This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ICASSP49660.2025.10890046, ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

A Risk Prediction Model for Real Estate Corporations Using High-Target Semantic BERT and Improved GRU

Xiaoyan Ma[1*], Peng Zhu[1*], Qinyuan Liu[1†], Zidong Wang[2]
[1]*Department of Computer Science and Technology, Tongji University, Shanghai, China*
[2]*Department of Computer Science, Brunel University London, United Kingdom*

***Abstract*—Accurately predicting real estate enterprise risk is crucial for the national economy. Although some initial works have been made on this topic such as Z-score, support vector machines, and logistic regression, there remains a gap in comprehensive models that can effectively capture the dynamic risk fluctuations from real estate-specific data. As such, a novel prediction model called HRAGRU is proposed for real estate enterprises to forecast potential risk through multimodal data including news reports, policy updates, and stock information in this paper. We first extract the semantic information from news text by using a BERT model optimized for high-target semantic density. Then we investigate the relationships among various data types through a graph neural network (GNN) model with randomly masked edges or nodes. Finally, we establish an improved gated recurrent unit (GRU) model to capture the interactions between new and historical data. The effectiveness of the proposed HRAGRU model is validated using data from A-share and Hong Kong-listed real estate companies, demonstrating its superior performance in forecasting corporate risk indices. Our sources are released at https://github.com/maxiaoyan290/HRAGRU**

***Index Terms*—Real Estate Enterprise Risk, BERT, Graph Neural Network, Gated Recurrent Unit**

I. Introduction

The real estate sector serves as a cornerstone in the development of China's national economy, standing as one of its pivotal industries [1]–[3]. Consequently, accurately predicting its potential future trajectories is critically important. Stock prices, as financial instruments shaped by supply and demand dynamics, reflect public assessments of a company's risk management and overall risk profile [4]–[6]. In addition, corporate financial statements and news data are integral to comprehensive risk assessment. Financial statements provide insights into a company's profitability, liquidity, and solvency, while news data influences investor sentiment and corporate performance [7]–[9]. Furthermore, government policies and macroeconomic controls play a crucial role in determining the susceptibility of businesses to various risks [10]–[12].

Deep learning techniques for stock sequence prediction have attracted considerable research interest due to their superior accuracy and effectiveness in capturing complex patterns in financial data [19]–[21]. Techniques including Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), and Gated Recurrent Unit (GRU) are extensively used to analyze stock data and predict future price movements [13]–[15]. Additionally, pre-trained models like BERT are widely employed to extract the pertinent features from stock-related news, thereby augmenting the predictive efficacy of financial forecasting models [23]–[25].

However, existing methods struggle to effectively analyze multimodal data, including news, policy, and stock information, for comprehensive risk assessment in the real estate domain, as they often overlook the interplay among diverse data modalities. Some models rely solely on stock data, neglecting the impact of corporate news and national policies on stock price fluctuations [16], [17]. Others incorporate news data but do not fully exploit the connections between news and stock data for enhanced prediction accuracy [18], [23]. Moreover, many aspect-based semantic representation models used in news text processing lack a focus on target information, reducing their effectiveness in downstream tasks.

To address these challenges, we introduce HRAGRU, a novel corporate risk index prediction model comprising High-Target-BERT (HTBERT), Random-Masked-GNN (RMGNN), and Attention-GRU (AGRU). Firstly, HTBERT utilizes BERT to generate semantically rich sentence representations tailored to specific targets, with an attention mechanism enhancing their relevance. Secondly, RMGNN applies random masking of graph edges and nodes to clarify interrelationships within the data, effectively mitigating the issue of node scarcity. Finally, AGRU uses an attention mechanism in the GRU gating process, enhancing the model's ability to capture historical and new data correlations.

In summary, the main contributions of our work can be summarized as follows: (1) We propose RMGNN to effectively incorporate stock-related data, combined with a novel HT-BERT model to address the issue of sparse target semantics; (2) We design a real estate enterprise risk index and propose the AGRU to enhance the model's ability to consider correlations within series data; (3) We extensively test our model on real estate enterprise datasets, achieving performance that surpasses baseline models.

