

FinMultiTime: A Four-Modal Bilingual Dataset for Financial Time-Series Analysis

Wenyan Xu*

Central University of Finance and
Economics
2022211032@email.cufe.edu.cn

Xiyu Wang

The University of Sydney
xiuywang.usyd@gmail.com

Liang Zhang

The Hong Kong University of Science
and Technology (Guangzhou)
liangzhang@hkust-gz.edu.cn

Dawei Xiang*

University of Connecticut
ieb24002@uconn.edu

Yanxiang Ma

The University of Sydney
yama9404@uni.sydney.edu.au

Chang Xu

The University of Sydney
c.xu@sydney.edu.au

Yue Liu

National University of Singapore
yliu@u.nus.edu

Shu Hu

Purdue University
hu968@purdue.edu

Jiaheng Zhang

National University of Singapore
jhzhang@nus.edu.sg

Abstract

Pure time-series forecasting tasks typically focus exclusively on numerical features; however, real-world financial decision-making demands the comparison and analysis of heterogeneous sources of information. Recent advances in deep learning and large-scale language models (LLMs) have made significant strides in capturing sentiment and other qualitative signals, thereby enhancing the accuracy of financial time-series predictions. Despite these advances, most existing datasets consist solely of price series and news text, are confined to a single market, and remain limited in scale. In this paper, we introduce **FinMultiTime**, the first large-scale, cross-market multimodal financial time-series dataset. FinMultiTime temporally aligns four distinct modalities—financial news, structured financial tables, K-line technical charts, and stock price time series—across both the S&P 500 and HS 300 universes. Covering 5,586 stocks from 2009 to 2025 in the United States and China, the dataset totals 112.6 GB and provides minute-level, daily, and quarterly resolutions, thus capturing short-, medium-, and long-term market signals with high fidelity. Our experiments demonstrate that (1) scale and data quality markedly boost prediction accuracy; (2) multimodal fusion yields moderate gains in Transformer models; and (3) a fully reproducible pipeline enables seamless dataset updates. The data for this paper can be found at¹.

CCS Concepts

- **Do Not Use This Code → Generate the Correct Terms for Your Paper;** *Generate the Correct Terms for Your Paper;* Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

Keywords

multimodal learning, bilingual datasets, financial time series forecasting

ACM Reference Format:

Wenyan Xu, Dawei Xiang, Yue Liu, Xiyu Wang, Yanxiang Ma, Shu Hu, Liang Zhang, Chang Xu, and Jiaheng Zhang. 2018. FinMultiTime: A Four-Modal Bilingual Dataset for Financial Time-Series Analysis. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 10 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 Introduction

Time-series regression models have long been the cornerstone of financial valuation and forecasting. Traditional statistical approaches [2, 3, 34, 43] focus exclusively on numerical features and overlook open-domain knowledge from diverse modalities [8]. Intuitively, integrating multiple modalities enables richer, multidimensional representations that often outperform unimodal models [12]. In equity investment, for example, investors draw on historical price series and real-time multimodal data to inform buy, sell, or hold decisions: structured tables provide financial indicators, social media captures market sentiment, and technical charts reflect long-term price trends [45, 46].

Moreover, according to the Efficient Market Hypothesis [27], prices absorb information with a lag, which provides a theoretical basis for exploiting multi-source signals not yet fully reflected in stock prices to predict future movements. Consequently, robust and reliable predictive models must assimilate heterogeneous data to capture the full complexity of price dynamics [4, 10].

Recently, the natural language processing (NLP) models enable sentiment analysis of financial news, event extraction from disclosures, table parsing in earnings reports, and automated chart

*Equal contribution

¹https://huggingface.co/datasets/Wenyan0110/Multimodal-Dataset-Image_Text_Table_TimeSeries-for-Financial-Time-Series-Forecasting

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference acronym 'XX, Woodstock, NY

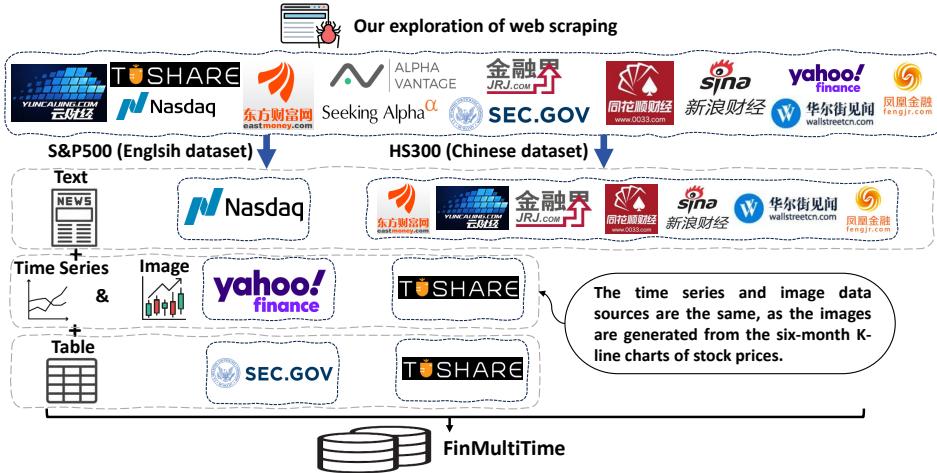
© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-XXXX-X/2018/06

<https://doi.org/XXXXXXX.XXXXXXX>

Table 1: Comparison of existing multimodal financial time-series datasets.

