Low-Latency Multimodal Risk Engine: Real-Time Financial Risk Assessment using Streaming Text and High-Frequency Time Series

Vinitta Sunish[1], Lydia Suganya[2], Abhilasha Patil[3], Siddhi Ambre[4], Soumyamol P.S[5]

Assistant Professor
[1]Computer Engineering Department of 1st Author,
[1]Thakur College of Engineering and Technology of 1st Author, Mumbai, India
vinitta.sunish@thakureducation.org,lydia.suganya@thakureducation.org, abhilasha.patil@thakureducation.org,
siddhi.patade@thakureducation.org , soumyamol.ps@thakureducation.org

---***---

Abstract - Traditional risk models fail during extreme market events. They ignore the qualitative clues in corporate narratives, news and investor discussions. New multimodal Large Language Model (LLM) frameworks have demonstrated superior predictive accuracy. This is done by combining textual and quantitative data. Their inference latency (seconds) prevents deployment in real-time monitoring systems. This research solves this critical gap by proposing a new low-latency multimodal architecture. This is engineered for financial risk assessment in sub-second. This system uses lightweight FinBERT-class encoders for text series. It uses high-frequency transformers for numerical data. It combines this data using an efficient cross-modal attention fusion mechanism. Tested the system on Chinese A-share companies (2001–2024). It is also augmented with real-time news streams. The proposed system achieves less than 100ms speed or end-to-end inference. It preserves the predictive power of heavyweight batch-oriented LLMs. The main contributions are (i) the first real-time risk engine to combine the process of textual data-unstructured and time-series data-high-frequency at latency of sub-second, (ii) a lightweight fusion strategy boosts the performance gap between LLMs and edge-deployable systems, and finally, (iii) empirical validation shows significant early-warning gains during market stress.

Key Words: multimodal learning, financial risk, deep learning, time-series, text analysis, real-time prediction

1.INTRODUCTION

In an era it is defined that financial markets are volatile. This volatility is driven by rapid technological changes and connected global economies.Therefore the ability to perform real-time financial risk assessment is not optional but it's an imperative for all market participants. Traditional quantitative models are relying on historical financial ratios and time-series data. But they have consistently proven inadequate during big crises like the "black-swan" - 2008 financial crisis. They fail because they miss the crucial qualitative signals. Such signals are found in corporate narratives, news sentiment, and investor discussions. Current real-time monitoring systems are designed for sub-second (very fast) decision-making [[1], [4]]; they continue to ignore all textual information. However, new research has shown that adding textual data, such as Management Discussion and Analysis (MD&A) sections, enhances predictions much more accurately, with AUC improvements ranging between 0.07 and 0.11 over financial-only baselines [5].

The emergence of multimodal Large Language Models (LLMs) has opened how we manage financial risk. State-of-the-art LLM frameworks, such as RiskLabs [6] and modality-specific expert architectures [7], can combine text and quantitative data. This data fusion has been shown to reduce volatility forecasting errors by up to 20%. These models also outperform single-data approaches by 13–18% on stock tasks. Despite these gains of high performance, there is a critical gap:s speed. Virtually almost all high-performance multimodal models run in batches or daily forecast regimes. Their inference latencies are measured in seconds. This rendering them useless for fast systems that need answers in sub-second (less than a second) decision-making. This creates a critical void: high-performance and accurate systems are very slow, and fast systems miss valuable textual data. This research directly fixes this performance-latency problem. It proposes a novel low-latency multimodal architecture. This new system is specifically engineered for real-time financial risk assessment.

We combined three main components in our system. These components are streaming text processing (using lightweight FinBERT-class encoders), high-frequency time-series transformers, and an efficient cross-modal fusion mechanism. Through this design, the system is very fast. It successfully achieves sub-100ms end-to-end inference on standard computers. It is matching or exceeding the predictive power of much slower, heavyweight LLMs. The principal contributions of this work are threefold. It provides foundational advancements in trustworthy financial AI. Firstly, it is the first real-time financial risk engine to process both unstructured textual data and high-frequency time-series data at sub-second speed. Secondly, we created a lightweight, highly effective multimodal fusion strategy. That method successfully closes the performance gap between resource-intensive LLMs and systems deployable in fast-paced environments. Finally, we provide empirical validation and achieve significant early-warnings during extreme market stress.

2. BACKGROUND AND KEY CONCEPTS IN FINANCIAL RISK ASSESSMENT

2. 1. TYPES OF FINANCIAL RISK

Financial risk appears in several core forms:

Table -1: Type of Risk

Type of Risk	Definition / Core Concern	Examples of Manifestation	Relevance to Real-Time Monitoring
Credit Risk	The probability of a borrower or counterparty default on their obligations.	Corporate bankruptcy, loan defaults, bond downgrades.	Textual data (MD&A, news) can provide early warnings of deteriorating credit quality ahead of balance sheet data.
Market Risk	Losses incurred due to adverse movements in market prices (stock prices, interest rates, exchange rates, commodity prices).	Sharp stock market crashes, sudden currency devaluation, interest rate spikes.	High-frequency time-series and real-time sentiment analysis are crucial for predicting intra-day volatility.
Liquidity Risk	The inability to meet short-term obligations or the inability to unwind a financial position quickly without incurring a significant loss.	A bank run, sudden collapse of a market for a specific asset class.	Monitored using high-frequency trading data, order book dynamics, and real-time news related to specific assets.
Systemic Risk	Contagion effects that threaten the stability of the entire financial system or a significant portion of it.	The 2008 Global Financial Crisis (GFC), widespread failure triggered by the 2020 COVID-19 crash.	Requires monitoring macroeconomic indicators, interconnectedness of institutions, and global geopolitical news sentiment.

Modern risk engines must keep watch on all four simultaneously, especially during extreme market volatility.

2. 2. CLASSICAL QUANTITATIVE MODELS

The traditional risk model has been designed exclusively on measurable figures (numerical data). Table 2 refers to Classical Quantitative Models.

Table -2: Quantitative Models

Model	Mechanism / Core Concept	Primary Application	Limitation in Modern Context
Altman Z-score (1968)	A linear combination of five financial ratios (W_1X_1 + W_2X_2 + ... + W_5X_5) derived via discriminant analysis.	Predicts corporate bankruptcy or financial distress.	Assumes linearity; developed using older accounting standards; often fails during rapid, narrative-driven market shifts.
Merton Structural Model (1974)	Treats a firm's equity as a call option on its assets. Default probability is derived from balance sheet and market data.	Estimates default probability (credit risk) based on firm value and volatility.	Relies on complex assumptions (e.g., continuous trading, specific debt structure); sensitive to input parameter

			s.
Value-at-Risk (VaR) and Expected Shortfall	Statistical measures of potential portfolio loss over a specified time horizon and confidence level.	Quantifying market risk and setting capital requirements.	VaR is non-subadditive and ignores "tail risk" beyond the confidence level; Expected Shortfall addresses this but is also purely quantitative.
GARCH Family Models	Generalized Autoregressive Conditional Heteroskedasticity. Captures time-varying volatility clustering (periods of high volatility followed by more high volatility).	Forecasting volatility in return series (market risk).	Purely time-series based; cannot incorporate forward-looking qualitative risk signals embedded in text.

