J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 8273--8291

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Design of Cloud Collective Financial Robot Based on Artificial Intelligence Technology

Qigao Wu[1,✉], Xiaoyan Huang[2]

[1] School of Finance & Economics, Wuxi Institute of Technology, Wuxi, Jiangsu, 214012, China
[2] School of Economics & Management, Yancheng Institute of Technology, Yancheng, Jiangsu, 224007, China

ABSTRACT
In this paper, we design a cloud-integrated financial robot based on artificial intelligence technology to provide advanced financial analysis and decision support for financial institutions. The robot platform is embedded with a large amount of financial domain knowledge and data, which can provide uninterrupted financial services to customers using a dialog engine. At the same time, it is equipped with the attention mechanism - long and short-term memory neural network model, in investment transactions and credit risk prediction, which can bring a new digital intelligence experience for financial institutions. The standard and similar sentence recognition accuracy of the article robot dialog engine can be stabilized at more than 90%, and the average access time of the user's access request is about 0.15s. The importance distribution of financial credit risk indicator features is 5~24, and when the number of features takes the value of 10, the risk prediction accuracy of the robot in this article is the highest, 97.98%. When the prediction model is trained to 50~70 epochs, the Loss value of the financial robot converges to 0.15~0.17. The accuracy of the model chosen in this paper for risk prediction as well as stock prediction is 95.35 and 96.2% respectively. And the absolute difference between the predicted stock price and the true value of the model in this paper is 0~0.28 yuan. Combined with DMI strategy for stock trading, the return is 30.7%. The financial robot improves the user experience and increases the value of risk control as well as stock prediction for financial institutions.

Keywords: Artificial Intelligence, Cloud Integration, Financial Robot, Attention Mechanism, Long and Short-term Memory Neural Networks

1. Introduction

In recent years, the artificial intelligence technology represented by ChatGPT big model has realized leapfrog development, and all kinds of general and vertical domain big models have blossomed. The

✉ Corresponding author.
E-mail address: 13961705852@163.com (Q. Wu).
Received 03 March 2024; Revised 10 May 2024; Accepted 25 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-451

high intensity of capital, data and decision-making in the financial industry determines that the financial industry is one of the industries suitable for the first landing of big models, and the financial big model will undoubtedly become the most popular industry vertical field big model, which is now taking root and blossoming.

Artificial intelligence big models play a crucial role in the development of financial high quality. On the one hand, the horizontal expansion of financial service scenes [1]. In the past, the integration of digital technology and finance solved relatively simple problems, such as credit, customer service, etc. The addition of the big model helps to improve the quality and efficiency of the solutions in the original scene, and at the same time, it has the ability to solve relatively complex problems, such as risk prevention and control, predictive decision-making, etc. [2-5]. For example, it can be applied to a customer service robot to provide customers with 24/7 service [6]. It can also be applied to a financial advisor robot to provide investment advice and risk assessment for customers [7]. On the other hand, the integration of industrial elements has been deepened vertically [8]. In the process of digitization of the real economy, the digital platform strategy of "technology + finance + industry" promotes the integration and development of finance with major national strategies and the modernized industrial system, and also promotes the provision of financial resources to small and medium-sized micro-enterprises and grass-roots people to provide accurate services, thus promoting the integration of different components of the industrial chain [9-12]. AI models can also be applied to a trading robot that automatically executes trading orders given by clients [13]. For the financial industry, in order to accelerate the return on investment of AI analytics, the migration of AI analytics to the cloud through "fully managed services" will be an important strategic action that cannot be ignored [14-16].

The application of artificial intelligence technology can automate the processing of financial business, reduce the operating costs of financial institutions and improve the efficiency of financial services [17]. Met, İ. et al. showed that artificial intelligence and robotics can analyze the business model of financial institutions in the digital age, and the development of these two technologies has transformed the customer model and the internal operation process in the traditional financial industry, so as to enable it to be competitive in the internally and externally changing business environment [18]. Golić, Z. explored the significance of the value of AI technology applied to the financial sector, where AI in the financial sector is able to fully understand the business needs within the industry and make changes to the industry processes through powerful technological capabilities to achieve effective operations and cost savings, in addition to its important role in job creation, etc. [19]. Ravi, V. et al. clarified the digital transformation experienced by the financial services industry, which has evolved from a pure journal and book entry model to a data and analytics driven financial business, it is really this shift from a traditional model to a complex digital transactional approach that has given rise to artificial intelligence robotics in the financial sector [20]. Sabharwal, C. L. et al. elaborated on the role played by robo-advisors in the financial sector, which combines intelligent algorithms with graphical vision to provide enhanced customer service to users and also analyze data analytics to improve business models and increase customer retention, in addition, robo-advisors have great potential in addressing data noise and behavioral elements in the financial domain [21]. Wei, Q. combined big data algorithms with robotic process automation (PRA), and proposed PRA financial robot task allocation applicable to different financial tasks by adjusting the function optimization algorithm, which improved the accuracy and efficiency of task allocation, and then enhanced the overall efficiency of enterprises [22]. Ma, J. et al. analyzed the necessity of the RPA financial robots in the financial management of the enterprises, and cloud computing, artificial intelligence and other technologies provide a good technical platform for robotic process automation technology to help enterprises reduce costs, improve efficiency and realize digital transformation in the development process [23].

The use of artificial intelligence technology for real-time financial market monitoring can improve the identification and early warning ability of regulators for market risks [24]. Bisht, D. et al. used cloud computing, artificial intelligence and other technologies to innovate enterprise financial management tools, realize enterprise risk management and assessment by analyzing real-time financial data, as well as realize the functions of virtual assistant based on AI and IoT, which improves the financial service quality and accessibility [25]. Wang, L. et al. designed an intelligent analysis and processing model of financial big data based on deep belief network to analyze enterprise financial data using AI and deep mining technology to provide financial risk prediction for the enterprise, and the PRA financial robot equipped with the model significantly improves the work efficiency and accuracy [26]. Kothandapani, H. P. Introduced PRA technology in financial risk management from the four aspects of market risk, credit risk, comprehensive risk, and operational and regulatory compliance, which effectively and accurately solves the risk assessment modeling problem for complex financial data, and improves the efficiency and accuracy of financial institutions' risk assessment [27]. Shi, S. et al. emphasized on the design of financial robots as an important means of realizing the PRA, and proposed an financial robot with cloud integration design and big data analysis to provide corporate decision makers with reference analysis of financial status [28]. Lomakin, N. I. et al. based on the computational theory of financial risk, used neural network method to design a time-series financial risk model to hedge financial risk, integrated and analyzed data samples in securities trading with the help of an artificial intelligence system, and ensured that the predicted securities trading prices with high accuracy and minimal risk of loss [29]. Pan, H. introduced the system composition of intelligent financial global monitoring, including trading quantitative investment, financial risk monitoring, economic posture monitoring and environmental risk monitoring, which provides a good observation platform for the randomly changing market prices and global economic trends and other aspects [30].

