ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Research on corporate governance information expression and investor behavior empowered by digital media art

Qianwen Lei[1,*], *Tingting Wei*[1], *Zixuan Yang*[1] *and Chen Chen*[1]

[1] Art Design College, ChongQing JiaoTong University, ChongQing 400000, ChongQing, China

SUMMARY: *This study proposes a computational analysis-oriented framework for analyzing how digital media art can enhance corporate governance information expression and influence investor behavior. A multi-modal data set containing 3240 corporate governance disclosure samples and 25600 investor interaction records was constructed, covering the governance chapter of the annual report, the ESG governance page, the investor relations webpage and the governance theme visualization page. A structured coding module is designed to convert visual rhetoric and governance semantics into a unified feature tensor, and a cross-modal fusion analysis module for investor behavior recognition is constructed to incorporate governance text nodes, visual structure nodes and interaction logs into a unified discriminant space. The experimental results show that the proposed method achieves 87.41% classification accuracy, 0.861 F1 value, 0.912 AUC and 0.793 MCC. The results show that the expression of corporate governance information empowered by digital media art can be effectively transformed into computable behavior representation, and provide support for the analysis of the expression effect of governance information and the identification of investor behavior.*

KEYWORDS: *Multi-modal representation learning; Cross-modal fusion; Governance information computing; Investor behavior recognition*

1 Introduction

Digital disclosure scenarios are reshaping how corporate governance information is generated, organized, and delivered. The visual page of the annual report, the ESG thematic interface, the investor relations platform, the short video road show and the interactive announcement together constitute the governance expression space. The information is no longer limited to static text, but turns to a composite media structure composed of text, graphics, color, layout, timing interaction and feedback trajectory. After digital media art enters the communication chain of corporate governance, the readability, emotional directivity and attention allocation mode of governance content change, and the path for investors to receive information changes from linear reading to dynamic perception across interfaces, cross-modalities and cross-contexts. Facing this change of expression, it is difficult to describe the influence process of governance information on investor judgment by only relying on financial indicators or traditional text analysis. Combining the visual narrative mechanism in digital media art with the multimodal representation, behavior modeling and fusion analysis in the computer field can help to explain how the governance expression affects investors' click, stay,

*632314050401@mails.cqjtu.edu.cn
https://doi.org/10.65102/is2026743
Published: 30 April, 2026

evaluation and trading tendency from a computational perspective, and then improve the accuracy of information disclosure analysis and the stability of behavior recognition.

Focusing on information perception and behavior prediction in financial markets, existing research has gradually shifted from single numerical modeling to joint analysis of multi-source heterogeneous data. Liu et al. [1] studied the stock index prediction mechanism driven by news, and constructed a prediction model combining grid network and emotional attention, indicating that external information expression forms can be used as an important input for market identification. Lee et al. [2] proposed a deep learning prediction method combining ESG sentiment and technical indicators, which strengthened the explanatory role of non-financial semantic signals in market judgment. Deng et al. [3] studied stock price crash risk prediction based on multi-emotion fusion, and constructed an interpretable ensemble learning framework to promote the linkage analysis of emotion structure and financial risk identification. Sheng et al. [4] proposed a stock price crash prediction model based on multi-modal data, and proved that the joint representation of text, image and market variables can improve the prediction performance. Du et al. [5] introduced reinforcement learning and multimodal models into portfolio management tasks, extending the application boundaries of behavioral response modeling in investment decision scenarios. Ferraro et al. [6] studied the influence of social media user-generated content on stock prediction, indicating that data correlation has been formed between platform-based communication and investor reaction. Leippold et al. [7] proposed a joint analysis path of technology pattern and news sentiment, which reflects the dependence of market behavior recognition on collaborative expression of heterogeneous information. An et al. [8] constructed a multi-modal integrated prediction framework based on social media, showing the adaptation ability of cross-source feature fusion in price prediction. Dahal et al. [9] compared the role of news sentiment in deep learning stock price prediction and verified the stable contribution of semantic features in different model structures. Zhao et al. [10] proposed a deep learning ensemble framework for stock price volatility prediction, which provided a unified modeling idea for behavior trend recognition under complex information input.

Existing results provide a solid foundation for the calculation of investor behavior, but the research focuses on news text, market sequence and social media sentiment. For the expression structure, visual rhetoric and interactive forms of corporate governance information empowered by digital media art, there is still a lack of a set of computational analysis paths that can simultaneously describe content semantics, interface styles and investor feedback. Corporate governance information has institutional attributes and signal attributes, and its expression not only affects the information accessibility, but also affects the investor's perception intensity of governance transparency, management credibility and enterprise risk posture. Based on this, this paper constructs a corporate governance information expression and investor behavior computational analysis framework powered by digital media art, introduces multi-modal visual representation and structured coding mechanism into the governance disclosure scenario, and designs a cross-modal fusion analysis module for investor behavior recognition to realize the unified modeling among governance information expression characteristics, investor response signals and behavior results. The goal of this paper was to enhance the computability of governance information expression and the interpretability of investor behavior recognition, and provide technical support for intelligent analysis, interface generation and behavior prediction in the digital disclosure environment.

2 Related Research

After digital media art entered the field of corporate governance communication, related research gradually shifted from single text interpretation to multi-source information collaborative modeling. The existing results are mainly distributed in the directions of news-driven market prediction, financial sentiment recognition, domain semantic representation, investor sentiment quantification, and interpretable behavior analysis. These studies provide a methodological foundation for computational analysis of governance information expression, as well as technical references for visual encoding, semantic organization, and behavioral feedback modeling in digital disclosure interfaces.

Focusing on news-oriented market volatility recognition, Xu et al. studied the process of stock price movement prediction oriented to news information, and proposed a multi-channel cross-residual deep learning framework to map multi-source news features into a unified discriminant space, which enhanced the coupling ability of time series signals and semantic cues [11]. Costola et al. studied the connection between machine learning sentiment analysis, epidemic news and stock market reaction, and proposed an analysis path with the linkage of news sentiment intensity and market reaction as the core, indicating that external narrative structure can change investors' risk perception and trading feedback [12]. Agarwal studied the utilization of domain knowledge in financial text emotion recognition, and proposed a financial emotion analysis model combining knowledge base and domain-specific representation, so that the semantic representation can be close to the reference, rhetoric and implicit attitude in the financial context [13]. Karanikola et al. compared the performance of classical methods and deep learning models in financial emotion analysis, and proposed that the recognition framework should be evaluated from three levels: feature organization, model adaptation and interpretation granularity, so as to improve the transferability of emotion calculation results [14]. Du et al. systematically sorted out the financial sentiment analysis technology and its application, and proposed a complete technical chain from text mining, representation learning, fusion computing to scene deployment, which provided a clear research boundary for multimodal governance information modeling [15].

Another category of studies puts more emphasis on the dynamic link between news structure, platform interaction, and investor feedback. Ashtiani et al. studied the intelligent prediction method of financial market based on news, proposed a review framework of market prediction driven by text mining and machine learning, and further clarified the corresponding relationship between news structure, sentiment intensity and model performance [16]. Groß-krußmann studied deep news sentiment representation learning in macro finance, and proposed that deep news sentiment representation can improve variable interpretation and trend judgment in complex financial environments, which reflects the value of high-level semantic coding in market perception tasks [17]. Divernois et al. studied the relationship between classified emotions of StockTwits and stock returns, and proposed that the fine-grained correlation between emotional tags of social platforms and returns changes could enhance the immediacy and stratification of investor group behavior recognition [18]. Cai et al. studied the effect of real-time investor sentiment on high-frequency return prediction, and proposed a mixed-frequency rolling decomposition prediction method, which dynamically aligned sentiment indicators with high-frequency time series and enhanced the recognition accuracy of short-term behavioral fluctuations [19]. Deng et al. studied the direction prediction of stock index with the participation of investor sentiment, and proposed an interpretable extreme gradient boosting framework to strengthen the interpretation path of results while maintaining the prediction effect [20].

From the above literature, it can be seen that the existing research has formed a continuous technology chain from text emotion recognition to market behavior prediction, but the input objects, modeling center of gravity and application boundaries of different methods are not consistent. In order to more clearly present the differences in technical paths and expression objects of these achievements, Table 1 summarizes the research content, method characteristics, main contributions and expression boundaries of related studies.

Table 1: Summary table of relevant studies

Reference	Research Content	Method Characteristics	Main Contribution	Expression Boundary
[11]	News-oriented stock price movement prediction	Multi-channel cross-residual deep learning	Strengthens the coupling between news features and market sequences	Limited attention to visual expression
[12]	Analysis of news sentiment and market reaction	Machine learning-based sentiment analysis	Connects narrative intensity with market feedback	Focused mainly on news event scenarios
[13]	Semantic representation of financial sentiment	Combination of knowledge base and domain representation	Improves the relevance of financial semantics	Lacks interface-level modeling
[14]	Comparison of financial sentiment methods	Comparison between classical models and deep models	Enhances the clarity of framework evaluation	Visual modality is not included
[15]	Review of financial sentiment analysis	Systematic summarization of the technical pipeline	Clarifies the directions of representation learning and fusion	Mainly a methodological review
[16]	Review of news prediction research	Collaboration of text mining and machine learning	Summarizes prediction processes and variable relationships	Insufficient coverage of governance disclosure scenarios
[17]	Sentiment representation of macro-financial news	Deep news sentiment representation learning	Improves interpretability in complex financial environments	Focuses on macro-level semantics
[18]	Relationship between social platform sentiment and returns	Fine-grained association of sentiment labels	Strengthens real-time identification of group behavior	Limited focus on corporate governance dimensions
[19]	Real-time investor sentiment and high-frequency returns	Mixed-frequency rolling decomposition prediction	Improves the accuracy of short-term fluctuation recognition	Lacks visual structural information
[20]	Stock index direction prediction under investor sentiment	Explainable XGBoost	Balances predictive performance and interpretability	Expression style factors are not explored

The results shown in Table 1 show that most of the current studies take news texts, social media phrases and price sequences as the main input, and pay insufficient attention to the layout organization, graphic narrative, color guidance, interface rhythm and interactive feedback of corporate governance information empowered by digital media art. Investor perception in the governance disclosure scenario is not only determined by the content of the sentence, but also by the visual hierarchy, the ratio of graphics and text, the dynamic presentation and the structural jump. It is difficult to describe the composite influence of the expression of governance information on investors 'browsing, staying, clicking, commenting and decision-making tendencies by simply relying on the text sentiment model.