Dataset Benchmarks	Venue & Time	Domain	Language	Text	Time Series	Image	Table	Span	Time Interval
Time-MMD [21]	NeurIPS 2024	Multi-domain (Economics)	English	✓	✓	✗	✗	1989-2024	Monthly
CiK [35]	arXiv 2024		English	✓	✓	✗	✗	2024	Monthly
NewsForecast [33]	NeurIPS 2024	Multi-domain (Bitcoin)	English	✓	✓	✗	✗	2019-2021	Daily
TimeCAP [20]	AAAI 2025	Multi-domain (Finance)	English	✓	✓	✗	✗	2019-2023	Daily
TSQA [16]	arXiv 2025		English	✓	✓	✗	✗	—	—
FNSPID Nasdaq [9]	KDD 2024		English, Russian	✓	✓	✗	✗	2009-2023	Minute-Level
FTS-Text-MoE Nasdaq [40]	arXiv 2025		English	✓	✓	✗	✗	2009-2025	Minute-Level
StockNet Dataset [41]	ACL 2018		English	✓	✓	✗	✗	2014-2016	Minute-Level
CH-RNN Dataset [37]	CIKM 2018		English	✓	✓	✗	✗	2017	Minute-Level
DOW30 [7]	arXiv 2023	Finance	English	✓	✓	✗	✗	2020-2022	Daily
TS-FF [17]	EMNLP 2024		English	✓	✓	✗	✓	2010-2020	Quarterly
SEP [15]	WWW 2024		English	✓	✓	✗	✗	2020-2022	Minute-Level
FinBen [38]	NeurIPS 2024		English, Spanish	✓	✓	✗	✗	—	—
FinMultiTime (Ours)	—		English, Chinese	✓	✓	✓	✓	2009-2025	Minute-Level

**Figure 1: Data Collection Pipeline for the Bilingual Four-Modal FinMultiTime Dataset**

summarization [1, 5, 19, 31, 42]. Despite rapid advances in NLP models, existing multimodal datasets remain constrained. Most integrate only price and sentiment within a single market, risking information loss (Table 1); recent efforts [17] incorporate quarterly tables but suffer from limited temporal coverage and low update frequency. Such datasets are too small to train large models or to validate generalization across market regimes [23, 24, 28], and they amplify the tendency of large language models (LLMs) to generate hallucinations in dynamic financial contexts [11, 44].

To address these limitations, we introduce FinMultiTime, a large-scale bilingual dataset that spans from 2009 to 2025. FinMultiTime aligns temporal data across four modalities: text, tables, images, and time series. Our dataset includes 4694 S&P 500 constituents and 892 HS 300 constituents. After rigorous cleaning and preprocessing, FinMultiTime comprises 112.6 GB of minute, daily and quarterly level data covering both U.S. and Chinese markets (Table 2). Real-time updates ensure the dataset reflects the latest market conditions, providing a comprehensive foundation for developing

and validating multimodal forecasting models. Experimental results demonstrate that incorporating large-scale multimodal data significantly reduces prediction error and improves trend-direction accuracy, with high-quality sentiment and long-term trend information proving especially critical. Our contributions lie in three aspects:

- We first present **FinMultiTime**, a large-scale bilingual, cross-market, four-modality dataset for AI-driven stock market prediction (Table 1).
- We provide reproducible and extensible usage guidelines to facilitate rapid adoption and expansion. The complete data collection, preprocessing pipeline, and example code will be available soon.
- Empirical results demonstrate that increasing dataset size significantly improves prediction performance (Table 9), while enhancements in data quality and diversity further boost model accuracy (Table 7 and 8).