II. Method

For each stock, given an input sequence $X = \{x_{t-T+1}, x_{t-T+2}, \ldots, x_t\}$, where T is the sequence length and $x_t = \{x_t^1, x_t^2, x_t^3, x_t^4, x_t^{4+1}, \ldots, x_t^{4+j}\}$ represents five data

*These authors contributed equally to this work.
†Corresponding Author.

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ICASSP49660.2025.10890046, ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

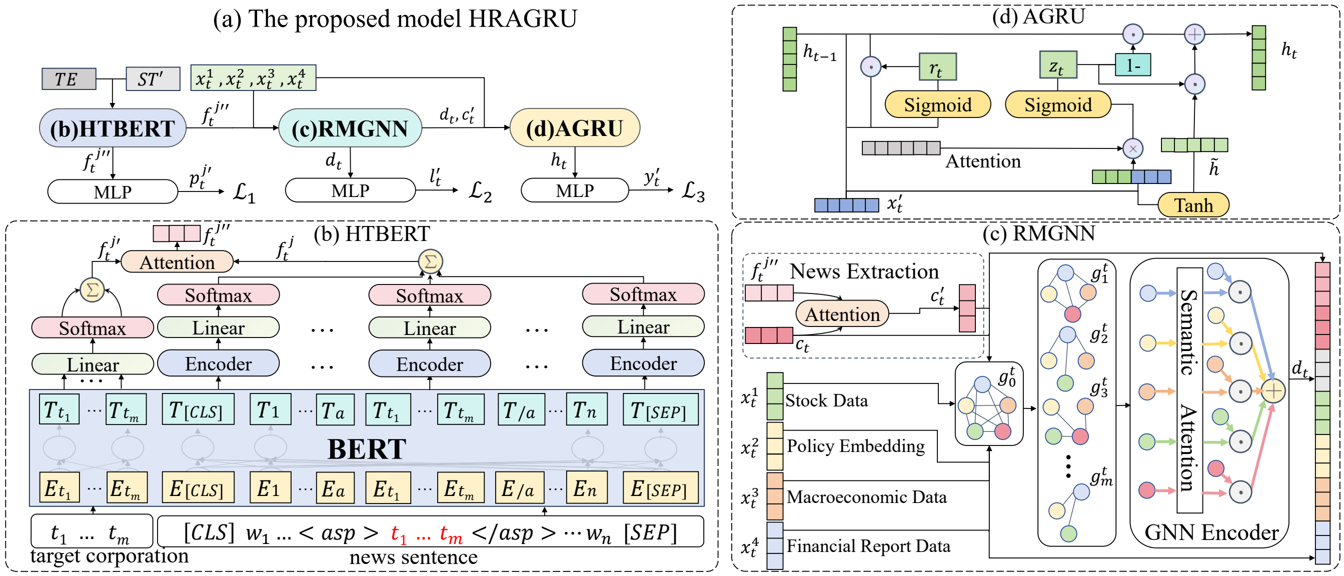

Fig. 1. The structure diagram of model HRAGRU. (a) illustrates the overall architecture of HRAGRU. The HTBERT (b) employs an attention mechanism, intending to denser semantic representations for the target. The RMGNN (c) utilizes a heterogeneous graph neural network with random masking for data integration, enhancing data comprehension. The AGRU (d) employs attention matrices for gate computation to enhance prediction accuracy.

categories. $x_t^1 \in \mathbb{R}^{sd}$, $x_t^2 \in \mathbb{R}^{pd}$, $x_t^3 \in \mathbb{R}^{md}$, and $x_t^4 \in \mathbb{R}^{fd}$ represent the daily stock data, current policy data, national macroeconomic data, and corporate financial data on day t, respectively. Additionally, x_t^{4+j} is the j-th news item on day t. sd, pd, md, and fd denote the vector dimensions of x_t^1, x_t^2, x_t^3, and x_t^4, respectively. We aim to predict the enterprise risk sequence over the next T' trading days, denoted as $Y = \{y_{t+1}, y_{t+2}, \ldots, y_{t+T'}\}$, where y_t represents the real estate enterprise risk index on the t-th day. This index is calculated as $y_t = \text{rank}(\frac{close_t - close_{t-10}}{close_{t-10}})$ with $close_t$ being the 5-day exponential moving average of the closing price on the t-th day. The function rank ranks each real estate enterprise's stock volatility, yielding percentage-based results. The value y_t ranges from 0 to 1, with values above 0.5 indicating higher risk and below 0.5 indicating lower risk.