Based on the above research vein, it can be seen that the combination of digital media art and governance information computational analysis needs to continue to promote multi-modal visual representation, structured coding and cross-modal fusion modeling on the basis of the existing financial emotion computing. In this paper, a computational framework for corporate governance information expression and investor behavior recognition is constructed, which integrates visual rhetoric, semantic features and interaction logs into a unified representation space, and enhances the consistency between governance information expression effect analysis and investor behavior recognition through a cross-modal association mechanism.

3 The framework of corporate governance information expression and investor behavior calculation and analysis empowered by digital media art

3.1 Multi-modal visual representation and structural coding method of corporate governance information

In the digital disclosure environment, corporate governance information has changed from a single text description to a composite expression object containing words, charts, colors, graphic symbols, layout blocks and interactive records. After digital media art enters governance communication, investors' judgments on transparency, robustness and credibility will be continuously affected by the visual hierarchy and interface rhythm. Accordingly, this section constructs the multi-modal visual representation and structural coding method of corporate governance information, so that governance content, visual rhetoric and contact trajectory can be uniformly described in the same computational space.

To illustrate how the original disclosure page enters the unified coding process, Fig. 1 illustrates the initial parsing structure of the multi-modal visual representation of corporate governance information. The left input includes the annual report page, governance announcement, ESG visual card and investor relations interface. The middle part completes the layout slicing, color tone extraction, chart category recognition, text block segmentation and hot spot positioning in turn. The figure emphasizes that governance information is gradually transformed from page objects into structural units.

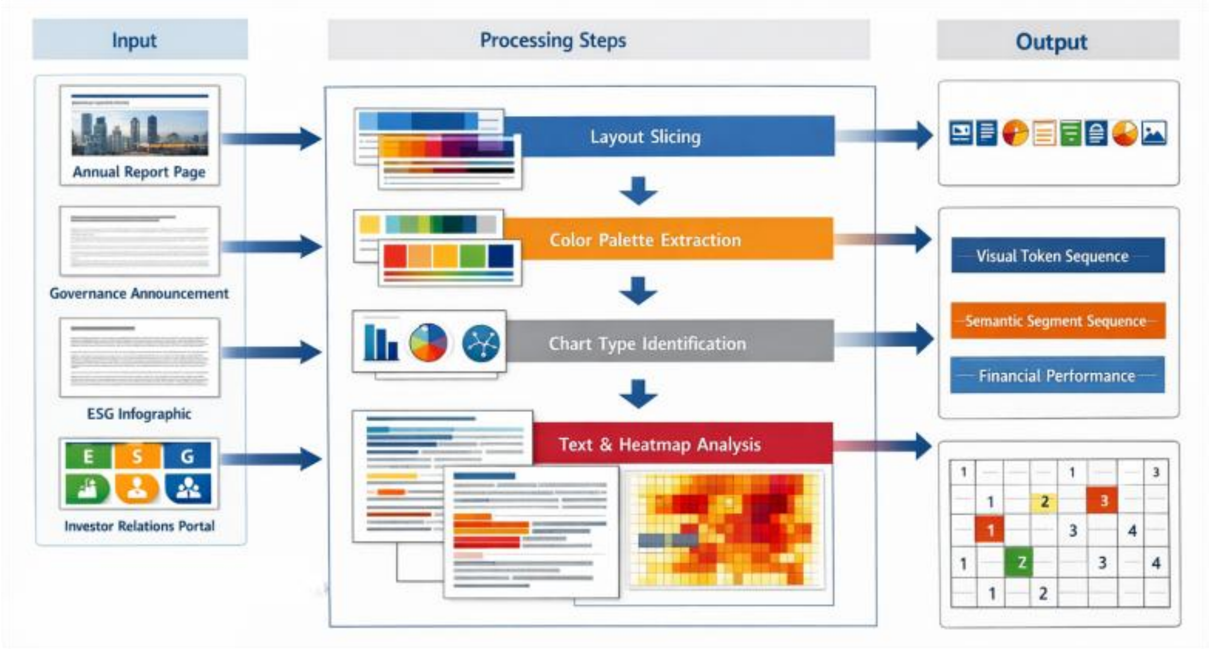

Figure 1: Initial parsing structure of corporate governance information multimodal visual representation

In order to map the visual hierarchy, semantic focus and media cues in the governance disclosure page to the computational space, the cross-modal base embedding representation is defined as follows.

$$E_i = \tanh\left(\Theta_t T_i + \Theta_v V_i + \Theta_p P_i + \Theta_m M_i + b_e\right) \tag{1}$$

Here, E_i represents the base embedding of the i governance expression unit, T_i represents the text semantic vector, V_i represents the visual encoding vector, P_i represents the position encoding, M_i represents the modal labeling, Θ_t, Θ_v, Θ_p, Θ_m are the corresponding mapping matrix, and b_e is the bias term. This formula retains the differences of title, diagram description, color block hint and text statement in a unified dimension, so that the subsequent relationship calculation has a common starting point, and the visual expression and governance semantics no longer belong to two isolated channels.

In order to ensure that the spatial adjacency, visual subordination and information jump relationships in the governance page can be completely retained, the weight function of the layout association graph is further constructed as follows.

$$A_{ij} = \sigma\left(\frac{\lambda_1}{1 + d_{ij}} + \lambda_2 o_{ij} + \lambda_3 e^{-h_{ij}} + \lambda_4 r_{ij}\right)) \tag{2}$$

Here, A_{ij} represents the layout association weight between node i and node j, d_{ij} represents the normalized distance, o_{ij} represents the region overlap proportion, h_{ij} represents the hierarchical difference coefficient, r_{ij} represents the jump or link relationship, λ_1 to λ_4 are learnable parameters, and σ is the compression function. In this formula, subordination, image-text adjacency, data annotation and hint links in the layout are all written into the graph structure to avoid visual expression being cut into disjoint local segments during the encoding stage.

In order to quantify the joint effect of color tone, graphic density and layout rhythm on the governance of expression intensity, the formula of visual rhetorical intensity estimation function is defined as follows.

$$R_i = \alpha \|c_i\|_2 + \beta \ln(1 + g_i) + \gamma l_i + \eta f_i \tag{3}$$

Here, R_i represents the visual rhetorical intensity of the i unit, c_i represents the color contrast vector, g_i represents the graphic complexity, l_i represents the typography rhythm coefficient, f_i represents the font level feature, and α, β, γ, η are the learnable weights. In this formula, the highlight band, key icon, hierarchical title and prompt icon are transcribed into continuous variables, so that the visual organization can be used as an important input to govern the expression strength to participate in the subsequent behavior analysis.

After completing the page-level basic parsing and layout relationship modeling, the computational expression of governance information cannot stay at the element recognition level, but also needs to further characterize the synergy between visual rhetoric and governance semantics. As shown in Fig. 2, the system first extracts visual expression factors such as color distribution, graphic density, font level, block contrast, and chart semantic labels from the governance disclosure page, and encodes these visual variables into continuous style vectors. Then, the text semantic flow and the visual style flow enter the dual processing channel respectively. The text channel is responsible for maintaining the semantic boundaries of the governance statement, the explanatory paragraph, and the indicator annotation, and the visual channel is responsible for quantifying the perceptual strength formed by the highlighted color blocks, the title hierarchy, the graphic composition, and the local emphasis area in the page. After the dual-channel features enter the unified space, the semantic gating module will suppress the irrelevant visual decorations according to the matching degree between the text nodes and the visual nodes, and retain the charts, ICONS and typography hints with explanatory value. After the gated screening, the results and the layout relationship matrix participate in weighted fusion, and finally form the governance node representation with visual rhetorical strength.

