Enterprise Cash Flow and Liability Management Forecast Optimization Implemented by Temporal Fusion Transformer

Minqing Liu

Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong 999077, China

Corresponding author: Minqing Liu (e-mail: katrinaliu8808@163.com)

ABSTRACT To improve the interpretability of enterprise cash-flow and liability-management forecasting, this study applies a Temporal Fusion Transformer model incorporating financial rule constraints. Static and dynamic feature embedding and bidirectional LSTM encoding are first used to extract long-term evolutionary representations of enterprise financial status . A Gated Residual Network then screens key variables, and prior financial rules, including “debt ratio > 70%” and “ cash flow < 3 months of operating expenses,” are embedded into the multi-head attention layer. A learnable gain coefficient adjusts attention weights for high-risk historical periods, thereby improving the financial rationality and traceability of model decisions. A sequence-to-sequence decoding structure is finally used to jointly forecast multi-period cash-flow and liability indicators. Experimental results show that the proposed method achieves strong predictive accuracy, with an average cash-flow MAPE of 6.3%, an average debt-structure Theil’s U of 0.20, and an average rulecompliance rate of 89.7%. The model can be integrated into enterprise financial decision-support systems to support dynamic debt management, liquidity-risk monitoring, and explainable engineering-oriented financial forecasting.

INDEX TERMS cash flow forecasting, liability management, temporal fusion transformer, financial rule constraints, engineering decision support

I. INTRODUCTION

AS corporate financial environments, particularly in capital-intensive sectors like the industry, becom e increasingly complex, accurate forecasting of cash flow and liability management has become a key component in ensuring corporate liquidity security and preventing financial risk [1-3]. Traditional statistical models struggle to capture the nonlinear relationships in multivariate time series data, which is common when tracking complex supply chain and inventory cycles. While deep learning methods have improved forecasting accuracy, they suffer from a crucial lack of interpretability, making it difficult for financial managers—especially those managing the tight liquidity cycles of firms—to trust their forecasts and incorporate them into actual decision-making processes [4,5]. In dynamic liability adjustments and risk early warning scenarios, the model’s forecasting behavior is often disconnected from corporate financial management, failing to effectively reflect the inherent logical connections between debt levels, cash flow conditions, and risk exposure, limiting its deep integration into enterprise-level decision-making systems [6,7]. Companies of different industries and sizes also differ significantly in their financial structures, operating cycles, and risk characteristics, leading to inconsistent forecasting performance and highlighting the bottleneck of balancing accuracy with interpretability and transparency in financial intelligent forecasting.

Corporate cash flow and liability management involve multi-dimensional financial indicators. Its dynamic changes are affected by both internal operations and the external economic environment. Mazouei N R [8] explored the impact of cash flow management on the performance of listed companies on the Tehran Stock Exchange. The change in the maturity of accounts receivable has a significant negative impact on the company’s performance. Luo Y [9]’s research showed that short-term liabilities can meet the company’s sustainable development and increase the company’s operating income. A bad capital structure can hurt the company’s finances. Septyanto D [10] showed that the current ratio, debt-to-asset ratio, and return on assets significantly affect financial distress, and some current ratios and return on assets have a positive impact on financial distress. Akhtar M [11] used the generalized method of moments to analyze and showed that financial leverage has a significant negative effect on performance. Yousef H H [12]’s research showed that free cash flow had a statistically significant impact on investment decision-making efficiency. These studies underscore the critical need for accurate prediction of cash flow and liabilities and for an intelligent system that combines high precision with high credibility.

Ling T [13] used a long short-term memory neural network to build a financial early warning model and used the wolf pack algorithm to optimize the initial weights and bias parameters, with a goodness of fit of 94.2%. Yang H [14] constructed a financial distress early warning risk feature extraction framework based on the XGBoost (eXtreme Gradient Boosting) model and the SHAP (SHapley Additive exPlanations) algorithm, which can provide a reasonable explanation for the important risk features that affect financial distress and their impact paths. Li K [15] used ARIMA (Autoregressive Integrated Moving Average) and Prophet models to model and predict the strong volatility of financial capital. de Jesus Jr L C [16] explored how to integrate features derived from topological data analysis to improve the performance of univariate time series forecasting models. Yu H [17] proposed an adaptive time fusion transformer based on the Sharpe ratio, integrating a risk-aware prediction mechanism to optimize financial forecasts while considering return volatility. Traditional models represented by ARIMA and Prophet and ensemble learning methods such as XGBoost have certain interpretability, but are difficult to model complex nonlinear time series dependencies and have limited prediction accuracy; deep models represented by LSTM can capture time series patterns, but lack inherent interpretability mechanisms; when dealing with corporate financial forecasting tasks, they fail to effectively balance the three core requirements of prediction accuracy, model interpretability, and management logic consistency. This paper proposes an improved TFT model that incorporates financial rule constraints for the joint forecasting of corporate cash flows and liabilities. First, static covariates, historical dynamic financial indicators, and known future information are classified, embedded, and standardized. A bidirectional LSTM is then used to encode historical sequences to extract temporal dependencies. Secondly, applying the debt ratio warning and cash flow safety line financial rule functions into the multi-head attention mechanism.

Through the learnable gain coefficient, explicitly embedding them into the multi-head attention layer of TFT in the form of soft constraints to directly correct the attention weights of high-risk historical periods.

Finally, a sequence-to-sequence decoder autoregressively generates multi-dimensional financial indicators of cash flows and liabilities for multiple steps into the future. This method innovatively embeds prior financial management rules as soft constraints into the attention mechanism of the TFT. This approach achieves a deep fusion of "data-driven and knowledge-guided" forecasting, which is particularly valuable for stabilizing financial operations in the high-risk, high-volatility environment.

II. ALGORITHM DESIGN

A. OVERALL MODEL ARCHITECTURE DESIGN

The study constructs an improved TFT model with a four-stage architecture (“input embedding–temporal encoding–feature fusion–multi-step forecasting”) for multivariate, multi-step, and interpretable joint forecasting of cash flow and liabilities, as shown in Figure 1 [18,19]. The model input comprises three feature channels: static covariates (e.g., industry, registered capital), historical dynamic features (past T quarters’ financial indicators), and known future features (e.g., financing plans). The three types of input enter the model through independent embedding paths to avoid feature confusion.

During the encoding phase, the historical dynamic feature sequence is normalized and embedded before being fed into a bidirectional LSTM network for temporal encoding. The forward LSTM captures temporal dependencies from the past to the present, while the backward LSTM models the impact of future states on historical representations. The concatenation of the two hidden states forms a context-rich encoding sequence, which serves as the key and value inputs for the subsequent attention mechanism. The feature fusion stage applies a variable selection network (VSN). Built on a gated residual network (GRN), this network implements nonlinear transformations of static and dynamic embedding vectors and adaptively selects key features. On this basis, a multi-head attention mechanism is constructed, modeling cross-timestep dependencies through a query, key, and value structure. This stage reserves interfaces for the subsequent application of financial rule constraints, ensuring that attention weights can be modulated by domain knowledge. The prediction phase employs a sequence-to-sequence (Seq2Seq) decoding architecture. The decoder uses an autoregressive method to gradually generate multidimensional financial indicator outputs, including operating cash flow, short-term debt, long-term debt, net cash flow, current ratio, and debt-to-asset ratio [20,21]. Each decoding step integrates the current context vector, the historical encoding state, and known future inputs, ensuring that the prediction process accounts for both historical patterns and external interventions. The overall model structure remains end-to-end differentiable and supports gradient-based joint optimization.

Figure 1 shows the overall architectural process of this paper: first, static covariates, historical dynamic features, and known future features are embedded separately; historical dynamic features are encoded via bidirectional LSTM to generate a hidden state sequence containing forward and backward temporal dependencies, which serves as the key and value of the attention mechanism; the three types of embedding vectors are fused and filtered through GRN to form the query input; the debt ratio and cash flow safety line financial rule functions are applied into the multi-head attention layer to modify the attention weights and improve the consistency of the prediction logic; finally, multi-step decoder autoregression is used to generate future multi-dimensional financial indicators, achieving high-precision and explainable joint prediction.