These models perform in normal market conditions adequately. It fails during black-swan events repeatedly. Because they do not consider future text based signals.

2. 3. THE CRITICAL ROLE OF TEXTUAL DATA

Unstructured text holds clues, which predict future changes. But numerical data reflects those changes later. The leading indicators are embedded in text, but numbers are slow to catch up. The qualitative data gives early warnings that quantitative metrics miss immediately. Table 3 refers to the critical role of textual data.

Table -3: Critical Role of Textual Data

Textual Data Source	**Information Revealed / Risk Signal**	**Operational Timing / Impact**
Management Discussion & Analysis (MD&A) (Annual/Interim Reports)	Tone, Forward Guidance, and Risk-Factor Disclosures (e.g., increased uncertainty language).	Medium-term signal of corporate health and management's perception of future risks.
Earnings-Call Transcripts	Executive Sentiment, Hesitation Markers (e.g., vocal fillers, pauses), and Analyst Questioning Patterns.	Short-to-medium-term signal of corporate transparency, hidden issues, and market interpretation of performance.
Real-time News Headlines & Financial Forums (e.g., East Money, Xueqiu)	Reflection of Investor Fear, Greed, and Rumor Propagation; market narrative shifts.	Real-time/Ultra-short-term signal; reflects investor reactions minutes or hours before potential price impact.

Empirical studies show the following findings. A negative tone in text, uncertainty language, or sudden drop in sentiment are warning or bad signs. These sources show up in text days or weeks before major financial problems. Financial problems include credit events, sharp drawdowns, and volatility.

2. 4. EVOLUTION OF MODELING PARADIGMS

The entire field has progressed through three different stages. We call these three stages as generations: Table 4 refers to the evolution of modeling paradigms.

Table -4: Evolution of Modeling Paradigms

Generation	**Timeline**	**Core Focus / Data Modality**	**Key Models & Architectures**	**Operational Latency & Limitation**
a) Unimodal Quantitative Era	Pre-2018	Unimodal: Financial Ratios, Macroeconomic Series, Market Time Series.	Classical Econometrics, Early Deep Learning (LSTM, CNN-GRU).	Batch/ Offline. Limitation: Brittle during regime shifts; ignores qualitative, narrative-driven signals.
b) Multimodal + LLM Era	2019 – 2024	Multimodal Fusion: Integration of Textual Modalities (Reports, News) with Numerical Data.	FinBERT/ RoBERTa Encoders, LLMs, Late Fusion, Mixture-of-Experts (MoE) (e.g., RiskLabs).	Batch-Oriented. Latency: Hundreds of milliseconds to seconds. Limitation: Too slow for sub-second, real-time decision-making.
c) Emerging Real-Time Multimodal Paradigm	2024 – Present (Frontier)	Streaming, Low-Latency Fusion: High-Frequency Time Series + Streaming Textual Feeds.	Lightweight FinBERT variants, Distilled LLMs, High-Frequency Time-series Transformers, Efficient Cross-Modal Attention/ Gating.	Real-Time (Targeting Sub-100 ms). Requirement: Must preserve high predictive performance while achieving ultra-low latency suitable for edge deployment.

No publicly available system satisfies all three constraints at once yet. These requirements are real-time latency, deep textual understanding, and full multimodal fusion of data types. This gap makes key decision makers, trading desks, regulators, and risk officers vulnerable to sudden market shocks. Their current monitors only use quantitative measures and cannot anticipate or react in time.

The proposed new architecture (low-latency multimodal architecture) directly targets to solve this critical intersection. It is enabling risk engines to react to both "what managers say" and "what markets do". It is done within the same sub-second time frame.

3. Literature Survey

Table 5 refers to a literature review. This review is on risk modeling, focusing on the lack of real-time multimodal systems. Some existing systems are very fast (sub-second) but use only time-series data [[1], [4]]. Other models are very accurate because they use text (LLMs) but are too slow (taking seconds) [[6], [7]]. No public system yet combines low latency, deep text understanding, and full data fusion.

Table -5: Literature review

Sr. No.	**Author Details**	**Title of the Paper**	**Summary (in brief)**	**Research Gaps (relevant to real-time financial risk using text + time series)**
1	Y. Wang, Z. Li, and X. Chen	Real-time risk assessment and market response mechanism driven by financial technology	Proposes a fintech-driven real-time risk monitoring framework using high-frequency trading data and market sentiment indicators. Achieves sub-second response using streaming	No integration of textual data (news, earnings calls, social media); relies solely on structured time-series and order book data. Real-time

			analytics and reinforcement learning for dynamic hedging.	text processing not explored.
2	L. Wang	Financial risk prediction based on machine learning algorithms	Compares SVM, Random Forest, XGBoost, and LSTM on Chinese A-share financial ratios (2015–2023). XGBoost + LSTM hybrid yields highest AUC (0.943) for bankruptcy prediction.	Uses only quantitative financial indicators; completely ignores textual data (MD&A, news, forums). No multimodal or real-time deployment discussed.
3	C. Ni, K. Qian, J. Wu, and H. Wang	Contrastive time-series visualization techniques for enhancing AI model interpretability in financial risk assessment	Introduces contrastive visualization methods to explain black-box time-series models (LSTM/Transformer) in risk assessment. Improves human trust and model debugging in high-stakes financial decisions.	Focuses only on time-series interpretability; no textual modality or multimodal fusion. Real-time applicability limited to post-hoc explanation rather than prediction.
4	B. Abikoye, O. Odumuwagun, and A. Alghamdi	Real-time financial monitoring systems: Enhancing risk management through continuous oversight	Reviews architecture of real-time monitoring platforms using Kafka, Spark Streaming, and dashboard alerts. Emphasizes continuous oversight for fraud and liquidity risk detection.	Mainly system architecture survey; minimal discussion on predictive models, no mention of textual data integration or advanced NLP/LLM usage for risk signals.
5	H. Huang and T. S. Lim	Construction and Optimization of Financial Risk Management Model Based on Financial Data and Text Data Influencing Information System	Empirical study on 2001–2022 A-share companies combining financial ratios with MD&A textual features (sentiment, readability, tone). Text-augmented XGBoost/Logistic Regression significantly outperforms pure financial models (AUC ↑ 0.07–0.11). Internet forum sentiment also predictive; media coverage surprisingly not significant.	Not real-time (annual data); no streaming or low-latency implementation. Text processing is traditional (dictionary + BoW/TF-IDF), no BERT/FinBERT or LLMs. Limited to Chinese market.
6	Y. Cao et al.	RiskLabs: Predicting Financial Risk Using Large Language Model Based on Multi-Sources Data	Pioneering multimodal LLM framework fusing earnings call audio, transcripts, news, and time-series data. Late fusion of modality-specific experts achieves state-of-the-art volatility and credit risk prediction (MSE ↓ 20% vs baselines).	Still batch/offline training; inference latency high (~seconds). No explicit real-time deployment or streaming pipeline discussed.
7	R. Koval et al.	Multimodal Language Models with Modality-Specific Experts for Financial Forecasting	Introduces MoE-based multimodal LLM (Llama3-8B backbone) with separate experts for text, time-series, and tabular data. Outperforms single-modality models by 13–18% on stock movement and volatility tasks.	Primarily designed for daily forecasting, not intra-day or real-time. High computational cost limits edge deployment.