The study adopted cloud integration (HTTP network communication protocol, code completion, containerization technology, etc.) to design the general framework of the financial robot. By collecting financial related data and combining the artificial intelligence technology based on Attention-LSTM model, the robot realizes the intelligent dialogue on financial knowledge and can accurately predict the credit risk of enterprises as well as stocks based on the characteristics of financial indicators. At the same time, through the performance test and other experiments, the robot in this paper is evaluated for its application effect in the financial field.

2. Artificial intelligence techniques for financial forecasting

The combination of Artificial Intelligence (AI) technology with financial robots is changing the way the financial industry operates and this paper utilizes the following AI technologies to achieve the predictive application of financial robots in institutions.

2.1. Long and short-term memory neural networks

LSTM is an improved recurrent neural network [31], which solves the long-term dependency problem of recurrent neural networks to a certain extent. LSTM consists of three parts: input layer, hidden layer and output layer, and its structure is shown in Figure 1.

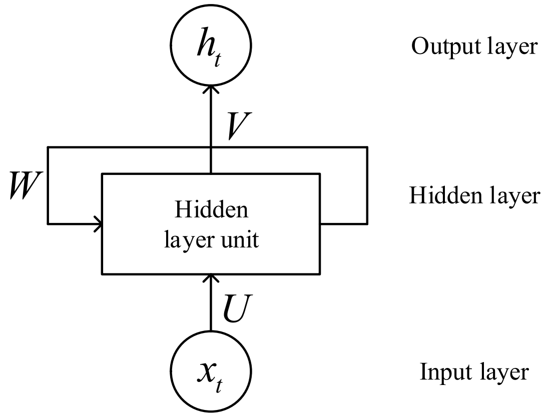

Fig. 1. LSTM structure diagram

x_t denotes the input vector at moment t, h_t denotes the output hidden state vector at moment t. U denotes the connection weight between the input layer and the hidden layer, w denotes the connection weight between the hidden layer and the hidden layer, and v denotes the connection weight between the hidden layer and the output layer.

2.2. *Attention mechanisms*

Figure 2 illustrates the structure of the attention mechanism. The nature of the attention mechanism [32-33] is derived from human visual attention and was first proposed in the field of visual images. The basic idea of the attention mechanism is to select from the input information the information that is more critical to the current task goal. It captures more valuable information by assigning corresponding weights to each input information according to its importance.

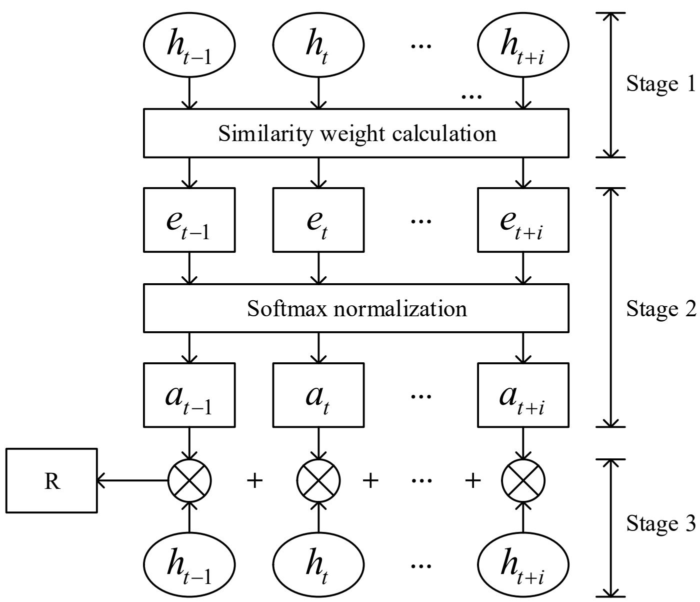

Fig. 2. Structure of the self-attention mechanism

h_t denotes the hidden state vector at moment t, e_t denotes the similarity weight of h_t, α_t denotes the normalized similarity weight of h_t, and R denotes the self-attention output matrix obtained by weighted summation. The computational process can be divided into three steps as follows:

1) The similarity weights are computed for each moment of data using a multilayer perceptron (MLP) with the following formula:

$$e_t = W_b \tanh(W_a[h_t; H]) \tag{1}$$

where W_a denotes the connection weights between the input and hidden layers of the MLP, W_b denotes the connection weights between the hidden and output layers of the MLP, and $H = (h_t - 1, h_t, \ldots, h_{t+i})$ denotes the hidden state matrix.

2) These similarity weights are normalized using the softmax function with the following formula:

$$a_t = soft\,max(e_t) = \frac{exp(e_t)}{\sum_{j=t-1}^{t+i}(e_j)} \tag{2}$$

where $exp()$ denotes the exponential function with the natural constant e as the base.

3) The normalized similarity weights and the corresponding data are weighted and summed to obtain the self-attention output matrix R, calculated as follows:

$$R = \sum_{j=t-1}^{d} \alpha_j \overline{h}_j \tag{3}$$

where h_j denotes the hidden state vector at moment j and α_j denotes h_j the normalized similarity weights.

2.3. Attention-LSTM modeling

In this paper, Attention-LSTM model [34] is used for predictive analysis of financial data. The Attention layer is connected to the output layer of the LSTM model, and different weights are assigned to the output information, in which the information more relevant to the current output receives a higher weight, so that the system pays more attention to the information related to the current output. Figure 3 shows the Attention-LSTM model structure.

