Design and Implementation of FARPM-Net Model for Financial Risk Prediction and Automated Auditing in Enterprises

R. Sudha[1*], R. Saranya[2], Payal R Kothari[3]

[1*]Associate Professor, Department of Commerce, PSG College of Arts & Science, Coimbatore, India;

Email: r_sudha@psgcas.ac.in.

[2]Assistnat Professor, Department of Computer Science, PSG College of Arts & Science, Coimbatore, India;

Email:saranya_r@psgcas.ac.in

[3] Research Scholar -PSG College of Arts & Science, Coimbatore. Practicing Cost Accountant & Proprietor-Kothari & Co. Email Id: - pragati169@gmail.com

*Corresponding Author: r_sudha@psgcas.ac.in

DOI: https://doi.org/10.30211/JIC.202503.018

Submitted: Oct. 14, 2025 Accepted: Dec. 17, 2025

ABSTRACT

This paper addresses the complex demands of enterprise financial risk prediction and automated auditing by proposing an innovative deep learning model—FARPM-Net—based on multimodal fusion and multi-level temporal modeling. The model integrates a multi-level Mamba module, temporal convolutional network (TCN), and an attention-based cross-modal fusion module to achieve integration of structured financial data, unstructured textual information, and external market factors, while capturing dynamics across multiple time scales. Experimental validation on two public datasets, SEC EDGAR 10-K and Yahoo Finance, demonstrates that FARPM-Net attains accuracies of 91.5% and 90.2%, respectively, representing a 4.7% improvement over mainstream models; F1 scores increase by up to 6.1%, and mean absolute error decreases by more than 16%, showcasing excellent capabilities in risk identification. Ablation studies confirm the contributions of each key module to the performance, verifying the synergistic advantages of multimodal fusion and multi-level temporal modeling. This work enhances the accuracy and stability of financial risk prediction and provides technical support for intelligent analysis of multimodal financial data. Future research will focus on model optimization, cross-modal fusion strategies, and interpretability to promote practical applications in auditing and risk management.

Keywords: Financial risk, Automated auditing, Hierarchical modeling, Mamba, Cross-modal fusion

1. Introduction

With the advent of the digital age, the volume of corporate financial data has been steadily increasing, presenting challenges to financial auditing and risk prediction. Traditional financial auditing methods often rely on manual reviews and static financial statement data, which struggle to cope with the dynamic, large-scale nature of financial data. This issue is particularly pronounced with

the development of complex business environments such as cross-border e-commerce and globalized transactions, where traditional methods are inadequate for real-time response to changing financial risks[1]. Therefore, leveraging advanced technologies, especially deep learning techniques, to improve the efficiency of financial data auditing and the accuracy of risk prediction has become a focal point of current research[1].

However, existing financial risk prediction and auditing methods have several limitations[2]. Traditional rule-based auditing systems are unable to handle real-time financial data streams and unstructured data, and they rely on static financial statements, making it difficult to capture dynamic changes in financial data[3]. In recent years, sequential modeling methods such as LSTM and GRU have been able to capture temporal dependencies in financial data; however, these methods often neglect the fusion of multimodal data and fail to comprehensively consider the impact of heterogeneous data, such as text and images, on financial risk[4]. While deep learning methods like Transformer and BERT have shown remarkable performance in handling textual data, effectively integrating multimodal data, especially in multi-target risk prediction tasks, remains an unresolved challenge. Additionally, the application of cutting-edge technologies such as Graph Neural Networks (GNN) and self-supervised learning in financial auditing is gradually increasing, yet challenges in computational efficiency and scalability persist[5].

To address these issues, this paper proposes a financial risk prediction and automated auditing method based on the FARPM-Net model. FARPM-Net combines the advantages of Temporal Convolutional Networks (TCN) and the Mamba network, effectively handling temporal dependencies in financial data. Additionally, through multimodal feature cross-fusion techniques, it enhances the model's ability to understand heterogeneous data. The contributions of this paper are primarily reflected in the following three aspects:

- The FARPM-Net model is introduced, combining Temporal Convolutional Networks (TCN) and the Mamba network to improve financial risk prediction accuracy.
- A cross-fusion module for multimodal features is developed to enhance the model's capability to handle complex data.
- An adaptive mechanism is incorporated, boosting the model's stability and flexibility in predicting multiple financial risk targets.

The structure of this paper is organized as follows: Section 2 reviews related research, focusing on the limitations of existing financial auditing and risk prediction methods as well as the applications of deep learning techniques. Section 3 introduces the design and implementation of the FARPM-Net model, including the functions and architecture of its core modules. Section 4 validates the effectiveness of the FARPM-Net model through experiments, demonstrating its advantages in financial risk prediction. Finally, Section 5 summarizes the work presented in this paper and discusses future research directions.

2. Related work

2.1 Deep Learning Applications in Financial Auditing and Risk Detection

With the widespread application of deep learning, significant progress has been made in the fields of financial auditing and risk detection. Graph Convolutional Networks (GCN) have been applied to model the complex financial relationships between enterprises, capturing the interconnections and capital flows between different companies, thereby improving the accuracy of financial risk prediction[6]. However, GCN methods typically require the pre-construction of graph structures between enterprises, and when handling large-scale financial data, they incur high computational costs, making it difficult to efficiently scale to real-time auditing and prediction tasks. Transformer Networks have gained widespread use due to their powerful sequence modeling capabilities, especially in handling temporal dependencies in long-term financial data. They have shown remarkable performance in text data analysis and natural language processing tasks, but transformer models generally rely on large-scale data training and expensive computational resources, and they have limited capability in handling nonlinear and non-stationary features in financial data[7]. In the domain of anomaly detection, Contrastive Learning has been employed to identify potential financial risks by learning the differences between normal and anomalous financial behaviors. However, this method often faces the issue of sample imbalance, particularly in financial data, where anomalous data is relatively scarce, leading to challenges in effective generalization[8]. The Deep Forest model, as an unsupervised learning method, is capable of addressing financial datasets that lack labeled data, but it has limitations when it comes to multimodal data fusion and handling complex long-term dependencies[9]. Adaptive Generative Models have been used to generate realistic financial data to simulate risk scenarios, but the quality and authenticity of generated data remain challenges when processing real, dynamically changing financial data[10].

Although these methods have made progress, they face limitations in managing long-term dependencies, real-time computation, and multiobjective optimization of financial data. The FARPM-Net model overcomes these challenges by combining TCN and Mamba, and integrating multimodal feature cross-fusion techniques. This approach efficiently addresses multimodal data processing, time-dependent modeling, and multitarget risk prediction, providing a more accurate solution.

2.2 Application of Temporal Modeling Techniques in Risk Prediction

With the rapid development of financial data analysis, an increasing number of studies have begun to explore how multimodal data fusion can be applied to risk assessment. Multimodal Convolutional Neural Networks (MC-CNN) is a common fusion method that integrates both image and structural data in financial reporting and graphical analysis. This method learns common features from image and text data to improve risk forecasting performance[11]. Deep learning by cooperation is particularly suitable for managing the information from multiple sources of financial data, improving the performance of each modal through information exchange between various modalities[12]. Multimodal autonomous learning uses the unsupervised learning mechanism to learn marmoidal features from automatically unidentified data[13]. Cross-modal Generative Adversarial Networks (GANs) is used for financial data for conversion between images and texts, generates more diverse synthetic data, and improves the adaptability of the model in the processing of various data formats[14]. The change autocoder (AEV) maps the data in various ways in the latent space using its

generative capability, and merges the various data features to improve the efficiency and accuracy of the financial risk assessment[15].

Unlike these methods, FARPM-Net combines TCN and Mamba networks with a cross-fusion module for multimodal functions. This approach addresses time dependence and multimodal characteristics of financial data, while exploring complex relationships between modalities, greatly enhancing the accuracy and adaptability of financial risk forecasting.