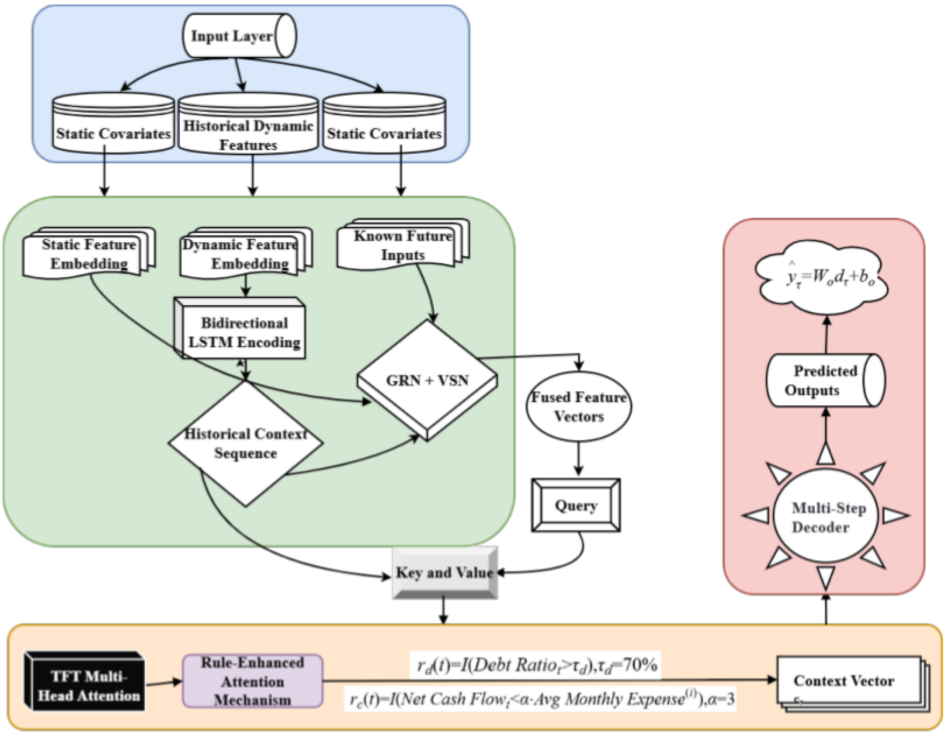

FIGURE 1. Overall Framework Diagram

B. MULTI-SOURCE FINANCIAL DATA PREPROCESSING AND ALIGNMENT

The original data covers the core fields in the balance sheet, income statement, and cash flow statement, such as net operating cash flow, short-term loans, long-term liabilities, operating income, debt-to-asset ratio, etc. The Shenwan secondary industry index yield, the quarterly GDP (Gross Domestic Product) year-on-year growth rate announced by the National Bureau of Statistics, and the central bank's one-year loan market benchmark rate are simultaneously obtained as external covariates. Let $x_t^{(i)} \in \mathbb{R}^D$ be the original financial vector of the i-th enterprise in the t-th quarter, and D be the feature dimension. The Z-score standard is used for the continuous financial variable $x_{t,d}$:

$$\hat{x}_{t,d}^{(i)} = \frac{x_{t,d}^{(i)} - \mu_d}{\sigma_d} \tag{1}$$

$$\mu_d = \frac{1}{N}\sum_{i=1}^{N} x_{i,d}^{(i)} \tag{2}$$

$$\sigma_d = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(x_{t,d}^{(i)} - \mu_d\right)^2} \tag{3}$$

μ_d and σ_d in formulas (1), (2), and (3) are the cross-sectional mean and standard deviation across enterprises, ensuring that the variables are in comparable dimensions. For categorical variables, one-hot encoding is used to map them into a binary vector $e_t \in \{0,1\}^K$, where K is the total number of industry categories [22]. In terms of time alignment, all variables are uniformly sampled to quarterly frequencies, and missing values are filled using linear interpolation. Assume that the observation value of a certain enterprise at time point t is missing, and the most recent valid observations before and after are t- and t+, respectively. Then the interpolation formula is:

$$x_t = x_- + \frac{t - t_-}{t_+ - t_-}(x_+ - x_-) \tag{4}$$

For known future variables, valid values are only filled in when $t \geq t_{\text{current}}$. All other periods are set to zero or marked as unavailable to ensure consistent input across model training and inference. Finally, a structured data tensor $X \in \mathbb{R}^{N\times T\times(D_h+D_s+D_f)}$ is constructed, where N is the number of sample companies, $T = 24$ is the time step, D_h is the historical dynamic feature dimension, D_s is the static feature dimension, and D_f is the known future feature dimension. This tensor, used as model input, meets TFT's modeling requirements for multi-source, multi-type, and aligned time series data, providing a standardized foundation for subsequent embedding and encoding.

C. STATIC AND DYNAMIC FEATURE EMBEDDING AND LSTM TEMPORAL ENCODING

This paper designs embedding paths for static covariates, historical dynamic features, and known future variables. After embedding, the historical dynamic features are fed into a bidirectional long-short-term memory network to extract the long-term dependency structure of the time series, generating a hidden state sequence rich in contextual information as the input of the subsequent attention mechanism. Static features and known future features participate in feature fusion after being independently embedded, but do not participate in the recursive encoding process [23,24]. Let the static feature vector of the i-th enterprise at time step t be $s^{(i)} \in \mathbb{R}^{D_s}$, which includes invariant attributes such as the industry to which the enterprise belongs, the scale of registered capital, and the years of establishment. It is mapped to a low-dimensional dense space through a learnable linear transformation matrix $W_s \in \mathbb{R}^{d_s \times D_s}$ and a bias term $b_s \in \mathbb{R}^{d_s}$:

$$h_s^{(i)} = W_s s^{(i)} + b_s \tag{5}$$

In formula (5), $h_s^{(i)} \in \mathbb{R}^{d_s}$ is a static embedding vector that remains unchanged throughout the entire time series and is shared by subsequent time steps. For the historical dynamic feature sequence $\{x_1^{(i)}, x_2^{(i)}, \ldots, x_T^{(i)}\}$, $x_t^{(i)} \in \mathbb{R}^{D_h}$ contains time-varying variables such as operating cash flow, short-term liabilities, and debt-to-asset ratio, and is mapped to a dense representation using a learnable embedding function $f_h : \mathbb{R}^{D_h} \to \mathbb{R}^{d_h}$:

$$e_t^{\text{hist}} = f_h(x_t) = W_h x_t + b_h \tag{6}$$

The known future feature sequence is mapped by an independent embedding function f_f:

$$e_t^{\text{fut}} = f_f(u_t) = W_f u_t + b_f \tag{7}$$

All embedding vectors are processed by layer normalization (LayerNorm) to stabilize the training process [25]. The historical embedding sequence $\{e_1^{\text{hist}}, \ldots, e_T^{\text{hist}}\}$ is input into the bidirectional LSTM encoder. The forward LSTM processes the input in time order, and its hidden state $\overrightarrow{h}_t \in \mathbb{R}^{d_l}$ is updated by the following gating mechanism:

$$\overrightarrow{i}_t = \sigma\left(W_{ii} e_t^{\text{hist}} + W_{hi} \overrightarrow{h}_{t-1} + b_i\right) \tag{8}$$

$$\overrightarrow{f}_t = \sigma\left(W_{if} e_t^{\text{hist}} + W_{hf} \overrightarrow{h}_{t-1} + b_f\right) \tag{9}$$

$$\overrightarrow{o}_t = \sigma\left(W_{io} e_t^{\text{hist}} + W_{ho} \overrightarrow{h}_{t-1} + b_o\right) \tag{10}$$

$$\begin{aligned} \overrightarrow{c}_t = {} & \overrightarrow{f}_t \odot \overrightarrow{c}_{t-1} \\ & + \overrightarrow{i}_t \odot \tanh\left(W_{ic} e_t^{\text{hist}} + W_{hc} \overrightarrow{h}_{t-1} + b_c\right) \end{aligned} \tag{11}$$

$$\overrightarrow{h}_t = \overrightarrow{o}_t \odot \tanh\left(\overrightarrow{c}_t\right) \tag{12}$$

The backward LSTM processes the sequence in reverse to generate the hidden state $\overleftarrow{h}_t$. Finally, the forward and backward hidden states are concatenated to form a context-aware encoding sequence:

$$h_{\text{enc}} = \left[\overrightarrow{h}_t; \overleftarrow{h}_t\right] \in \mathbb{R}^{2d_l} \tag{13}$$

The bidirectional LSTM encoder serves as a crucial temporal feature extractor, providing a context-rich representation for the subsequent TFT attention mechanism. Its functional role is distinct from the multi-head self-attention in the TFT core, offering three key advantages in modeling long-term dependencies: Robust modeling of sequence noise and short-term fluctuations; Explicit forward and backward temporal dynamics modeling; and Acting as a “knowledge distillation” pre-processing step for the TFT.