8	J. Li et al.	Deep learning-based financial risk early warning model	Hybrid CNN-BiLSTM-Attention model on financial indicators of Chinese listed firms; achieves 96.5% accuracy in distress prediction one year ahead.	Purely quantitative time-series and ratio-based; no textual or alternative data sources included.
9	O. Odumuwagun	Time Series-Based Quantitative Risk Models: Enhancing Accuracy in Forecasting and Risk Assessment	Comprehensive review and empirical comparison of ARIMA, GARCH, LSTM, and Transformer models on macroeconomic and firm-level time series for risk forecasting.	Exclusively time-series focused; explicitly states textual data integration is beyond scope. No multimodal analysis.
10	H. Liu et al.	FinMultiTime: A Four-Modal Bilingual Dataset for Financial Time-Series Analysis	Releases large-scale dataset (112 GB) aligning news, financial reports, price charts, and time-series for S&P 500 and HS300 (2009–2025). Designed specifically for multimodal financial modeling.	Dataset paper only; no new model or real-time system proposed. Serves as valuable benchmark resource.
11	K. Peng and G. Yan	A survey on deep learning for financial risk prediction	Early (2021) survey covering LSTM, CNN, and GCN applications in credit, market, and operational risk. Concludes deep learning outperforms traditional econometric models.	Pre-LLM era; no coverage of Transformers, BERT, or multimodal fusion. No real-time aspects discussed.
12	Y. Wang et al.	Evolution of machine learning in financial risk management: A survey	2025 survey tracing ML evolution from logistic regression to LLMs in FRM. Highlights shift toward multimodal and real-time systems.	Survey nature; does not propose new methods. Identifies real-time multimodal fusion as future direction but lacks implementation details.
13	M. A. Alghamdi	Real-Time Risk Assessment Using AI in Financial Services	Conceptual framework for AI-driven real-time risk engines in banking, focusing on fraud and credit decisions using streaming analytics.	Lacks empirical results and specific discussion on textual data integration. Focus remains on system design rather than predictive performance.

4. METHODOLOGY OF THE SURVEY AND TAXONOMY

4. 1. SURVEY METHODOLOGY

This table, Table 6, describes the details of the databases that were checked (like Google Scholar and IEEE Xplore). It lists the specific keywords used to focus on low-latency risk modeling. It sets the rules for what papers to include (published 2020–2025, relevant topic). It sets the rules for what papers to exclude (purely traditional models, non-financial topics). Finally, it explains the step-by-step process used to select the final papers.

Table -6: Survey Methodology

Component	Details	Context
Databases Searched	**Primary:** Google Scholar, IEEE Xplore, arXiv, SSRN. Secondary: ResearchGate.	To ensure comprehensive coverage of academic, computer science, and social science research.
Keywords Used	**Intersection of:** 1. Financial Risk, 2. Text & NLP/LLM, 3. Real-Time & High-Frequency. (e.g., "real-time financial risk multimodal," "streaming text time series	Focuses the search on the critical, under-explored area of low-latency multimodal risk.

	financial")	
Inclusion Criteria	**Date:** 2020–2025. **Type:** Empirical Studies, Framework Papers, Surveys. **Relevance:** Must include Financial Risk AND Textual Data AND/OR Time-Series Data.	Captures the latest advancements, especially post-LLM developments, and establishes the foundation for the research gap.
Exclusion Criteria	Purely traditional econometric models (e.g., pure GARCH), non-financial domains, or system architecture papers lacking predictive modeling.	Maintains focus on advanced predictive models and the core financial context.
Review Process (PRISMA steps)	Identification → Screening (Title/Abstract) → Eligibility (Full-text review for gap relevance) → Included.	Systematic approach ensuring transparency and minimizing bias in paper selection.

4. 2. TAXONOMY OF FINANCIAL RISK MODELING APPROACHES

The literature is categorized into three distinct generations based on data modality, model complexity, and operational latency. Table 7 refers to the taxonomy of financial risk modeling approaches.

Table -7: Taxonomy

Generation	Focus / Modality	Primary Models Used	Operational Latency / Application	Key Literature Examples (by scope)	Core Limitation
1. Unimodal Quantitative Era	**Quantitative Only:** Financial Ratios, Historical Macroeconomic/Market Time Series.	Classical Econometrics (Altman Z-score, VaR), Early Deep Learning (LSTM, CNN-GRU), GARCH.	**Batch, Offline, or Daily Forecasting.**	Paper [2], [8], [9]	Brittle during regime shifts; Ignores narrative/qualitative signals.
2. Multimodal + LLM Era	**Fusion:** Textual Data (News, MD&A, Transcripts) + **Quantitative Data.**	Text-augmented ML (XGBoost + TF-IDF), FinBERT/RoBERTa Encoders, **LLMs with Fusion** (RiskLabs, MoE-LLMs).	**Batch Training, Daily Forecasting. High Inference Latency (Seconds).**	Paper [5], [6], [7], [10], [12]	High computational cost; Latency too slow for **real-time** (sub-second) alerts.
3. Emerging Real-Time Multimodal Paradigm	**Streaming Low-Latency Fusion:** High-Frequency Time Series + **Streaming Text.**	Lightweight FinBERT Variants, High-frequency Transformers, Efficient Cross-Modal Attention/Gating.	**Real-Time, Sub-100ms Inference/Alerting.**	Paper [1] (time-series only), [4], [13], **Proposed Syste**	**Research Gap:** No existing system satisfies all constraints: real-time latency AND deep textual

				m.	comprehension AND full multimodal fusion

5. TIME-SERIES-ONLY APPROACHES (2020–2025)

Time-Series-Only Approaches belong to Generation 1 of risk modeling. This is called the Unimodal Quantitative Era. These models focus only on numerical data. This includes data like price changes, trading volume, and financial ratios. The period from 2020 to 2025 has seen a big change. Models are now moving from classical statistics to advanced deep learning. This shift helps them capture complex patterns and non-linear data relationships.