3. Method

3.1. Overview of Our Model

The proposed FARPM-Net model is designed for automated corporate financial risk prediction, integrating multimodal recognition and hierarchical time modeling. It uses a two-stage hybrid architecture to leverage complementary relationships between heterogeneous data sources and model both short-term and long-term financial dynamics. Figure 1 shows the FARPM-Net architecture, consisting of a front-end multimodal feature extraction and fusion module, and a back-end multi-level Mamba modeling module. The model effectively captures the heterogeneity and temporal dependencies of financial data, enabling end-to-end risk prediction with high scalability and precision.

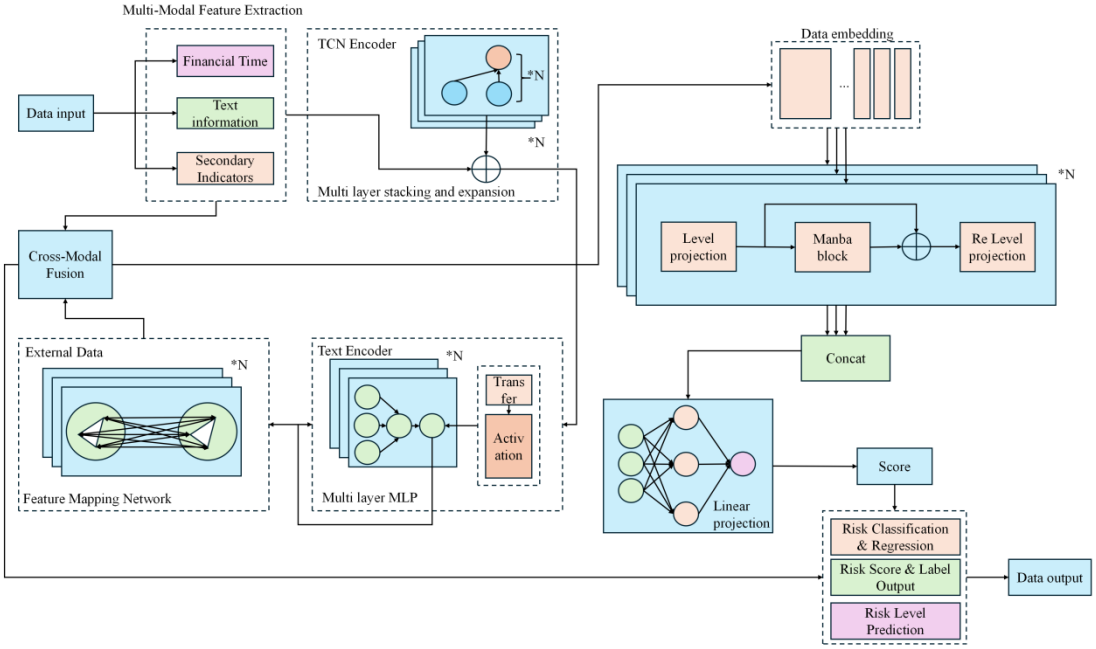

Figure 1. Overall architecture of the FARPM-Net model. The model consists of a multimodal feature extraction and cross-fusion module, and a multi-level Mamba modeling module.

The FARPM-Net model integrates three primary data sources: structured financial time series, unstructured textual information, and external auxiliary data. These are processed through three separate encoding channels: a TCN encoder for financial time series, a lightweight semantic encoder for textual data, and a feature mapping network for external factors. The encoded representations are then fused through a cross-modal fusion module using an attention mechanism, generating a unified representation. This is fed into the multi-level Mamba module, which captures both short-term (fine-

grained) and long-term (coarse-grained) temporal dependencies via state-space modeling. The fine-grained branch models rapid fluctuations, while the coarse-grained branch focuses on quarterly and annual trends. The fused outputs are projected and concatenated for risk prediction. FARPM-Net's architecture preserves deep temporal modeling capabilities while incorporating multimodal feature fusion, significantly improving the accuracy and robustness of financial risk prediction. It supports end-to-end training and can be extended to accommodate more modalities and fine-grained tasks, offering strong application potential and generalization ability.

3.2. Temporal Convolutional Network Encoding Module

The TCN module in FARPM-Net for modeling structured financial time series is illustrated in Figure 2. This module processes time series data of core enterprise operational indicators, such as balance sheets, income statements, and cash flow statements, focusing on capturing local fluctuations and mid-term trend changes. The TCN model, based on causal and dilated convolutions, employs a multi-layer stacking and dilation mechanism to efficiently model locally significant information within long historical sequences. Unlike the Mamba module, which emphasizes global modeling, the TCN module prioritizes fine-grained expression of short-term dynamics and boundary signals, complementing the subsequent temporal path's ability to model abrupt changes.

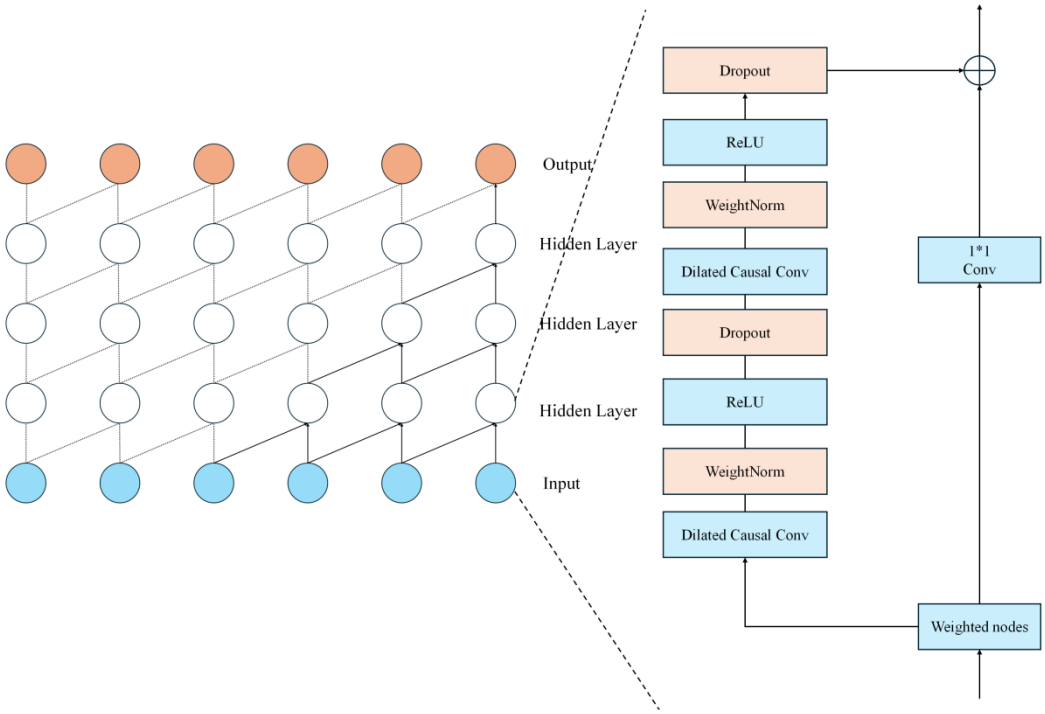

Figure 2. Architecture of the TCN encoding module for structured financial time series in FARPM-Net.