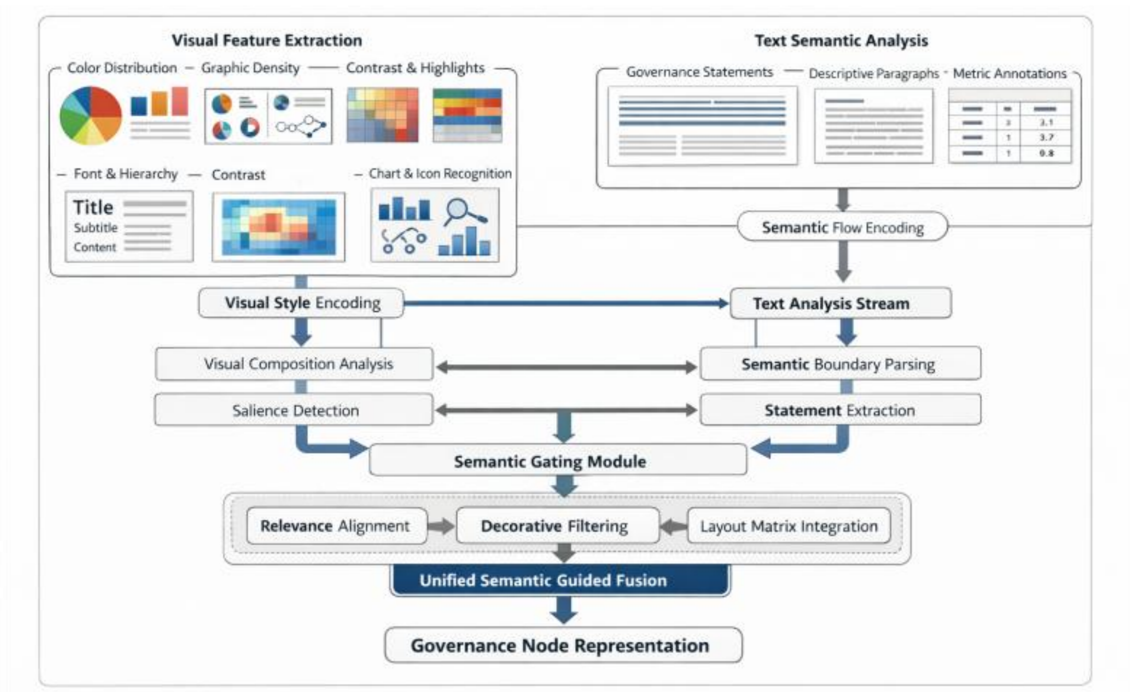

Figure 2: Flowchart of coupling visual expression factor and governance semantics

In order to suppress the interference caused by visual style deviation on governance semantic judgment, the gated alignment calculation formula of visual nodes and text nodes is given below.

$$G_{ij} = \mathrm{sigmoid}\left(\frac{(\Theta_q Q_i)^\top (\Theta_k K_j)}{\sqrt{d}} + \mu_1 A_{ij} + \mu_2 R_j\right) \tag{4}$$

where G_{ij} represents the semantic gating coefficient between text node i and visual node j, Q_i and K_j represent the text query vector and visual key vector respectively, d represents the latent space dimension, R_j is the strength of visual rhetoric, A_{ij} is the weight of layout association, Θ_q, Θ_k are mapping parameters, μ_1, μ_2 are adjustment coefficients. This formula filters the visual elements that really participate in the interpretation according to the semantic proximity, and suppresses the noise caused by invalid decorations through the page relationship and style strength.

In order to compress the local node relationship into a page-level governance expression, it is necessary to perform a hierarchical aggregation calculation on the cross-modal graph structure, which is specifically expressed in the following equation.

$$X_i^{(\ell+1)} = \rho\left(\sum_{j\in\Omega(i)} \frac{A_{ij} G_{ij}}{\sum_{k\in\Omega(i)} A_{ik}\, G_{ik} + \varepsilon}\, \Theta^{(\ell)} \left[X_j^{(\ell)} \| R_j\right]\right) \tag{5}$$

Here, $X_i^{(\ell+1)}$ represents the $\ell + 1$ layer node representation, $\Omega(i)$ represents the neighborhood set of node i, A_{ij} represents the layout association weight, G_{ij} represents the semantic gating coefficient, R_j represents the visual rhetorical strength, $\Theta^{(\ell)}$ represents the ℓ layer transformation matrix, ρ is the nonlinear mapping function, and ε is the stable term. This formula uses spatial, semantic and style constraints at each level of propagation, so that nodes not only retain the meaning of the text, but also absorb the information of diagram composition and emphasis.

After the coupling of visual factors and governance semantics, it is still necessary to integrate local nodes, page structures and investor contact trajectories into a unified coding result. Otherwise, the governance expression features can only stay at the level of scattered intermediate variables, which is difficult to directly serve investor behavior recognition and effect analysis. As shown in Fig. 3, the generation of the unified encoding tensor consists of three parts. The first part is a node-level input layer, which receives the governance node embeddings, layout adjacency relations, semantic gated weights, and visual rhetorical strength obtained in the previous stage to form a cross-modal graph structure with multiple attributes. The second part is the hierarchical aggregation layer, which performs information propagation in the graph neighborhood, and integrates neighboring text nodes, chart nodes, and hint nodes by weighting, so that semantic relations, spatial relations, and visual relations in the local expression can be transferred synchronously, and gradually compressed into more stable intermediate representations. The third part is the page-level output layer, which aligned and fused the aggregated page structure representation with the interaction trajectories such as investor stay time, scroll depth, click hot spot and jump order, and finally output the unified encoding tensor.

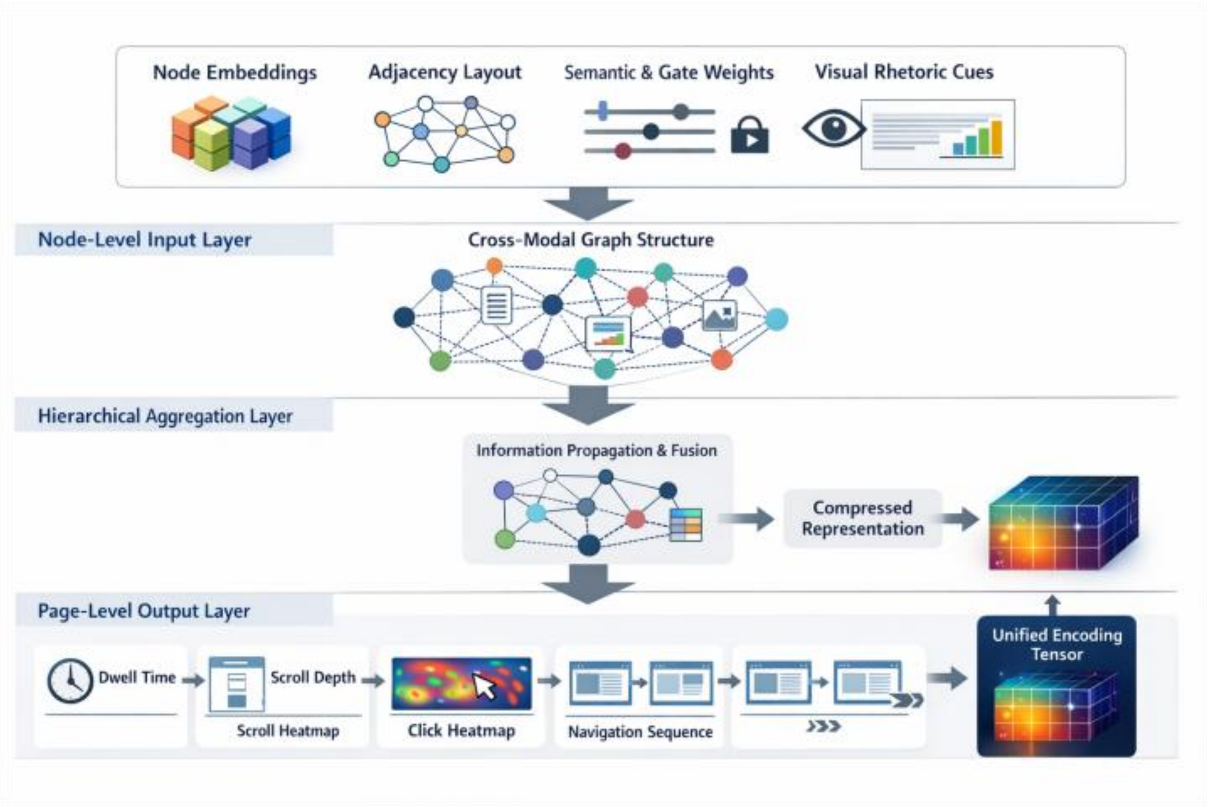

Figure 3: Flowchart of unified encoding tensor generation

In order to form the final structured coding result for investor behavior recognition, the page-level representation and interaction trajectory should be jointly mapped into the output tensor expression as follows.

$$Z = \tanh\left(\Theta_z \mathrm{Pool}\left(X^{(L)}\right) + \Theta_s S + \Theta_u U + b_z\right) \tag{6}$$

where Z represents the final output governance expression tensor, $\mathrm{Pool}\left(X^{(L)}\right)$ represents the global pooling of the highest level node representation, S represents the page-level visual statistics vector, U represents the interaction trajectory encoding formed by stay, scroll, click and jump, Θ_z, Θ_s, Θ_u are the mapping matrix, and b_z is the bias term. This formula integrates the expression of governance content and the contact method of investors into the output space, so that the structured coding naturally has behavior interpretability, and provides direct input for subsequent identification and effect analysis.

Combining the above processes, this paper establishes a complete coding chain in corporate governance information processing, which consists of page parsing, visual rhetorical modeling, semantic gated screening, graph structure aggregation and unified tensor output. This method transforms color organization, graphic emphasis, layout rhythm and interface hierarchy in digital media art into computable variables, while preserving the institutional semantics of the governance text itself and the behavioral clues in the investor contact trajectory.

3.2 Cross-modal fusion analysis module for investor behavior recognition

After completing the multi-modal visual representation and structural coding of corporate governance information, the system also needs to further deal with the heterogeneous differences between text semantics, visual rhetoric and interaction trajectories, otherwise the expression intensity, reading path and investor response in the governance page are difficult to be uniformly mapped into the same discriminant space. To this end, this paper constructs a cross-modal fusion analysis module for investor behavior recognition, which synchronously sends governance text nodes, visual structure nodes, page browsing sequences and operation logs into the hierarchical fusion network to compress redundant visual disturbances and invalid interaction signals while maintaining the continuity of governance semantics. This module does not stop at feature splicing, but integrates governance disclosure content, artistic interface design and investor feedback into a computable response chain.