Table 2: Overview of Bilingual Financial Dataset Specifications for the HS300 (Chinese) and S&P 500 (English) Indices

Bilingual Dataset	Type	Size	Format	Stocks	Records	Frequency
HS300 (Chinese)	Image	2.43 GB	PNG	810	52,914	Semi-Annual
	Table	568 MB	JSON/JSONL	810	2,430	Quarterly/Annual
	Time series	345 MB	CSV	810	810	Daily
	Text	652.53 MB	JSONL	892	1,420,362	Minute-Level
	All	3.96 GB	—	—	1,476,516	—
SP500 (English)	Image	8.67 GB	PNG	4,213	195,347	Semi-Annual
	Table	84.04 GB	JSON/JSONL	2,676	8,028	Quarterly/Annual
	Time series	1.83 GB	CSV	4,213	4,213	Daily
	Text	14.1 GB	JSONL	4,694	3,351,852	Minute-Level
	All	108.64 GB	—	—	3,559,440	—

2 Constructing FinMultiTime

The construction of the FinMultiTime dataset begins with the systematic acquisition and processing of multi-source information. In this section, we detail the sources and procedures involved in assembling all modalities of FinMultiTime as shown in Figure 1.

2.1 Data Mining

We collect data from two of the major financial markets, as shown in Table 2. For the U.S. stock **numerical** data, we first retrieve daily OHLCV (*Open, High, Low, Close, and Volume*) data for S&P 500 constituent stocks via the Yahoo Finance API². We segment the data into semi-annual windows and visualize it using candlestick **charts** generated with the `mplfinance` library. The upper panel plots daily prices, where red candlesticks indicate rising prices (*close > open*), and green candlesticks indicate falling prices. The lower panel shows daily trading volume, with bar heights representing volume in millions and colors aligned with price direction, illustrating the relationship between price movement and trading activity (see Figure 2 Tesla 2024 H2 chart). For the **news** sentiment data, we initially explored several platforms (e.g., Investing.com, Seeking Alpha, and Alpha Vantage), but strict usage restrictions limited their accessibility. Inspired by the FNNSPID project [9], we adopted a strategy of scraping publicly available news from Nasdaq. Building on the original FNNSPID scripts, we developed a more robust, continuously running pipeline with several enhancements, including improved handling of abnormal pages, refined auto-pagination, cookie popup filtering, and compatibility with multiple ChromeDriver versions. The scraping process comprises two phases: the first uses Selenium to collect news headlines and corresponding URLs for each stock; the second retrieves the full article content from these URLs. The extracted texts constitute the news modality of our dataset. Structured financial **tables** are obtained primarily via the Securities and Exchange Commission (SEC) Submissions and Company Facts APIs³. From 10-K and 10-Q filings of S&P 500 companies since 2000, we automatically extract key indicators from XBRL facts in balance sheets, cash flow statements, and statements of shareholders' equity, while removing irrelevant fields such as announcement dates and filing types. For details on the U.S. financial data, see Table 3.

For the Chinese market, daily **numerical** OHLCV data for HS 300 constituent stocks is retrieved through the Tushare API⁴ and used to generate technical candlestick **charts** consistent with the U.S. market. **News** sentiment data is also collected from multiple Chinese financial media sources—including Sina Finance, Wallstreetcn, iFinD⁵; Eastmoney, YunCaijing⁶; Ifeng News, JRJ⁷—covering the period from March 31, 2020, to March 31, 2025. A detailed overview of bilingual news is provided in Table 4. Structured financial **table** data for the HS 300 is acquired via Tushare API, including quarterly and annual balance sheets, income statements, and cash flow statements for the period spanning 2005 to 2024.

Data Ethics To ensure ethical compliance, we strictly adhere to robots.txt directives during news scraping, collecting only publicly accessible content that requires no payment or subscription.

2.2 Data Preprocessing

To construct FinMultiTime, we extract and align four distinct data modalities—technical charts, structured tables, normalized price series, and news—across mostly constituent stocks of the HS 300 and S&P 500 indices, as of April 2025. The pipeline is designed to maximize temporal coverage while maintaining diversity in model inputs and ensuring comparability across the data sources.

Technical Charts For each stock, we segment daily OHLCV data—Open, High, Low, Close, and Volume—into semi-annual windows and generate candlestick charts with corresponding volume bars. The original RGB charts are converted to 8-bit grayscale to reduce input dimensionality. We then prompt GPT-4.1 with a fixed instruction to assign one of five long-term trend categories to each chart: 1 (Slightly Up), 2 (Significantly Up), 3 (Flat), 4 (Slightly Down), or 5 (Significantly Down). This approach compresses multi-month price movement into a single trend label, serving as a visual indicator to enhance subsequent short-term price forecasting (Figure 2).

Structured Tables For the structured-financial-table modality, we select six representative accounting variables that reflect profitability, liquidity, and capital structure. For HS300 stocks, we extract quarterly and annual values from the income statement and cash flow statement, including net profit, operating cash flow, and free cash flow. For S&P 500 stocks, we collect analogous metrics from

²<https://finance.yahoo.com/>

³<https://www.sec.gov/search-filings-edgar-application-programming-interfaces>

⁴<https://tushare.pro/>

⁵Sina Finance: <https://finance.sina.com.cn>; Wallstreetcn: <https://wallstreetcn.com>; iFinD: <https://www.ifind.com.cn>

⁶Eastmoney: <https://www.eastmoney.com>; YunCaijing: <https://www.yuncaijing.com>

⁷Ifeng News: <https://finance.ifeng.com>; JRJ: <https://www.jrj.com.cn>

Table 3: Comparison of Financial Tables for HS300 and S&P 500. The 10-Q is a quarterly financial report filed by publicly traded companies, while the 10-K is a comprehensive annual report. Both provide detailed information on a company's financial position, operating performance, and cash flow at the end of the reporting period.