D. ATTENTION WEIGHT MODIFICATION BASED ON FINANCIAL RULES

This paper applies a domain knowledge-driven attention correction mechanism into the multi-head attention mechanism, and explicitly constrains the attention weight distribution by embedding a priori financial rule functions [26,27]. Let the query, key, and value matrices of the standard multi-head attention be $Q \in \mathbb{R}^{H \times d_k}$, $K \in \mathbb{R}^{T \times d_k}$, and $V \in \mathbb{R}^{T \times d_v}$, respectively. H is the decoding step size, T is the encoding sequence length, and d_k and d_v are the key and value dimensions. The standard attention score is calculated as:

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \tag{14}$$

On this basis, two types of financial status indicator functions are applied to identify high-risk time points in historical series. Debt ratio warning function is defined as:

$$r_d(t) = I(\text{Debt Ratio}_t > \tau_d), \quad \tau_d = 70\% \tag{15}$$

In formula (15), $I(\cdot)$ is an indicative function, which is activated when the company’s debt-to-asset ratio exceeds the industry warning threshold. The setting of the debt ratio warning threshold $\tau_d = 70\%$ refers to the reference value for monitoring corporate debt ratios proposed by the State Administration of Financial Supervision in the "Core Indicators for Risk Supervision of Commercial Banks", as well as multiple studies on financial risks of manufacturing enterprises. This threshold is widely used as a critical point for financial risk warning in capital intensive industries such as electronics. Defining the cash flow safety line function:

$$\begin{aligned} r_c(t) = I(&\text{Net Cash Flow}_t < \\ &\alpha \cdot \text{Avg Monthly} \\ &\text{Expense}^{(i)}), \quad \alpha = 3 \end{aligned} \tag{16}$$

When the company’s net cash flow is insufficient to cover three months of operating expenses, an early warning is triggered. The setting of the cash flow safety line parameter α=3 months is based on the common practice standard of enterprise liquidity management, which means that enterprises usually need to maintain cash reserves covering at least three months of operating expenses to cope with short-term payment risks. This standard is widely adopted in cash flow management literature and has been confirmed by relevant industry financial guidelines. The above rules are combined into a composite financial risk indicator function:

$$g(t) = r_d(t) \vee r_c(t) \tag{17}$$

When calculating the attention weight, for the historical time step t that satisfies $g(t) = 1$, a learnable gain coefficient $\gamma \in \mathbb{R}^+$ ($\gamma > 1$) is applied before the softmax input to enhance the model's attention to high-risk states [28,29]. The modified attention score matrix is:

$$\widetilde{A} = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + \gamma \cdot 1_H \otimes g^T\right) \tag{18}$$

In formula (18), $1_H \in \mathbb{R}^H$ denotes an all-ones vector with dimension H, where H is the decoding horizon (forecast step count). The symbol $\otimes$ represents the outer product, which ensures that additional attention bias is given to high-risk historical states in all decoding steps. This operation yields a matrix in $\mathbb{R}^{H\times T}$, where each row (corresponding to a decoding step) is identical and contains the composite risk indicator $g^T \in \mathbb{R}^{1\times T}$ across all historical time steps T. In order to control the influence of the rule and maintain the trainability of the model, the parameter is transformed into a gated form:

$$\gamma = 1 + \sigma\left(w_\gamma^T h_s + b_\gamma\right) \tag{19}$$

In formula (19), h_s is the static feature embedding, and $\sigma(\cdot)$ is the Sigmoid function, which makes the gain coefficient dynamically adapt to the individual characteristics of the enterprise. The final attention output is:

$$\text{Attention}(Q, K, V) = \widetilde{A}V \tag{20}$$

While maintaining the end-to-end differentiability of Transformer, this mechanism embeds financial management experience into the model decision path in the form of soft constraints, so that attention distribution not only depends on data-driven relevance but also obeys the rationality of domain logic, providing subsequent predictions with context-weighted representations that comply with the principle of financial prudence, as shown in Figure 2.

Figure 2 shows the complete process of the rule-embedded attention mechanism: the query and key are used to calculate the basic attention score, and the value is used to generate the final output; at the same time, the two types of financial rules, debt ratio warning and cash flow safety line, are combined into a composite risk indicator function; this signal is combined with the learnable gain coefficient, and the attention bias matrix across decoding steps is constructed through the all-one vector and the outer product; this bias is injected into the original attention score to form a revised score matrix, which guides the model to pay more attention to historical high-risk periods when making predictions.

E. INTERPRETABLE GATED FUSION AND MULTI-STEP PREDICTION OUTPUT

This paper applies a GRN and multi-step decoding mechanism at the end of the TFT architecture and builds an interpretable output module to support attribution analysis of prediction results [30,31]. In the feature fusion stage, the static embedding vector, the historical encoding state, and the known future feature embedding are spliced and aligned. The input features are weighted using the Variable Selection Network (VSN):

$$X_{\text{fused}} = \sum_{k=1}^{K} \text{GRN}_k(E_k) \odot w_k \tag{21}$$

In formula (21), $E_k \in \mathbb{R}^{T\times d_e}$ is the embedding matrix of the k-th class feature across T time steps with embedding dimension d_e, w_k is the learnable weight vector, and $\odot$ represents element-by-element multiplication. The GRN module contains a gating mechanism and residual connections, and its calculation process is:

$$\text{GRN}(x) = x + \text{GLU}\left(W_1\eta(W_2x + b_2) + b_1\right) \tag{22}$$

In formula (22), $\eta(\cdot)$ is the ReLU activation function, and GLU (Gated Linear Unit) is defined as:

$$\text{GLU}(z) = z_{:d} \odot \sigma(z_{d:}) \tag{23}$$

A gating mechanism automatically suppresses non-essential features, improving the model's focus on key financial drivers. During the forecasting phase, a sequence-to-sequence decoding architecture is employed, with the decoder generating multivariate financial indicators for the next H = 6 quarters using an autoregressive method. Given a decoding time step $\tau \in \{T+1, \ldots, T+H\}$, the decoder inputs include the context vector c_τ (derived from the attention output), known future features e_τ^{fut}, and the previous moment's predicted value $\hat{y}_{\tau-1}$. The decoded hidden state d_τ is updated by the LSTM unit:

$$d_\tau = \text{LSTMCell}\left(d_{\tau-1}, [c_\tau; e_\tau^{\text{fut}}; \hat{y}_{\tau-1}]\right) \tag{24}$$

The final prediction output is a six-dimensional vector $\hat{y}_\tau \in \mathbb{R}^6$, corresponding to operating cash flow, short-term liabilities, long-term liabilities, net cash flow, current ratio and debt-to-asset ratio:

$$\hat{y}_\tau = W_o d_\tau + b_o \tag{25}$$

To support model interpretability, the attention weight matrix $A^{(h)} \in \mathbb{R}^{H\times T}$ of each head $h \in \{1, \ldots, N_h\}$ in the multi-head attention layer is recorded, and averaged along the head dimension and weighted aggregated with the feature weight to generate a feature-time importance heat map:

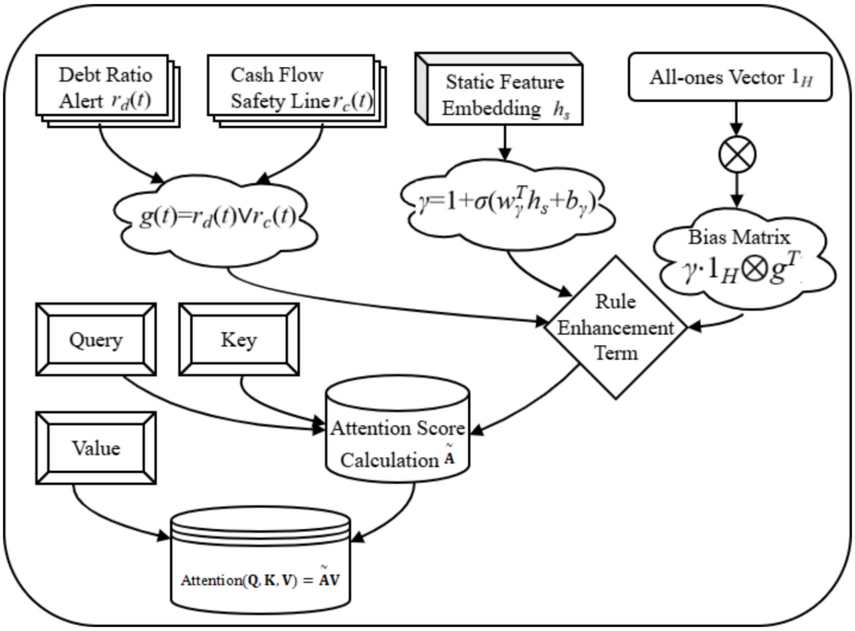

FIGURE 2. Schematic Diagram of the Rule Embedding Attention Mechanism

$$M \in \mathbb{R}^{H\times T}, \quad M_{\tau,t} = \frac{1}{N_h}\sum_{h=1}^{N_h} A_{\tau,t}^{(h)} \cdot w_{\text{feat}}(t) \tag{26}$$

The loss function uses the weighted mean squared error (WMSE) to balance the differences in the dimensions and volatility of different financial indicators [32,33]:

$$L = \frac{1}{H}\sum_{\tau=T+1}^{T+H}\sum_{k=1}^{6} w_k(y_{\tau,k} - \hat{y}_{\tau,k})^2 \tag{27}$$

The weight w_k is set according to the Coefficient of Variation (CV) of each indicator in the training set:

$$w_k = \frac{\sigma_k/\mu_k}{\sum_{j=1}^{6}(\sigma_j/\mu_j)} \tag{28}$$

In formula (28), σ_k and μ_k are the standard deviation and mean of the k-th indicator. The model is trained end-to-end using the AdamW optimizer, using a learning rate warmup and cosine annealing scheduling strategy to ensure stable convergence during the training process [34,35]. The model parameter settings are shown in Table 1.

TABLE 1. Model Parameter Settings

Category	Parameter	Symbol	Value/Setting
Model architecture	Decoder steps	H	6
Model architecture	Encoder sequence length	T	12
Model architecture	Key dimension	d_k	64
Model architecture	Value dimension	d_v	64
Training setup	Optimizer	–	AdamW
Training setup	Initial learning rate	η	1×10^{-3}
Training setup	Learning rate warm up steps	–	1,000
Training setup	Learning rate schedule	–	Cosine annealing
Training setup	Batch size	B	256
Training setup	Maximum epochs	–	200
Training setup	Early stopping patience	–	20

III. EXPERIMENTS AND VERIFICATION

A. EXPERIMENTAL DESIGN

The quarterly data underwent rigorous cleaning, alignment, and a chronological split into training, validation, and test sets to prevent time leakage. All hyperparameters of baseline models are determined through grid search on the validation set to ensure fair comparison. The specific settings are as follows: the LSTM model adopts 2-layer stacking, with 128 hidden units, and uses Adam optimizer; The maximum depth of the tree in the XGBoost model is set to 6, the learning rate is 0.1, and the number of subtrees is 100; The Prophet model uses an additive model and includes quarterly seasonality; The N-BEATS model uses a universal configuration, stacking 30 blocks with a hidden layer size of 512; The standard TFT model follows the recommended architecture in the original paper, with 4 attention heads and 160 hidden units in the encoder/decoder LSTM. These settings are consistent with typical configurations of various models in relevant time series prediction literature. All model input features remain consistent, and hyperparameters are determined through a grid search on the validation set, as shown in Table 2.

TABLE 2. Experimental Settings

Item	Configuration
Study period	24 quarters
Number of sample enterprises (N)	480
Industry distribution	Electronics (160 firms), Chemicals (160 firms), Machinery (160 firms). All firms are publicly listed manufacturing companies from China.
Financial metrics (dynamic)	Operating cash flow, short-term debt, long-term debt, revenue, debt-to-asset ratio, current ratio, accounts receivable, net profit
Data sources	Corporate financial data: CSMAR database; External variables: National Bureau of Statistics of China (GDP YoY, CPI YoY), People's Bank of China (1-year LPR)
Static attributes	Industry, registered capital scale (small/medium/large), years since establishment
External variables	Quarterly YoY GDP growth, CPI YoY, 1-year LPR rate
Dataset split	Training set: first 70% of time series (chronologically); validation set: next 10%; test set: final 20%. Split is applied uniformly across all enterprises to prevent temporal leakage.
Experimental environment	Python 3.9, PyTorch 1.13, NVIDIA A100 GPU, 64GB RAM

B. CASH FLOW FORECAST ACCURACY ASSESSMENT

Cash flow forecast accuracy is evaluated using MAE, RMSE, and MAPE, defined as:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{29}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{30}$$

$$\mathrm{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{31}$$

Figure 3 shows the cash flow forecasting accuracy of the proposed method, TFT, N-BEATS (Neural Basis Expansion Analysis for Interpretable Time Series Forecasting), Prophet, XGBoost, and LSTM in the electronics, chemical, and machinery manufacturing industries. The horizontal axis represents industry category, while the vertical axis represents the MAE, RMSE, and MAPE error metrics. Across all industries and metrics, the proposed method significantly outperforms the other compared methods. In terms of MAE, the proposed method achieves a loss of only 1.1 million yuan in the electronics industry, a reduction of 58%, 63%, and 42% compared to LSTM, XGBoost, and TFT, respectively. The proposed method also consistently maintains the lowest error levels in the chemical and machinery industries. The MAPE metric further shows that the the proposed method achieves performance of 5.9%, 6.6%, and 6.4% in the three industries, respectively, with an average of 6.3%. This is significantly lower than Prophet (21.8%-23.1%) and XGBoost (17.5%-19.2%), demonstrating its significant superiority in relative error control. This consistent, cross-industry performance demonstrates that proposed model possesses both high accuracy and good generalization capabilities.

The performance gain originates from two mechanisms: the BiLSTM and gated selection network capture cash flow cycles and structural shifts, while the rule-constrained attention proactively focuses on historical high-risk periods (e.g., debt >70% or low cash flow), aligning predictions with financial rationality. This 'data-driven + knowledge-guided' fusion ensures both logical consistency and robustness.

C. ASSESSMENT OF THE FORECAST STABILITY OF LIABILITY STRUCTURE

For short-term and long-term debt trend forecasts, Theil's U coefficient and Direction Accuracy (DA) are used to assess stability:

$$U = \frac{\sqrt{\frac{1}{H}\sum_{t=1}^{H}(y_t - \hat{y}_t)^2}}{\sqrt{\frac{1}{H}\sum_{t=1}^{H}y_t^2} + \sqrt{\frac{1}{H}\sum_{t=1}^{H}\hat{y}_t^2}} \tag{32}$$

In formula (32), both y_t and $\hat{y}_t$ represent the original, non-standardized values of the debt indicators. The denominator uses the square roots of the average sums of squares of these original values.

$$\mathrm{DA} = \frac{1}{H}\sum_{t=1}^{H} I\left((y_t - y_{t-1})(\hat{y}_t - \hat{y}_{t-1}) > 0\right) \tag{33}$$

Figure 4 compares the stability of debt structure prediction in the electronics, chemical, and machinery manufacturing industries using the proposed method, TFT, N-BEATS, Prophet, XGBoost, and LSTM methods. The left figure uses Theil's U to measure the degree of fit between the predicted value and the true value fluctuation, while the right figure uses DA to reflect the model's ability to judge the trend of debt changes. The proposed method performs best in all industries: in terms of Theil's U index, the electronics industry is 0.20, the chemical industry is 0.19, and the machinery industry is 0.22, with an average of 0.20, which is significantly lower than LSTM (0.38, 0.35, 0.39) and Prophet (0.48, 0.46, 0.49), indicating that its modeling of liability fluctuations is smoother and more accurate; in terms of directional accuracy, the proposed method reaches 83.6%, 84.1% and 82.5% in the three industries, respectively, which is better than N-BEATS (76.8%, 78.2%, 75.9%) and TFT (80.3%, 81.7%, 79.5%), indicating that it can more reliably capture the turning points of rising or falling liabilities, highlighting the robustness and universality of the model in predicting complex liability dynamics.