The input raw financial time series is denoted as $(X_{ts}=\{x_1^{ts},x_2^{ts},...,x_T^{ts})$, where X_t represents the financial feature at the t-th time step with dimensionality d_{ts}, and the sequence length is T. At the l-th layer of the TCN, the sequence is modeled using a dilated convolution structure. The output at the t-th time step of the l-th layer is $h_t^{(l)}$, where $W_i^{(1)}$ denotes the weights of the i-th convolutional kernel,

$r^{(l)}$ is the dilation rate of this layer, k is the kernel size, and $b^{(l)}$ is the bias term. The input to the current layer is $h_t^{(l-1)}$, with the initial input defined as in (1).

$$h_t^{(l)} = \sum_{i=0}^{k-1} W_i^{(l)} \cdot h_{t-r^{(l)} \cdot i}^{(l-1)} + b^{(l)} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 1]}$$

As the number of network layers increases, the model's receptive field gradually expands. Let R denote the overall receptive field of the model, L be the total number of TCN layers, and $r^{(l)}$ the dilation rate at the l-th layer. The receptive field determines the historical time-step range the model can utilize at the current moment and serves as a key metric for assessing its modeling capacity as in (2).

$$R = 1 + (k - 1) \cdot \sum_{l=0}^{L-1} r^{(l)} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 2]}$$

After each convolutional layer, batch normalization and nonlinear activation operations are applied to enhance model stability, and residual connections are used to preserve lower-layer input information. The tensor H_{TCN} represents the time-series embedding generated by the TCN encoder, with dimensions $T \times d'$, where $h^{(L)}$ denotes the output of the final layer. Here, BN represents the batch normalization operation, and ReLU denotes the activation function as in (3).

$$H_{TCN} = \mathrm{ReLU}(\mathrm{BN}(h^{(L)})) + h^{(0)} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 3]}$$

The feature sequence output by this module serves as the temporal modality embedding, which is input to the modality fusion module for joint modeling with textual and external information. The TCN module excels at capturing sudden changes, short-term fluctuations, and anomalies, complementing the Mamba module. While the TCN provides high-resolution, local risk perception, the Mamba module models long-term dependencies, enhancing the accuracy and robustness of financial risk prediction.

3.3. Multi-Modal Feature Cross Fusion Module

The primary task of the multimodal feature cross-fusion module in FARPM-Net is to effectively integrate features from three distinct modalities: financial time series, textual information, and external auxiliary data. Since these data originate from different sources and exhibit heterogeneous representations, fully exploiting their interrelations and complementarities is crucial for improving the accuracy of financial risk prediction. The merge module uses an attention mechanism that facilitates the exchange of information between modalities and creates an explicit correlation matrix that generates a uniform functional expression rich in multisource information. This expression forms a solid basis for later time modeling. Figure 3 shows the core of this module.

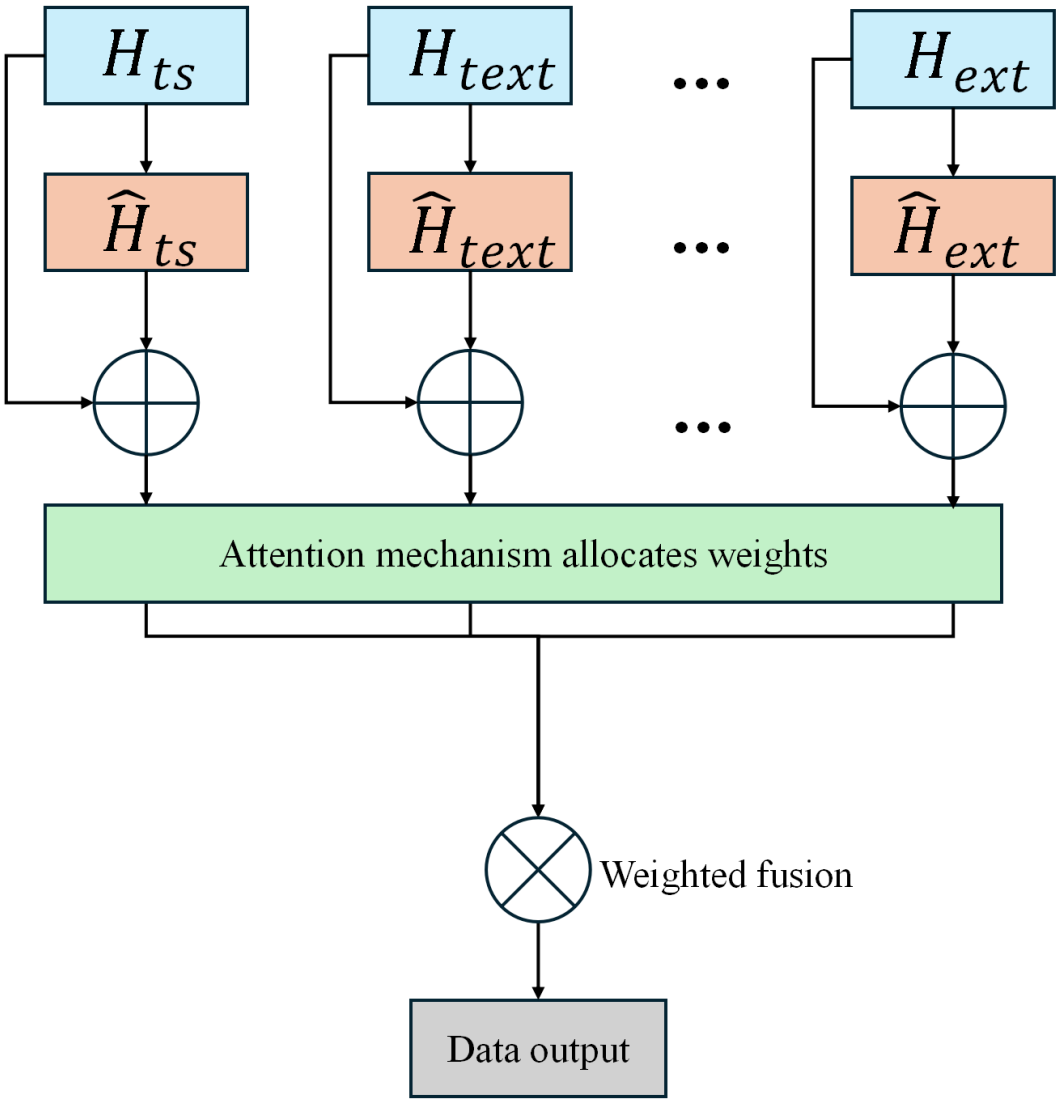

Figure 3. Architecture of the multimodal feature cross-fusion module in FARPM-Net.

Specifically, the features from the three modalities are denoted as financial time series features H_{ts}, textual features H_{text}, and external auxiliary features H_{ext}. Due to differences in their dimensionalities and distributions, the module first applies linear transformations to map these feature sets into a common feature space. The corresponding weight matrices are W_{ts}, W_{text}, and W_{ext}, with bias terms b_{ts}, b_{text}, and b_{ext}, respectively. This operation unifies the feature dimensions of the three modalities, facilitating subsequent fusionas in (4)(5)(6).

$$\widetilde{H}_{ts} = H_{ts}W_{ts} + b_{ts} \cdots\cdots \text{[Formular 4]}$$

$$\widetilde{H}_{text} = H_{text}W_{text} + b_{text}b_{ts} \cdots\cdots \text{[Formular 5]}$$

$$\widetilde{H}_{ext} = H_{ext}W_{ext} + b_{ext}b_{ts} \cdots\cdots \text{[Formular 6]}$$

Subsequently, the fusion module computes the similarity between different modalities and employs an attention mechanism to automatically assign weights, thereby quantifying the influence

among modalities. A_{ij} represents the correlation weight matrix between modality i and modality j, where d' denotes the feature dimension. The softmax operation ensures normalization of the weightsas in (7).

$$A_{ij} = \text{softmax}\left(\frac{\tilde{H}_i \tilde{H}_j^T}{\sqrt{d'}}\right) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 7]}$$

Based on the weight matrices, the module performs weighted fusion of features from each modality to form richer feature representations. Finally, the fusion module concatenates the weighted features from all modalities to obtain the final multimodal fused representation as in (8)(9).

$$\hat{H}_i = A_{ij}\tilde{H}_j \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 8]}$$

$$H_{fusion} = \text{Concat}(\hat{H}_{ts}, \hat{H}_{text}, \hat{H}_{ext}) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 9]}$$

Through this cross-fusion mechanism, FARPM-Net is able to more comprehensively and accurately understand the financial condition of the enterprise and changes in the external environment, thereby significantly improving the effectiveness of financial risk prediction. This module not only facilitates collaborative modeling across modalities but also provides the model with a more representative and robust multimodal feature representation, serving as a critical component for achieving efficient financial risk perception.