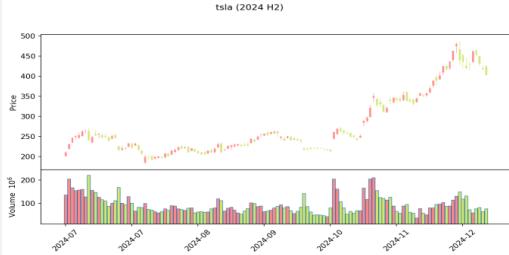
Table Type	HS300 (Chinese)			S&P500 (English)		
	Balance Sheet	Cash Flow Statement	Income Statement	Balance Sheet	Cash Flow Statement	Equity Statement
Format	JSONL	JSON	JSONL	JSON	JSONL	JSON
Field Count	147	31	92	28	80	33
10-Q nums	48,537	81,070	45,257	81,070	47,526	81,070
10-K nums	24,551	27,793	18,260	27,793	18,636	27,793
Time span	2001/12/31-2024/09/30			2000/01/03-2025/04/25		

Table 4: Comparison of Two News Sources and Data Attributes

Source	Nasdaq News	Sina Finance	WallstreetCN	10jqka	Eastmoney	Yuncaijing	Fenghuang	Jinrongjie
Time Period	2009-04-08 to 2025-04-08				2020-03-31 to 2025-03-31			
Stock Symbol	Yes	No	No	No	No	No	No	No
Headline	Yes	No	Yes	Yes	Yes	Yes	No	Yes
URL	Yes	No	No	No	No	No	No	No
Text Type	Article			Flash News				
Filter Rate	–	18.12%	14.83%	22.51%	21.20%	53.39%	19.57%	24.35%
Summarization	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Language	English	Chinese	Chinese	Chinese	Chinese	Chinese	Chinese	Chinese

System: Now you are a financial expert analyzing candlestick charts. Based on the candlestick chart provided below, determine the stock price trend. Please output only one of the following numeric values: 1 for Significantly Up, 2 for Slightly Up, 3 for No Change, 4 for Slightly Down, 5 for Significantly Down. 10 gray images will be passed in each time, you will give score in format as shown below in the response from assistant.

User: "Images to Stock Symbol -- TSLA: Tesla (TSLA)



Assistant: "1"

User: ### Images to Stock Symbol – {symbol}: {img}

System: Now you are a financial expert with stock recommendation experience. Based on a specific stock, score for range from 1 to 5, where 1 is negative, 2 is somewhat negative, 3 is neutral, 4 is somewhat positive, 5 is positive. 10 summarized news will be passed in each time, you will give score in format as shown below in the response from assistant.

User: "News to Stock Symbol -- TSLA: Tesla (TSLA) increases production by 22% ## News to Stock Symbol -- TSLA: Tesla (TSLA) faces a 30% drop in deliveries ## News to Stock Symbol -- TSLA: Tesla (TSLA) stock remains stable"

Assistant: "5, 2, 3"

User: "News to Stock Symbol -- TSLA: Tesla (TSLA) unveils new electric vehicle model ## News to Stock Symbol -- TSLA: Tesla (TSLA) faces lawsuit over autopilot feature"

Assistant: "4, 1"

User: ### News to Stock Symbol – {symbol}: {text }

Figure 2: Prompt–Response Example for Candlestick Chart Six-Month Trend Scoring

Figure 3: Prompt–Response Example for Assigning 1–5 Sentiment Scores to News Items

Table 5: Chinese (HS300) / English (S&P500) Stock Time Series Data

Date	Open	High	Low	Close	Volume	Dividends	Stock Splits
2025-03-27 00:00:00	11.3700	11.4100	11.3500	11.3900	55334940	0	0
2025-03-28 00:00:00	11.3900	11.4000	11.3400	11.3500	64494555	0.1275	1.2
2025-03-31 00:00:00	11.3600	11.3800	11.2600	11.2600	111612564	0	0
...