To illustrate the overall processing sequence of the cross-modal fusion analysis module from multi-source input to behavioral output, Fig. 4 shows the main flow of this module. In the figure, the left input includes governance text coding, visual expression tensor, hot spot click sequence, scrolling depth sequence and stay time series. In the middle, the modal alignment, gated screening, hierarchical recursive aggregation and cross-modal attention diffusion are completed in turn. The governance text flow firstly provides the system semantics and disclosure focus, the visual tensor flow is responsible for providing the layout level, chart composition and color emphasis intensity, and the interaction sequence flow records the user's stay, scroll and click behavior on different page nodes. After the three channels enter the unified timeline, the system first performs local alignment, and then screens out the weak response regions that do not participate in the judgment through the gating unit. Then, the hierarchical recursion and cross-modal attention are used to connect the semantic focus, visual focus and behavioral hot areas.

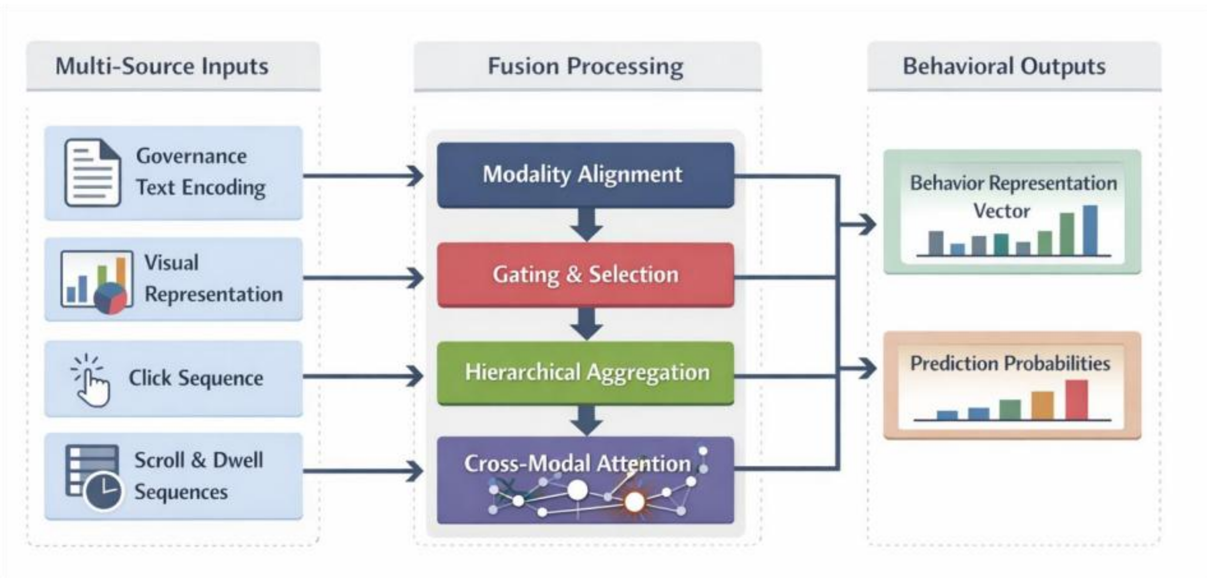

Figure 4: The overall flowchart of the cross-modal fusion analysis module

In order to unify the inputs of different modalities into comparable time coordinates and node indices, this paper first constructs the modal alignment mapping function, whose unified expression form is shown in the following equation.

$$H_t = \phi\left(\Gamma_g G_t + \Gamma_v V_t + \Gamma_b B_t + \Gamma_\tau \tau_t + b_h\right) \tag{7}$$

Here, H_t represents the basic fusion state at the t time, G_t represents the governance semantic node vector, V_t represents the visual expression vector, B_t represents the interaction behavior vector, τ_t represents the time position coding, Γ_g, Γ_v, Γ_b, Γ_τ are mapping parameters, b_h is the bias term, ϕ is a nonlinear function. This formula pulls governance content, visual style, and behavioral trajectory into a unified space.

In order to characterize the selection relationship between governance expression focus and investor contact intensity, the modal gated response function is introduced in the following to form an adjustable information retention mechanism.

$$M_t = \mathrm{sigmoid}(W_m[H_t \| R_t \| C_t] + b_m) \tag{8}$$

Here, M_t represents the gating weight at time t, R_t represents the page hot area response strength, C_t represents the local context consistency vector, W_m and b_m are learnable parameters, and $\|$ represents vector concatenation. This formula adjusts the information retention ratio according to whether the governance node is really contacted and whether the contact is continuous, so that the model gives priority to the expression units directly related to behavior transformation.

In order to ensure that different modes will not be unbalanced due to local abnormal operations in the dynamic propagation phase, this paper constructs a hierarchical recursive update mechanism, whose core calculation form is as follows.

$$S_t^{(\ell)} = \omega_t^{(\ell)} \odot \tilde{S}_t^{(\ell)} + \left(1 - \omega_t^{(\ell)}\right) \odot S_{t-1}^{(\ell)}, \quad \omega_t^{(\ell)} = \sigma\left(U_\ell H_t + Z_\ell S_{t-1}^{(\ell)} + b_\ell\right) \tag{9}$$

Here, $S_t^{(\ell)}$ denotes the recursive state of the ℓ th layer at time t, $\tilde{S}_t^{(\ell)}$ denotes the current candidate state, $\omega_t^{(\ell)}$ denotes the update gate, U_ℓ,Z_ℓ, and b_ℓ are layer parameters, and $\odot$ denotes element-wise multiplication. This formula adjusts the update ratio according to the compatibility degree between the current information and the historical state.

To further illustrate how the hierarchical recursion results enter the high-level association computation, Fig. 5 illustrates the internal structure of cross-modal attention diffusion and action discrimination. The left side of the figure receives the multi-layer recursive state, gated weight and page-level visual focus distribution, the middle part generates the global association graph through query mapping, cross-modal attention allocation and residual integration, and the right side outputs the investor behavioral intention vector, risk preference score and final discrimination result. The recursive states from different layers are first reorganized into a unified set of queries, keys, and values, and the system assigns cross-modal attention weights according to the degree of dependency of governance semantic nodes on visual focus and behavioral hot spots. Subsequently, the attention diffusion module propagates high weight information in the scope of the page, and the residual integration unit re-superimposes the high-level state after diffusion and the original recursive state to finally obtain the comprehensive behavior state.

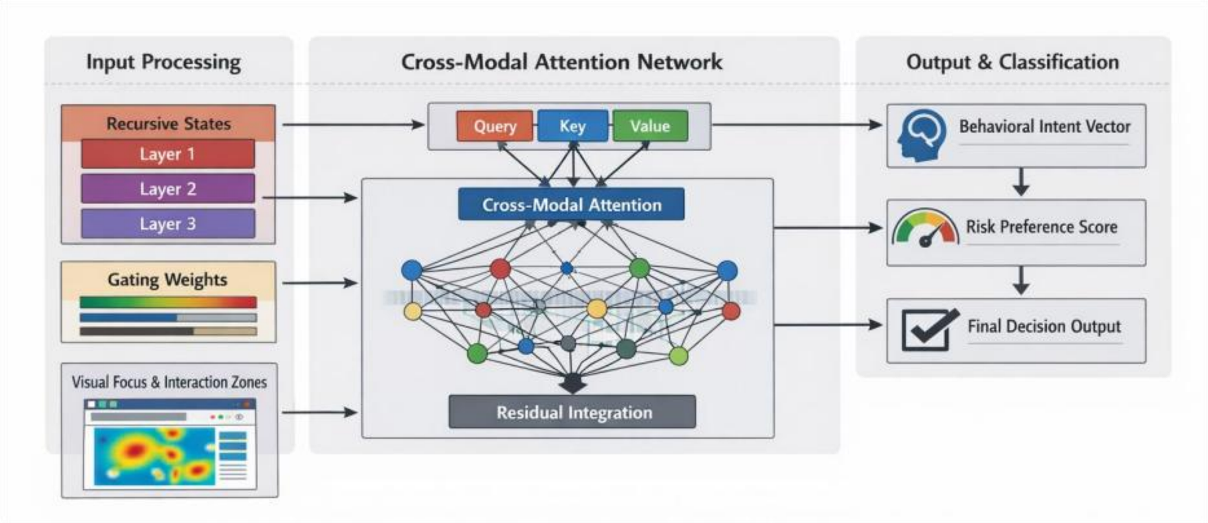

Figure 5: Cross-modal attention diffusion and action discrimination structure diagram

In order to establish the high-order connection between governance nodes, visual focus and interactive hot spots in the global scope, the formula of cross-modal attention diffusion is defined as follows, which is expressed as follows.

$$A_{ij}^{*}=\frac{\exp\left((Q_iK_j^{T})/\sqrt{d}+\kappa_1M_j+\kappa_2P_{ij}\right)}{\sum_{u=1}^{N}\exp\left((Q_iK_u^{T})/\sqrt{d}+\kappa_1M_u+\kappa_2P_{iu}\right)} \tag{10}$$

Here, A_{ij}^{*} represents the cross-modal attention weight of node i to node j, d represents the latent space dimension, M_j represents the gating weight, P_{ij} represents the page structure prior, and κ_1 and κ_2 are the adjustment coefficients. This formula takes semantic relevance, contact intensity and page prior relations into the attention allocation, so that the high-value governance regions maintain stronger communication ability in the global diffusion.

In order to transcribe the correlation information after diffusion into a global behavior state that can be directly discriminated by the recognition model, the residual fusion expression is given below, which is calculated as follows.

$$F=\mathrm{LayerNorm}\left(\sum_{j=1}^{N}A_{ij}^{*}W_fV_j+\Lambda\bar{S}+b_f\right) \tag{11}$$

Here, F represents the global fusion behavior state, V_j represents the value vector, W_f is the linear mapping matrix, $\bar{S}$ represents the equilibrium convergence result of the multi-layer recursive state, Λ is the residual adjustment matrix, b_f is the bias term, and LayerNorm represents the layer normalization. This formula integrates the long-term dependence information obtained by deep diffusion with the local continuous information in the recursive state, so that the final state not only has a global view, but also does not lose the stage change in the real contact process.