A. High-Target-BERT

We utilize the pre-trained BERT model to extract semantic information from each entry. Formally, the original news sentence is represented as *ST* = {[CLS], w_1, ..., t_1, ..., t_m, ..., w_n, [SEP]}, where *ST* comprises *n* words, with each w_i denoting a word in *ST*. The target enterprise, *TE* = {t_1, ..., t_m}, consists of *m* words, where each t_i is a specific word in *ST* that belongs to *TE*. Following Ma et al. [26], we introduce explicit markers <asp>and </asp>to denote aspect boundaries: *ST'* = {[CLS], w_1, ..., <asp>, t_1, ..., t_m, </asp>, ..., w_n, [SEP]}.

Inspired by the work of Karimi et al. [27], we use a semantic sequential layer to acquire more profound linguistic knowledge, using ST' as input, and ultimately apply an attention mechanism to obtain the final representation by incorporating the target company's semantics. As is shown in Figure 1, after applying a transformer encoder [31] to each hidden layer, a softmax layer is added to integrate global features. The preliminary representation $f_t^j \in \mathbb{R}^d$ of x_t^{4+j} is computed as

$$f_t^j = \frac{\sum i = 1^n (softmax(W_s o_i + b_s))}{n} \tag{1}$$

where $W_s \in \mathbb{R}^{d \times b}$ and $b_s \in \mathbb{R}^d$ are trainable parameters. Here, $o_i \in \mathbb{R}^b$ is the output for the *i*-th word from the encoder, with b as the output dimension of the encoder and d as the dimension of the semantic representation vector input to the attention layer. The representation of the target entity within x_t^{4+j}, denoted as $f_t^{j'} \in \mathbb{R}^d$, is similarly defined by:

$$f_t^{j'} = \frac{\sum_1^m (softmax(W_t o_i^{'} + b_t))}{m} \tag{2}$$

where $o_i^{'} \in \mathbb{R}^b$ represents the feature representation output for the *i*-th word, $W_t \in \mathbb{R}^{d \times b}$ and $b_t \in \mathbb{R}^d$ are trainable parameters.

Having obtained f_t^j and $f_t^{j'}$, we apply the attention mechanism, calculating the normalized scores for the news data x_t^{4+j} and the target entity as $s_t^j = \frac{W_q f_t^{j'} \odot W_k f_t^j}{\sqrt{a}}$ and $a_t^j = \frac{W_q f_t^{j'} \odot W_k f_t^{j'}}{\sqrt{a}}$, respectively. Finally, the semantic dense representation $f_t^{j''}$ is formulated as follows:

$$f_t^{j''} = \frac{e^{s_t^j} \cdot W_v f_t^j + e^{a_t^j} \cdot W_v f_t^{j'}}{e^{s_t^j} + e^{a_t^j}} \tag{3}$$

where W_q, W_k, and $W_v \in \mathbb{R}^{a \times d}$ are trainable transformer attention parameters, a is the dimension of the semantic representation vector output from the attention layer. $\odot$ represents the Hadamard product, e denotes the base of the natural logarithm, $\cdot$ signifies scalar multiplication with matrices.