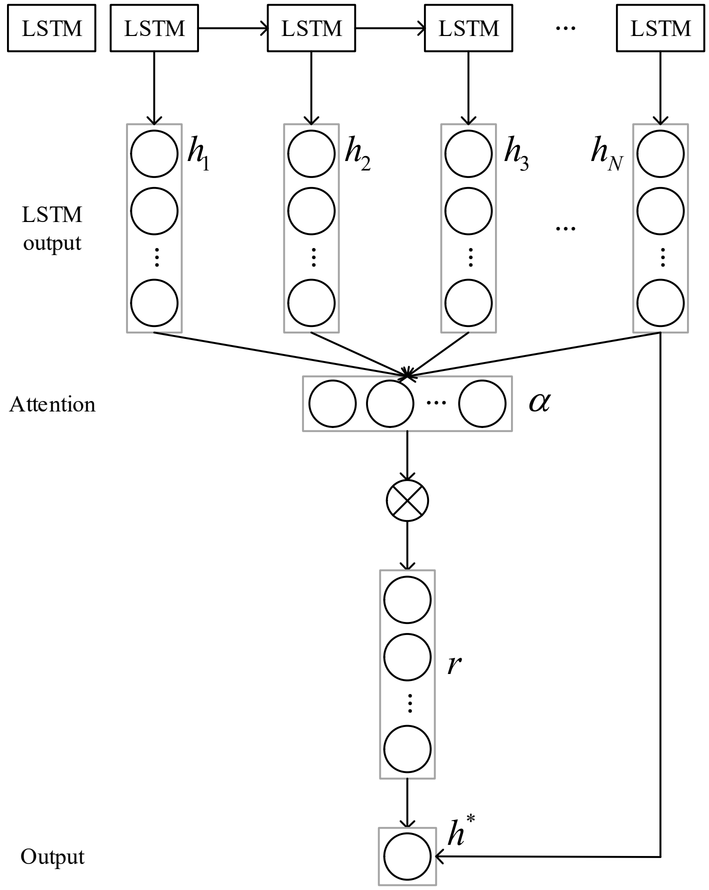

Fig. 3. Attention-LSTM mechanism

Let $H = [h_1, \cdots, h_N] \in R^{d \times N}$ be the output vector of the LSTM model and N be the length of the memory window, and use the data of day N to predict the data of day $N+1$. The Attention mechanism is implemented as follows:

$$M = H^T W h_N = (h_i^T W h_N)_{i=1,2,\ldots,N} \tag{4}$$

where, $W \in R^{d \times d}$ is the parameter matrix to be trained and $M \in R^N$ is the relevance vector.

The Softmax function is chosen to normalize the processing vector M to obtain the attention weight vector $\alpha \in R^N$:

$$\alpha = soft\max(M) \tag{5}$$

$$\alpha_i = \frac{\exp(hT_i^T W h_N)}{\sum_{i=1}^{N} \exp(h_i^T W h_N)}, i = 1, 2, \ldots, N \tag{6}$$

where α_i is the attention probability of the i nd node to the N rd node.

Thereby the output vector H of the LSTM model, weighted by the attention weights α, obtains the following expression for the output eigenvalue (environment vector) r:

$$r = H\alpha = \sum_{i=1}^{N} \alpha_i h_i \tag{7}$$

where $r \in R^d$.

The final output is:

$$h^{\bullet} = \tanh(W_r^T r + W_h^T h_N) \tag{8}$$

where $W_r, W_h \in R^d$ is the parameter vector to be trained.

3. Research on technologies related to cloud integration development environments

3.1. HTTP network communication protocol

Hypertext Transfer Protocol (HTTP) is an application layer protocol that is widely used for Internet transmission [35]. Among them, HTTP/2.0 is widely used with HPACK algorithm for message header compression, new binary format, multiplexing techniques and server-side push.

The two most common request methods used in HTTP are GET and POST; GET requests connect parameters directly to the URL, are separated by delimiters, and therefore have a length limit; GET requests can be cached, bookmarked, and retained in the history; POST requests are relatively hidden, with the content being placed in the body of the message, and are therefore often used for transmitting usernames, passwords, and files, among other things.

3.2. Code Completion

1) Language Server. Modern integrated development environments (IDEs) usually need to support multiple languages, the programming process of editing, debugging, running and other tasks between a variety of languages there is a relatively large degree of similarity, so you can take advantage of the language server, client-side structure, the use of JSON-RPC communication, decoupling, while this structure can be better utilized in the integration of microservices structure.

2) Language Service Protocol. The LSP protocol contains two parts, header and content, and the header and content are distinguished by \r\n. The header part contains two fields, content length and content type. The content part is the actual message load. The content part of the message uses JSON-RPC to describe the request, response and notification. The content part is encoded using the character set provided in the content type field.

3) Completion Request. Completion requests are part of the LSP protocol and are sent from the client to the server to compute a request for code completion at a specified cursor position. Completion of the content of the parsing and calculation of the work is often required to consume a relatively large amount of computing resources, so the request can be received to create a separate program for the processing of the code.

4) Clangd. Clangd is part of the LLVM open-source project , according to the Clang compiler for code parsing , LSP support for the C language provides syntax errors and warnings checking , code repair , automatic code completion , code signing help , code navigation and other auxiliary functions .

3.3. Containerization technology

Docker is a container virtualization platform. It is built on publicly available open source technology.Docker virtualization implementation is based on a set of Linux kernel mechanism, process resource separation provided by the kernel namespace. Docker does not need Hypervisor to realize the virtualization of hardware resources, in the CPU, memory utilization Docker will have an advantage in terms of efficiency. And has a strengthened image and container control mechanism. At the same time, Docker also has obvious advantages in terms of startup speed, community support, and so on. Therefore, Docker is chosen as the basis of the virtualized execution environment.

3.4. Microservice artifacts

Load Balancing Load balancing provides the ability to evenly distribute workloads over available resources.The components of a single instance of a system are often difficult to demand in terms of performance, so there are often multiple instances to share the task.The API gateway is responsible for request forwarding, API combining, and protocol conversion, and the API gateway is the only entry point to the microservices.The presence of the API gateway simplifies the adaptation of the mobile, web, and embedded systems as well as the The existence of API gateway simplifies the adaptation of mobile, web, embedded systems and some other complex situations. The existence of API gateway makes the front-end only need to deal with API gateway without having to care about the internal communication details, so API gateway is an encapsulation of the back-end system (microservices) to improve the transparency of the whole system.

3.5. Vue-based user front-end interface technology

Front-end development is the use of hypertext (HTML) for the construction of the basic page, style selector (CSS) for page styling, scripting language for more complex user interactions and other functions, the scripting language is mainly JavaScript and its derivatives.Vue uses the template syntax of HTML to declaratively bind the DOM to the underlying Vue instance of the data, which can be constructed with a user-interactive software interface, while utilizing technologies such as AJAX to dynamically obtain remote data.

3.6. Transformer Intelligent Code Modeling Technology

Transformer employs a similar structure to self-encoders, introducing the mechanism of multi-head attention through encoder-decoder-decoder connections, first designed to handle translation tasks and subsequently achieving good results in natural language processing.

Encoder: The encoder has 6 identical layers, internally it is composed of a multi-head attention layer and a fully-connected layer, residual connectivity and normalization are added between layers, and the model inputs are encoded for words as well as positional encoding.

Decoder: The decoder also has six identical layers, and in addition to the two sub-layers in each encoder layer, the decoder inserts a third sub-layer that performs multi-head attention on the output of the encoder stack. Unlike the encoder, MASK is performed among the multi-head attention to prevent direct inference using the results.

4. Overall design of financial robots

4.1. Robot framework model

The overall architecture of the robot system is shown in Figure 4. According to the architectural design, the project is divided into access layer, display layer, interface layer, business logic layer, data access layer, and database layer.

The overall functional structure of the system can be obtained from the analysis of the functional requirements of the system, which is responsible for the recall module that recalls the user's questions in the knowledge base, the reordering module that reorders the data recalled by the recall module again for fine ranking, the knowledge base management module that manages the knowledge base, the statistical analysis module that analyzes the data, and the spelling error correction module.