In order to form interpretable discrimination results for investor behavior tendency, this paper further constructs a multi-task joint output function, whose structural form is given in the following equation.

$$\hat{y}=\mathrm{softmax}(W_yF+b_y),\quad \hat{r}=\tanh(W_rF+b_r) \tag{12}$$

Here, $\hat{y}$ represents the investor behavior category probability distribution, W_y and b_y are the classification head parameters, $\hat{r}$ represents the risk appetite strength regression results, and W_r and b_r are the regression head parameters. The formula outputs discrete behavior categories and continuous risk preference values at the same time, so that the model can not only identify the behavioral tendencies of continuing reading, collecting, consulting or trading, but also measure the risk acceptance under different governance expression styles.

In order to coordinate the training direction between classification recognition, intensity regression and structural consistency constraints, the module finally adopts a joint loss function, whose comprehensive form is shown below.

$$\mathcal{L} = \alpha\mathcal{L}_{cls} + \beta\mathcal{L}_{reg} + \gamma\mathcal{L}_{con} + \delta\|\Theta\|_2^2 \tag{13}$$

Here, $\mathcal{L}_{cls}$ represents the behavior classification loss, $\mathcal{L}_{reg}$ represents the risk preference regression loss, $\mathcal{L}_{con}$ represents the cross-modal consistency constraint term, $\|\Theta\|_2^2$ represents the parameter regularization term, and α, β, γ, δ are the weight coefficients. This formula ensures that the model maintains the consistency of each modal representation while pursuing recognition accuracy, and avoids the improvement of classification performance at the expense of interpretation structure.

In order to keep the training results stable against page structure perturbations and user behavior fluctuations, this paper continues to introduce a robustness correction term, whose constraint form is as follows.

$$\mathcal{R} = \frac{1}{N}\sum_{i=1}^{N}\left\|F_i - \tilde{F}_i\right\|_2^2 + \xi\sum_{i=1}^{N}\sum_{j=1}^{N}\left(A_{ij}^* - \tilde{A}_{ij}^*\right)^2 \tag{14}$$

Here, $\mathcal{R}$ represents the robustness correction loss, F_i and $\tilde{F}_i$ represent the global fusion state under the original input and perturbed input, A_{ij}^* and $\tilde{A}_{ij}^*$ represent the corresponding attention weights, and ξ is the balance coefficient. By constrainting the state difference and attention structure difference before and after the perturbation, this formula enhances the adaptability of the model to local page revision, graphic replacement and slight user operation noise.

Based on the above process, the cross-modal fusion analysis module for investor behavior recognition further completes the joint modeling of governance semantics, visual expression and interaction trajectory on the basis of unified coding results. Through modal alignment, gated screening, hierarchical recursion and attention diffusion, the module compresses the scattered page content, artistic expression and investor contact behavior into the same discriminant space, so that the content structure, presentation structure and behavior response structure of corporate governance information can be described synchronously.

4 Research results of corporate governance information expression and investor behavior empowered by digital media art

4.1 Investor behavior recognition results based on multimodal governance information modeling

In order to verify the effectiveness of multimodal governance information modeling in investor behavior recognition, this paper selects 3240 corporate governance disclosure samples and 25600 interaction records to carry out experiments. The sample covers the governance chapter of the annual report, the ESG governance page, the investor relations webpage and the governance theme visualization page. The interactive data includes five types of behaviors: stay time, scroll depth, hot click, chart expansion and inquiry jump. The recognition tasks were divided into five categories: deep reading, quick browsing, chart focusing, risk-averting feedback and active participation, to observe whether the expression of governance information empowered by digital media art could be stably mapped to investor behavior outcomes. In order to ensure that the comparison results have a consistent basis, each model is tested under the same training rounds, the same hardware environment and the same data partition conditions.

In order to first illustrate the actual differences of different investor behaviors in the process of governance information contact, and provide behavioral layer reference for subsequent analysis of identification results, Table 2 statistics the mean results of five types of investor behaviors in terms of stay time, rolling depth, number of chart clicks, coverage rate of key blocks and follow-up interaction rate.

Table 2: Statistics of key interaction characteristics for different investor behaviors

Behavior Category	Average Dwell Time / s	Average Scroll Depth / %	Number of Chart Clicks	Key Section Coverage Rate / %	Follow-up Interaction Rate / %
In-Depth Reading	86.4	91.2	3.8	88.7	26.5
Quick Browsing	18.7	37.5	0.6	29.4	4.8
Chart-Focused Viewing	61.3	74.8	5.1	81.6	19.7
Risk-Avoidance Feedback	49.6	68.1	2.4	73.8	14.2
Active Engagement	79.8	88.6	4.7	85.9	31.4

As can be seen from Table 2, deep reading and active participation are at a high level in terms of average length of stay, scrolling depth and coverage of key blocks, indicating that these two types of investors are more inclined to continuously contact with governance information and can complete the complete process from page browsing to key block identification. The chart focus class has the highest number of chart clicks with 5.1, which indicates that the diagram, structure card and summary modules have obvious attention absorption effects. The fast browsing class has the lowest in all indicators, especially the coverage rate of key blocks and the subsequent interaction rate are only 29.4% and 4.8%, indicating that its contact stays at the surface layer. Although the duration of the risk-averse feedback class is higher than that of quick browsing, it is lower than that of deep reading, reflecting that such investors are more focused on risk tips and negative governance information.

In order to more clearly observe the discrimination boundary and error distribution of the proposed model on the five types of behaviors, Fig. 6 shows the heat map of the confusion matrix on the test set.

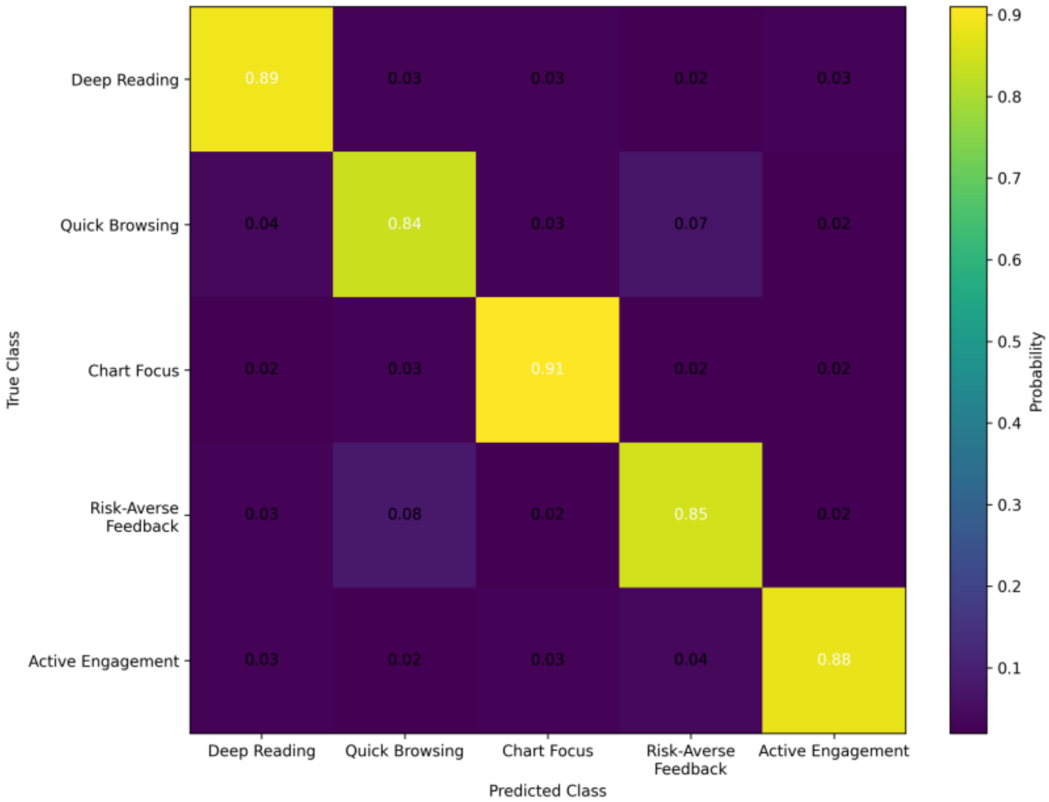

Figure 6: Heat map of confusion matrix for five categories of investor behavior recognition

Fig. 6 shows that the main diagonal recognition rates of the proposed model on five types of behaviors are 0.89 for deep reading, 0.84 for quick browsing, 0.91 for chart focusing, 0.85 for risk-averse feedback, and 0.88 for active participation. Among them, the chart focus class has the highest recognition rate, which indicates that the governance chart, summary card and structured prompt area form clearer behavioral boundaries after multimodal modeling. Deep reading and active participation also remain stable above 0.88, reflecting that the model can distinguish between continuous browsing and high-intensity interaction. There is a certain crossover between quick browsing and risk-averse feedback, and the mutual misclassification ratio of them is 0.07 and 0.08 respectively, which is significantly lower than the average misclassification level of the same type of the comparison model of 0.13, indicating that the joint representation effectively compreses the category overlap area.

After the behavioral feature statistics are completed, it is necessary to further compare the differences between the proposed model and different baseline methods in the main recognition indicators. To this end, Table 3 presents the accuracy, F1 value, AUC, and MCC results of each method on the test set. Here, BERT-Only, Vision-only, Text-Image Concat and Sequence-Log Model are selected as the comparison objects, corresponding to the recognition path of single Text, single Vision, simple image-text splicing and relying Only on behavior timing, respectively.

