X. Costly features classification using monte carlo tree search. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–8.

10. Zhibin, Z.; Liping, S.; Xuan, C. Labeled box-particle CPHD filter for multiple extended targets tracking. *J. Syst. Eng. Electron.* **2019**, *30*, 57–67.

11. Wei, X.; Saha, D. KNEW: Key Generation using NEural Networks from Wireless Channels. In Proceedings of the 2022 ACM Workshop on Wireless Security and Machine Learning, San Antonio, TX, USA, 19 May 2022; pp. 45–50.

12. Zou, Z.B.; Song, L.P.; Song, Z.L. Labeled box-particle PHD filter for multi-target tracking. In Proceedings of the 2017 third IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1725–1730.

13. Zou, Z.; Careem, M.; Dutta, A.; Thawdar, N. Joint spatio-temporal precoding for practical non-stationary wireless channels. *IEEE Trans. Commun.* **2023**. Available online: https://ieeexplore.ieee.org/document/10034681 (accessed on 22 March 2023). [CrossRef]

14. Zou, Z.; Wei, X.; Saha, D.; Dutta, A.; Hellbourg, G. SCISRS: Signal Cancellation using Intelligent Surfaces for Radio Astronomy Services. In Proceedings of the GLOBECOM 2022-2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 4–8 December 2022 ; IEEE: Piscataway, NJ, USA, 2022; pp. 4238–4243.

15. Zou, Z.; Careem, M.; Dutta, A.; Thawdar, N. Unified characterization and precoding for non-stationary channels. In Proceedings of the ICC 2022-IEEE International Conference on Communications, Seoul, Republic of Korea, 16–20 May 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 5140–5146.

16. Peng, H.; Huang, S.; Chen, S.; Li, B.; Geng, T.; Li, A.; Jiang, W.; Wen, W.; Bi, J.; Liu, H.; et al. A length adaptive algorithm-hardware co-design of transformer on fpga through sparse attention and dynamic pipelining. In Proceedings of the 59th ACM/IEEE Design Automation Conference, Francisco, CA, USA, 10–14 July 2022; pp. 1135–1140.

17. Zhang, Y.; Mu, L.; Shen, G.; Yu, Y.; Han, C. Fault diagnosis strategy of CNC machine tools based on cascading failure. *J. Intell. Manuf.* **2019**, *30*, 2193–2202. [CrossRef]

18. Shen, G.; Zeng, W.; Han, C.; Liu, P.; Zhang, Y. Determination of the average maintenance time of CNC machine tools based on type II failure correlation. *Eksploatacja i Niezawodność* **2017**, *19*, 604–614. [CrossRef]

19. Shen, G.; Han, C.; Chen, B.; Dong, L.; Cao, P. Fault analysis of machine tools based on grey relational analysis and main factor analysis. In *Proceedings of the Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2018; Volume 1069, p. 012112.

20. Shen, G.-X.; Zhao, X.Z.; Zhang, Y.-Z.; Han, C.-Y. Research on criticality analysis method of CNC machine tools components under fault rate correlation. In *Proceedings of the IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2018; Volume 307, p. 012023.

21. Nazari, M.; Alidadi, M. Measuring credit risk of bank customers using artificial neural network. *J. Manag. Res.* **2013**, *5*, 17. [CrossRef]

22. Song, Z.; Johnston, R.M.; Ng, C.P. Equitable Healthcare Access During the Pandemic: The Impact of Digital Divide and Other SocioDemographic and Systemic Factors. *Appl. Res. Artif. Intell. Cloud Comput.* **2021**, *4*, 19–33.

23. Song, Z.; Mellon, G.; Shen, Z. Relationship between Racial Bias Exposure, Financial Literacy, and Entrepreneurial Intention: An Empirical Investigation. *J. Artif. Intell. Mach. Learn. Manag.* **2020**, *4*, 42–55.

24. Teles, G.; Rodrigues, J.; Rabê, R.A.; Kozlov, S.A. Artificial neural network and Bayesian network models for credit risk prediction. *J. Artif. Intell. Syst.* **2020**, *2*, 118–132. [CrossRef]

25. He, F.; Ye, Q. A bearing fault diagnosis method based on wavelet packet transform and convolutional neural network optimized by simulated annealing algorithm. *Sensors* **2022**, *22*, 1410. [CrossRef]

26. Han, Y.; Wang, B. Investigation of listed companies credit risk assessment based on different learning schemes of BP neural network. *Int. J. Bus. Manag.* **2011**, *6*, 204–207. [CrossRef]

27. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.

28. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.

29. Zhang, L. The Evaluation on the Credit Risk of Enterprises with the CNN-LSTM-ATT Model. *Comput. Intell. Neurosci. CIN* **2022**, *2022*, 6826573. [CrossRef] [PubMed]

30. Han, C.; Lin, T. Reliability evaluation of electro spindle based on no-failure data. *Highlights Sci. Eng. Technol.* **2022**, *16*, 86–97. [CrossRef]

31. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D Nonlinear Phenom.* **2020**, *404*, 132306. [CrossRef]

32. Xu, F.; Zheng, Y.; Hu, X. Real-time finger force prediction via parallel convolutional neural networks: A preliminary study. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 3126–3129.

33. Yin, W.; Kann, K.; Yu, M.; Schütze, H. Comparative study of CNN and RNN for natural language processing. *arXiv* **2017**, arXiv:1702.01923.

34. Elton, E.J.; Gruber, M.J.; Blake, C.R. A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases. *J. Financ.* **2001**, *56*, 2415–2430. [CrossRef]

35. Berger, A.N.; Udell, G.F. A more complete conceptual framework for SME finance. *J. Bank. Financ.* **2006**, *30*, 2945–2966. [CrossRef]

36. Zhang, B.T.; Joung, J.G. Time series prediction using committee machines of evolutionary neural trees. In Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406), Washington, DC, USA, 6–9 July 1999; IEEE: Piscataway, NJ, USA, 1999; Volume 1, pp. 281–286.

37. Yu, H.; Chen, R.; Zhang, G. A SVM stock selection model within PCA. *Procedia Comput. Sci.* **2014**, *31*, 406–412. [CrossRef]

38. Zhou, Q.; Wang, L.; Juan, L.; Zhou, S.; Li, L. The study on credit risk warning of regional listed companies in China based on logistic model. *Discret. Dyn. Nat. Soc.* **2021**, *2021*, 1–8. [CrossRef]

39. Halteh, K.; Kumar, K.; Gepp, A. Using cutting-edge tree-based stochastic models to predict credit risk. *Risks* **2018**, *6*, 55. [CrossRef]

40. Su, E.D.; Huang, S.M. Comparing firm failure predictions between logit, KMV, and ZPP models: Evidence from Taiwan's electronics industry. *Asia-Pac. Financ. Mark.* **2010**, *17*, 209–239. [CrossRef]

41. Zhu, Y.; Zhou, L.; Xie, C.; Wang, G.J.; Nguyen, T.V. Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. *Int. J. Prod. Econ.* **2019**, *211*, 22–33. [CrossRef]

42. Chen, Y.C.; Huang, W.C. Constructing a stock-price forecast CNN model with gold and crude oil indicators. *Appl. Soft Comput.* **2021**, *112*, 107760. [CrossRef]

43. Li, M.; Zhang, Z.; Lu, M.; Jia, X.; Liu, R.; Zhou, X.; Zhang, Y. Internet Financial Credit Risk Assessment with Sliding Window and Attention Mechanism LSTM Model. *Tehnički vjesnik* **2023**, *30*, 1–7.

44. Vidal, A.; Kristjanpoller, W. Gold volatility prediction using a CNN-LSTM approach. *Expert Syst. Appl.* **2020**, *157*, 113481. [CrossRef]

45. Chen, B.R.; Liu, Z.; Song, J.; Zeng, F.; Zhu, Z.; Bachu, S.P.K.; Hu, Y.C. FlowTele: Remotely shaping traffic on internet-scale networks. In Proceedings of the 18th International Conference on emerging Networking EXperiments and Technologies, Rome, Italy, 6–9 December 2022; pp. 349–368.

46. Zhang, R.; Zeng, F.; Cheng, X.; Yang, L. UAV-aided data dissemination protocol with dynamic trajectory scheduling in VANETs. In Proceedings of the ICC 2019—2019 IEEE International Conference on Communications (ICC), Shanghai, China, 20–24 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.

47. Georgios, K. Credit risk evaluation and rating for SMES using statistical approaches: The case of European SMES manufacturing sector. *J. Appl. Financ. Bank.* **2019**, *9*, 59–83.