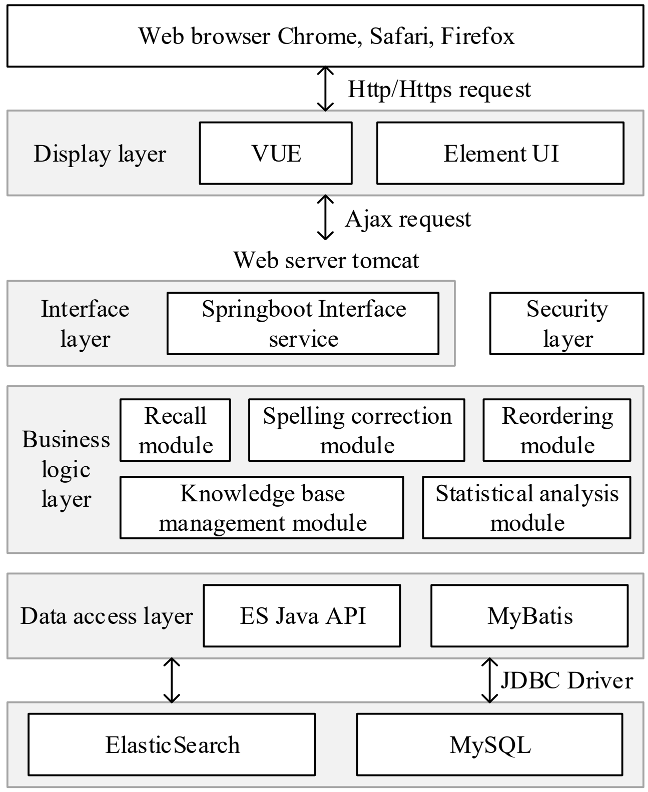

Fig. 4. Overall system architecture

4.2. Robot function module design

4.2.1. Knowledge base management module design. The Knowledge Base Management module is mainly for adding questions and answers, deleting questions and answers, modifying questions and answers, and querying questions and answers to the knowledge base.

1) Add questions and answers: This function is to add questions and answers to the knowledge base for funds or stocks that do not exist in the knowledge base.

2) Delete questions and answers: This function is for funds that are not sold in this product center or funds that have been taken offline, the related questions and answers should not exist in the knowledge base, so they should be deleted.

3) Modify questions and answers: When editing questions, it is not possible to consider them perfectly, and this function is for modifying questions and their related answers to make them standardized.

4) Query questions and answers: this function is an auxiliary function, when you need to add a question or delete a question, you need to consider whether the question exists or not, and then carry out the relevant operations.

4.2.2. Recall module design. The system architecture diagram of the recall template is divided into four layers in total:

1) Input layer: the input layer is divided into two parts, the first part is the result of the disambiguation, i.e., a sentence is disambiguated into individual words, and the disambiguator is a customized disambiguator. The second part of the sentence encoding part, similarly, the sentence input by the user is first divided into words to remove the deactivation operation.

2) Recall layer: the two parts obtained from the first part of the processing are fed into different interfaces in the elastic search framework.

3) Rough ranking layer: assuming that the final text to be obtained is n, according to the flow direction of the data, the BM25 algorithm is used for literal recall, and the preliminary design is to recall n/2.

4) Output layer: the results of the coarse ranking layer are counted up to become n articles, because the literal-based recall and semantic-based recall together will have duplicate parts, and need to carry out de-weighting operation before outputting to the reordering layer.

4.2.3. Spelling correction design. The basis of spelling error correction is done based on a language model that uses a binary language model to determine the probability that a sentence constitutes a sentence, a sentence is represented as $S = s_1, s_2, \ldots, s_m$, where s_i is each word that represents the sentence, then the probability of the sentence is:

$$P(s) = p(s_1)p(s_2 \mid s_1)\ldots p(s_m \mid s_1, s_2, \ldots, s_{m\text{-}1}) \tag{9}$$

Perplexity is the magnitude of the probability that a sequence of words forms a sentence and is calculated as shown in equation (10):

$$pp(S) = P(S)^{-\frac{1}{m}} = \sqrt[m]{\frac{1}{P(S)}} \tag{10}$$

4.2.4. Reordering design. The design of reordering consists of 4 layers.

1) Input layer: the input layer is divided into two parts consisting of two parts, the first part is the question input by the user, and the second part is the content that is recalled according to the recall system.

2) Rewriting layer: the rewriting layer is to rewrite the user's input, and the rewriting layer is designed to standardize the user's input sentences to facilitate the post order sort matching model.

3) Pre-processing layer: the same fine-tuning of the input to the sorting model, but one of the biggest roles of this layer is to check the user's input for the presence of historical questions.

4) Re-ordering layer: belongs to the core content of the fine-ordering layer, the role of this layer is mainly to recall each sentence and the user's question to do text matching, to get the corresponding similarity score, and then according to the score to sort.

4.2.5. Module design for financial statistics analysis. The statistical analysis module has these three main functions: user analysis, behavioral analysis financial forecasting.

1) User analysis. In the database stores the user's basic personal information, such as birth date, gender, occupation, registration time and so on. In addition, there is a table for recording logs, recording a certain point in time, a certain location of a user to do what the query.

2) Behavior Analysis. From the system level, more macro data can be obtained based on the log table. For example, every day, every week, every month, how often the system is used, making full use of the time control, from a variety of dimensions, evaluating user viscosity and observing the active situation. In addition, by counting the content of user queries in a cycle, we can get the content that people pay more attention to in that time period, and realize the function similar to hotspot. Through the hotspot and the content of user queries, the query results can be optimized to provide users with better answers.

3) Financial Forecasting. During the operation of the robot, by analyzing the information of relevant individuals and financial institutions, and using the deep learning model based on artificial intelligence technology mentioned above to mine the correlation between them, it can realize the functions of credit risk assessment of financial institutions as well as stock price prediction.

4.3. System security design

1) A secure encoding method is used for the user's personal information, and when a query is required, a parameterized query is used instead of direct splicing parameter execution. When the user inputs, the security check of the input range needs to be carried out.

2) The system needs to control the restrictions on user access to resources. When users access specific resources, they need to carry out operations such as identity verification and comparison of operating privileges, so as to prevent users from directly modifying the form, thus skipping the privilege checking and directly carrying out ultra-authorized operations.

3) When a user has been accessing a certain resource for a certain period of time, the user's connection should be cut off.

4.4. Financial Bot Performance Analysis and Dialog Engine Testing

Dialogue engine module is the core module of this financial robot system, which can give corresponding answers according to different questions, which also has the function of similar sentence recognition, even if the question sentence is different from the standard sentence in expression, the dialogue engine can recognize and feedback the correct answer to the user.