3.4. Multi-Level Mamba Modeling Module

To more fully capture the dynamic changes in financial data across different temporal scales, this paper designs the multi-level Mamba modeling module, as shown in Figure 4, within FARPM-Net. The module consists of two parallel Mamba branches, each handling data patterns at fine-grained and coarse-grained temporal scales. The goal is to simultaneously capture local fluctuations and macro trends in financial data, thereby enhancing the model's ability to perceive and represent multi-period financial risks. Through the dual-scale state-space modeling mechanism, FARPM-Net explicitly models different temporal dependency patterns within a unified representation space, providing rich and multi-level semantic feature support for subsequent risk prediction tasks.

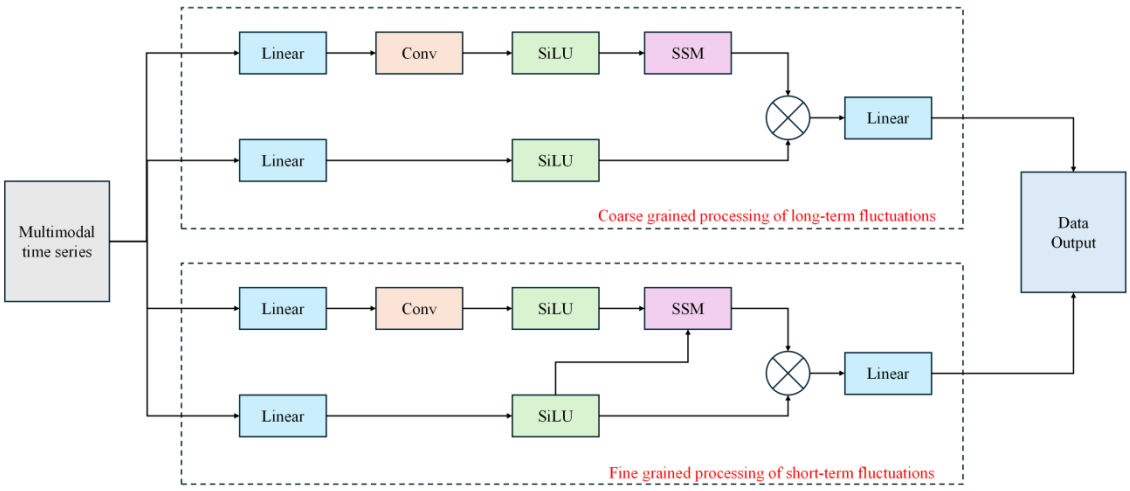

Figure 4. Architecture of the multi-level Mamba modeling module in FARPM-Net. The module

consists of two parallel Mamba branches: one for modeling short-term fluctuations and the other for long-term trends in financial data.

The model input is the multimodal time series ($X=\{x_1,x_2,...,x_T\}$) obtained after the previous stage of cross-modal fusion, where ($x_t \in R^d$) represents the input vector at the t-th time step. To accommodate the state-space modeling structure, the input is first transformed into an intermediate representation via a linear mapping, where L denotes the mapped embedding vector, $W_e \in R^{h \times d}$ is the weight matrix, and $b_e \in R^h$ is the bias term, both of which are learnable parameters as in (10).

$$u_t = W_e x_t + b_e \quad \cdots\cdots \text{[Formular 10]}$$

Subsequently, the model establishes temporal evolution relationships in the latent space, where s_t represents the hidden state vector. A and $B \in R^{h \times h}$ are the state transition matrix and input mapping matrix, respectively, while σ denotes the nonlinear gating function used to enhance the model's ability to select key state changes. Through this mechanism, the model fuses and updates information from each time step with historical states, thereby constructing a dynamic dependency chain. This design differs from the chain structure of traditional RNN models, with its core advantage being the support for parallel state propagation and global modeling capability, which significantly alleviates gradient vanishing and performance degradation, especially when handling long sequences as in (11)(12).

$$s_t = As_{t-1} + Bu_t \quad \cdots\cdots \text{[Formular 11]}$$

$$h_t = \sigma(s_t) + u_t \quad \cdots\cdots \text{[Formular 12]}$$

With respect to implementation details FARPM-Net Creates two independent Mamba branches to manage granularity data at different times. The fine grain office is focused on detecting original anomalies such as cash flow fluctuations, increased demand, and unexpected expenditure that function as sudden risk signals during short cycles. On the other hand, the coarse grained field uses an array of glide window aggregations or sample representations to handle an array and focus on mid - and long-term trends such as quarterly or annual changes, including macro risk features such as changes in asset structure, increased debt ratios, or decreased profitability.

The two branches ultimately generate feature representations, which are then merged into a unified temporal representation vector. H_f and H_c represent the sequence outputs along the time dimension from the fine-grained and coarse-grained Mamba branches, respectively. W_f and W_c are the linear projection weight matrices, while b_f and b_c are the corresponding bias terms. These generate the global embedding vectors for the two branches in a unified dimension as in (13)(14).

$$z_f = W_f H_f + b_f, \quad z_c = W_c H_c + b_c \quad \cdots\cdots \text{[Formular 13]}$$

$$z = \text{concat}(z_f, z_c) \quad \cdots\cdots \text{[Formular 14]}$$

The Mamba module, using state-space modeling, efficiently captures the semantics of financial data, handling periodicity, lag, and trends in corporate financials. It overcomes the limitations of traditional neural networks by incorporating dynamic system mechanisms, allowing for the modeling of both current and historical states. This enables FARPM-Net to track long-term corporate state evolution and capture risk signals more effectively. The module's global convolution and attention

characteristics further enhance its ability to weigh information across time steps, improving prediction accuracy and robustness.

4. Experiment

4.1 Datasets

This experiment utilized two publicly available and widely recognized financial datasets, which include both textual and structured time-series data, meeting the multimodal fusion requirements of FARPM-Net. The SEC EDGAR 10-K dataset, released by the U.S. Securities and Exchange Commission, includes detailed financial statements and management discussion reports that publicly traded companies must submit annually[16]. It contains structured data such as balance sheets, income statements, and cash flow statements, along with in-depth descriptions of the company's operational performance and risk factors. This combination of textual and numerical data offers valuable input for multimodal models, facilitating the exploration of the underlying relationships and potential risk signals in corporate finances. It is widely used in financial analysis, risk assessment, and automated auditing research. The Yahoo Finance dataset provides historical stock prices, trading volumes, and financial summaries for numerous publicly traded companies worldwide[17]. It is extensive, covering a long time span, and excels at capturing dynamic financial metric changes, especially short-term fluctuations and mid-term trends. Combined with the textual and structured data from SEC EDGAR, the Yahoo Finance time series enhances FARPM-Net's capability to analyze and predict risks across different time scales.