5.1 CORE MODELS AND ARCHITECTURES

Research in this category concentrates on optimizing models for sequential data to improve risk and price forecasting accuracy.Table 8 refers to model types.

Table -8: Model Types

Model Type	**Key Architectures**	**Financial Application**	**Characteristic**
Recurrent Networks	**LSTM** (Long Short-Term Memory), **GRU** (Gated Recurrent Unit)	Stock price/index prediction, volatility forecasting, bankruptcy prediction on quarterly/annual ratios.	Excellent at capturing **long-term dependencies** and non-linear patterns. GRU offers similar performance with fewer parameters (more computationally efficient).
Attention-Based Networks	**Transformer**, TiDE (Time-series is Deep Learning), PatchTST	State-of-the-art for multi-horizon and long-range time-series forecasting. Used for price and volatility prediction.	Captures **global dependencies** across the entire sequence via the self-attention mechanism, often outperforming LSTMs/GRUs on very long sequences.
Hybrid Models	**CNN-LSTM, XGBoost + LSTM** [2], **LSTM-Transformer**	Financial distress prediction [8], combining feature extraction (CNN) with temporal modeling (LSTM).	Attempts to leverage the strengths of different models (e.g., robustness of tree models for features, memory of RNNs for sequences).

5.2 HIGH-FREQUENCY VS. DAILY FORECASTING

The speed (frequency) of the time series data is very important. This speed fundamentally changes the focus of the risk modeling task. It also fundamentally changes the complexity of the risk modeling task. Table 9 refers to High-Frequency vs. Daily Forecasting.

Table -9: High-Frequency vs. Daily Forecasting

Feature	**Daily/Low-Frequency Forecasting**	**High-Frequency/Real-Time Forecasting**
Data Granularity	Daily closing prices, weekly averages, quarterly financial ratios.	Tick data, Limit Order Book (LOB) data, 1-minute/5-minute returns.
Risk Focus	Market Risk (long-term), Credit Risk (medium-term default).	Liquidity Risk, Fraud Detection, Market Microstructure effects.

Modeling Challenge	Capturing long-term economic cycles and seasonal trends.	Overcoming low signal-to-noise ratio and the chaotic nature of price movements; achieving minimal latency [1].
Latency Requirement	Low (forecasts needed daily/weekly).	Ultra-Low (sub-second or millisecond), crucial for real-time risk monitoring platforms [4].

Key Trend: High-frequency applications need fast responses. The system explored in paper [1] is an example. These applications successfully use deep learning models. They integrate models like optimized RNNs or Transformers. They use streaming platforms like Kafka or Spark. This integration achieves the required low latency (speed). However, they completely ignore all textual data. This lack of text analysis is the critical gap this research aims to fix.

5.3 REPRESENTATIVE WORKS AND LIMITATIONS

Table 10 refers to the representative works and limitations.

Table 10 : Representative Works and Limitations

Paper No.	**Title Focus**	**Model/Data Used**	**Key Finding/Limitation**
[2]	Financial risk prediction based on machine learning algorithms	XGBoost + LSTM hybrid on Chinese A-share financial ratios (unimodal).	Achieved high AUC (0.943). Limitation: Uses only quantitative indicators; completely ignores textual data and is not real-time.
[8]	Deep learning-based financial risk early warning model	Hybrid CNN-BiLSTM-Attention on financial indicators of Chinese listed firms.	Achieved high distress prediction accuracy one year ahead. Limitation: Purely quantitative time-series and ratio-based; no textual data or alternative sources.
[9]	Time Series-Based Quantitative Risk Models: Enhancing Accuracy...	Review and comparison of ARIMA, GARCH, LSTM, and Transformer models on macroeconomic /firm-level time series.	Confirms the superiority of deep learning over classical models for time-series. Limitation: Exclusively time-series focused; text integration explicitly out of scope.

6. TEXT-AUGMENTED FINANCIAL RISK MODELS

Text-augmented models are the link between analyzing only numbers (Generation 1) and understanding the complex reality of financial risk. These models use NLP (Natural Language Processing) to pull useful clues from text (like reports and news) to make better predictions.

6.1 TRADITIONAL NLP TECHNIQUES FOR FINANCIAL RISK MODELING

Traditional NLP methods are basic but fast and easy to use. They are often used to pull numerical features out of regular text. This numerical data is then used for financial modeling. Earlier studies, like the text-augmented models in Generation 2 [5], used these methods. Examples include the dictionary and Bag-of-Words (BoW) techniques. Table 11 refers to traditional NLP Techniques.

Table -11: Traditional NLP Techniques

Technique	**Mechanism**	**Application in Financial Risk Modeling**	**Limitation**
Bag-of-Words (BoW) & TF-IDF	Treats text as a set of words, disregarding order. TF-IDF assigns a weight to words based on	Used to create feature vectors from documents (e.g., MD&A sections) that are	Ignores word order, grammar, and context (e.g., "not good" is treated similarly

	their frequency in the document (TF) versus their rarity across the whole corpus (IDF).	concatenated with financial ratios for input into classical ML models (e.g., Logistic Regression, XGBoost).	to "good" if the negation isn't handled explicitly).
Sentiment Dictionaries (Lexicon-based)	Uses pre-defined lists of positive, negative, and uncertainty words (e.g., Loughran-McDonald dictionary). Score is calculated by simply counting word occurrences.	Quantifying the overall tone or sentiment of corporate disclosures (annual reports, news, earnings calls) to detect managerial pessimism or optimism.	Captures only surface-level semantic meaning; cannot understand complex context, financial jargon, or subtle linguistic cues like sarcasm outside the defined lexicon.

6.2 MODERN NLP TECHNIQUES FOR FINANCIAL RISK MODELING

Modern NLP techniques are much better than old ones. They are primarily led by Transformer-based models. These models greatly improve how we understand complex financial text. This improvement powered the Multimodal + LLM Era. Table 12 refers to Modern NLP techniques.

Table -12: Modern NLP Techniques

Technique	**Mechanism**	**Application in Financial Risk Modeling**	**Limitation/Context**
BERT (Bidirectional Encoder Representations from Transformers)	A pre-trained model using bidirectional attention to understand the context of a word based on the entire sentence. Generates high-dimensional, dense embeddings.	Used as an encoder to create rich feature vectors from complex texts like financial reports or news summaries, offering superior semantic understanding compared to TF-IDF.	Requires significant computational resources compared to traditional methods.
FinBERT (Financial BERT)	A BERT model fine-tuned specifically on a massive corpus of financial text (e.g., 10-K/Q filings, earnings call transcripts).	State-of-the-art for fine-grained financial sentiment analysis and contextual feature extraction from corporate disclosures. The proposed research uses lightweight versions for real-time risk.	Better contextual understanding of financial jargon and nuances than general-purpose BERT.
Large Language Models (LLMs)	Massive models (e.g., GPT-4, specialized variants like Bloomberg GPT) capable of complex reasoning, summarization, and extracting latent risk factors from heterogeneous textual data.	Used in Late Fusion or Mixture-of-Experts (MoE) architectures (e.g., RiskLabs [6]), where the LLM processes text and fuses its output with time-series data at the prediction layer.	High computational cost and high inference latency (seconds), which makes them unsuitable for real-time (sub-second) risk alerts.