A total of 46,542 corpus of this system is categorized into 10 groups for testing, and each group is tested 200 times respectively to test whether the dialog engine can feedback the correct results. Figure 5 shows the test results of standard and similar sentences of the robot dialog engine in this paper.

From the figure, it can be seen that when using the sentences that exist in the knowledge base to test the dialogue engine, the success rate is high, and the average success rate of the 200 test results can be stabilized at more than 97%, and in the actual use of the situation, if the user directly hits the standard sentence, it can get high-quality feedback from the dialogue engine. When using similar sentences that do not exist in the knowledge base to test the dialog engine, the accuracy of the 200 test results can be stabilized at more than 90%, and in actual use, even if the questions posed by the user are somewhat different from the standard sentences in the corpus, the dialogue experience can be high in most cases.

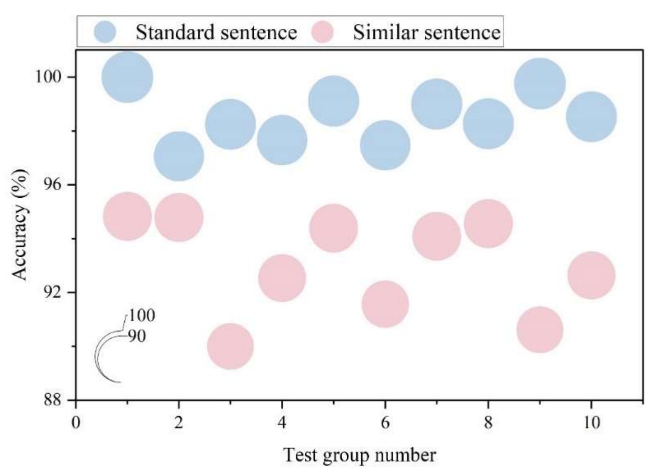

Fig. 5. Test results of the dialog engine standard and similar sentence

In performance testing, this article utilizes the Postman tool to test the speed of the server's interface response.Postman is an HTTP debugging software that can visually display and simulate HTTP requests under various conditions.Postman puts the request body in the body and sends data in Json format.The backend server receives the request and handles the request as if it were a normal request, returning information such as data and status codes. Postman puts the request body in the body and sends the data in Json format.

The tests batch-enter the questions to simulate concurrent client access requests at the same time. In order to control the variables to explore the impact of different data volumes on the server, the test divided the data volume into 10, 50, 100, 300 and 500 to test the response speed of the system dialog respectively. The recorded response time includes the time spent on network requests. The performance test results are shown in Table 1.

From the table, it can be seen that the average response time of the system is around 0.15 seconds for each article after the user has asked a question, which can give the user a good lag-free dialog experience. From the figure, it can be seen that the curve changes smoothly with the increase of the number of test entries, which proves that the system stability is good.

Table 1. Performance test results

Test number	Mean response time(s)	Memory occupancy rate(%)	CPU occupancy rate(%)
10	0.156	13.4	24.6
50	0.174	14.8	26.1
100	0.173	15.3	25.4
300	0.186	15.6	28.3
500	0.167	15.7	29.5

5. Financial Robotics Credit Risk Assessment and Stock Forecasting

Experimental environment:

Windows 10 operating system 16GB RAM, development tool: PyCharm 2022 TensorFlow 2.0, graphics card: NVIDIA Ge Force MX450, CPU: Intel(R)Core(TM)i5-1135G7@2.40 GHz.

In order to enrich the functions of the financial robot designed in this paper, the following 15 enterprise credit risk assessment index features are initially constructed by obtaining enterprise credit risk evaluation data.

Cash Ratio, Gross Sales Margin, Earnings per Share, Year-on-Year Growth Rate of Operating Income, Year-on-Year Growth Rate of Total Assets, Asset Liability Ratio, Cash Inflow, Total Asset Turnover Ratio, Year-on-Year Growth Rate of Net Cash Flows, Current Debt Ratio, Payment Ratio of Total Assets, Net Cash Flows from Operating Activities, Growth Rate of Current Assets, Year-on-Year Growth Rate of Net Profit, and Liquidity Ratio, which are denoted in order as R1~R15.In this section, we will go through the feature importance for indicator screening, laying the foundation for predicting corporate credit risk using the robot in this paper below.

5.1. Indicator Importance Assessment and Feature Subset Screening

Using the feature screening model for the initial screening of the indicators, through the model, the importance score of each feature indicator is output, and at the same time, the feature subset algorithm is performed through the model algorithm, and the features are selected from the feature importance by testing multiple thresholds. The importance of each feature and the relationship between the number of features and the accuracy rate are finally obtained as shown in Figure 6.

From the figure, it can be seen that the importance of features from strong to weak are R10, R7, R2, R5, R1, R3, R6, R4, R9, R8, R15, R11, R13, R12, R14, and the importance takes values between 5 and 24. According to the results of the relationship between the number of features and the accuracy of the model, when the number of features in the subset of features is selected as 10, the prediction accuracy of this model is higher, 97.98%, and the results are more reliable. Therefore, the top 10 indicators of feature importance are screened as credit risk evaluation indicators of financial enterprises. That is, cash ratio, gross sales margin, earnings per share, year-on-year growth rate of operating income, year-on-year growth rate of total assets, gearing ratio, cash inflow, total asset turnover, year-on-year growth rate of net cash flow, and current liability ratio.

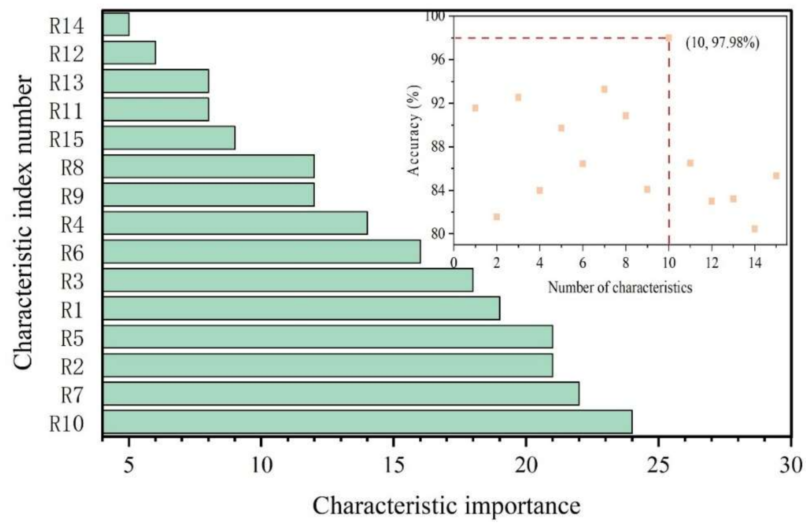

Fig. 6. The importance of each characteristic and its relationship with accuracy

5.2. Evaluation of Predictive Model Training Effectiveness

The training set is trained using this model, and two datasets are selected, one is the credit risk prediction dataset of financial firms and the other is the stock prediction dataset. The experiment allocates the training set and test set in the ratio of 5:5. Figure 7 shows the training results of the model of this paper on the training set and test set.