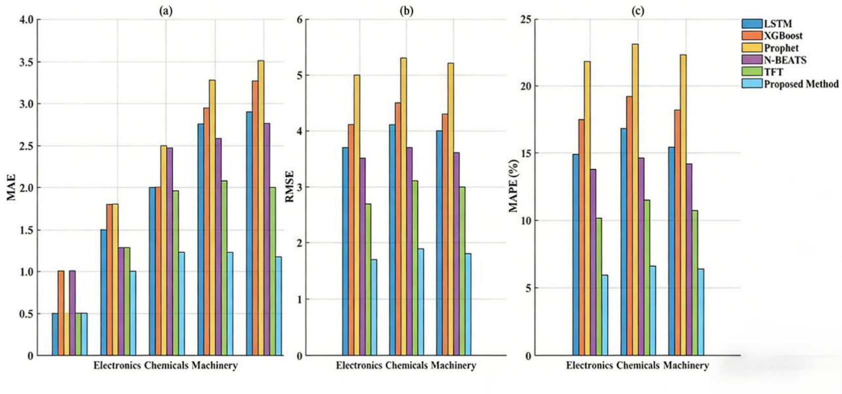

FIGURE 3. Cash Flow Forecast Accuracy Evaluation. Figure 3(a): MAE (millions); Figure 3(b): RMSE (millions); Figure 3(c): MAPE

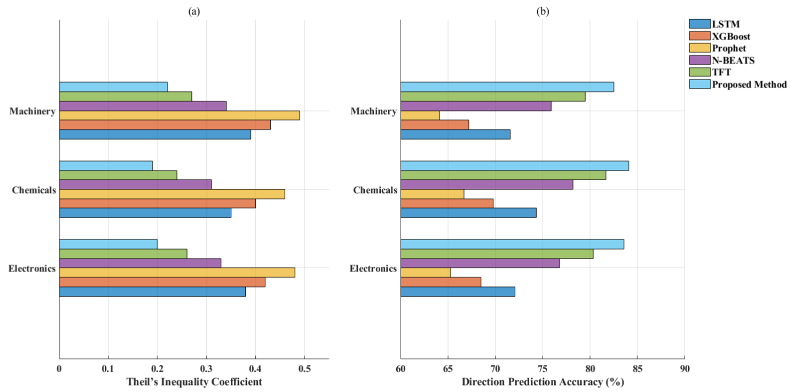

FIGURE 4. Debt Structure Forecast Stability Assessment. Figure 4(a): Theil's U; Figure 4(b): DA

XGBoost and Prophet rely on statistical patterns in data, making it difficult to identify nonlinear management behaviors such as high debt ratios triggering refinancing constraints or cash flow constraints leading to a surge in short-term debt, leading to misjudgment of trends. This method embeds rules, such as debt ratios $> 70\%$, at the attention layer, enabling the model to automatically increase attention during historically high-risk periods, more accurately simulating real-world behaviors such as passive leverage increases or debt rollovers. The model uses rules to identify combinations of "high debt + low cash flow," anticipating rising debt repayment pressures and significantly improving the debt-paying ratio (DA). This mechanism, which transforms management logic into attention biases, enables the model to not only fit historical data but also better simulate corporate financial decision-making processes.

D. MODEL INTERPRETABILITY EVALUATION

The "rule compliance rate" is defined to quantify the consistency between the attention mechanism and the financial logic. If a sample assigns a higher attention weight than the mean to the time step that satisfies the debt ratio $> 70\%$ or cash flow < 3 months of expenditure during prediction, it is considered to comply with the rule. The rule compliance rate is calculated as:

$$\text{Rule Compliance Rate} = \frac{1}{N}\sum_{i=1}^{N} I(\text{Attention Focus on High-Risk Period}_i) \tag{34}$$

Figure 5 compares the rule-conformance rates of six methods: the proposed method, TFT, N-BEATS, Prophet, XGBoost, and LSTM, in the electronics, chemical, and machinery industries. The horizontal axis represents industry category, and the vertical axis represents the proportion of models assigning high attention weights to high-risk periods during the forecasting process (rule-conformance rate). The rule-conformance rates of LSTM and XGBoost are generally low, averaging 61.5% and 59.3%, respectively, indicating that their attention allocation relies more on statistical data correlation than financial logic.

Prophet and N-BEATS show slight improvements, but remain below 69%. TFT, due to its attention mechanism, averages 75.7%, demonstrating a certain degree of logical alignment. The proposed method achieves rule-conformance rates of 89.1%, 90.2%, and 89.8% in the three industries, respectively, with an average of 89.7%, significantly higher than the other methods. It maintains a consistent lead across all industries, demonstrating that its forecasting process is highly consistent with the prudent principles of corporate financial management. An 89.7% rule compliance rate demonstrates that the model adheres to financial prudence principles in the vast majority of cases.

This advantage stems from the fact that the model explicitly embeds the guiding constraints of financial rules into the attention mechanism, rather than relying solely on data-driven implicit learning. Traditional models' attention or feature weights are automatically learned through backpropagation, which may focus on statistically significant but management-irrelevant noise patterns and misjudge short-term revenue fluctuations as financial risk signals. The proposed method applies a learnable rule gain function into the multi-head attention calculation. When the historical state meets risk combinations such as "high debt + low cash flow," the attention weight at the corresponding time step is automatically enhanced, ensuring that the model's decision-making basis is consistent with the risk identification logic of financial personnel when predicting future debt pressure or cash flow tension. This not only improves the transparency of the model but also makes it an auditable and explainable decision support tool, truly realizing the transition from "black box prediction" to "logically consistent intelligent assistance."

E. DECISION SUPPORT EFFECTIVENESS EVALUATION

Financial warning rules are set based on forecast results: When the forecasted net cash flow for the next two periods is negative and the debt-to-asset ratio increases, a "liquidity risk" warning is triggered. The warning signals are compared with actual corporate defaults, rating downgrades, or emergency financing events to calculate the matching accuracy.

Figure 6 shows the effectiveness of six methods— the proposed method, TFT, N-BEATS, Prophet, XGBoost, and LSTM—in decision support across the electronics, chemicals, and machinery industries. Darker colors correspond to the warning accuracy (%). Higher values indicate a closer match between the model's predicted "liquidity risk" warning and actual corporate defaults, rating downgrades, or emergency financing events. Prophet (68.5%-70.2%) and XGBoost (72.1%-74.0%) perform relatively poorly overall, demonstrating limited warning capabilities. LSTM and N-BEATS show slight improvements, reaching 74.8%-76.5% and 77.6%-78.3%, respectively. TFT performs relatively well, reaching 81%-82.4%, demonstrating the auxiliary role of the attention mechanism in risk identification. The proposed method achieves warning accuracy of 91.6%, 92.3%, and 90.7% in the three industries, respectively, with an average of 91.5%, the highest among all methods, demonstrating its superior decision support capabilities in predicting real financial risks.

The heatmap displays the warning accuracy (%) of six methods across three industries. The gradient color axis on the right indicates the warning accuracy percentage, ranging from 65% (dark blue) to 95% (bright yellow). Darker blue colors represent lower warning accuracy, while brighter yellow colors indicate higher warning accuracy. Higher values demonstrate a closer match between the model's predicted "liquidity risk" warnings and actual corporate defaults, rating downgrades, or emergency financing events.