the balance sheet, cash flow statement, and statement of changes in equity, including shareholders' equity, operating cash flow, and

retained earnings (or accumulated deficit). All financial variables

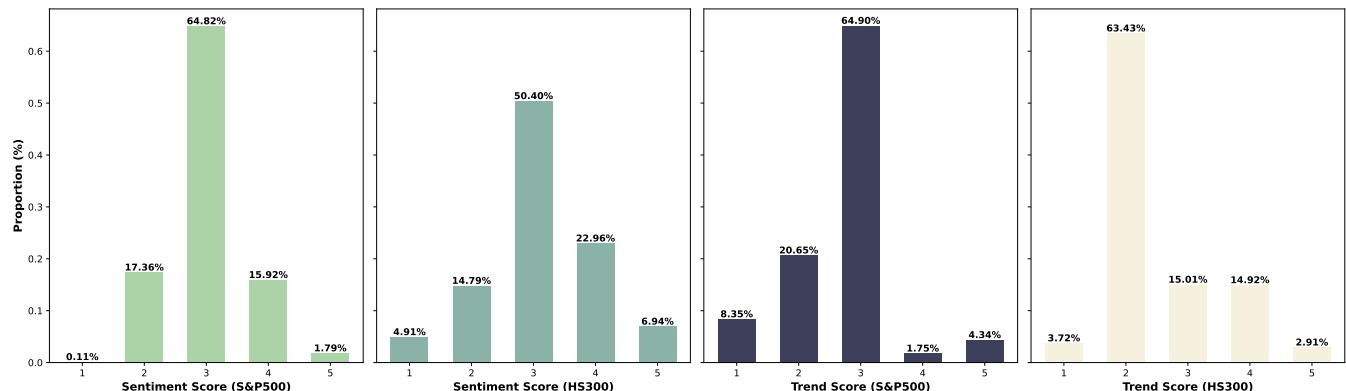


Figure 4: Figures (a) and (b) show the proportions of LSA-generated news sentiment scores (1 = negative, 2 = somewhat negative, 3 = neutral, 4 = somewhat positive, 5 = positive) for S&P 500 and HS300 stocks, respectively. Figures (c) and (d) display the corresponding six-month candlestick-chart trend scores using the same 1–5 scale (1 = negative trend, 5 = positive trend).

Date	2022-05-02 00:00:00 UTC	2025-03-29 17:22:52
Symbol	STSA (Tesla)	002594.SZ(比亚迪)
Headline	Elon Musk Twitter Co-investors Will Be Rare Birds	王传福：我国新能源汽车产品和产业链均领先全球3-5年
Text	When there's a \$44 billion merger in the offing, it's natural that lots of investors like to think Elon Musk's personal fortune will grow. After all, he's tough manager of other people's money, like private equity firms, and he's known to justify investing alongside the Tesla chief executive. Morgan Stanley has spearheaded a \$13 billion debt package for the acquisition. It's highly leveraged, at times. Twitter is forecast to break even next year, using data compiled by Refinitiv, which is only a billion dollars of equity capital already committed by Musk alone, and an additional loan of \$1.25 billion backed by the time value of Tesla stock, owned by Musk. [Read more]	我国新能源汽车无论是技术产品还是产业链都领先全球大概3-5年。应把握这个窗口期，坚持开放创新，以更高的层次的绿色技术和标准，推动新能源汽车产业高质量发展。 “3·29”中国电动汽车百人会论坛上，比亚迪董事长兼总裁王传福在“中国电动汽车百人会论坛（2025）”高层论坛上表示。 [Read more]
URL	https://www.nasdaq.com/articles/elon-musk-twitter-co-investors-will-be-rare-birds	-
LSA Sum	But Musk isn't necessarily offering co-investment opportunities beyond Twitter, and his comments have been met with some skepticism by both with automotive governments. If these investors could buy a large enough stake, state could still be a hefty single outlay even for a large investment institution, never mind for individuals and family offices... Elon Musk, who has agreed to buy Twitter for \$44 billion, has been inundated with offers from potential equity partners to join him in the deal for the social media group. [Read more]	我国新能源汽车技术、产品和产业链均领先全球3-5年。 [Read more]

Figure 5: LSA-Generated Summaries of English and Chinese Stock News

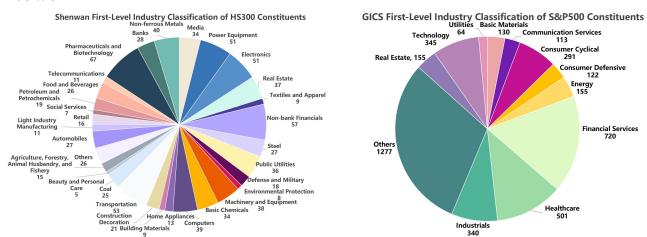


Figure 6

are aligned with the company's reporting calendar. Period-end financial figures are matched to the closing price on the last trading day of each quarter or year, then forward-filled to cover all trading days within that reporting window, ensuring synchronization with daily price series.