Observing the change of the loss function in the figure, it can be seen that when the number of iterations of this model is 50~70 times, the Loss value of the model on the credit risk dataset as well as the stock prediction dataset tends to be stable, and the curves of the predicted value and the real value are close to each other, and the model's prediction results have a small error and high degree of fit, and the Loss value of this model on the two sets of datasets converges to the range of 0.15~0.17.

All kinds of operation-related information of financial enterprises will be constructed as credit risk data set or stock data set, and then the model in this paper will be used for prediction, to get the credit risk of each of its enterprises as well as the stock prediction value, so as to provide a reference research basis for the operation of enterprises.

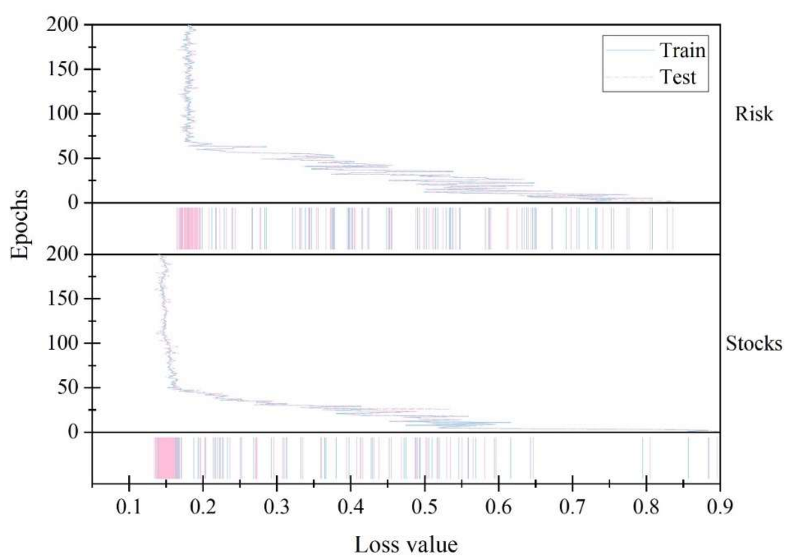

Fig. 7. The training results of this model in the training set and the test set

5.3. Comparative analysis of model prediction performance

In order to verify the superiority of the Attention-LSTM model constructed in this paper, the model is experimentally compared with SVM, RF, CNN, and LSTM, BP, a total of six sets of prediction models, to compare the precision, recall, F1-score, and accuracy of each model on the test set. The datasets are the credit risk dataset and stock prediction dataset used above. The specific comparison results of the evaluation metrics are shown in Figure 8.

By comparing and analyzing the evaluation metrics of the above multiple single prediction models, on the same sample dataset, the improved Attention-LSTM model constructed in this study has a better prediction effect in terms of precision rate, recall rate, F1-score value and accuracy rate.

The SVM model and RF model, as traditional prediction models, have relatively low prediction accuracy for data with large amount of data and do not take the temporal characteristics of data into account. Since a single CNN model cannot memorize historical information and apply it to subsequent prediction, its judgment of time series features is relatively weak, and the prediction accuracy is correspondingly low. Using the strong characteristics of the attention mechanism for feature extraction, it can be improved and optimized with the LSTM model using residual connection, which can avoid falling into the situation of optimal solution while considering the time series problem.

In this paper, after the preliminary feature screening of data, the accuracy of the Attention-LSTM model constructed for corporate credit risk prediction and stock prediction reaches 95.35 and 96.2% respectively, which is obviously due to other models.

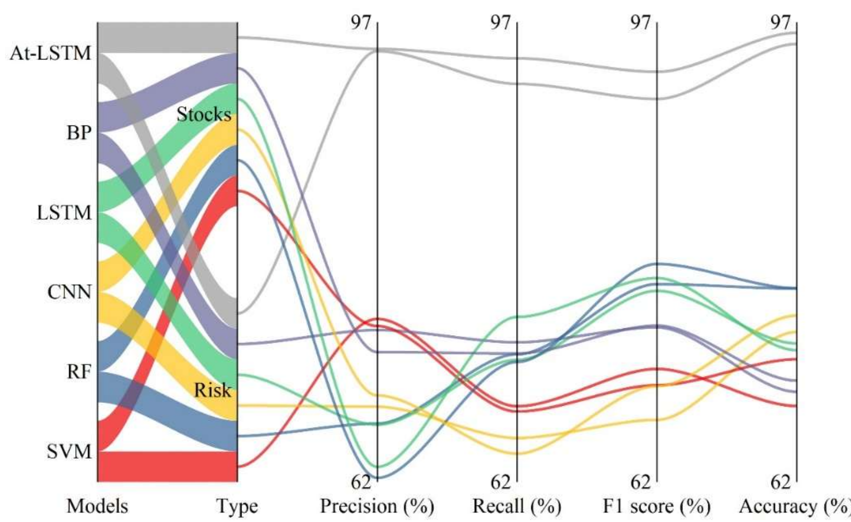

Fig. 8. The evaluation index is the actual comparison result

5.4. Forecasting the value of a company's stock and evaluating buying and selling strategies

In this section, the stock data of a financial firm is selected from January 1, 2023 to December 31, 2023, and various models are chosen to forecast the stock price of the firm as a means of evaluating the value of the financial robots designed in this paper in providing investment advice. Figure 9 shows the results of each model's prediction of the firm's stock price within one year.

The data in the figure show that, for the fluctuation change of the enterprise's stock, the model selected by the financial robot in this paper can get the minimum error stock price prediction, and the absolute difference with the real stock price is 0~0.28 yuan, which is much lower than the prediction error of the comparative model, which indicates that this paper's financial robot has a certain degree of reliability in the field of stock market prediction.