.4.2 Experimental Setup and Configuration

The experiments were conducted on a high-performance computing platform with an NVIDIA Tesla A100 GPU (40GB VRAM), Intel Xeon Gold 6338 CPU (32 cores), 256GB memory, and 4TB NVMe SSD storage. The powerful GPU accelerated FARPM-Net's training, particularly for multi-level temporal modeling and multimodal fusion. The CPU supported data preprocessing and parallelism, improving efficiency. The system ran on Ubuntu 22.04 LTS with PyTorch 2.0, CUDA 11.7, and cuDNN 8.4, while Python 3.9 ensured compatibility. The SEC EDGAR 10-K and Yahoo Finance datasets were cleaned and preprocessed, including denoising, tokenization, and normalization. Data splitting was done with an 80/20 ratio for SEC EDGAR and 70/30 for Yahoo Finance. The Adam optimizer with an initial learning rate of 0.0005 was used for training, with a cosine annealing scheduler to enhance stability. Hyperparameters were controlled to ensure model generalization and experimental reproducibility.

4.3 Evaluation Metric

In financial risk forecasting tasks FARPM-Net In order to evaluate the performance of a model in detail, five major indicators widely used in the sector are selected and the model's performance can be reflected in several dimensions. These metrics not only assess the model's classification accuracy and error control ability but also cover its performance in handling sample imbalance and risk score regression accuracy, thereby fully demonstrating the comprehensive effect of multimodal, multi-level

risk identification[18][19].

Accuracy is the most intuitive classification performance indicator that measures the correctly predicted sample ratio between all samples. TP represents the number of true positives, TN represents the number of true negatives, FP represents the number of false positives, and FN represents the number of false negatives. Thus, high accuracy shows that most normal and divergent risk cases can be accurately identified in complex multi-source financial data that reflect the overall effectiveness of the model decision-making process as in (15).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 15]}$$

Precision reflects the accuracy of the model in the prediction of high risk cases. High accuracy FARPM-Net It suggests that there is little to generate false positive values in the risk warnings, which contributes to minimizing regular business interruptions and improving the reliability of models in practical applications of financial auditas in (16).

$$\text{Precision} = \frac{TP}{TP+FP} \cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 16]}$$

Recall measures the model's ability to capture true high-risk samples. This metric is particularly important for financial risk prediction, as failing to identify risks could lead to significant economic losses. FARPM-Net improves the identification rate of latent risks through multimodal fusion and long-term/short-term modeling, thereby enhancing recall performance as in (17).

$$\text{Recall} = \frac{TP}{TP+FN} \cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 17]}$$

F1-score, as the harmonic mean of precision and recall, balances the risks of false positives and false negatives and is an important metric for the comprehensive evaluation of the model's risk classification performance. When addressing the class imbalance issue, the F1-score of FARPM-Net more comprehensively reflects the model's practical application value as in (18).

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 18]}$$

Mean Absolute Error (MAE) is used to evaluate the model's prediction accuracy for continuous risk scores. $\hat{y}_i$ represents the predicted risk score for the i-th sample, y_i is the true score, and N is the total number of samples. As a supplementary metric for FARPM-Net's output risk levels, MAE reflects the model's precise control over financial risk quantification, revealing its performance in fine-grained risk assessment as in (19).

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \cdots\cdots\cdots\cdots\cdots\cdots \text{[Formular 19]}$$

4.4 Comparative Experimental Results and Analysis

To comprehensively validate the effectiveness and advanced nature of FARPM-Net in the domain of financial risk prediction, this paper designs a systematic comparative experiment. Five prominent models from the current multimodal and temporal modeling fields, which have demonstrated strong performance, are selected as benchmarks. The experiments are conducted on two publicly available and representative financial datasets—SEC EDGAR 10-K and Yahoo

Finance—covering both textual information and structured time series data, fully reflecting the diversity and complexity of real-world application data. Through a multi-metric evaluation, the experiment accurately assesses the model's overall performance in risk classification, score prediction, and handling sample imbalance, thereby thoroughly verifying FARPM-Net's ability to integrate multisource heterogeneous data, capture both short-term and long-term dynamics, and enhance risk prediction accuracy.Table 1 shows the experimental results:

Table 1. Performance comparison of FARPM-Net with other models on the SEC EDGAR 10-K and Yahoo Finance datasets.

Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MAE
AuditBERT[20]	FRED	82.1	79.5	74.3	76.8	0.134
	Eurostat	80.7	77.8	72.9	75.2	0.141
Fi-GNN[21]	FRED	84.3	81.2	76.5	78.7	0.126
	Eurostat	83.0	79.7	74.8	77.1	0.132
TimeMAE[22]	FRED	85.7	82.5	78.1	80.2	0.118
	Eurostat	84.4	81.0	76.3	78.5	0.125
Mamba[23]	FRED	87.4	84.1	80.2	82.1	0.111
	Eurostat	86.1	83.3	78.7	80.9	0.118
Autoformer[24]	FRED	86.8	83.7	79.6	81.5	0.113
	Eurostat	85.5	82.7	78.0	80.2	0.120
FARPM-Net	FRED	91.5	88.9	85.4	87.1	0.093
	Eurostat	90.2	87.2	83.9	85.5	0.101

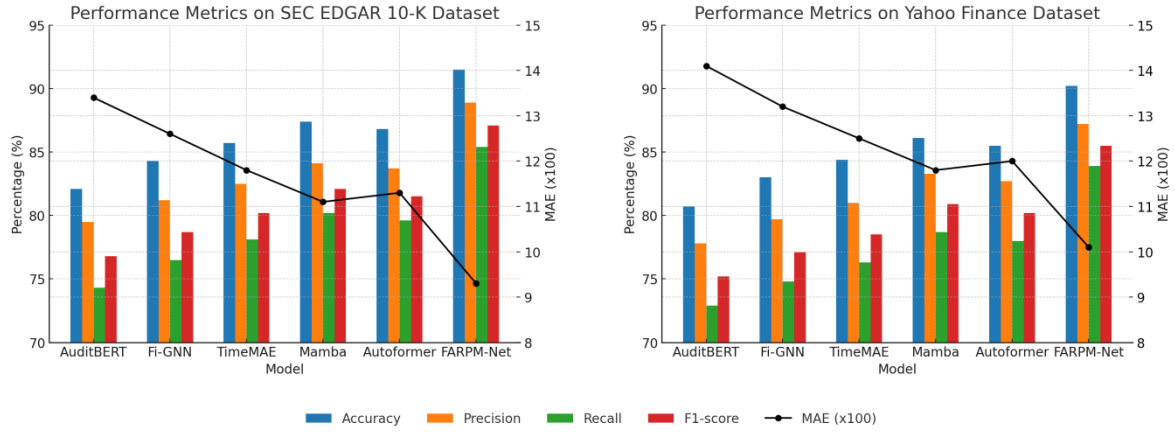

Figure 5. Comparative Experimental Results of Various Models on SEC EDGAR 10-K and Yahoo Finance Datasets.

Figure 5 clearly demonstrates FARPM-Net's systematic advantages across key evaluation metrics. On the SEC EDGAR 10-K dataset, FARPM-Net achieved an accuracy of 91.5%,

significantly outperforming advanced models such as Mamba 87.4% and Autoformer 86.8%. In terms of F1-score, FARPM-Net reached 87.1%, representing a 6.1\% improvement over Mamba and surpassing Fi-GNN 78.7% and TimeMAE 80.2% by 8.4% and 6.9%, respectively. For precision and recall, FARPM-Net achieved 88.9% and 85.4%, the highest among all comparison models, exceeding Mamba by approximately 5.7% and 6.4%, respectively. In the regression task, the model attained a mean absolute error (MAE) of 0.093, outperforming Mamba (0.111) and Autoformer (0.113), with a relative error reduction of over 16%.On the Yahoo Finance dataset, FARPM-Net also delivered outstanding performance, achieving 90.2% accuracy approximately 4.7% higher than the second-best model, Mamba 86.1%. Its F1-score improved from Mamba's 80.9% to 85.5%, a 5.7% gain. Precision and recall reached 87.2% and 83.9%, respectively, exceeding baseline models such as Fi-GNN and TimeMAE by more than 3%. The MAE decreased from Mamba's 0.118 to 0.101, representing a 14.4% reduction in error.These improvements not only demonstrate superior performance in individual metrics but also highlight FARPM-Net’s well-balanced architecture in both classification and regression tasks.The result of the two datasets is the integration of highly dimensional and heterogeneous inputs from several sources. FARPM-Net The excellent capability is still confirmed. In various indicators, performance increased by 3 to 6%. The reduction is more than 10% and has double advantages in terms of accuracy and robustness. In addition, FARPM-Net is an indispensable feature of real high-risk financial applications and data complexity that show the smallest variations in performance over multiple experiments, showing strength in the face of input confusion and sample variability.