Table 3: Comparison of investor behavior recognition performance of different methods

Method	Accuracy	F1	AUC	MCC
BERT-Only	79.84%	0.781	0.846	0.662
Vision-Only	75.63%	0.742	0.801	0.598
Text-Image Concat	82.57%	0.816	0.874	0.711
Sequence-Log Model	84.23%	0.832	0.889	0.736
Proposed Model	87.41%	0.861	0.912	0.793

Table 3 shows that the proposed model achieves the best results on the four indicators, with an accuracy of 87.41%, an F1 value of 0.861, an AUC of 0.912, and an MCC of 0.793. Compared with BERT-Only, the accuracy is improved by 7.57 percentage points, indicating that it is difficult to explain the interface differences after digital media art empowerment by relying only on governance texts. Compared with Vision-Only, the F1 score is improved by 0.119, indicating that the visual expression must be co-modeled with the governance semantics to form stable recognition. Compared with Text-Image Concat, the proposed model improves the AUC by 0.038 and the MCC by 0.082, which indicates that the simple concatenation method is not enough to deal with the deep correlation between page structure and behavior trajectory. The performance of the Sequence-Log Model is already strong, but it is still lower than that of the proposed model, indicating that it is also difficult to maintain higher recognition accuracy if the interaction Sequence is separated from the governance content and visual structure.

Looking at a single indicator is not enough to tell the overall degree of balance of the model in multi-dimensional performance. In order to compare the comprehensive performance of different methods in terms of precision, recall, stability and consistency, Fig. 7 shows the radar chart results of each method.

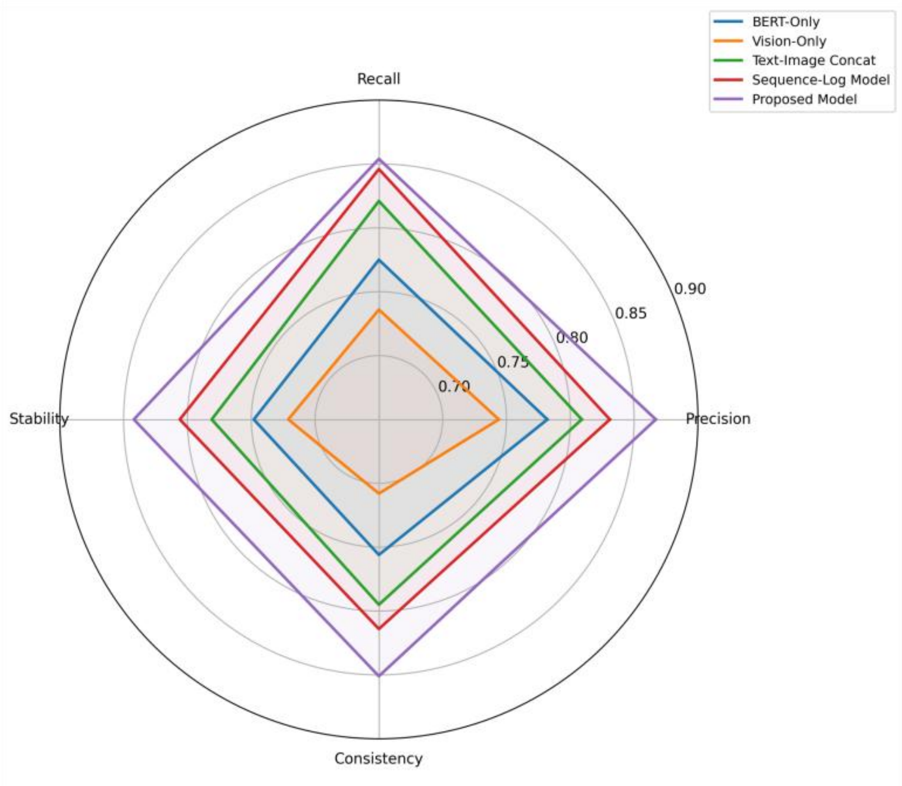

Figure 7: Radar plot of multi-index recognition performance of different methods

Fig. 7 shows that the values of the proposed model in the four dimensions of precision, recall, stability and consistency reach 0.867, 0.854, 0.842 and 0.851 respectively, and the range of the four indicators is only 0.025, indicating that the overall performance distribution is relatively balanced. In contrast, the four indicators of the Sequence-Log Model are 0.831, 0.846, 0.806 and 0.814, respectively. Although the recall rate is close to that of the proposed model, the stability and consistency are still lower by 0.036 and 0.037, respectively. The precision and recall of Text-Image Concat are 0.809 and 0.821, respectively, and the stability of Concat is only 0.781 under complex browsing paths. BERT-Only has the lowest agreement of 0.756. The results show that the proposed method does not rely on the local improvement of a certain index, but forms a more coordinated recognition structure in multi-dimensional discrimination ability.

In order to show the difference in the prediction confidence distribution of the model on different investor behavior categories, Fig. 8 further presents the boxplots of the prediction scores of the five categories of behavior.

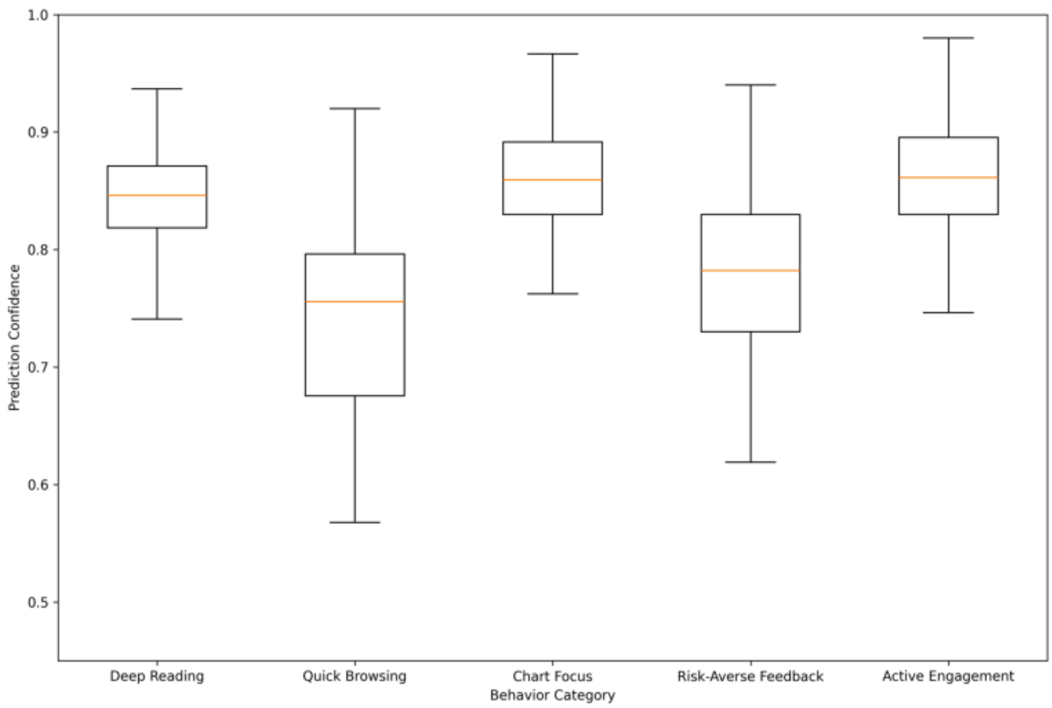

Figure 8: Boxplot of confidence in prediction of investor behavior for five categories

Fig. 8 indicates that there is a clear stratification in the prediction confidence for different behavior categories. The median confidence of the active participation class was 0.87, and the upper and lower interquartile range was 0.83 to 0.91. The median for deep reading was 0.85, and the interquartile range was 0.81 to 0.89; The median of the focus class of the diagram reaches 0.86, indicating that the diagram area, key labels and interactive hotspots have strong identification directivity. The median of the fast browsing class is only 0.74, and the span of the upper and lower bounds reaches 0.18, which is significantly higher than that of other categories. The risk averse feedback class has a median of 0.78, which is lower than deep reading and active participation, but higher than quick browsing. The median ranking of the five types of behaviors is basically consistent with the actual interaction strength, indicating that the confidence distribution of the model output has good behavior explanatory power.

Based on the above results, it can be seen that the proposed method maintains high accuracy and strong stability in the recognition of five types of investor behaviors. The average recognition rate of the main diagonal reached 0.874, and the lowest value of the four types of core indicators remained above 0.842, indicating that the multimodal governance information modeling did not show obvious performance collapse due to the increase of categories. Combined with the statistics of interaction characteristics in Table 2, it can be found that the model has the most stable grasp on the three high-intensity behaviors of deep reading, chart focus and active participation, while the two behaviors with closer boundaries of quick browsing and risk-aversion feedback also maintain a controllable misclassification range. The results show that the expression of corporate governance information empowered by digital media art can be effectively transformed into computable behavioral representation, and further support the subsequent expression effect analysis and structural difference evaluation.

4.2 Analysis on the expression effect of corporate governance information based on cross-modal fusion mechanism

After completing investor behavior identification, this paper further analyzes the role of cross-modal fusion mechanism on the expression effect of corporate governance information. The expression effect referred to here is not only the aesthetic degree of page appearance, but also the information arrival rate, key recognition rate, reading completion rate and subsequent interactive conversion rate formed by the governance content after digital media art is empowered. In order to ensure consistency with the experiment in the previous section, this paper still uses 3240 governance disclosure samples and 25600 interaction records, and divides governance pages into four categories: text-oriented, chart-complementary, narrative-enhanced and fusion expression, so as to investigate the actual differences of different expression structures under the effect of cross-modal fusion mechanisms.

In order to first present the overall performance of the four types of governance pages in terms of expression efficiency, Table 4 organizes the results of different page structures in terms of information arrival rate, key node identification rate, reading completion rate and inquiry conversion rate.