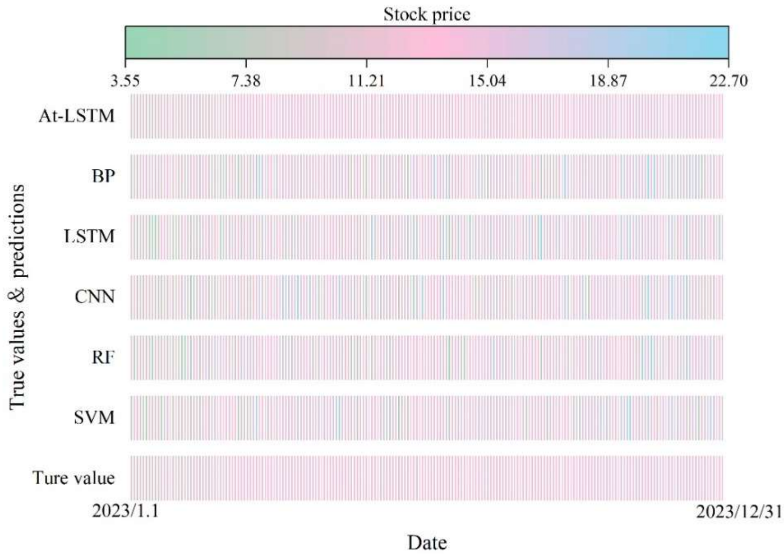

Fig. 9. The results of each model for the price of the stock in a year

Because the stock price prediction result in this paper is the closing price, the DMI indicator is selected for the detailed study of buying and selling strategies. In this paper, the above financial enterprise stocks are selected to verify that unexpected events are contingent, and it is difficult to accurately predict the occurrence of unexpected events in stock trading. Given the initial investment amount of $20,000, according to the strategy algorithm of DMI indicators for buying and selling operations. And in the final calculation of stock returns and yield to verify the actual feasibility of the model.

The DMI indicator consists of four curves, which are long/short indicator positive DI, negative DI, tendency indicator ADX, ADXR. The trading strategy of the DMI indicator is mainly based on the cross situation of the four curves, as well as the position and running direction of the positive DI curve, and the buying and selling strategy of the DMI indicator is as follows: first of all, when the DMI's four curves oscillate narrowly in the region around 20, if the positive DI curve crosses the negative DI curve upwards, the positive DI curve will be able to move up to the negative DI curve. DI curve up through the negative DI, ADX, ADXR curve, indicating that the market is relatively strong in the main force, if the stock price at the same time upward breakthrough in the medium and long term averages, predicting that in the short term the stock price will enter a strong upward phase, you can consider buying at this time. When the four curves of the DMI is higher than 20, carried out between 20-40, if the positive DI curve down below the negative DI, ADX, ADXR curve, indicating that the market in the short side of the main force is relatively strong, if at this time the stock price at the same time sticking below the average of the medium and long term, predicting that the stock price short-term potential risk of decline, at this time can be appropriate to sell. According to the DMI indicator algorithm strategy to buy and sell stocks, Figure 10 shows some of the K-line and buy and sell points of the financial enterprises tips chart. Table 2 shows in more detail the time of buying and selling, transaction price, stock balance and fund balance.

The data in the table shows that the data derived from the buying and selling operations in strict accordance with the buying and selling points calculated by the DMI buying and selling strategy algorithm shows a total fund balance of 26,140, a net income of 6,140 and a return of 30.7%.

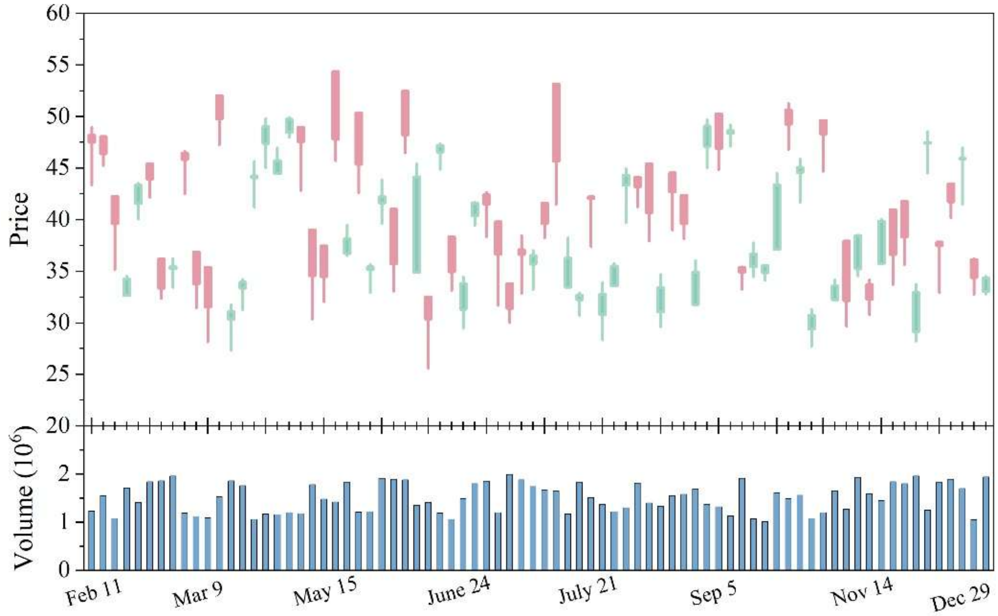

Fig. 10. Some tips for the k and the trading points

Table 2. DMI index sales and strategy benefits

Date	Buy/sell	Transaction price (Yuan)	Hold	Money balance (Yuan)
2023 Feb 06	Buy	41.15	480	248
2023 Feb 11	Sell	45.24	0	21740
2023 Mar 24	Buy	50.63	420	475.4
2023 May 04	Sell	54.37	0	22835.4
2023 June 12	Buy	43.54	500	1065
2023 June 20	Sell	50.26	0	26140
Yield				30.7%

6. Conclusion

This paper introduces the artificial intelligence technology based on financial prediction, including the long and short-term memory neural network, attention mechanism, combined with the cloud integration development environment technology to work together to complete the overall design of the financial robot. The value of its application in the financial field is realized by building robot function modules. The financial robot can recognize the standard interrogative sentences and the acquaintance interrogative sentences in the financial knowledge base on time, and the average access time to the user's access request is between 0.156~0.186s. By filtering the top ten financial indicators of importance, the financial robot can predict credit risk with an accuracy of 97.98%. The Attention-LSTM model equipped with the robot can achieve loss convergence on two datasets with convergence values between 0.15 and 0.17 after training to about 70 rounds. And its performance on credit risk as well as stock price prediction is better than the comparison model. Meanwhile, combining this robot with the DMI stock buying and selling strategy can achieve a high rate of return.

References

[1] Wenjing, C. (2025). Simulation application of virtual robots and artificial intelligence based on deep learning in enterprise financial systems. Entertainment Computing, 52, 100772.

[2] Lamberton, C., Brigo, D., & Hoy, D. (2017). Impact of Robotics, RPA and AI on the insurance industry: challenges and opportunities. Journal of Financial Perspectives, 4(1).

[3] Sabharwal, C. L. (2018). The rise of machine learning and robo-advisors in banking. IDRBT Journal of Banking Technology, 28.

[4] Asparouhova, E., Bossaerts, P., Cai, X., Rotaru, K., Yadav, N., & Yang, W. (2024). Humans in charge of trading robots: The first experiment. Review of Finance, rfae007.

[5] Dirican, C. (2015). The impacts of robotics, artificial intelligence on business and economics. Procedia-Social and Behavioral Sciences, 195, 564-573.

[6] Doguc, O. (2021). Applications of robotic process automation in finance and accounting. Beykent Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 14(1), 51-59.