Price Series and News Daily closing prices are normalized on a per-stock basis to enforce stationarity across both markets (Table 5). After collecting the raw URLs, headlines, and full texts, we use the Sumy library’s latent semantic analysis (LSA) algorithm to summarize each article into 3–4 sentences, which is approximately 16% of the original length (Figure 5). A relevance weight W_f (see Appendix A) is applied to prioritize sentences that mention the target stock ticker. To manage the volume of intraday news, we aggregate all summaries for a given stock on the same day, rank them by ticker frequency, and retain only the top entry as that day’s representative

news. Each GPT-4.1 request includes at most ten entries ($temperature = 0$), ensuring deterministic sentiment inference. The model returns a sentiment score from 1 (negative) to 5 (positive). These scores are min-max normalized before multimodal fusion (Figure 3).

Summary Figures 2 and 3 illustrate the five-level rubric for images and news, while Figure 4 shows that the resulting sentiment distributions are approximately Gaussian. Mild skewness is observed: S&P 500 scores are slightly left-skewed (mildly negative), while HS300 scores lean right (neutral to positive). This reflects broader market dynamics—U.S. softness versus a sustained rally in China during the sampling period.

3 FinMultiTime Properties

With data mining and preprocessing complete, FinMultiTime is now ready for in-depth analytical evaluation. This section highlights key insights drawn from a range of analytical perspectives.

3.1 Dataset Overview

FinMultiTime is a comprehensive and heterogeneous dataset exceeding 112.6 GB in total size. Its multidimensional structure highlights the dataset's richness and diversity. The assembly process consumed approximately 5 TB of computing resources over a 60-day period, underscoring the complexity of the task and our commitment to ensuring high-quality, robust data for downstream analysis.

3.2 Data Statistics

Language Distribution As shown in the first three panels of Figure 8, we compare the proportions of Chinese and English news articles, tabular records, and charts to illustrate FinMultiTime’s multilingual coverage and its relevance for global financial research.

Temporal Distribution Figures 9 and 10 show the yearly volume of U.S. stock market news (1999–2025) and Chinese market news (2000–2025), respectively. Figures 11 and 12 present the number of K-line charts for the U.S. market (2006–2025) and the Chinese market (2000–2025). These temporal trends provide insights into the historical development of financial news coverage and visual analysis tools across markets.

Industry Distribution Figures 6 and 7 compare the industry compositions of the HS300 constituents (classified by Shenwan Level-1) and the S&P 500 constituents (classified by GICS Level-1). Under the more granular Shenwan classification, HS300 stocks are concentrated in sectors such as Pharmaceuticals & Biotechnology, Non-Bank Financials, Transportation, Electrical Equipment, and Electronics. In contrast, the broader GICS classification highlights the dominance of large-scale sectors like Financial Services, Healthcare, Information Technology, and Industrials within the S&P 500.

Together, these analyses illustrate FinMultiTime’s distinct value as a benchmark dataset for advanced financial analysis and time-series forecasting—thanks to its extensive market coverage, strong multilingual foundation, and broad temporal span.

4 Experiments

To validate the effectiveness of *FinMultiTime*, we assessed the dataset’s overall performance through quantitative and qualitative evaluations. We outline our experimental strategy and demonstrate the dataset’s robustness in real-world applications.

4.1 Experiment Settings

We evaluate stock prediction performance on a FinMultiTime subset of 70 representative stocks, comparing various approaches.

4.1.1 Datasets. We conduct our experiments using a subset of the FinMultiTime dataset, specifically selecting the 35 most influential constituents of the 2025 S&P 500 index and the 35 most influential constituents of the HS300 index, resulting in a total of 70 representative stocks. These samples were processed through our annotation pipeline, producing 77,702 sentiment-annotated news articles, 599,846 semiannual K-line charts, and 3,853 quarterly or annual structured financial records. For detailed statistics, please refer to Table 6.

4.1.2 Compared Methods. We conducted a comparative study of existing unimodal and bimodal approaches for stock price prediction.

Traditional time-series models: These methods transform each modality—historical prices, news articles, financial tables, and K-line charts—into temporal sequences. This includes price histories, sentiment scores derived from news, financial variables extracted from tables, and visual features obtained from charts. Each resulting sequence is then processed individually by standard architectures such as RNN, LSTM, GRU [30], 1D CNN [6], and a 4-layer TimeNet [36].

Recent bimodal models combining text and time series: We also evaluate three representative LLM-based approaches that jointly utilize text and time-series data: CALF [22], ChatTime [32], and FTS-Text-MoE [40]. CALF enhances cross-modal fusion by aligning data distributions; ChatTime treats time series as language tokens to enable unified LLM modeling; and FTS-Text-MoE leverages a sparse, lightweight MoE Transformer decoder for efficient forecasting. All three methods take raw news text and structured time-series data as joint inputs.