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ICASSP49660.2025.10890046, ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

B. Random-Masked-GNN

In most cases, the number of daily news items for real estate companies far exceeds other types of data. Therefore, we preprocess the news data by averaging all articles for the day to capture overall sentiment. An attention mechanism then weights each article based on its correlation with this sentiment, as shown below.

$$c_t = \frac{\sum_{j=1}^{n'}(f_t^{j''})}{n'}, c_t^{'} = c_t || \sum_{j=1}^{n'} \left(\frac{e^{\frac{W_Q c_t \odot W_K f_t^{j''}}{\sqrt{a}}}}{\sum_{j=1}^{n'} e^{\frac{W_Q c_t \odot W_K f_t^{j''}}{\sqrt{a}}}} W_V f_t^{j''} \right) \tag{4}$$

where n' represents the total number of news articles on the t-th day, $c_t \in \mathbb{R}^a$ is the clustered news representation, and $c_t^{'} \in \mathbb{R}^{2a}$ is the final feature representation for that day. W_Q, W_K, $W_V \in \mathbb{R}^{a \times a}$ are the trainable parameters. $||$ denotes vector concatenation.

The graph comprises up to five nodes (x_t^1, x_t^2, x_t^3, x_t^4, $c_t^{'}$), forming a fully connected structure at each time step t, denoted as g_0^t. Random masking is applied to both edges and nodes, resulting in m' generated subgraphs, $G = \{g_1^t, g_2^t, \ldots, g_{m'}^t\}$. The original graph g_0^t, along with m' masked graphs, are utilized in the feature extraction process by the GNN encoder. The encoder computation for each node is as follows:

$$\begin{aligned} o_{ij} &= q^T \odot tanh(W_g(\sigma(W_{\gamma_j}\eta_j)||\sigma(W_{\gamma_i}\eta_i)) + b_g) \\ \beta_{ij} &= \frac{e^{o_{ij}}}{\sum_{j \in N_i} e^{o_{ij}}}, \eta_i' = \sum_{j \in N_i} \beta_{ij} \cdot W_{\gamma_j}\eta_j + W_{\gamma_i}\eta_i \end{aligned} \tag{5}$$

where γ_i represents the node category, σ is the activation function, $q^T \in \mathbb{R}^O$, $W_g \in \mathbb{R}^{O \times 2P}$, and $b_g \in \mathbb{R}^O$ are all learnable parameters. O and P are the dimensions of the attention vectors and graph nodes, respectively. W_{γ_i} and W_{γ_j} are learnable parameters for node categories i and j, respectively. η_i and η_j are nodes vectors, N_i is the neighbor set of node i, o_{ij} is the attention value, β_{ij} is the attention coefficient, and $\eta_i' \in \mathbb{R}^P$ is the new node vector after GNN transformation.

The final integrated data is obtained by concatenating the hierarchical graph representations from g_0^t and $G = \{g_1^t, g_2^t, \ldots, g_{m'}^t\}$, as shown in the following equation:

$$d_t = \frac{\sum_{i \in g_0^t} \eta_i'}{\mathrm{Num(g_0^t)}} || \frac{\sum_{i \in g_1^t} \eta_i'}{\mathrm{Num(g_1^t)}} || \ldots || \frac{\sum_{i \in g_{m'}^t} \eta_i'}{\mathrm{Num(g_{m'}^t)}} \tag{6}$$

where d_t denotes the ultimate representation of stock-related data on day t. The function Num calculates the total number of nodes in the graph $g_{m'}^t$.

C. Attention-GRU

The traditional GRU computes the update gate using linear transformations and activation functions, neglecting dynamic correlations between input data and hidden states. We address this challenge by enhancing the GRU with an attention mechanism that evaluates correlations between input data and previous hidden states and improves the accuracy of the update gate calculation, while maintaining the overall structure of the GRU except for the update gate computation.