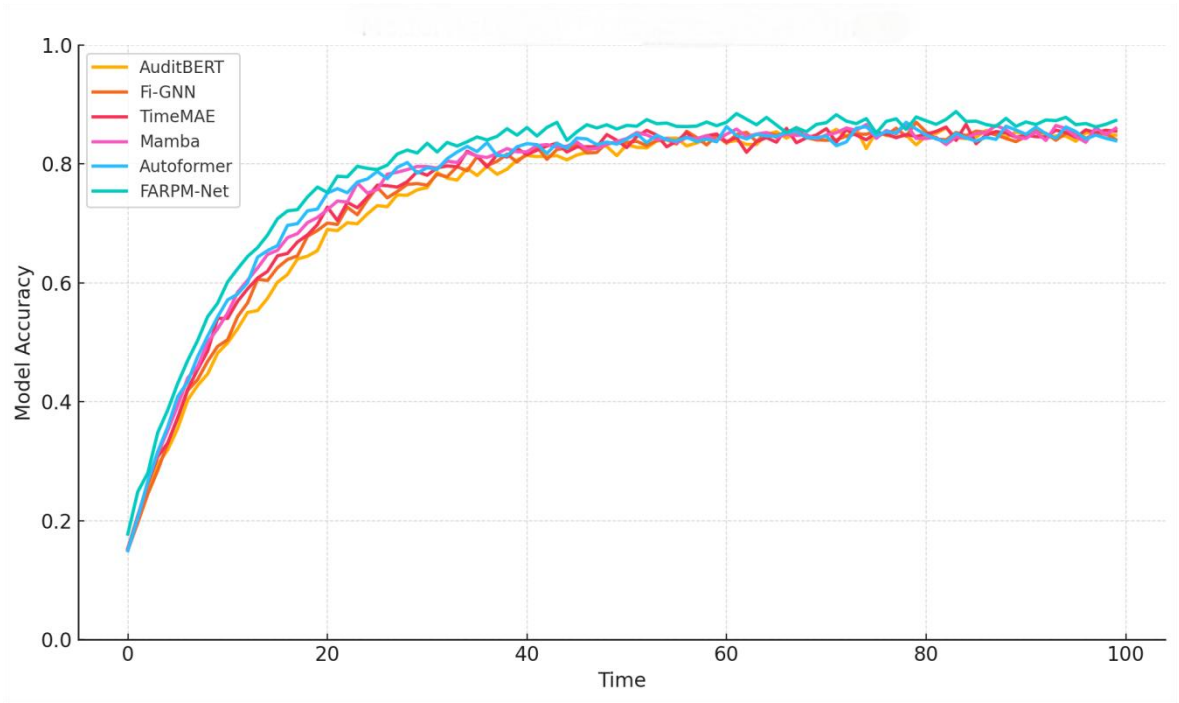

Figure 6. Accuracy comparison of FARPM-Net and baseline models over time, highlighting FARPM-Net’s superior performance due to effective multi-level temporal modeling and intermodal fusion strategies.

As shown in Figure 6, FARPM-Net consistently outperforms the other baseline models across key metrics such as classification accuracy, recall, and F1-score, while also exhibiting significantly lower regression error in risk scoring. This figure intuitively reflects the time modeling at multiple levels and the effectiveness of the intermodal fusion strategy in real applications. We confirm FARPM-Net that we have a stable and effective ability to identify and predict risks in complex financial environments. These results not only confirm the rigidity of the model design, but also demonstrate its applicability and value for automated corporate assessment and broader applications in financial risk management.

4.5 Ablation experimental results and analysis

To study the FARPM-Net specific contribution to the overall performance of individual core modules, this work is ablation It is designed for research. Mamba Module, TCN Module that phases out the accuracy of module-module merger for financial risk forecasting tasks, F1-score, MAE assessment of the effects of individual elements on important indicators. The experiment was conducted with both data sheets and the results are summarized in Table 2.

Table 2. Single-module ablation study of key FARPM-Net components, used to evaluate the individual contribution of each module to financial risk prediction performance

Model Configuration	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MAE
FARPM-Net	FRED	91.5	88.9	85.4	87.1	0.093
	Eurostat	90.2	87.2	83.9	85.5	0.101
w/o Multi-level Mamba	FRED	87.2	84.3	81.0	83.4	0.105
	Eurostat	86.0	82.7	79.5	81.0	0.114
w/o Replaced by Conv Layer	FRED	88.8	85.2	82.1	84.7	0.102
	Eurostat	87.4	83.8	80.4	82.0	0.109
w/o Cross-modal Fusion	FRED	87.9	84.1	80.7	83.3	0.104
	Eurostat	86.5	82.9	79.1	80.8	0.112

The ablation experiments conducted on the SEC EDGAR 10-K and Yahoo Finance datasets systematically validate the significant impact of the key components in FARPM-Net on the overall model performance. When the multi-level Mamba module was removed, the model's accuracy and F1-score on the SEC EDGAR dataset dropped by approximately 4.7\% and 3.7\%, respectively, and by 4.2\% and 4.5\% on the Yahoo Finance dataset. Meanwhile, the mean absolute error (MAE) increased by more than 11\%, clearly demonstrating the Mamba module's core role in capturing long-term temporal dependencies and multi-scale dynamics in enterprise financial data. It serves as a cornerstone for accurate modeling of complex financial risks.On the other hand, when the TCN module was replaced with a traditional convolutional layer, the model experienced a performance decline on both datasets.Specifically, In the SEC EDGAR dataset, accuracy and F1-score of the result

decreased by 2.7\% and 2.4\%, respectively, The TCN module is important for capturing short-term local changes in financial data and shows that the model's sensitivity to sensitive risk signals is significantly increased. In addition, the removal of the transmodular fusion module limits the possibility of integrating several source information from the model. SEC EDGAR and Yahoo Finance Accuracy and F1-score in both datasets by approximately 3.8\% and MAE has increased by more than 10\%. These results highlight the important role that attention-based fusion mechanisms play in enabling effective interaction and cooperative representation between modalities.As shown in Figure~\ref{figure7} FARPM-Net basic elements play an important role in model performance.