4.2 Qualitative Tests

Impact of Multimodal Data on Prediction Performance: Tables 7 and 8 demonstrate clearly that incorporating additional modalities such as news sentiment, fundamental data, and image trends substantially enhances model prediction performance. For instance, in predicting HS300 stocks with a 24-hour horizon using the GRU model, the mean squared error (MSE) drops significantly from 0.2523 (time-series only) to 0.0727 upon including the image trend modality. Similarly, for S&P500 stocks with a 48-hour horizon using the CNN model, the MSE improves notably from 0.0151 with pure time-series data to 0.0053 when news sentiment data is added. These examples illustrate that leveraging multimodal data effectively enhances prediction accuracy for stock prices.

Comparative Analysis of Model Performance: A comparison of the two tables highlights distinct differences in performance among models across varying prediction horizons. For example, in predicting HS300 stocks with a 96-hour horizon, the TimeNet model achieves a mean absolute error (MAE) of 0.1762 using only time-series data, whereas the CNN model performs less effectively, showing an MAE of 0.2543. This indicates TimeNet’s superior capability for longer-term forecasting. Conversely, in the 24-hour horizon prediction task for S&P500 stocks, the CNN model exhibits an MSE of 0.0045 using only time-series data, significantly outperforming both the RNN model (MSE of 0.1215) and the TimeNet model (MSE of 0.3199). This clearly indicates the CNN model’s superior performance in short-term forecasting scenarios. Therefore, selecting an appropriate model based on the forecast horizon and data modality is critical for achieving optimal prediction performance.

4.3 Quantitative Tests

To investigate the impact of dataset scale on performance, we trained models using bilingual datasets (HS300/S&P 500), each comprising three subsets with 5, 15, and 35 stocks.

As shown in Table 9, model performance consistently improves as the number of stocks increases from 5 to 35, underscoring the benefits of larger training datasets for enhancing predictive accuracy. Furthermore, among the three evaluated models (Chat Time, CALF, and FTS-Text-MoE), the FTS-TEXT-MoE model achieves the best performance in most cases. Specifically, for the S&P 500 dataset with 5 stocks, FTS-TEXT-MoE attains a significantly lower mean absolute error (MAE) of 0.0894 and mean squared error (MSE) of 0.2263, compared to Chat Time’s MAE of 0.4753 and MSE of 0.4432, and CALF’s MAE of 0.3763 and MSE of 0.2688. Similarly, in the HS300 dataset with 35 stocks, the FTS-TEXT-MoE model records the lowest MAE (0.3817) and MSE (0.5010), clearly surpassing the performance of Chat Time (MAE 0.5643, MSE 0.5117) and CALF (MAE 0.5301, MSE 0.5718).

These results highlight the superior capability of the FTS-TEXT-MoE model in effectively integrating textual and time-series data to capture complex patterns, thereby yielding more accurate short-term stock price predictions. All experiments utilized 96 days of historical data to forecast prices for the subsequent 24 days. Each model was trained for 100 epochs, evaluated on the 5-stock split after removing one outlier, and the reported results represent mean values.

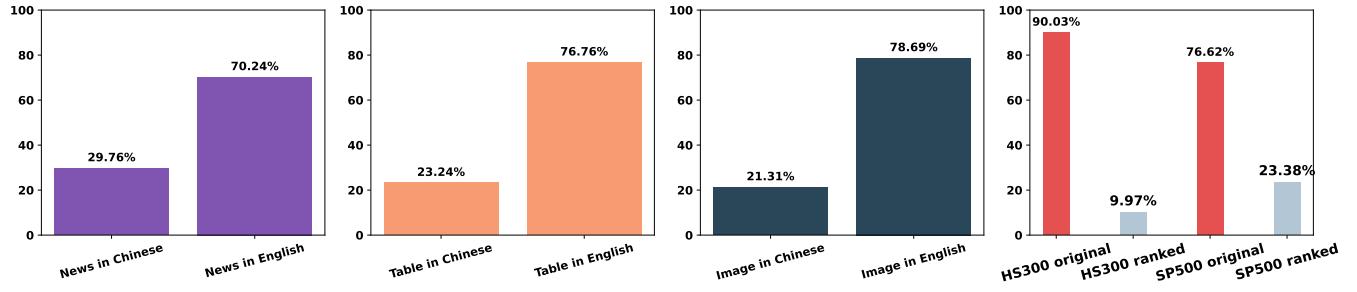


Figure 8: Proportions of Chinese vs. English Modalities (News, Tables, Images) and Coverage Ratios of Ranked vs. Original Daily News for HS300 and S&P 500.