$$\begin{aligned} x_t^{'} &= c_t^{'}||d_t||x_t^1||x_t^2||x_t^3||x_t^4 \\ r_t &= \sigma(U_r \odot x_t^{'} + W_r \odot h_{t-1} + b_r) \\ z_t &= \sigma(Att(Tanh(W_x x_t^{'} + b_x)||Tanh(W_h^{'} h_{t-1} + b_h)) + b_z) \\ \tilde{h} &= Tanh(U_h x_t^{'} + W_h(r_t \cdot h_{t-1}) + \tilde{b}) \\ h_t &= (1 - z_t) \cdot \tilde{h} + z_t \cdot h_{t-1} \end{aligned} \tag{7}$$

where $x_t^{'} \in \mathbb{R}^I$, h_{t-1}, $h_t \in \mathbb{R}^H$ represent the input, previous hidden state, and currently hidden state vectors, respectively. σ is the activation function sigmoid. r_t denotes the internal reset vector, z_t represents the update vector, and $\tilde{h} \in \mathbb{R}^H$ indicates the new memory content. $U_r \in \mathbb{R}^I$, $W_r \in \mathbb{R}^H$, b_r, $U_h \in \mathbb{R}^{H \times I}$, $W_h \in \mathbb{R}^{H \times H}$, $\tilde{b} \in \mathbb{R}^H$ are learnable parameters corresponding to the reset gate and new memory content. $I = 2a + m' \times P + sd + pd + fd + md$ and H denote the input and hidden sizes. $Att \in \mathbb{R}^{2D}$ represents the learnable attention function, where D denotes the dimension of the attention vectors. $W_x \in \mathbb{R}^{D \times I}$, $W_h^{'} \in \mathbb{R}^{D \times H}$, $b_x \in \mathbb{R}^D$, $b_h \in \mathbb{R}^D$, and b_z are learnable parameters.

D. Loss Function

To ensure the interpretability of the proposed HRAGRU, we design the unique loss functions for the respective module. As is shown in Figure 1, the loss function can be computed as follows:

$$\begin{aligned} \mathcal{L}_1 &= \sum_{t=1}^{T} \frac{\sum_{j \in S_t}(p_t^{j'} - p_t^j)^2}{\mathrm{Num'(S_t) + e^{-5}}}, \mathcal{L}_2 = \sum_{t=1}^{T}(l_t^{'} - l_t)^2 \\ \mathcal{L}_3 &= \sum_{t=T+1}^{T+T'}(y_t^{'} - y_t)^2, \mathcal{L} = \lambda_1\mathcal{L}_1 + \lambda_2\mathcal{L}_2 + \lambda_3\mathcal{L}_3 \end{aligned} \tag{8}$$

where S_t denotes the set of all news items on day t and the function $\mathrm{Num'}$ retrieves individual news items from S_t. The predictions $p_t^{j'}$, $l_t^{'}$, and $y_t^{'}$ are obtained from $f_t^{j''}$, d_t, and h_t through a multilayer perceptron (MLP). $p_t^{j'}$ and p_t^j represent the predicted sentiment polarity and the corresponding sentiment label of the j-th news item on the day t, with positive and negative sentiments labeled as 1 and 0, respectively. Similarly, $l_t^{'}$ and l_t denote the predicted and actual real estate enterprise risk indices for day t, derived from the graph-level representation. $y_t^{'}$ and y_t represent the time series model's predicted and actual real estate enterprise risk indices on the day t. Data from days 1 to T are used to predict the risk index for days $T+1$ to $T+T'$. λ_1, λ_2, and λ_3 adjust the weights of the different loss functions.

III. EXPERIMENTS

A. Datasets

Our dataset originates from the quantitative data platform Akshare[1]. We have curated a sample of 125 real estate enterprises, encompassing daily data spanning from January

[1]akshare.akfamily.xyz

3, 2018, to June 30, 2024, as AKshare data had only been updated through the end of 2024 at the time of the experiment. These enterprises are publicly listed on China's A-shares and Hong Kong stock exchanges. The dataset is partitioned into a training period from January 3, 2018, to December 31, 2021, and a testing period from January 3, 2022, to June 30, 2024. The training subset comprises approximately 120,000 samples, while the testing subset consists of 75,000 samples. Each temporal sequence in the dataset involves forecasting the corporate risk index at the 31st time step using the preceding 30 time steps as input.