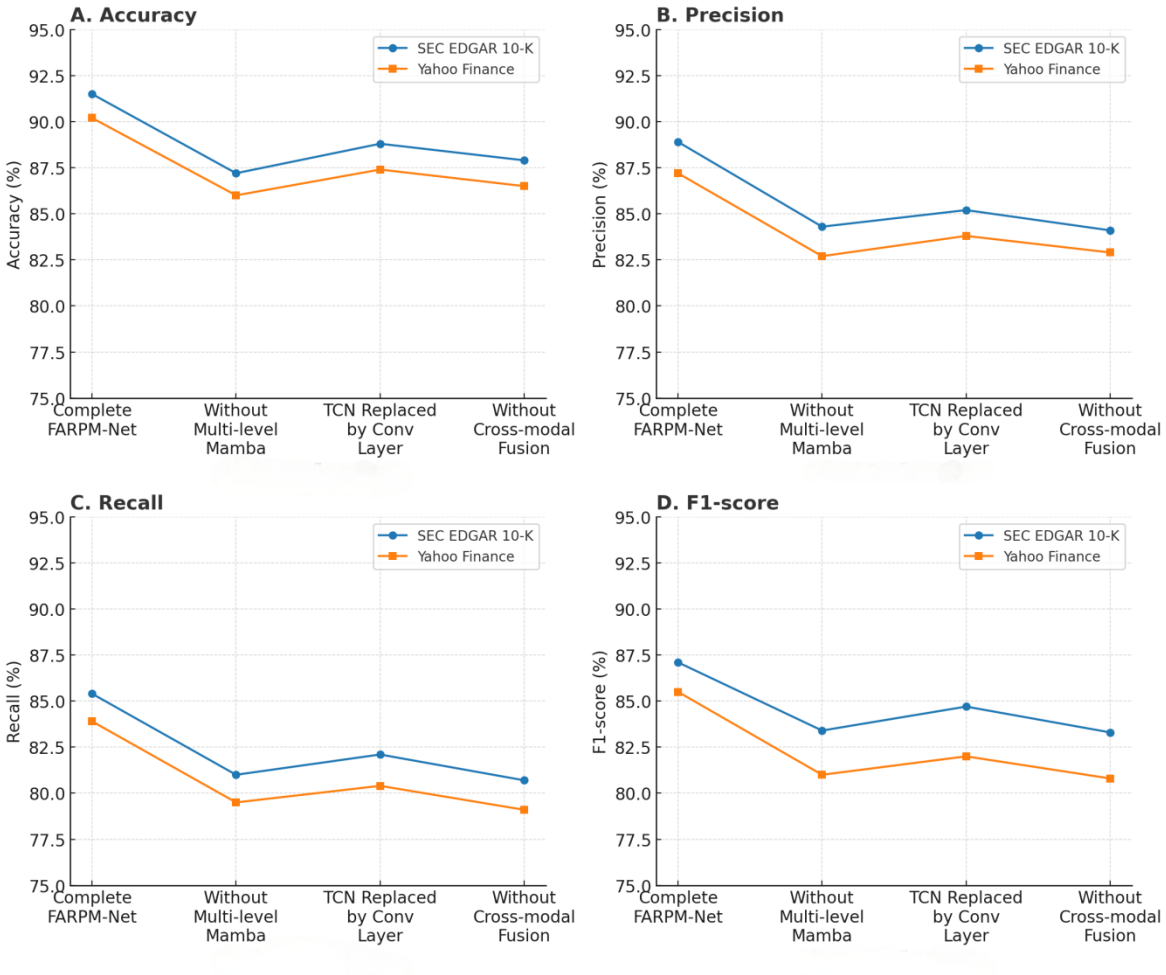

Figure 7. Performance Comparison of Model Configurations in Single-Module Ablation Experiments on SEC EDGAR 10-K and Yahoo Finance Dataset.

However, experience confirms the individual contribution of the individual main components, the interactions and interdependencies between the modules have not been fully discovered. To cope with this, the article further designs a multi module joint ablation experiment. By simultaneously removing different combinations of modules and cross-fusion modules, this study analyzes in detail the impact of the module combination on the overall model performance. Table 3 summarizes the

detailed results of the joint ablation experiments.

Table 4. Joint ablation study of key FARPM-Net modules, used to evaluate the synergistic effects of module combinations and their impact on financial risk prediction performance.

Model Configuration	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MAE
w/o Mamba & TCN	FRED	85.6	82.3	79.2	80.7	0.110
	Eurostat	84.7	81.0	78.0	79.4	0.117
w/o Mamba & Fusion	FRED	85.0	81.7	78.8	80.1	0.112
	Eurostat	84.1	80.5	77.3	78.7	0.120
w/o TCN & Fusion	FRED	86.2	83.1	79.7	81.3	0.108
	Eurostat	85.3	81.5	78.1	79.7	0.115
w/o Mamba, TCN & Fusion	FRED	83.4	79.5	77.1	78.3	0.119
	Eurostat	82.9	78.0	75.4	76.6	0.126
Model Configuration	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MAE
w/o Mamba & TCN	FRED	85.6	82.3	79.2	80.7	0.110
	Eurostat	82.9	78.0	75.4	76.6	0.126

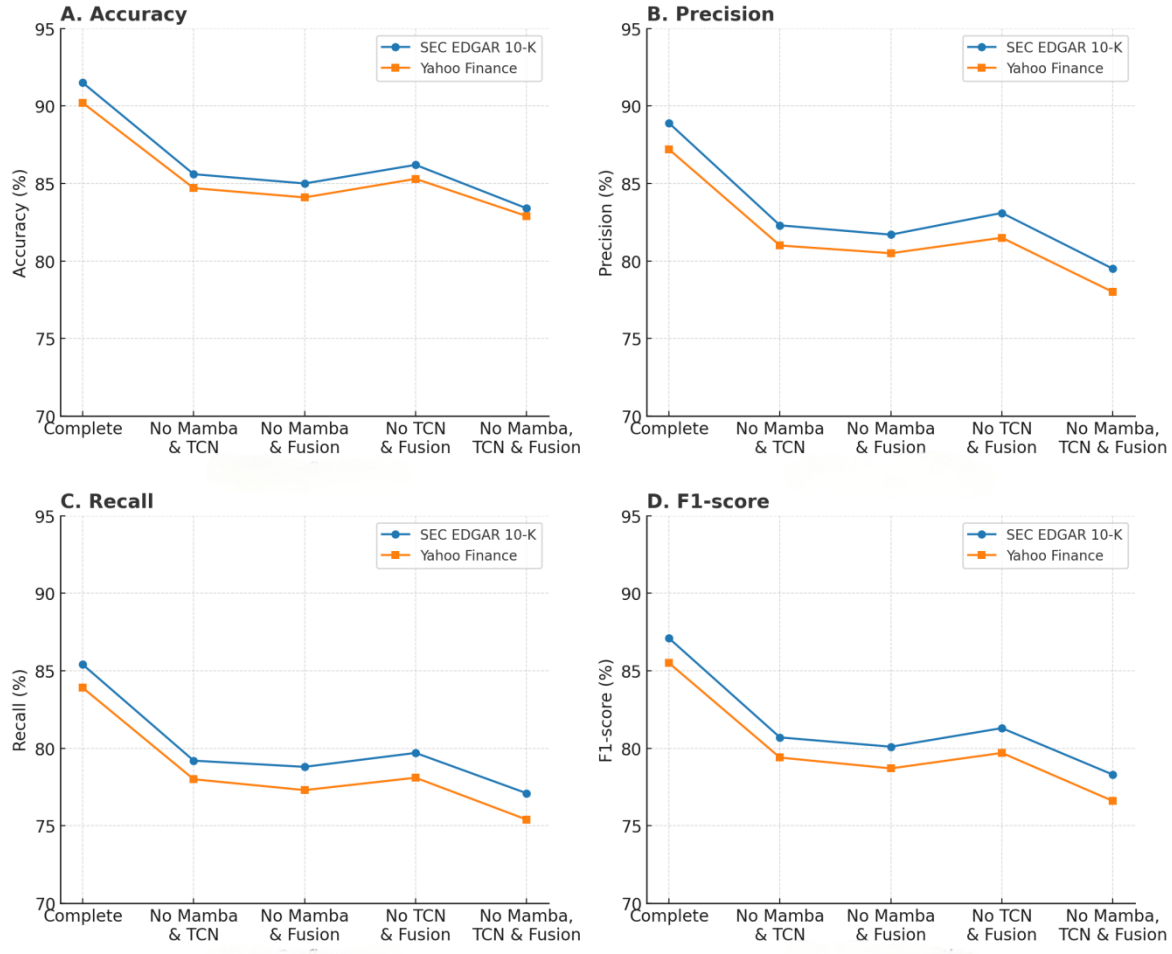

Figure 8. Performance Comparison of Model Configurations in Multi-Module Ablation Experiments on SEC EDGAR 10-K and Yahoo Finance Datasets.