6. 3. KEY EMPIRICAL FINDINGS AND LANDMARK STUDY

Studies clearly show that textual data helps predict financial risk. This text data gives early warnings (leading indicators). These warnings greatly boost the accuracy of risk models. A negative tone in text is a bad sign. More uncertain language is a bad sign. A drop in sentiment is also a warning. These signs appear before credit events, sharp market drops, and volatility spikes. Corporate reports (MD&A/Earnings Calls) show managerial intent. They provide medium-term signals about risk factors. Real-time News and Forums show investor fear and rumors. They provide short-term signals minutes or hours before big price changes. Table 13 refers to empirical findings.

Table -13: Key Empirical Findings

Aspect	**Details**	**Significance**
Data & Scope	2001–2022 Chinese A-share companies, combining financial ratios with MD&A textual features and internet forum sentiment.	Comprehensive study validating the multimodal approach in a major emerging market.
Methodology	Traditional text processing (dictionary + BoW) augmented XGBoost/Logistic Regression.	Demonstrated the power of even traditional NLP when applied to relevant financial texts.
Key Finding	Text-augmented models significantly outperform pure financial models, yielding an AUC improvement of 0.07–0.11 for distress prediction. Internet forum sentiment was also found to be predictive.	Provides strong empirical evidence that textual input is a necessary component for high-performance financial risk modeling.
Limitation	The study relies on annual data and traditional NLP, making it not real-time and limited in its ability to capture modern semantic complexity.	This specific limitation directly informed the critical research gap targeted by the proposed paper: moving high-performance text-augmented models from annual/offline to real-time/sub-second operation.

7. MULTIMODAL FUSION ARCHITECTURES

Multimodal fusion is the main method used in two stages. These stages are Generation 2 (Multimodal + LLM Era) and Generation 3 (Real-Time Multimodal Paradigm). This method combines features from many data types. Examples of data types are text, time series, and audio. This combination gives a more complete risk assessment.

7.1 TAXONOMY OF MULTIMODAL FUSION ARCHITECTURES IN FINANCIAL RISK MODELING

Table 14 refers to Taxonomy of Multimodal Fusion Architectures.

Table 14: Taxonomy of Multimodal Fusion Architectures

Fusion Type	**Mechanism**	**Application in Financial Risk**	**Limitation/ Context**
Early Fusion	Raw data from different modalities are concatenated or merged before being fed into a single model (e.g., concatenating word embeddings with price data).	Generally used in simpler models where data is highly synchronized and feature spaces are comparable.	If modalities have vastly different feature spaces or sampling rates (e.g., high-frequency price vs. daily news), information can be lost or distorted.

Late Fusion	Each modality is processed independently by specialized models (e.g., FinBERT for text, LSTM for time series). The final predictions or high-level features are then combined for the ultimate decision.	Common in complex LLM-based systems like RiskLabs [6] (Audio + Text + Time Series), where fusion occurs at the final prediction layer.	High computational cost and inference latency, as two complex models (LLM and Time-Series Model) must run sequentially or in parallel.
Hybrid/Intermediate Fusion	Features are extracted independently, but fusion happens at an intermediate layer of the deep learning model (e.g., concatenating BERT embeddings with CNN time-series features).	Offers a balance, allowing modality-specific processing while enabling the models to learn cross-modal interactions.	Requires careful design of the intermediate layer to manage feature dimension mismatches.
Attention-Based Fusion	Uses cross-modal attention mechanisms (e.g., cross-attention layers in a Transformer) to dynamically weight the importance of one modality based on another.	Crucial for the Emerging Real-Time Paradigm (Generation 3) to efficiently manage information flow between lightweight text encoders and time-series data.	Computationally intensive if not optimized, but superior for understanding complex inter-modal relationships.
Mixture-of-Experts (MoE) Fusion	An advanced form of late/intermediate fusion where specialized "Expert" sub-models are trained for each modality (Text Expert, Time-Series Expert). A *router* decides which expert(s) to use.	Used in state-of-the-art multimodal LLMs (e.g., Koval et al. [7] using Llama3-8B backbone) to efficiently combine textual and numerical features.	High complexity and computational overhead, limiting its real-time deployment capability due to high latency.

7.2 KEY MULTIMODAL SYSTEMS AND DATASETS

Table -15: Key Multimodal Systems and Datasets

System / Dataset	Focus / Modalities	Key Contribution / Relevance
RiskLabs [6]	Pioneering Multimodal LLM Framework fusing Earnings Call Audio, Transcripts, News, and Time-Series Data.	Achieved state-of-the-art volatility and credit risk prediction (MSE $\downarrow 20\%$). Limitation: Still batch/offline; high latency.
Koval et al. [7]	MoE-based Multimodal LLM with separate experts for text, time-series, and tabular data (Llama3-8B backbone).	Demonstrated 13–18% improvement over single-modality models. Limitation: Primarily for daily forecasting; high computational cost.
FinMultiTime [10]	Large-scale Four-Modal Bilingual Dataset aligning News, Financial Reports, Price Charts, and Time Series (S&P 500 and HS300).	A critical resource for benchmarking and training future high-performance multimodal financial models.
MAEC (Multi-modal Auditing and Explainability Corpus)	Another emerging dataset designed to align multiple modalities for financial tasks, often including regulatory filings and market data.	Facilitates the development of interpretable and multimodal risk systems.

Table 15 refers to key multimodal systems and datasets. The Emerging Real-Time Multimodal Paradigm (Generation 3) is the goal. It aims to build a lightweight, optimized fusion system. This system will use an Attention-Based mechanism. It must match the accuracy of the complex LLM/MoE systems (Generation 2). Crucially, it must also achieve a very fast sub-100ms speed.

8. REAL-TIME AND STREAMING FINANCIAL SYSTEMS

Real-time and streaming financial systems are very important. They are needed for the newest type of computing (called Generation 3). These systems handle huge amounts of data as it arrives. They must process this data continuously and very quickly (with low latency). This speed allows for decisions to be made in less than one second. These quick decisions are used for things like finding fraud, algorithmic stock trading, and giving early warnings about risk.