B. Baselines and Settings

We compare our proposed HRAGRU model with the following baselines: **CNN** [29], **LSTM** [28], **GRU** [30], **NGCU** [32], **Transformer** [31], **CTTS** [22], **FactorVAE** [33], **TRA** [34], **AlphaStock** [35], **DeepTrader** [36], **Mamba** [37], **xLSTM** [38], and **PatchTST** [40], incorporating **BERT** [39] for news representation extraction as part of the input. In the baseline model comparison experiment, we normalized and concatenated the various input data types used in HRAGRU and fed them into the baseline models, adjusting them to utilize multimodal data.

In our experiments, we set the $batch\_size$ to 16 and the number of epochs $epoch$ to 100 to ensure stable training. For the HTBERT module, parameters d and b are set to 10 and 768, respectively. In the RMGNN task, two masked graph versions are generated with a 50% probability of masking nodes and edges, with O and P set to 15 and 20, respectively. The AGRU module is trained with a 0.2 dropout rate and D and H are set to 10 and 256, respectively. The loss function weights, $\lambda_1, \lambda_2, \lambda_3$, are 0.2, 0.3, and 0.5. To ensure the statistical significance of the performance improvements, we trained each model in both the baseline and ablation experiments 8 times. After confirming model convergence, we averaged the predictions from these 8 runs to obtain the final results.

C. Prediction Performance

In our experiments, model performance is assessed using Mean Squared Error (MSE), Accuracy (ACC), Recall, and F1-score (F1) metrics applied to the testing set. Enterprises are categorized as low or high risk based on a 0.5 threshold for predicted risk indices, with values below 0.5 classified as high risk and those above 0.5 classified as low risk. As shown in Table I, although traditional models (CNN, LSTM, GRU, NGCU, TRA, and Transformer) perform well, they fail to fully capture the complex relationships between new and historical data. Advanced models (AlphaStock, DeepTrader, CTTS, FactorVAE, Mamba, xLSTM, and PatchTST) offer improved capabilities in capturing dynamic temporal relationships, yet they still fall short of optimal performance due to challenges in fully leveraging data interrelationships. Our model, as shown in Table I, surpasses these baseline models, achieving the lowest MSE and the highest ACC, Recall, and F1-score, highlighting the benefits of integrating intrinsic temporal relationships in forecasting corporate risk indices.

TABLE I
THE INDEX OF THE BASELINE EXPERIMENTS AND ABLATION EXPERIMENTS

model	MSE	ACC	Recall	F1
CNN	0.1618	0.5073	0.0961	0.1001
LSTM	0.0850	0.6378	0.5479	0.6137
GRU	0.0775	0.6450	0.6001	0.6225
NGCU	0.1361	0.4023	0.6193	0.5102
TRA	0.1166	0.4350	0.2287	0.2668
Transformer	0.1193	0.4672	0.1425	0.1968
AlphaStock	0.1122	0.4753	0.0556	0.2120
DeepTrader	0.1350	0.5048	0.3494	0.3992
CTTS	0.0746	0.6597	0.6912	0.6832
FactorVAE	0.0738	0.6771	0.6802	0.6793
Mamba	0.0755	0.6730	0.5425	0.6728
xLSTM	0.0760	0.6714	0.5601	0.6711
PatchTST	0.0734	0.6766	0.6941	0.6842
HRAGRU-(b)+BERT	0.0730	0.6755	0.6959	0.6864
HRAGRU-(c)	0.0728	0.6742	0.6923	0.6850
HRAGRU-(d)+GRU	0.0743	0.6735	0.6901	0.6821
HRAGRU	**0.0726**	**0.6798**	**0.7028**	**0.6876**

D. Ablation Study

Furthermore, we conduct ablation experiments to assess the efficacy of the three proposed modules. Table I summarizes the results, showing that the full model, which integrates all three modules, delivers the best predictive accuracy for forecasting corporate risk indices. This underscores the critical roles of HTBERT (b), RMGNN (c), and AGRU (d). HTBERT excels at extracting key information from textual news data, while RMGNN effectively identifies and analyzes complex interrelationships within the dataset. When combined with AGRU, these modules significantly enhance the model's ability to uncover intrinsic data relationships. Moreover, AGRU demonstrates a clear advantage over traditional GRU within the same architecture, owing to its enhanced capability to capture relationships across varying time intervals, further solidifying its contribution to the model's overall performance.