[7] Nagarajah, E. (2016). Hi robot. What does automation mean for the accounting profession, 34-37.

[8] Demiralay, S., Gencer, H. G., & Bayraci, S. (2021). How do Artificial Intelligence and Robotics Stocks co-move with traditional and alternative assets in the age of the 4th industrial revolution? Implications and Insights for the COVID-19 period. Technological Forecasting and Social Change, 171, 120989.

[9] Karaoglu, S., Arpaci, U., & Ayvaz, S. (2017). A deep learning approach for optimization of systematic signal detection in financial trading systems with big data. International Journal of Intelligent Systems and Applications in Engineering, 2017(Special Issue), 31-36.

[10] Gotthardt, M., Koivulaakso, D., Paksoy, O., Saramo, C., Martikainen, M., & Lehner, O. (2020). Current state and challenges in the implementation of smart robotic process automation in accounting and auditing. ACRN Journal of Finance and Risk Perspectives.

[11] Han, C., Zhang, X., Jiao, H., Wang, M., & Han, T. (2021, December). A review on financial robot process auto-mining based on reinforcement learning. In International Forum on Digital TV and Wireless Multimedia Communications (pp. 471-484). Singapore: Springer Singapore.

[12] Xiong, R., & Zhang, X. (2019, May). The impact of financial robots on enterprise accounting in big data environment. In 1st International Conference on Business, Economics, Management Science (BEMS 2019) (pp. 227-229). Atlantis Press.

[13] Rane, N. L., Choudhary, S. P., & Rane, J. (2024). Artificial Intelligence-driven corporate finance: enhancing efficiency and decision-making through machine learning, natural language processing, and robotic process automation in corporate governance and sustainability. Studies in Economics and Business Relations, 5(2), 1-22.

[14] Mavlutova, I., & Volkova, T. (2019, October). Digital transformation of financial sector and challengies for competencies development. In 2019 7th International Conference on Modeling, Development and Strategic Management of Economic System (MDSMES 2019) (pp. 161-166). Atlantis Press.

[15] Kedziora, D., & KIVIRANTA, H. (2018). Digital Business Value Creation with Robotic Process Automation (rpa) in Northern and Central Europe. Management (18544223), 13(2).

[16] Kokina, J., & Blanchette, S. (2019). Early evidence of digital labor in accounting: Innovation with Robotic Process Automation. International Journal of Accounting Information Systems, 35, 100431.

[17] Lu, F. (2022). Influence of Financial Robot on Financial Management of Small and Medium-sized. International Journal of Management and Education in Human Development, 2(01), 223-225.

[18] Met, İ., Kabukçu, D., Uzunoğulları, G., Soyalp, Ü., & Dakdevir, T. (2020). Transformation of business model in finance sector with artificial intelligence and robotic process automation. Digital business strategies in blockchain ecosystems: Transformational design and future of global business, 3-29.

[19] Golić, Z. (2019). Finance and artificial intelligence: The fifth industrial revolution and its impact on the financial sector. Zbornik radova Ekonomskog fakulteta u Istočnom Sarajevu, (19), 67-81.

[20] Ravi, V., & Kamaruddin, S. (2017). Big data analytics enabled smart financial services: opportunities and challenges. In Big Data Analytics: 5th International Conference, BDA 2017, Hyderabad, India, December 12-15, 2017, Proceedings 5 (pp. 15-39). Springer International Publishing.

[21] Sabharwal, C. L., & Anjum, B. (2018, December). Robo-revolution in the financial sector. In 2018 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 1289-1292). IEEE.

[22] Wei, Q. (2024, April). Task Allocation of Financial Robots Based on RPA and Big Data Algorithms. In 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE) (pp. 01-05). IEEE.

[23] Ma, J., & Jia, H. (2022, December). Application of financial robots based on RPA technology in small and medium-sized enterprises. In 2022 International Conference on Knowledge Engineering and Communication Systems (ICKES) (pp. 1-7). IEEE.

[24] Jędrzejka, D. (2019). Robotic process automation and its impact on accounting. Zeszyty Teoretyczne Rachunkowości, (105), 137-166.

[25] Bisht, D., Singh, R., Gehlot, A., Akram, S. V., Singh, A., Montero, E. C., ... & Twala, B. (2022). Imperative role of integrating digitalization in the firms finance: A technological perspective. Electronics, 11(19), 3252.

[26] Wang, L., Shen, Y., Liang, S., Zhou, N., Jia, C., & Shang, P. (2023, April). Application Design of RPA Financial Robot Integrating Financial Big Data and Financial Sharing Services. In International Conference on Computational Finance and Business Analytics (pp. 63-72). Cham: Springer Nature Switzerland.

[27] Kothandapani, H. P. (2023). Applications of Robotic Process Automation in Quantitative Risk Assessment in Financial Institutions. International Journal of Business Intelligence and Big Data Analytics, 6(1), 40-52.

[28] Shi, S., Li, G., & Han, B. (2023, January). Cloud-client integrated designed financial robot based on artificial intelligence. In 2023 7th International Conference on Management Engineering, Software Engineering and Service Sciences (ICMSS) (pp. 41-45). IEEE.

[29] Lomakin, N. I., Popov, I. A., Shohneh, A. B., Maramygin, M. C., & Gorbunova, A. B. (2020, March). AI-system of stock exchange trading robot for financial risk hedging. In International Scientific Conference" Far East Con"(ISCFEC 2020) (pp. 3273-3282). Atlantis Press.

[30] Pan, H. (2019, May). Intelligent finance global monitoring and observatory: A new perspective for global macro beyond big data. In 2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS) (pp. 623-628). IEEE.

[31] Jiayun Li,Zhen Zhang,Hongyu Zhao,Honghu Li & Chao Zhai. (2025). ROV localization in the nuclear reactor pressure vessel using LSTM and an improved adaptive Kalman Filter. Measurement(PE), 116206-116206.

[32] Indira Linginani & Muddana A. Lakshmi. (2025). An intelligent model of pulmonary emphysema detection using adaptive image segmentation and multi-dilated densenet with attention mechanism. Applied Soft Computing112483-112483.

[33] Lu Chen,Xinwei Luo,Hanlu Zhou,Qifan Shen,Long Chen & Chuanming Huan. (2025). Underwater acoustic multi-target recognition based on channel attention mechanism. Ocean Engineering119841-119841.

[34] Xuanhao Huang,Yangfan Li & Xinwei Wang. (2024). Integrating a multi-variable scenario with Attention-LSTM model to forecast long-term coastal beach erosion. The Science of the total environment176257.

[35] Nathalie Wehbe,Hyame Assem Alameddine,Makan Pourzandi & Chadi Assi. (2025). Empowering 5G SBA security: Time series transformer for HTTP/2 anomaly detection. Computers & Security104114-104114.