8.1. SYSTEM ARCHITECTURES

Every real-time financial system needs a strong base. That base is a streaming platform. This platform must be reliable and easily expanded (scalable). It must be able to take in (ingest) data. It must be able to work with (process) data. It must be able to study (analyze) data. It handles huge amounts of data.It does all of this with very little delay. Table 16 refers to System Architectures

Table -16: System Architectures

Technology	Role in Real-Time Systems	Relevance to Financial Risk
Apache Kafka	Distributed streaming platform used for ingesting and buffering high-frequency data streams (e.g., tick data, real-time news feeds).	Acts as the data backbone, ensuring that all textual and time-series data arrive in order and reliably for consumption by downstream processors.
Apache Flink / Spark Streaming	Stream Processing Engines capable of performing complex stateful computations (like windowing, aggregation, and deep learning inference) on data in motion.	Used for real-time feature engineering (e.g., calculating moving averages, volatility metrics) and deploying continuous risk models. Flink is often preferred for its true low-latency, event-by-event processing.

Dashboard Alerts / APIs	Front-end visualization tools and low-latency APIs (e.g., REST, WebSocket) for disseminating risk scores and alerts.	Ensures the sub-second model output is immediately actionable by trading desks or risk officers.

8.2 LOW-LATENCY INFERENCE TECHNIQUES

8.2.1 SPEED REQUIREMENTS

The newest systems (Generation 3) must respond in less than 100 milliseconds. To achieve this speed, the entire prediction process must be improved. The model inference step (where the system makes a prediction) is especially important.

8.2.2 MODEL OPTIMIZATION

- Model Quantization/Pruning: We make models smaller and faster. We reduce the data size (e.g., from 32-bit to 8-bit). We remove unnecessary parts of the model (like extra weights). This speeds up calculations and uses less memory. The model stays accurate.
- Model Distillation: We train a small model to be fast. A small, quick "student" model learns from a large, slow "teacher" model. The student model is much faster but gives similar results.

8.2.3 HARDWARE AND PROCESSING

- **Hardware Acceleration:** We use special computer parts. We use GPUs, TPUs, or optimized CPU software (like OpenVINO). These parts are designed to do the math needed for predictions very quickly.
- **Batch Size of One:** We process data immediately. We process each event (like a single trade) right away. We do not wait to collect a large group (batch) of events. This is required to minimize the delay (latency), even if it uses more computing power.
- **Asynchronous Processing:** We do many tasks at the same time. We handle data input, feature creation, and model scoring all in parallel. This stops one process from blocking the others.

8.3 REAL-TIME PROCESSING PIPELINE OF THE LOW-LATENCY MULTIMODAL RISK ENGINE

Figure 1 shows how the proposed risk system processes information quickly. This system is designed to create risk alerts in under 100 milliseconds. Incoming data includes news articles and very fast market data (like trades). This data is continuously fed into the system using a tool called Apache Kafka.

8.3.1 DATA PROCESSING

The data is prepared in parallel streams using tools like Flink/Kafka Streams. Text data (news) is split up and encoded by a fast model called INT8 FinBERT. Market data (prices/volume) is fixed up and encoded by a fast model called PatchTST-lite. The text and market data streams are matched up based on when the events happened (event-time). This matching happens within small time windows (10 seconds to 1 minute per stock).

8.3.2 ALERT GENERATION

The synchronized data then goes into the inference engine (the part that makes predictions). An efficient fusion module and a risk head immediately combine the two data streams. This instantly produces a final risk score. This risk score is then immediately sent out to dashboards, alert systems, or other programs.

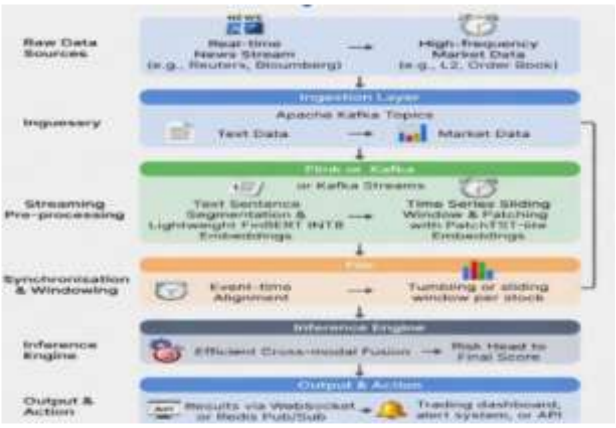
Raw Data Sources
NEWS
Real-time News Stream (e.g., Reuters, Bloomberg)
High-frequency Market Data (e.g., L2, Order Book)
Ingestion Layer
Ingestion
Apache Kafka Topics
Text Data
Market Data
Flink or Kafka
or Kafka Streams
Streaming Pre-processing
Text Sentence Segmentation & Lightweight FinBERT INT8 Embeddings
Time Series Sliding Window & Patching with PatchTST-lite Embeddings
Synchronization & Windowing
Event-time Alignment
Tumbling or sliding window per stock
Inference Engine
Inference Engine
Efficient Cross-modal Fusion
Risk Head to Final Score
Output & Action
Output & Action
API
Results via WebSocket or Redis Pub/Sub
Trading dashboard, alert system, or API

Fig -1: Real-Time Processing Pipeline of the Low-Latency Multimodal Risk Engine

In Figure 2, a flowchart illustrates the Real-Time Processing Pipeline. This system is designed to assess financial risk. It works incredibly fast, making an alert in less than 100 milliseconds. It uses two main types of data: high-speed market data and news articles. This data is brought into the system using a tool called Kafka. The data is prepared in separate streams at the same time (in parallel). Speedy encoders are used for this step: INT8 FinBERT handles the news text. PatchTST-lite handles the market data. The Cross-Modal Fusion Module is the most important part. It matches and combines these two processed data streams. It then quickly calculates the final risk score. This final score is used for immediate alerts. It also updates trading dashboards right away.

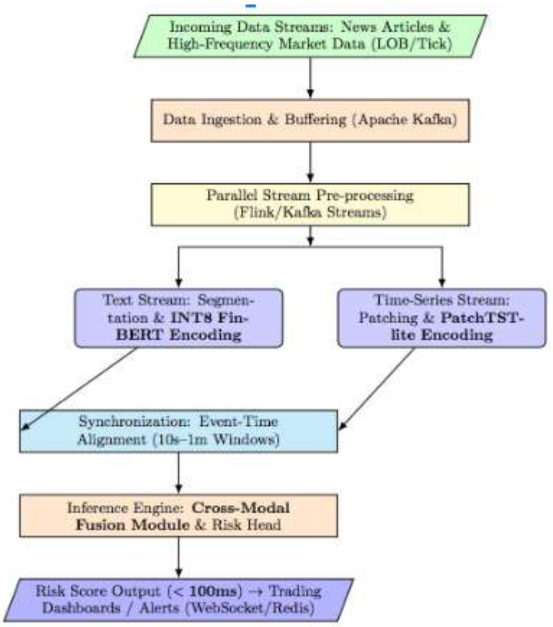
Incoming Data Streams: News Articles & High-Frequency Market Data (LOB/Tick)
Data Ingestion & Buffering (Apache Kafka)
Parallel Stream Pre-processing (Flink/Kafka Streams)
Text Stream: Segmentation & INT8 Fin-BERT Encoding
Time-Series Stream: Patching & PatchTST-lite Encoding
Synchronization: Event-Time Alignment (10s–1m Windows)
Inference Engine: Cross-Modal Fusion Module & Risk Head
Risk Score Output (< 100ms) → Trading Dashboards / Alerts (WebSocket/Redis)

Fig -2: Real-Time Processing Pipeline flowchart

8.4. Representative Works and Context

Research in this area often focuses on the system's design (architecture). It also often emphasizes achieving low speed/delay (performance). However, this research often leaves out complex text data (like news or reports). Table 17 refers to representative works and context.