Unlike models that rely solely on data fitting, the proposed method generates warnings based not only on numerical trends but also, through rule-guided focus, on the key historical pattern of "high debt + tight cash flow," ensuring management-logical consistency in the forecasts. The model can identify the combined signal of "rising debt-to-asset ratio + negative future net cash flow" and dynamically adjust the warning sensitivity based on industry characteristics. This "rule-driven + context-aware" decision-making mechanism ensures that warnings are not only accurate but also explainable and traceable, truly achieving the transition from "passive alerting" to "active risk control," providing highly reliable intelligent support for corporate financial managers.

F. MODEL GENERALIZATION ABILITY EVALUATION

Cross-industry tests are conducted in the food and beverage, apparel, and pharmaceutical manufacturing industries, respectively. The model is trained on the entire manufacturing industry without fine-tuning. The MAPE of the test sets of each industry is calculated and the average is taken.

Figure 7 shows the cross-industry generalization capabilities of six methods— the proposed method, TFT, N-BEATS, Prophet, XGBoost, and LSTM—in three unseen industries: food and beverage, apparel, and pharmaceutical manufacturing. Prophet achieves MAPEs of 23.8%, 24.6%, and 23.1%, respectively, significantly higher than the other methods, indicating weak generalization capabilities in new industry scenarios. LSTM and XGBoost also exhibit high errors, averaging over 17%. N-BEATS and TFT show some improvement, but remain in the 14.4%–15.9% and 11.2%–12% ranges, respectively. The proposed method achieves the lowest MAPE across all industries, with MAPEs of 6.8% for food and beverage, 7.4% for apparel, and 6.6% for pharmaceutical manufacturing. The average MAPE is only 6.9%, highlighting its remarkable cross-industry adaptability.

This powerful generalization capability stems from the model's collaborative modeling of static and dynamic features through a gated variable selection network (VSN). When faced with unseen industries such as food and beverage and pharmaceutical manufacturing, the model leverages static features to guide the adaptive adjustment of feature weights, rapidly matching the financial behavior patterns of these new industries. Furthermore, the time series representations extracted by the history encoder are transferable across industries, enabling the capture of universal financial dynamics such as cash flow cycles and evolving debt structures. Combined with the constraints imposed by financial rules on the attention mechanism, the model can infer the financial evolution path of a target industry based on existing knowledge, even in the absence of training samples from the

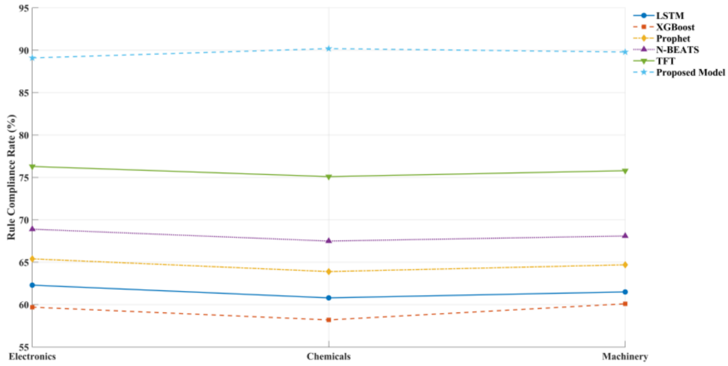

FIGURE 5. Model Interpretability Evaluation

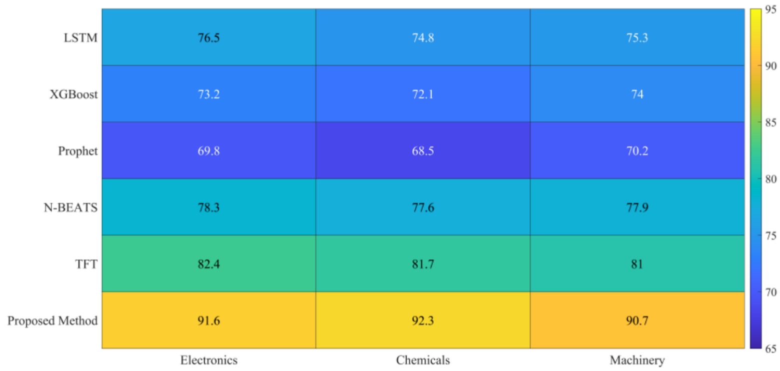

FIGURE 6. Decision Support Effectiveness Evaluation

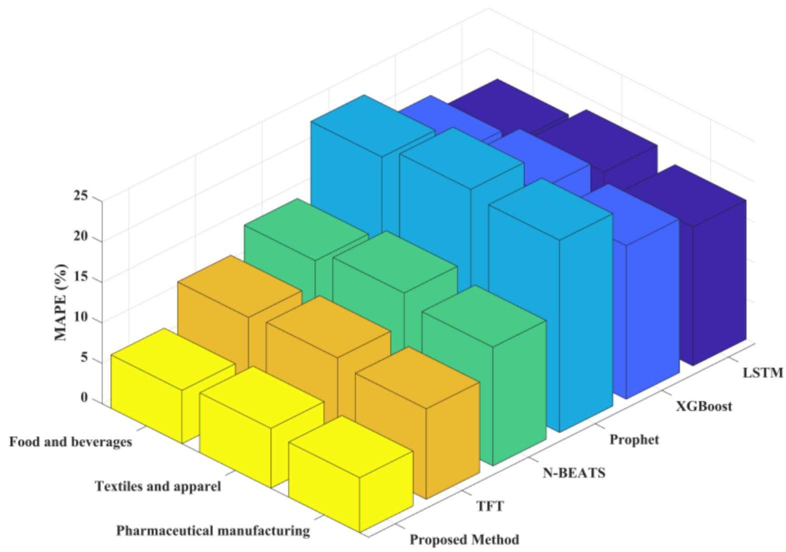

FIGURE 7. Model Generalization Ability Evaluation

target industry, demonstrating strong structural adaptability and extrapolation capabilities.

G. MODEL ROBUSTNESS EVALUATION

Simulating data quality degradation on the test set: randomly deleting 10% of the data points to simulate missing data, and adding $\pm 15\%$ uniform noise to simulate measurement error. Calculating the increase in MAPE after the disturbance:

$$\Delta \text{MAPE} = \frac{|\text{MAPE}_{\text{noisy}} - \text{MAPE}_{\text{clean}}|}{\text{MAPE}_{\text{clean}}} \tag{35}$$

Figure 8 shows the robustness performance of six methods: the proposed method, TFT, N-BEATS, Prophet, XGBoost, and LSTM, across the electronics, chemical, and machinery industries. The horizontal axis shows the different methods, while the vertical axis shows the rate of change in MAPE (ΔMAPE) under 10% data loss and $\pm 15\%$ noise, which measures the model's stability under data quality degradation. LSTM maintains a range of 41.2% to 43.5%, XGBoost reaches 45.8% to 47.3%, and Prophet reaches 47.1% to 48.9%, indicating its sensitivity to noise. Despite its stable structure, N-BEATS still exhibits a rate of change exceeding 40%. TFT exhibits anomalous performance, with a ΔMAPE as high as 51.3% to 54.6%, indicating that its complex structure is prone to overfitting under perturbations. The proposed method improves by only 32.5%, 33.8%, and 33.6% in the three industries, respectively, reaching an average of 33.3%. All scatter plots are at the lowest level, significantly outperforming the comparison methods and demonstrating strong noise tolerance.

This exceptional robustness stems from the model's application of a gated residual network and variable selection mechanism during the feature fusion phase, which dynamically filters and suppresses noise from input features. In the presence of missing data or measurement errors, traditional models struggle to distinguish between key signals and abnormal fluctuations, leading to an imbalance in feature weight distribution. The proposed method uses a gating mechanism to automatically downweight the contribution of damaged or unstable features while retaining rule-constrained, high-confidence feature pathways, forming a sparse representation that is resilient to interference. A multi-step decoding structure aligns known future variables with historical contextual feedback, enhancing the smoothness and consistency of the prediction path. This stability design, based on structural redundancy and dynamic adjustment, enables the model to maintain a reasonable financial evolution trajectory despite input perturbations, demonstrating its robust modeling capabilities in complex real-world environments.