IV. CONCLUSION

In this paper, we have proposed a novel time series model HRAGRU for predicting real estate enterprise risk, addressing limitations in extracting insights from real estate data by capturing interrelationships within the data. We have also introduced the real estate enterprise risk index as a metric to track evolving risk profiles. Empirical findings have confirmed the superior predictive performance of our proposed model relative to baseline methods, verifying its effectiveness in real estate enterprise risk prediction and supporting the use of advanced data mining techniques in risk assessment.

V. ACKNOWLEDGE

This work was supported by the National Key Research and Development Program of China under Grant 2022YFB4501704, the National Science Foundation of China under Grants 62222312 and 62473285, the Shanghai Science and Technology Innovation Action Plan Project under Grant 22511100700, and the Fundamental Research Funds for the Central Universities.

This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ICASSP49660.2025.10890046, ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

REFERENCES

[1] Y. Huang, J. Khan, E. Girardin, and U. Shad, "The role of the real estate sector in the structural dynamics of the Chinese economy: An input-output analysis," *China World Econ.*, pp. 61–86, 2021.

[2] C. Liu, Y. Zheng, Q. Zhao, *et al.*, "Financial stability and real estate price fluctuation in China," *Physica A: Statistical Mechanics and its Applications*, pp. 122980, 2020.

[3] D. Zhang, J. Cai, J. Liu, *et al.*, "Real estate investments and financial stability: Evidence from regional commercial banks in China," *Eur. J. Finance*, pp. 1388–1408, 2018.

[4] S. A. Ross, R. W. Westerfield, and B. D. Jordan, "Fundamentals of corporate finance," 2022.

[5] Y. Li and Z. Zhao, "The dynamic impact of intellectual capital on firm value: Evidence from China," *Appl. Econ. Lett.*, pp. 19–23, 2018.

[6] Q. Meng, X. Ni, and J. Zhang, "Share pledging and corporate risk-taking: Insights from the Chinese stock market,"*[EB/OL]*, 2019.

[7] F.-T. Tsai, H.-M. Lu, and M.-W. Hung, "The impact of news articles and corporate disclosure on credit risk valuation," *J. Bank. Finance.*, pp. 100–116, 2016.

[8] L. A. Smales, "News sentiment and bank credit risk," *J. Empirical Finance*, pp. 37–61, 2016.

[9] T. D. Huynh and Y. Xia, "Climate change news risk and corporate bond returns," *J. Financial Quant. Anal.*, pp. 1985–2009, 2021.

[10] M. Gertler, N. Kiyotaki, and A. Queralto, "Financial crises, bank risk exposure and government financial policy," *J. Monet. Econ.*, pp. S17–S34, 2012.

[11] H. Chenet, J. Ryan-Collins, and F. Van Lerven, "Finance, climate-change and radical uncertainty: Towards a precautionary approach to financial policy," *Ecological Economics*, pp. 106957, 2021.

[12] M. Das and H. Ordal, "Macroeconomic stability or financial stability: How are capital controls used? Insights from a new database," *J. Financial Stability*, vol. 63, p. 101067, 2022.

[13] S. Selvin, R. Vinayakumar, E. Gopalakrishnan, V. K. Menon, and K. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," *ICACCI*, pp. 1643–1647, 2017.

[14] Y. Gao, R. Wang, and E. Zhou, "Stock prediction based on optimized LSTM and GRU models," *Sci. Program.*, pp. 4055281, 2021.

[15] T. Muhammad, A. B. Aftab, M. Ibrahim, M. M. Ahsan, M. M. Muhu, S. I. Khan, and M. S. Alam, "Transformer-based deep learning model for stock price prediction: A case study on Bangladesh stock market," *Int. J. Comput. Intell. Appl.*, p. 2350013, 2023.

[16] S. Mehtab, J. Sen, and A. Dutta, "Stock price prediction using machine learning and LSTM-based deep learning models," *Mach. Learn. Metaheuristics Algorithms Appl.*, pp. 88–106, 2021.