Table 4: Comparison of effect indicators for different governance expression structures

Page Type	Information Reach Rate	Key Node Recognition Rate	Reading Completion Rate	Inquiry Conversion Rate
Text-Dominant Type	78.5%	71.2%	51.7%	9.6%
Chart-Supplemented Type	84.3%	79.8%	62.9%	13.1%
Narrative-Enhanced Type	86.7%	82.4%	66.1%	14.8%
Integrated Expression Type	91.8%	88.6%	73.4%	18.9%

Table 4 shows that the fusion expressive page ranks first in the four indicators, in which the information arrival rate reaches 91.8%, the key node recognition rate reaches 88.6%, the reading completion rate reaches 73.4%, and the inquiry conversion rate is 18.9%. Although text-dominated pages have complete information, the reading completion rate is only 51.7%, indicating that high-density text structure is not conducive to continuous contact. The information arrival rate of the chart supplement page reached 84.3%, 5.8 percentage points higher than that of the text-dominated page, but the query conversion rate was only 13.1%, indicating that the local chart enhancement did not fully form the subsequent interaction chain. The reading completion rate of narrative-enhanced pages reaches 66.1%, indicating that visual narrative can improve contact continuity, but it is still lower than that of integrated expression, and the integrated design of semantic hierarchy, visual organization and interactive entrance has more advantages.

In order to illustrate the overall differences of different methods in the expression effect evaluation task, Table 5 further compares the performance of various models in the three indicators of expression consistency score, visual explanation adequacy, and behavioral response matching.

Table 5: Evaluation results of the expression effect of governance information for different methods

Method	Expression Consistency Score	Visual Explanation Adequacy	Behavior Response Matching Degree
Text-Only Evaluation	0.764	0.681	0.792
Visual-Only Evaluation	0.738	0.756	0.721
Late Fusion Model	0.801	0.782	0.828
Proposed Model	0.842	0.817	0.865

Table 5 shows that the proposed model is superior to the other methods in the three expression effect indicators, and the expression consistency score reaches 0.842, the visual interpretation adequacy is 0.817, and the behavioral response matching degree is 0.865. Although Text-Only Evaluation retains the institutional semantics of governance content, due to the lack of visual organization information of the page, its visual interpretation adequacy is only 0.681. Visual-only Evaluation achieves 0.756 in Visual explanation adequacy, but Only 0.721 in behavioral response matching, indicating that relying on graphics, color and layout alone is still unable to accurately explain investor follow-up feedback. The Late Fusion Model can achieve a behavioral response matching degree of 0.828, which is still 0.037 lower than that of the proposed model, indicating that deep cross-modal fusion is more effective in governance expression evaluation.

To further examine the contribution of different components in the cross-modal fusion analysis module, Table 6 presents the results of independent ablation experiments.

Table 6: Ablation experimental results of the expression effect analysis module

Model Setting	Expression Consistency Score	Key Node Recognition Rate	Behavior Response Matching Degree
Full Model	0.842	88.6%	0.865
Without Visual Gating	0.811	84.9%	0.836
Without Interaction Trajectory Branch	0.804	85.7%	0.822
Without Cross-Modal Attention Diffusion	0.793	83.8%	0.818

As can be seen from Table 6, after removing cross-modal attention diffusion, the expression consistency score decreases from 0.842 to 0.793, with a drop of 4.9%, indicating that high-level association propagation plays a core role in governance expression evaluation. After removing the visual gating, the recognition rate of key nodes decreases from 88.6% to 84.9%, with a reduction of 3.7%, which reflects that the coupling control between visual hierarchy and governance focus is indispensable. After removing the interaction trajectory branch, the matching degree of behavior response decreases from 0.865 to 0.822, a decrease of 4.3%, indicating that the investor contact path is an important basis for evaluating the expression effect. The three sets of results collectively show that the expression effect gain of the proposed model does not come from a single module, but is supported by multiple structural units.

In order to observe the intensity of attention distribution on different regions of the four types of governance pages more intuitively, Fig. 9 shows the heat density maps of the title area, the chart area, the risk tip area and the interactive entry area.

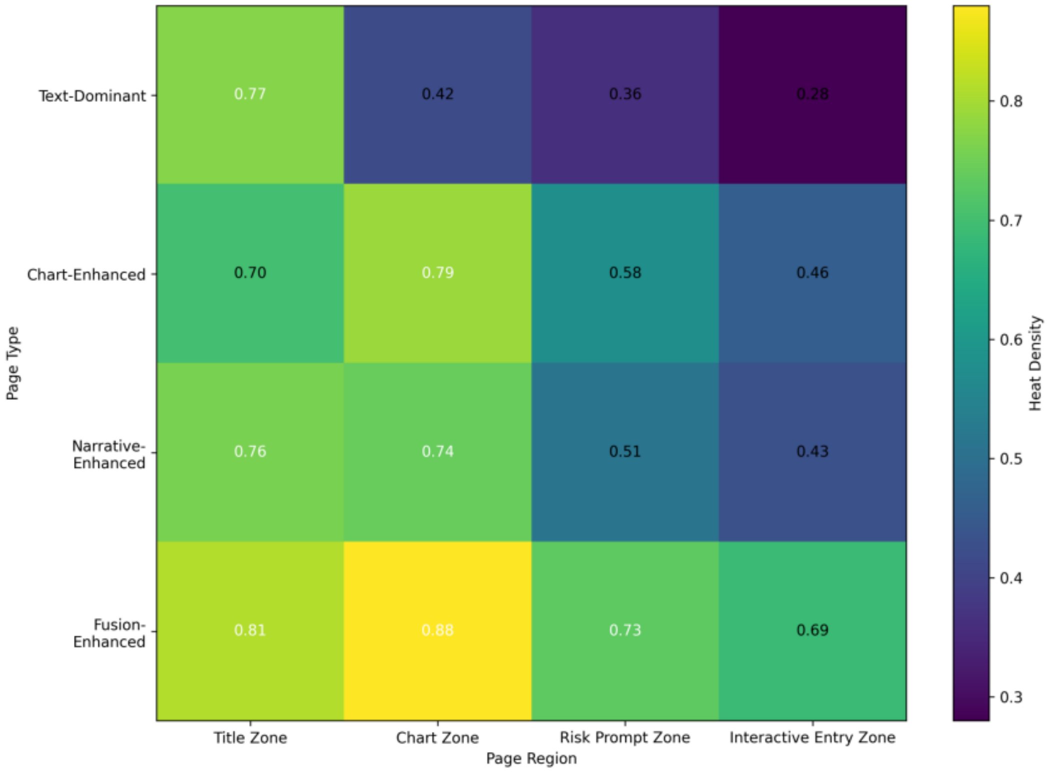

Figure 9: Density map of visual hotspots for four categories of governance pages

Fig. 9 shows that the fusion expression page has the most balanced density of hot areas in the four regions. The average heat values of the title area, the chart area, the risk warning area and the interactive entrance area are 0.81, 0.88, 0.73 and 0.69, respectively, and the attenuation range of heat between regions is controlled within 0.19. It indicates that investors' attention can move continuously along the structural hierarchy of governance information. For text-dominated pages, the hot value reaches 0.77 in the title area, but only 0.42 in the chart area, and 0.28 in the interactive entry area, indicating that the attention is focused on the first paragraph and quickly decays. The heat of the narrative enhanced page chart area reached 0.74, but the risk warning area was only 0.51, indicating that the lack of semantic coordination in visual enhancement would lead to uneven communication of governance focus. The chart area of the chart supplement page is 0.79, but the interactive entry area is only 0.46, indicating that it has not yet formed a complete interactive chain.

In order to further show the distribution pattern and stability degree of different page structures in the comprehensive expression score, Fig. 10 shows the violin plot of the comprehensive expression score of four types of governance pages.

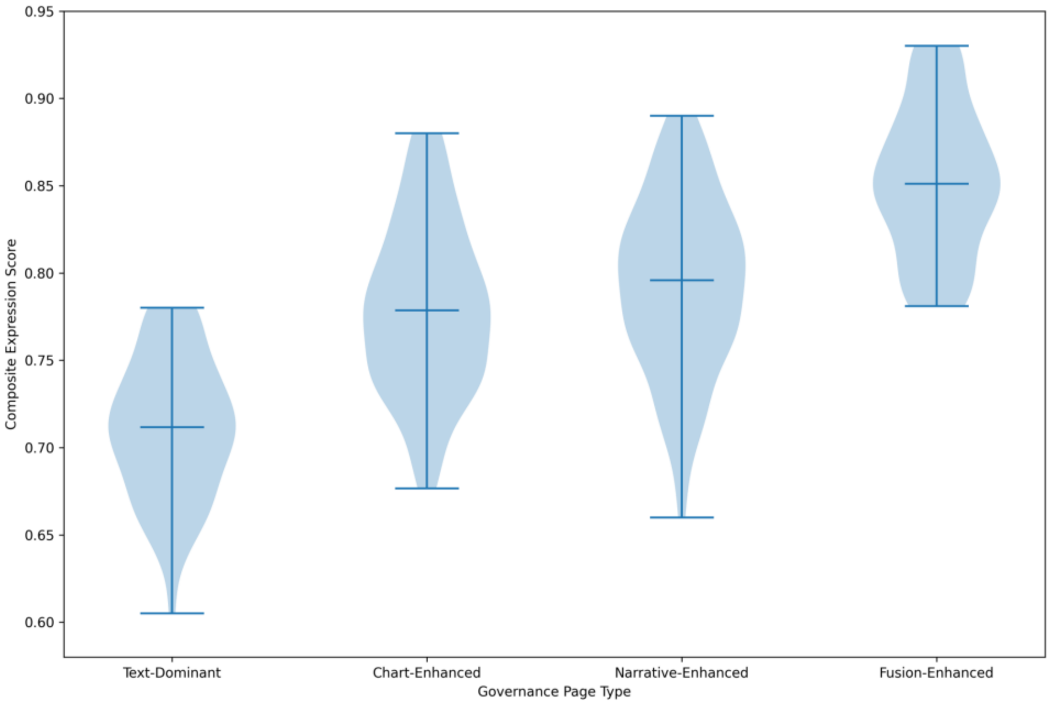

Figure 10: Violin plot of the comprehensive expression scores of the four categories of governance pages

Fig. 10 shows that the median comprehensive expression score of fusion expression pages is 0.846, the interquartile range is concentrated between 0.812 and 0.879, and the minimum value of the lower tail is still maintained at 0.781, indicating that this type of page can maintain a high and stable expression effect in different enterprise samples. The median of narrative-intensive pages is 0.791, and the upper and lower interquartile range is 0.742 to 0.826, which is higher than that of text-dominated pages as a whole, but the lower tail is longer, indicating that the pages have a strong dependence on local visual narrative. The median of chart supplement pages is 0.768, and the distribution span reaches 0.19, indicating that there is still a disconnect between the chart and the main text in some samples. The median of text-dominated pages is the lowest, only 0.713, and the upper quartile is still less than 0.75, indicating that a single text structure is difficult to support high-quality governance expression.