H. STABILITY UNDER DIFFERENT ECONOMIC FLUCTUATION CYCLES

To test the model's generalization and stability across different market environments, we divided the test set into "high volatility" and "low volatility" periods based on macroeconomic volatility. The specific classification criteria are as follows:

High volatility: When the standard deviation of the quarterly year-on-year GDP growth rate exceeds one standard deviation from the historical mean, this is defined as a high volatility economic environment. Low volatility: When the standard deviation of the quarterly year-on-year GDP growth rate is below the historical mean, this is defined as a low volatility economic environment. The MAPE of each model was calculated for these two subsets. As can be seen from Table 3, the proposed method not only achieves the best accuracy (MAPE=5.8 $\pm$ 0.3%) in the low volatility period, but also maintains the lowest absolute error (MAPE=7.1 $\pm$ 0.6%) in the most challenging high volatility period. The standard deviations are 0.3 and 0.6 respectively, which are much lower than those of other models, indicating that its forecasting performance is the most stable under different economic environments without drastic fluctuations. The performance gap between periods of high and low volatility (+1.3%) is the smallest of all models, demonstrating its strong adaptability and stability. The proposed method demonstrates remarkable robustness and stability. Even in an uncertain environment characterized by significant macroeconomic fluctuations, the model provides reliable and consistent forecasts, offering significant practical value to corporate finance managers who must make critical decisions amidst dynamic risk.

TABLE 3. Performance Stability Across Economic Cycles

Method	Low volatility period MAPE	High volatility period MAPE
Proposed method	5.8% $\pm$ 0.3%	7.1% $\pm$ 0.6%
TFT	7.5% $\pm$ 1.1%	10.2% $\pm$ 1.5%
N-BEATS	8.3% $\pm$ 1.3%	12.1% $\pm$ 1.8%
Prophet	19.5% $\pm$ 2.1%	26.8% $\pm$ 3.0%
XGBoost	16.2% $\pm$ 1.9%	21.4% $\pm$ 2.3%
LSTM	14.8% $\pm$ 1.7%	19.6% $\pm$ 2.1%

I. ABLATION EXPERIMENTS

To quantify the contribution of each core component to model performance, we constructed multiple ablation variants, as shown in Table 4. All variants use the same feature set, hyperparameters, and training process as the full model to ensure fair comparison.

Ablation experiments show that each core component significantly contributes to model performance.

After removing the financial rule constraint , the MAPE for cash flow forecasting increased from 6.3% to 8.7%, while the Theil's U for debt structure deteriorated from 0.20 to 0.28, showing the most significant performance degradation. This demonstrates that this mechanism, by guiding the model to focus on risky periods such as "high debt and low cash flow," is the most critical factor in improving forecast accuracy and stability. After removing the BiLSTM encoder, the MAPE increased to 7.5% and the Theil's U reached 0.23, demonstrating that it effectively extracts long-term temporal dependencies and suppresses noise. After removing the GRN feature filtering, the MAPE and Theil's U reached 7.0% and 0.22, respectively, demonstrating that this module adaptively focuses on key features and enhances model robustness. The complete model achieves optimal performance across all metrics, validating the effectiveness

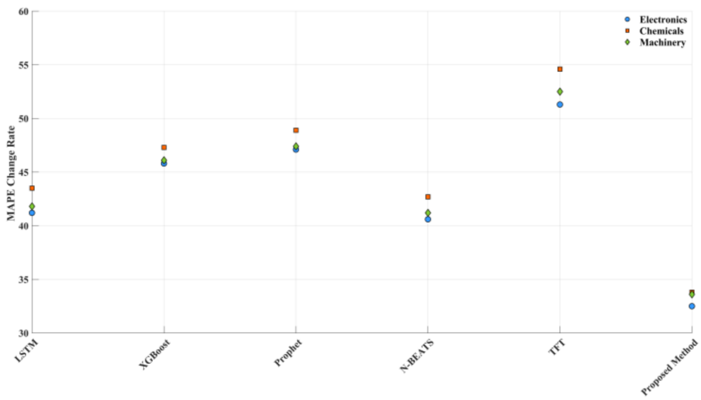

FIGURE 8. Model Robustness Evaluation

TABLE 4. Ablation experiments

Model variant	Rule constraint	BiLSTM encoder	GRN feature selection	Cash flow MAPE (%)	Debt structure Theil's U
Full model	√	√	√	6.3	0.20
Without rule constraint	×	√	√	8.7	0.28
Without BiLSTM encoder	√	×	√	7.5	0.23
Without GRN feature selection	√	√	×	7.0	0.22

of the "knowledge-guided" and "data-driven" fusion architecture.

IV. CONCLUSION

This study proposes a TFT model enhanced with financial-rule constraints for joint cash-flow and liability forecasting. The model integrates static, historical, and known-future feature embeddings, employs a BiLSTM for temporal encoding, and embeds rules (debt-ratio and cash-flow safety thresholds) into the attention mechanism to adjust weights on high-risk periods. Experiments show the model achieves an average MAPE of 6.3%, a Theil's U of 0.20 for debt structure, and an 89.7% rule-compliance rate, with over 90% early-warning accuracy, validating the 'data-driven + knowledge-guided' paradigm for robust, interpretable financial forecasting.

AUTHOR CONTRIBUTIONS

Minqing Liu designed, collected and analyzed the data, and drafted the manuscript. Minqing Liu conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Minqing Liu participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] L. Wei, "Optimizing cash flow management for ship financing leasing: Risk identification, risk assessment, and management strategies," Cogent Business & Management, vol. 12, no. 1, Art. no. 2464941, 2025, doi: 10.1080/23311975.2025.2464941.

[2] A. E. Nasimiyu, "Cashflow management practices and financial performance of small and medium business enterprises in Kenya," African Journal of Commercial Studies, vol. 4, no. 3, pp. 252-263, 2023, [Online]. Available: https://ijcsacademia.com/index.php/journal/article/view/182.

[3] B. H. Subhani, S. Zunhuan, and M. A. Khan, "Finance for a Greener future: Evolving the financial sector for ESG and sustainable corporate debt management," Borsa Istanbul Review, vol. 25, no. 2, pp. 337-349, 2025, doi: 10.1016/j.bir.2025.01.011.

[4] F. Fasano, C. Adornetto, I. Zahid, M. La Rocca, L. Montaleone, G. Greco, et al., "The dilemma of accuracy in bankruptcy prediction: A new approach using explainable AI techniques to predict corporate crises," European Journal of Innovation Management, vol. 28, no. 11, pp. 1-22, 2025, doi: 10.1108/EJIM-06-2024-0633.

[5] C. P. Vijay, "A Deep Learning Approach for Financial Risk Prediction in Enterprise Management Systems," EKSPLORIUM-BULETIN PUSAT TEKNOLOGI BAHAN GALIAN NUKLIR, vol. 46, no. 1, pp. 1579-1597, 2025, [Online]. Available: https://eksplorium.com/index.php/journal/article/view/184.

[6] H. Wei and X. Wang, "Financial Risk Management Early-Warning Model for Chinese Enterprises," Journal of Risk and Financial Management, vol. 17, no. 7, pp. 255, 2024, doi: 10.3390/jrfm17070255.

[7] H. Chu, "An Empirical analysis of corporate financial management risk prediction based on associative memory neural network," Computational Intelligence and Neuroscience, vol. 2021, no. 1, Art. no. 4383742, 2021, doi: 10.1155/2021/4383742.

[8] N. R. Mazouei, "Investigating the Effect of Cash Flow Management on Firm Performance: The Mediating Role of the Business Cycle in Companies Listed in the Tehran Stock Exchange," Accounting and Auditing with Applications, vol. 1, no. 2, pp. 103-114, 2024, doi: 10.22105/aaa.v1i2.39.

[9] Y. Luo and C. Jiang, "The impact of corporate capital structure on financial performance based on convolutional neural network," Computational Intelligence and Neuroscience, vol. 2022, no. 1, Art. no. 5895560, 2022, doi: 10.1155/2022/5895560.