[17] Y. Liu, Z. Wang, and B. Zheng, "Application of regularized GRU-LSTM model in stock price prediction," *ICCC*, pp. 1886–1890, 2019.

[18] X. Weng, X. Lin, and S. Zhao, "Stock price prediction based on LSTM and BERT," *ICMLC*, pp. 12–17, 2022.

[19] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, and A. Troncoso, "Deep learning for time series forecasting: A survey," *Big Data*, pp. 3–21, 2021.

[20] W. Jiang, "Applications of deep learning in stock market prediction: Recent progress," *Expert Syst. Appl.*, pp. 115537, 2021.

[21] Z. Hu, Y. Zhao, and M. Khushi, "A survey of forex and stock price prediction using deep learning," *Appl. Syst. Innov.*, pp. 9, 2021.

[22] Z. Zeng, R. Kaur, S. Siddagangappa, S. Rahimi, T. Balch, and M. Veloso, "Financial time series forecasting using CNN and transformer," *arXiv preprint arXiv:2304.04912*, 2023.

[23] J. Kim, H.-S. Kim, and S.-Y. Choi, "Forecasting the S&P 500 index using mathematical-based sentiment analysis and deep learning models: A FinBERT transformer model and LSTM," *Axioms*, pp. 1–22, 2023.

[24] Q. Chen, "Stock movement prediction with financial news using contextualized embedding from BERT," *arXiv preprint arXiv:2107.08721*, 2021.

[25] M. G. Sousa, K. Sakiyama, L. de Souza Rodrigues, *et al.*, "BERT for stock market sentiment analysis," in *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, 2019, pp. 1597–1601

[26] F. Ma, C. Zhang, B. Zhang, and D. Song, "Aspect-specific context modeling for aspect-based sentiment analysis," *CCF ICCN*, pp. 513–526, 2022.

[27] A. Karimi, L. Rossi, and A. Prati, "Improving BERT performance for aspect-based sentiment analysis," *arXiv preprint arXiv:2010.11731*, 2020.

[28] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, pp. 1735–1780, 1997.

[29] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," *IEEE*, pp. 156–165, 2017.

[30] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *NIPS*, pp. 1–9, 2014.

[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, pp. 6000–6010, 2017.

[32] J. Wang, X. Li, J. Li, Q. Sun, and H. Wang, "NGCU: A new RNN model for time-series data prediction," *Big Data Res.*, p. 100296, 2022.

[33] Y. Duan, L. Wang, Q. Zhang, and J. Li, "FactorVAE: A probabilistic dynamic factor model based on variational autoencoder for predicting cross-sectional stock returns," *AAAI*, pp. 4468–4476, 2022.

[34] H. Lin, D. Zhou, W. Liu, and J. Bian, "Learning multiple stock trading patterns with temporal routing adaptor and optimal transport," *SIGKDD*, pp. 1017–1026, 2021.

[35] J. Wang, Y. Zhang, K. Tang, J. Wu, and Z. Xiong, "AlphaStock: A buying-winners-and-selling-losers investment strategy using interpretable deep reinforcement attention networks," *SIGKDD*, pp. 1900–1908, 2019.

[36] Z. Wang, B. Huang, S. Tu, K. Zhang, and L. Xu, "DeepTrader: A deep reinforcement learning approach for risk-return balanced portfolio management with market conditions embedding," *AAAI*, pp. 643–650, 2021.

[37] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint arXiv:2312.00752*, 2023.

[38] M. Beck, K. Pöppel, M. Spanring, A. Auer, O. Prudnikova, M. Kopp, G. Klambauer, J. Brandstetter, and S. Hochreiter, "X-LSTM: Extended long short-term memory," *arXiv preprint arXiv:2405.04517*, 2024.

[39] J. Kenton, M.-W. Devlin, C. Chang, and L. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL-HLT*, pp. 2, 2019.

[40] N. Nie, N. H. Nguyen, P. Sinthong, et al., "A time series is worth 64 words: Long-term forecasting with transformers," *arXiv preprint arXiv:2211.14730*, 2022.