As shown in Figure 8 further reveal the synergistic effects of key modules within FARPM-Net and their overall impact on model performance. Compared to the complete model, removing both the multi-level Mamba module and the TCN module resulted in a performance decline on the SEC EDGAR dataset, with accuracy and F1-score dropping by approximately 6.5% and 7.3%, respectively. A similar performance drop was observed on the Yahoo Finance dataset, with accuracy and F1-score decreasing by 6.1% and 7.1%, respectively. This indicates that the absence of both long-term and short-term temporal modeling severely impairs the model's ability to capture dynamic financial features.When the multi-level Mamba module and the cross-modal fusion module were simultaneously removed, model performance further deteriorated, with accuracy on the SEC EDGAR and Yahoo Finance datasets decreasing by approximately 7.1% and 6.7%, respectively, and F1-scores dropping by over 7%.Additionally, removing both the TCN module and the cross-modal fusion module also led to a notable performance decline, highlighting the importance of both short-term dynamic modeling and multimodal information integration.The most significant degradation occurred when all three core modules the multi-level Mamba module, the TCN module, and the cross-

modal fusion module were removed. In this scenario, accuracy and F1-score on the SEC EDGAR dataset dropped by over 8.8%, and by more than 8.3% on the Yahoo Finance dataset. The MAE also increased significantly, indicating that the absence of all three modules critically undermines the model's ability to perform comprehensive risk identification.Ablation results in a single Transfusnet It is clarified that the modules are nested to each other and that the removal of one module significantly affects the prediction ability and stability of the model. Each module is important for overall performance. These results are useful in dealing with complex economic data Transfusnet Proving its superiority and proved its strong predictive ability and stability.

In summary, the joint ablation experiments clearly demonstrate the close collaborative relationship among the multi-level temporal modeling, multi-scale dynamic capture, and cross-modal fusion modules within FARPM-Net. The synergy among these components significantly enhances the model's capability to perceive and predict complex financial risks. The removal of any key combination leads to a substantial degradation in overall performance, underscoring both the irreplaceability of each component and the effectiveness of their coordinated optimization in the model's architectural design.

5. Conclusion and Discussion

This paper proposes the FARPM-Net model for enterprise financial risk prediction and automated auditing, addressing key challenges in the field by integrating multimodal fusion and multi-level temporal modeling. The model combines a multi-level Mamba module, TCN, and an attention-based cross-modal fusion module to achieve deep fusion of structured financial data, unstructured textual information, and external market factors. Experimental results on two authoritative datasets demonstrate FARPM-Net's superior performance in accuracy, recall, F1 score, and risk score regression, highlighting its strong capabilities in risk identification and quantification.

Ablation studies confirm the vital contributions of each module: the Mamba module captures long-term financial dependencies, the TCN enhances short-term dynamics sensitivity, and the cross-modal fusion module optimizes multi-source data representation. The synergy between these components significantly boosts the model's generalization and stability, ensuring reliable financial risk predictions. Future work will focus on further optimizing model architecture, exploring better intermodal integration strategies, and improving computational efficiency through lightweight construction. Additionally, the model could be extended to applications like fraud detection and credit rating, with the ultimate goal of improving transparency and reliability in financial decision-making.

Acknowledgements

This article received no financial or funding support.

Conflicts of Interest

The authors confirm that there are no conflicts of interest.

References

[1] An, G., Park, J.and Lee, K. Contrastive learning-based anomaly detection for actual corporate environments. Sensors, 2023, 23(10), 4764.

[2] Buch, R., Grimm, S., Korn, R.and Richert, I. Estimating the value-at-risk by temporal VAE. Risks, 2023, 11(5), 79.

[3] Cai, L., Zhu, L., Zhang, H.and Zhu, X. Da-GAN: dual attention generative adversarial network for cross-modal retrieval. Future Internet, 2022, 14(2), 43.

[4] Cao, Y., et al. RiskLabs: predicting financial risk using large language model based on multi-sources data. arXiv preprint arXiv:2404.07452, 2024.

[5] Chatterjee, P.and Das, A. Adaptive financial recommendation systems using generative AI and multimodal data. Journal of Knowledge Learning and Science Technology, 2025, 4(1).

[6] Chen, P.and Ji, M. Deep learning-based financial risk early warning model for listed companies: a multi-dimensional analysis approach. Expert Systems with Applications, 2025, 127746.

[7] Cui, Y.and Yao, F. Integrating deep learning and reinforcement learning for enhanced financial risk forecasting in supply chain management. Journal of the Knowledge Economy, 2024, 1–20.

[8] Dong, H., Liu, R.and Tham, A.W. Accuracy comparison between five machine learning algorithms for financial risk evaluation. Journal of Risk and Financial Management, 2024, 17(2), 50.

[9] Gao, G., Mishra, B.and Ramazzotti, D. Causal data science for financial stress testing. Journal of Computational Science, 2018, 26, 294–304.

[10] Kang, H.and Kang, P. Transformer-based multivariate time series anomaly detection using inter-variable attention mechanism. Knowledge-Based Systems, 2024, 290, 111507.

[11] Li, Z., Cui, Z., Wu, S., Zhang, X.and Wang, L. Fi-GNN: modeling feature interactions via graph neural networks for CTR prediction. Proceedings of the ACM International Conference on Information and Knowledge Management, 2019, 539–548.

[12] Liu, X.and Wang, W. Deep time series forecasting models: a comprehensive survey. Mathematics, 2024, 12(10), 1504.

[13] Lombardo, G., Pellegrino, M., Adosoglou, G., Cagnoni, S., Pardalos, P.M.and Poggi, A. Machine learning for bankruptcy prediction in the American stock market: dataset and benchmarks. Future Internet, 2022, 14(8), 244.

[14] Lu, S., Zhang, X., Su, Y., Liu, X.and Yu, L. Efficient multimodal learning for corporate credit risk prediction with an extended deep belief network. Annals of Operations Research, 2025, 1–38.

[15] Peng, K.and Yan, G. A survey on deep learning for financial risk prediction. Quantitative Finance and Economics, 2021, 5(4), 716–737.

[16] Sheela, S., Alsmady, A.A., Tanaraj, K.and Izani, I. Navigating the future: blockchain's impact on accounting and auditing practices. Sustainability, 2023, 15(24), 16887.

[17] Shi, S., Li, F.and Li, W. A hybrid long short-term memory–graph convolutional network model for enhanced stock return prediction: integrating temporal and spatial dependencies. Mathematics, 2025, 13(7), 1142.

[18] Shi, X., Zhang, Y., Yu, M.and Zhang, L. Deep learning for enhanced risk management: a novel approach to analyzing financial reports. PeerJ Computer Science, 2025, 11, e2661.

[19] Tan, X.and Kok, S. Explainable risk classification in financial reports. arXiv preprint arXiv:2405.01881, 2024.

[20] Wang, Z., Kong, F., Feng, S., Wang, M., Yang, X., Zhao, H., Wang, D.and Zhang, Y. Is Mamba effective for time

series forecasting? Neurocomputing, 2025, 619, 129178.

[21] Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J.and Sun, L. Transformers in time series: a survey. arXiv preprint arXiv:2202.07125, 2022.

[22] Wu, H., Xu, J., Wang, J.and Long, M. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting. Advances in Neural Information Processing Systems, 2021, 34, 22419–22430.

[23] Zhang, W., Yang, L., Geng, S.and Hong, S. Self-supervised time series representation learning via cross reconstruction transformer. IEEE Transactions on Neural Networks and Learning Systems, 2023.

[24] Zhou, Z.-H.and Feng, J. Deep forest. National Science Review, 2019, 6(1), 74–86.