[10] D. Septyanto, N. F. Welandasari, and N. Sayidah, "An empirical test of the financial ratio effect on financial distress in Indonesia (study in garment and textile industry)," Economics and Business Quarterly Reviews, vol. 5, no. 2, pp. 222-230, 2022, [Online]. Available: https://ssrn.com/abstract=4146069.

[11] M. Akhtar, K. Yusheng, M. Haris, Q. U. Ain, and H. M. Javaid, "Impact of financial leverage on sustainable growth, market performance, and profitability," Economic Change and Restructuring, vol. 55, no. 2, pp. 737-774, 2022, doi: 10.1007/s10644-021-09321-z.

[12] H. H. Yousef and H. K. Ojah, "The Effect of Free Cash Flow on the Efficiency of Investment Decisions in a Sample of Iraqi Companies Listed on the Iraqi Stock Exchange," American Journal of Economics and Business Management, vol. 5, no. 11, pp. 176-191, 2022, [Online]. Available: https://www.researchgate.net/profile/Hashim-Hayder-2/publication/365823160_The_Effect_of_Free_Cash_Flow_on_the_Efficiency_of_Investment_Decisions_in_a_Sample_of_Iraqi_Companies_Listed_on_the_Iraqi_Stock_Exchange.

[13] T. Ling and Y. Cai, "Financial Crisis Prediction Based on Long-Term and Short-Term Memory Neural Network," Wireless Communications and Mobile Computing, vol. 2022, no. 1, Art. no. 5728470, 2022, doi: 10.1155/2022/5728470.

[14] H. Yang, E. Li, Y. F. Cai, J. Li, and G. X. Yuan, "The extraction of early warning features for predicting financial distress based on XGBoost model and shap framework," International Journal of Financial Engineering, vol. 8, no. 03, Art. no. 2141004, 2021, doi: 10.1142/S2424786321410048.

[15] K. Li and Y. Zhou, "Improved financial predicting method based on time series long short-term memory algorithm," Mathematics, vol. 12, no. 7, pp. 1074, 2024, doi: 10.3390/math12071074.

[16] L. C. de Jesus Jr, F. Fernández-Navarro, and M. Carbonero-Ruz, "Enhancing financial time series forecasting through topological data analysis," Neural Computing and Applications, vol. 37, no. 9, pp. 6527-6545, 2025, doi: 10.1007/s00521-024-10787-x.

[17] H. Yu, "An Adaptive Sharpe Ratio-Based Temporal Fusion Transformer for Financial Forecasting," Frontiers in Humanities and Social Research, vol. 2, no. 1, pp. 54-65, 2025, doi: 10.71465/fhsr23.

[18] J. Yang, P. Li, Y. Cui, X. Han, and M. Zhou, "Multi-Sensor Temporal Fusion Transformer for Stock Performance Prediction: An Adaptive Sharpe Ratio Approach," Sensors, vol. 25, no. 3, pp. 976, 2025, doi: 10.3390/s25030976.

[19] A. Farooq, M. I. Uddin, M. Adnan, A. A. Alarood, E. Alsolami, and S. Habibullah, "Interpretable multi-horizon time series forecasting of cryptocurrencies by leverage temporal fusion transformer," Heliyon, vol. 10, no. 22, Art. no. e40142, 2024, [Online]. Available: https://www.cell.com/heliyon/fulltext/S2405-8440(24)16173-7.

[20] J. Liu, S. Wu, and V. T. Pham, "Design of Accounting Earnings Forecasting Model Based on Artificial Intelligence," Scientific Programming, vol. 2022, no. 1, Art. no. 8793596, 2022, doi: 10.1155/2022/8793596.

[21] C. Zhang, H. Zhang, T. Pu, and J. Pan, "Supply Chain Demand Forecasting Based on Data Mining Algorithm and Seq2Seq," International Journal of Control, Automation and Systems, vol. 23, no. 1, pp. 89-104, 2025, doi: 10.1007/s12555-024-0141-8.

[22] H. Niu, Q. Pan, and K. Xu, "Hybrid deep learning models with multi-classification investor sentiment to forecast the prices of China's leading stocks," Plos One, vol. 18, no. 11, Art. no. e0294460, 2023, doi: 10.1371/journal.pone.0294460.

[23] X. Zhang, Y. Ma, and M. Wang, "An attention-based Logistic-CNN-BiLSTM hybrid neural network for credit risk prediction of listed real estate enterprises," Expert Systems, vol. 41, no. 2, Art. no. e13299, 2024, doi: 10.1111/exsy.13299.

[24] E. Balat, C. T. Ekinci, H. Ş. Arlı, C. Ulus, and M. F. Akay, "Deep Learning Based Financial Forecasting Models," AINTELIA Science Notes Journal, vol. 1, no. 2, pp. 6-12, 2022, doi: 10.5281/zenodo.8071617.

[25] J. Li and C. Wang, "A deep learning approach of financial distress recognition combining text," Electronic Research Archive, vol. 31, no. 8, pp. 4683-4707. https://www.aimspress.com/article/doi/10.3934/era.2023240, 2023.

[26] H. Lin and Q. Sun, "Financial volatility forecasting: A sparse multi-head attention neural network," Information, vol. 12, no. 10, pp. 419, 2021, doi: 10.3390/info12100419.

[27] S. Knifo and A. Alzubi, "DR-Z2AN: dual-recurrent neural network with a tri-channel attention mechanism for financial management prediction," Complex & Intelligent Systems, vol. 11, no. 1, pp. 53, 2025, doi: 10.1007/s40747-024-01613-x.

[28] W. Xu, Y. Liu, W. Liu, H. Li, and G. Sun, "HDML: Hybrid data-driven multi-task learning for China's stock price forecast," Applied Intelligence, vol. 54, no. 23, pp. 12420-12438, 2024, doi: 10.1007/s10489-024-05838-8.

[29] Z. Li, Z. Jiang, and X. Pan, "Default risk prediction of enterprises based on convolutional neural network in the age of big data: Analysis from the viewpoint of different balance ratios," Complexity, vol. 2022, no. 1, Art. no. 5139562, 2022, doi: 10.1155/2022/5139562.

[30] X. Deng, "Research on Enterprise Financial Information Risk Identification and Dynamic Control Management Model Based on Genetic Algorithm," International Journal of High Speed Electronics and Systems, vol. 34, no. 03, Art. no. 2540019, 2025, doi: 10.1142/S0129156425400191.

[31] J. Laborda, S. Ruano, and I. Zamanillo, "Multi-Country and Multi-Horizon GDP Forecasting Using Temporal Fusion Transformers," Mathematics, vol. 11, no. 12, pp. 2625, 2023, doi: 10.3390/math11122625.

[32] Y. Pang, S. Shi, Y. Shi, and Y. Zhao, "A nonlinear dynamic approach to cash flow forecasting," Review of Quantitative Finance and Accounting, vol. 59, no. 1, pp. 205-237, 2022, doi: 10.1007/s11156-022-01066-8.

[33] B. Talebi, R. Abdi, Z. Hajiha, and N. Rezaei, "Modelling Optimal Predicting Future Cash Flows Using New Data Mining Methods (A Combination of Artificial Intelligence Algorithms)," Advances in Mathematical Finance and Applications, vol. 4, no. 3, pp. 1007, 2023, doi: 10.22034/AMFA.2021.1910900.1493.

[34] S. J. Delgado, N. J. Panganiban, K. A. Robles, A. D. Buenaventura, M. J. Vergara, and C. N. Evangelista, "Advancements in Deep Learning: A Comprehensive Survey on Architectures, Optimization Techniques, and Applications," International Journal of Technology and Modeling, vol. 3, no. 2, pp. 57-68, 2024, doi: 10.63876/ijtm.v3i2.141.

[35] Y. Shao, J. Yang, W. Zhou, H. Sun, L. Xing, Q. Zhao, et al., "An improvement of Adam based on a cyclic exponential decay learning rate and gradient norm constraints," Electronics, vol. 13, no. 9, pp. 1778, 2024, doi: 10.3390/electronics13091778.