Based on the above results, it can be seen that the cross-modal fusion mechanism has a substantial enhancement effect on the expression effect of corporate governance information. The information arrival rate, key node recognition rate, reading completion rate and query conversion rate of fused expression pages reach 91.8%, 88.6%, 73.4% and 18.9%, respectively. At the same time, the highest median and the shortest low score tail in the comprehensive expression score distribution are maintained, which indicates that the structure not only has better average effect, but also has stronger stability. Ablation results further show that cross-modal attention diffusion, visual gating and interaction trajectory branch support the three key links of expression consistency, focus recognition and behavior matching respectively, and the lack of one will lead to the decline of the effect. It can be seen that the expression of governance information empowered by digital media art can only form a sustainable expression gain when the semantic level, visual organization and interactive feedback are calculated collaboratively, and provide a reliable basis for the subsequent discussion and conclusion.

5 Discussion

After digital media art participates in the expression of corporate governance information, the disclosure page is no longer a simple text-bearing interface, but a composite calculation object with the joint action of semantic structure, visual rhetoric and investor contact behavior. The multi-modal visual representation and cross-modal fusion analysis framework constructed in this paper has achieved stable results in investor behavior recognition and governance information expression effect evaluation. In the recognition experiment, the accuracy of the model reaches 87.41%, the F1 value is 0.861, and the AUC is 0.912, indicating that the joint modeling of governance text, chart structure, color level and interaction trajectory can more completely capture the differences in investor responses. In the analysis of expression effect, the information arrival rate, the key node recognition rate and the reading completion rate of the fusion expression page reach 91.8%, 88.6% and 73.4%, respectively, which indicates that the artistic organization of digital media is not surface modification, but an important variable affecting the efficiency of governance information understanding. The advantages of this model mainly come from three aspects. Firstly, the structured encoding preserves the correspondence between governance semantics and page hierarchy. Secondly, cross-modal attention diffusion enhances the linkage between text focus, visual focus and behavior hotspots. Thirdly, the interaction trajectory branch makes the expression evaluation results have a stronger contact basis. At the same time, complex page styles, differences in enterprise disclosure habits, and low-quality interaction logs still cause disturbances to model stability.

6 Conclusions

Focusing on the task of corporate governance information expression and investor behavior calculation and analysis under the empowerment of digital media art, this paper constructs an overall method chain composed of multi-modal visual representation, structured coding and cross-modal fusion recognition, and completes the verification on the interaction data between the governance disclosure page and investors. The results show that after the governance text, chart structure, color hierarchy, layout rhythm and behavior trajectory enter the unified representation space, the communication state of governance information and its influence on investor contact, understanding and feedback can be more accurately described. This study illustrates that digital media art is not an additional decoration, but an important information dimension that can be computationally modeled and participate in behavior prediction. At the same time, this paper still has some limitations. The sample sources mainly focus on public disclosure pages and platform logs, and the corporate disclosure style and investor group structure are still not rich enough. Although the model improves the ability of expression recognition and effect evaluation, there is still room for further improvement in adaptability to low-quality pages, abnormal interactions, and cross-platform migration scenarios. The follow-up research can be promoted in three aspects. Firstly, cross-industry, cross-market and cross-terminal samples should be expanded to improve the generalization ability of the model. Secondly, the lightweight graph calculation and incremental update mechanism were introduced to reduce the deployment cost. Thirdly, the generative interface optimization and real-time feedback learning are combined to realize the closed-loop linkage between governance information expression, behavior recognition and page adjustment, so as to enhance the continuous application value of the method in the real digital disclosure environment. In addition, this paper has cross-validated the decision-making process through hot area distribution, attention weight and ablation results, but the explanation granularity still mainly stays at the page area and modal contribution level, and the corresponding relationship between fine-grained governance semantic units and investor cognitive paths can be further deepened.

Author's Profile

Qianwen Lei was born in Quanzhou, Fujian, China in 2005. I am currently studying at the School of Art and Design at Chongqing Jiaotong University. My main research focus is digital media art design interaction and user experience. 632314050401@mails.cqjtu.edu.cn

Tingting Wei was born in 2005 in Zhenfeng, Guizhou, China. I am currently studying at the School of Art and Design at Chongqing Jiaotong University. My main research focus is on branding, visual culture, and communication design. 632314050403@mails.cqjtu.edu.cn

Zixuan Yang was born in 2004 in Zhangshu, Jiangxi, China. I am currently studying at the College of Art and Design at Chongqing Jiaotong University. My main research focus is on the study of narrative and public cognition guidance in data visualization design. 632314050407@mails.cqjtu.edu.cn

Chen Chen was born in 2005 in Chongqing, China. I am currently studying at the College of Art and Design at Chongqing Jiaotong University.My main research direction is the study of the visual expression of local culture in brand design. 632314050421@mails.cqjtu.edu.cn

References

[1] Liu W J, Ge Y B, Gu Y C. News-driven stock market index prediction based on trellis network and sentiment attention mechanism[J]. Expert Systems with Applications, 2024, 250: 123966.

[2] Lee H, Kim J H, Jung H S. Deep-learning-based stock market prediction incorporating ESG sentiment and technical indicators[J]. Scientific Reports, 2024, 14(1): 10262.

[3] Deng S, Luo Q, Zhu Y, et al. Multi-sentiment fusion for stock price crash risk prediction using an interpretable ensemble learning method[J]. Engineering Applications of Artificial Intelligence, 2024, 135: 108842.

[4] Sheng Y, Qu Y, Ma D. Stock price crash prediction based on multimodal data machine learning models[J]. Finance Research Letters, 2024, 62: 105195.

[5] Du S, Shen H. Reinforcement learning-based multimodal model for the stock investment portfolio management task[J]. Electronics, 2024, 13(19): 3895.

[6] Ferraro A, Sperlì G. How does user-generated content on Social Media affect stock predictions? A case study on GameStop[J]. Online Social Networks and Media, 2024, 43: 100293.

[7] Leippold M, Wang Q, Yang M. Technical patterns and news sentiment in stock markets[J]. The Journal of Finance and Data Science, 2024, 10: 100145.

[8] An W, Wang L, Zeng Y R. Social media-based multi-modal ensemble framework for forecasting soybean futures price[J]. Computers and Electronics in Agriculture, 2024, 226: 109439.

[9] Dahal K R, Pokhrel N R, Gaire S, et al. A comparative study on effect of news sentiment on stock price prediction with deep learning architecture[J]. Plos one, 2023, 18(4): e0284695.

[10] Zhao Y, Yang G. Deep learning-based integrated framework for stock price movement prediction[J]. Applied Soft Computing, 2023, 133: 109921.

[11] Xu D, Xu Z, Chen S, et al. A multi-channel cross-residual deep learning framework for news-oriented stock movement prediction[J]. Economic research-Ekonomska istraživanja, 2023, 36(2).

[12] Costola M, Hinz O, Nofer M, et al. Machine learning sentiment analysis, COVID-19 news and stock market reactions[J]. Research in international business and finance, 2023, 64: 101881.

[13] Agarwal B. Financial sentiment analysis model utilizing knowledge-base and domain-specific representation[J]. Multimedia Tools and Applications, 2023, 82(6): 8899-8920.

[14] Karanikola A, Davrazos G, Liapis C M, et al. Financial sentiment analysis: Classic methods vs. deep learning models[J]. Intelligent Decision Technologies, 2023, 17(4): 893-915.

[15] Du K, Xing F, Mao R, et al. Financial sentiment analysis: Techniques and applications[J]. ACM Computing Surveys, 2024, 56(9): 1-42.

[16] Ashtiani M N, Raahemi B. News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review[J]. Expert Systems with Applications, 2023, 217: 119509.

[17] Groß-Klußmann A. Learning deep news sentiment representations for macro-finance[J]. Digital Finance, 2024, 6(3): 341-377.

[18] Divernois M A, Filipović D. StockTwits classified sentiment and stock returns[J]. Digital Finance, 2024, 6(2): 249-281.

[19] Cai Y, Tang Z, Chen Y. Can real-time investor sentiment help predict the high-frequency stock returns? Evidence from a mixed-frequency-rolling decomposition forecasting method[J]. The North American Journal of Economics and Finance, 2024, 72: 102147.

[20] Deng S, Huang X, Zhu Y, et al. Stock index direction forecasting using an explainable eXtreme Gradient Boosting and investor sentiments[J]. The North American Journal of Economics and Finance, 2023, 64: 101848.