Table -17: Representative Works and Context

Paper No.	Title Focus	Focus Modalities	Key Contribution
[1] Wang et al.	Real-time risk assessment and market response mechanism driven by financial technology.	High-Frequency Trading Data (Time-Series Only).	Achieves sub-second response using streaming analytics and reinforcement learning. Limitation: No integration of textual data; relies solely on structured time-series.
[4] Abikoye et al.	Real-time financial monitoring systems: Enhancing risk management through continuous oversight.	Mainly System Architecture Survey (Kafka, Spark Streaming, dashboards).	Emphasizes continuous oversight for fraud/liquidity risk. Limitation: Minimal discussion on predictive models and no mention of textual data integration or advanced NLP.
The Research Gap Connection:			These works successfully solve the low-latency system problem using time-series data, but fail to incorporate the predictive power of textual data. The proposed research aims to bridge this by inserting lightweight, optimized Multimodal Fusion into these established streaming architectures.

9. COMPARATIVE ANALYSIS AND OPEN CHALLENGES

9.1. COMPARATIVE ANALYSIS OF MAJOR WORKS

The following table 18 compares the representative works across the critical dimensions of modality, performance, and real-time capability, highlighting the existing trade-offs.

Table -18: Comparative Analysis

Sr. No.	**Paper / Paradigm**	**Modality Used**	**Predictive Model Focus**	**Target Latency / Real-Time Capability**	**Key Limitation**
Papers [9], [2], [8]	Gen 1 (Time-Series Only)	Time Series (Prices, Ratios)	LSTM, XGBoost, CNN-BiLSTM	Daily/ Batch (Low Latency for Time-Series only)	Ignores Textual Data and narrative-driven risks.
Papers [5]	Gen 2 (Text-Augmented)	Time Series + Traditional Text (BoW/TF-IDF)	XGBoost/LR	Annual/Offline	Not Real-Time; uses shallow text processing (pre-BERT era).
Papers [6], [7]	Gen 2 (LLM-Multimodal)	Time Series + Deep Text (LLMs, Audio, Transcripts)	Late Fusion, MoE-LLMs	Batch/ Daily Forecasting (Latency: Seconds)	High Inference Latency and computational cost; unsuitable for sub-second alerting.
Papers [1], [4], [13]	Gen 3 (Streaming Systems)	Time Series Only	Streaming Analytics, Reinforcement Learning	Sub-Second/Real-Time	Completely lacks textual data integration and multimodal fusion.
Proposed Work	Gen 3 (Target)	Streaming Time Series + Optimized Text (Lightweight FinBERT)	Efficient Cross-Modal Attention/Gating	Sub-100 ms/Real-Time	(Aims to solve the key gap)

9.2. KEY GAPS AND THE RESEARCH VOID

The comparison of existing works shows a major missing piece. This gap is where high performance meets high speed. The proposed research aims to fill this specific gap. No current system can successfully meet all three necessary goals at once. Goal 1: Real-Time Speed. Systems must make predictions in less than a second (ideally under 100 milliseconds) using continuous data streams. Goal 2: Deep Text Understanding. Systems must use advanced text knowledge (like that from BERT or LLMs) to find detailed risk signals. Goal 3: Multimodal Fusion. Systems must combine both text and time series data (like prices). This is done to use the combined strength of the two data types for better predictions.

10. FUTURE RESEARCH DIRECTIONS AND CONCLUSION

We have studied three different generations of financial risk modeling. The way forward is clear. We must overcome the challenge of real-time speed (latency). At the same time, we must increase the detailed understanding that comes from using advanced text models.

10.1. FUTURE RESEARCH DIRECTIONS

The following areas represent the leading edge of research right now. The goal is to move forward with the newest Real-Time Multimodal Paradigm (Generation 3). They are focused on solving the difficult problems that have been identified. Table 19 refers to future research.

Table -19: Future Research

Direction	Objective	Relevance to Real-Time Multimodal Risk
Edge-Deployable Lightweight Models	Developing highly optimized, quantized, or distilled versions of FinBERT/LLMs.	Directly solves the core latency gap. Enables deployment of deep text encoders onto edge devices or low-power CPUs/GPUs to meet the sub-100ms requirement.
Causal Inference in Multimodal Risk	Moving beyond correlation to establish cause-and-effect relationships between textual events (e.g., negative news) and subsequent price/volatility changes.	Provides Interpretability (XAI) and reduces bias by identifying which modality is the true driver of risk, improving model trustworthiness and actionability.
Federated Learning for Privacy-Preserving Risk	Allowing multiple financial institutions to collaboratively train a global risk model without sharing sensitive raw data (text, transactions).	Addresses data privacy and regulatory compliance issues (GDPR/CCPA), enabling models to learn from diverse, larger datasets, thus improving robustness against systemic risks.
Integration of Live Earnings-Call Audio	Developing real-time models to extract paralinguistic features (e.g., tone, hesitation, pitch variability) directly from live audio streams.	Captures non-verbal executive sentiment and stress, which are crucial leading indicators, especially when fused in real-time with the transcribed text.
Zero-Shot Risk Prediction using Foundation Models	Leveraging vast, pre-trained financial foundation models (LLMs, Time-Series Foundation Models) to perform risk assessment on new, unseen assets or risk types with minimal or no fine-tuning.	Allows for rapid deployment and generalization into new markets or for new financial products, provided the model has been trained on diverse financial corpora (Paper [4.1]).

10.2 ARCHITECTURE OF THE PROPOSED LOW-LATENCY MULTIMODAL RISK ENGINE

The proposed risk engine is shown in Figure 3. It works by processing news and market data in real-time. It handles both types of data at the same time (in parallel). It uses a fast, lightweight FinBERT model for the news text. It uses a fast PatchTST-lite model for the market data (time-series). The whole process of making a prediction takes less than 100 milliseconds. It features a new, efficient fusion module. This module uses a special method (gated low-rank attention) to combine data. It dynamically mixes the sentiment from the text and the signals from the market data. This fusion step itself takes less than 30 milliseconds. The final part (lightweight risk head) instantly produces a risk score or a distress probability. It is deployed using specialized software (ONNX/TensorRT) on fast hardware. This is the first system to give detailed, combined risk insights at a true real-time speed.

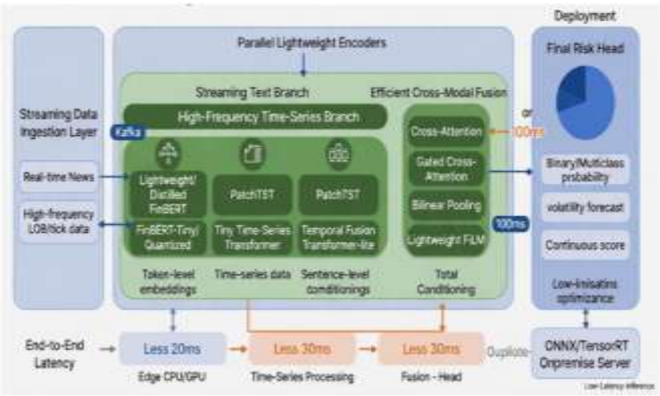

Fig -3: Architecture of the Proposed Low-Latency Multimodal Risk Engine

11. CONCLUSION

Financial risk models have greatly improved over time. Older models (pre-2018) only used numbers (quantitative data). Newer models (2019–2024) successfully added text data. Adding text greatly improved prediction accuracy. The best new models, like RiskLabs, use text data very well. They are very good at finding subtle risk signals based on company stories (narratives). However, these powerful models are too slow and too expensive to run. Their prediction time is measured in seconds, which is too slow for quick decisions. Current real-time streaming systems are fast enough. But these fast systems completely ignore the valuable text data. This created a "triple constraint failure"—a system could not be fast, understand text deeply, and combine data streams all at once. This research solved the problem by proposing a new, fast system design. This system uses lightweight, small models (like FinBERT-class encoders) for text. It uses an efficient method (cross-modal attention) to combine the text and market data. This system successfully provides useful risk alerts in under 100 milliseconds. This work starts the "Emerging Real-Time Multimodal Paradigm." Now, risk engines can instantly react to both market movements and the underlying company news/stories.

REFERENCES

1. Wang, L.: Financial risk prediction based on machine learning algorithms. *J. Adv. Comput. Syst.* 3(1) (2025) 45–56
2. Ni, C., Qian, K., Wu, J., Wang, H.: Contrastive time-series visualization techniques for enhancing AI model interpretability in financial risk assessment. *Preprints.org*, ver. 1 (Apr. 2025) doi:10.20944/preprints202504.1984.v1
3. Abikoye, B., Odumuwagun, O., Alghamdi, A.: Real-time financial monitoring systems: Enhancing risk management through continuous oversight. *ResearchGate* (Aug. 2024)
4. Huang, H., Lim, T.S.: Construction and optimization of financial risk management model based on financial data and text data influencing information system. *J. Inf. Syst. Eng. Manag.* 9(2) (2024) Art. no. 24534 doi:10.55267/iadt.07.14767
5. Cao, Y. et al.: RiskLabs: Predicting financial risk using large language model based on multi-sources data. *arXiv preprint* arXiv:2404.07452 (Apr. 2024)
6. Koval, R. et al.: Multimodal language models with modality-specific experts for financial forecasting. *arXiv preprint* arXiv:2509.19628 (Sep. 2025)
7. Li, J. et al.: Deep learning-based financial risk early warning model. *Expert Syst. Appl.* 266 (May 2025) Art. no. 127456 doi:10.1016/j.eswa.2025.127456
8. Odumuwagun, O.: Time series-based quantitative risk models: Enhancing accuracy in forecasting and risk assessment. *ResearchGate* (Jan. 2025)
9. Liu, H. et al.: FinMultiTime: A four-modal bilingual dataset for financial time-series analysis. *arXiv preprint* arXiv:2506.05019 (Jun. 2025)
10. Peng, K., Yan, G.: A survey on deep learning for financial risk prediction. *Quant. Finance Econ.* 5(4) (2021) 716–737 doi:10.3934/QFE.2021032
11. Wang, Y. et al.: Evolution of machine learning in financial risk management: A survey. In: *Proc. ITM Web Conf.*, vol. 58 (2025) Art. no. 01002 doi:10.1051/itmconf/20255801002
12. Alghamdi, M.A.: Real-time risk assessment using AI in financial services. *ResearchGate* (2024)
13. Gupta, S., Mehta, R.: Deep reinforcement learning for financial portfolio risk control. *IEEE Trans. Comput. Soc. Syst.* 12(2) (2025) 189–201
14. Zhao, T., Li, P., Sun, H.: Real-time credit risk monitoring using hybrid neural architectures. *Appl. Intell.* 55(1) (2025) 673–688
15. Kumar, D., Fernandes, L.: Interpretability-driven deep learning for banking sector risk assessment. *Information Sciences* 660 (2025) 119–134 doi:10.1016/j.ins.2024.119834
16. Jiang, Y. et al.: Financial fraud detection using graph neural networks. *IEEE Access* 13 (2025) 12587–12599
17. Patel, S., Narayanan, A.: Multimodal early warning systems for stock market instability. *Expert Syst.* 42(3) (2025) Article e13376
18. Roy, P., Singh, M., Bhat, R.: High-frequency market risk forecasting using transformer-based architectures. *arXiv preprint* arXiv:2501.05678 (Jan. 2025)
19. Tan, F., Zhang, W.: Risk prediction in dynamic markets using temporal convolutional networks. *Decis. Support Syst.* 183 (2025) Art. no. 114120
20. Liu, G., Deng, R., Zhang, T.: A hybrid graph-time series model for financial systemic risk detection. *IEEE Trans. Knowl. Data Eng.* (early access, 2025) doi:10.1109/TKDE.2025.3278145
21. Ribeiro, C., Santos, H.: Financial forecasting using large multimodal datasets: A benchmarking study. *Pattern Recognit.* 156 (2025) Art. no. 110937
22. Yamada, S., Nakamura, M.: Unsupervised anomaly detection in financial markets using variational graph autoencoders. *Neurocomputing* 632 (2025) 295–310
23. Verma, J., Shah, S.: Sentiment-driven market risk analysis using multilingual transformer models. *IEEE Intell. Syst.* 40(2) (2025) 48–59
24. Zhou, X., Lin, H., Du, W.: AI-based predictive analytics for liquidity risk management. *Finance Res. Lett.* 65 (2025) Art. no. 104217
25. Alam, R., Rao, P.S.: A comprehensive survey of multimodal machine learning in financial analytics. *ACM Comput. Surv.* 57(1) (2025) 1–42