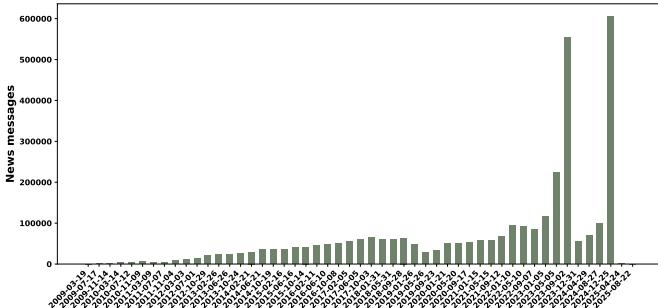


Figure 9: Number of S&P500 News Articles Over Time

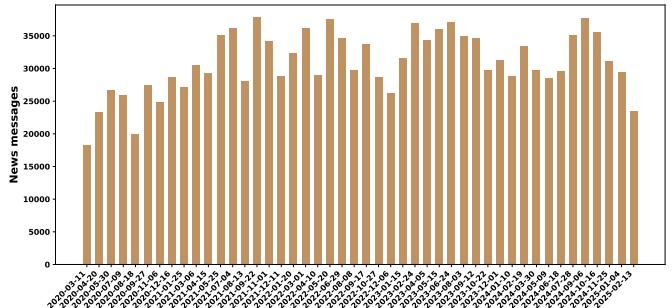


Figure 10: Number of HS300 News Articles Over Time

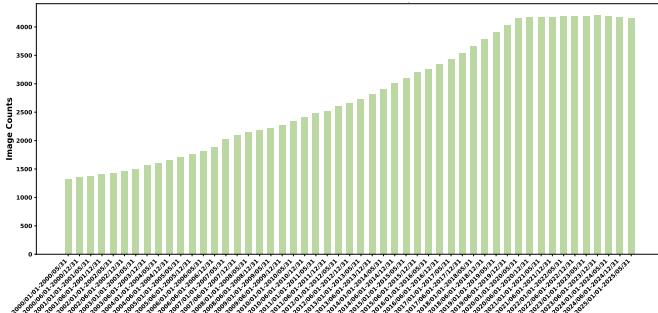


Figure 11: S&P500 Candlestick Chart Counts by Time Period

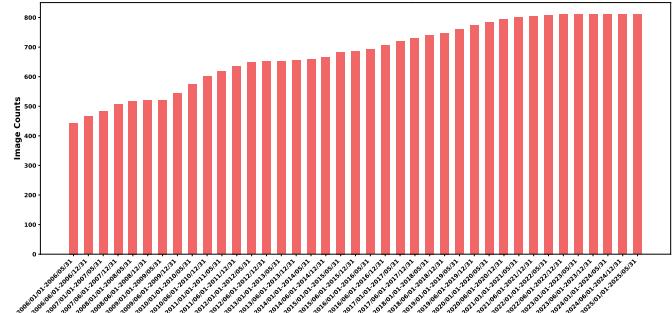


Figure 12: HS300 Candlestick Chart Counts by Time Period

Table 6: HS300 vs. S&P 500 – Multimodal Record Counts (35 stocks each)

	Semi-annual trend images	Quarterly / annual tables	Daily time-series points	News-sentiment scores
HS300	299,923	1,749	299,923	26,467
S&P 500	299,923	2,104	299,923	51,235
Total	599,846	3,853	599,846	77,702

5 Related Work

Financial time-series models Traditional time-series models like linear regression [34], ARIMA [2] and GARCH [3] depend on stationarity and strong assumptions, so they often miss complex dependencies or abrupt shocks. Recently, machine learning [13, 14, 18], deep learning [29] and NLP [26, 39] have tapped sentiment and other qualitative signals to enhance forecast accuracy. This trend mirrors Markowitz’s market correlation concept, linking sentiment from news, blogs, and social media to asset prices.

With growing data and compute, LLMs now enable finer sentiment quantification [25]. Moreover, TSMixer-MICM [17] turns quarterly financial-statement tables into time-series features, aligning them with price and text data for three-modal analysis.

Financial Multimodal time-series datasets. Financial Multimodal time-series datasets fall into two groups. General economic collections (e.g., Time-MMD [21], CiK [35]) pair macro-text with monthly indicators but are too coarse and small for fine-grained forecasting. Financial-specific sets target asset prices: NewsForecast links Bitcoin news to daily prices; TimeCAP[20], DOW30[7],

Table 7: Model performance across modalities and prediction horizons (24, 48, 96) with fixed input length of 96 for HS300 stocks. Red values indicate the worst time-series prediction for each modality at the given horizon. The table compares five models over four modalities: (1) Time Series Only (price data), (2) News Sentiment, (3) Image Trend, and (4) Fundamental Table. Metrics reported are MAE and MSE (lower is better).

Horizon	Modality	CNN		GRU		LSTM		RNN		TimeNet	
		MAE (↓)	MSE (↓)								
24	Time Series	0.1720	0.1097	0.2317	0.2523	0.0985	0.0281	0.2058	0.1197	0.0886	0.2808
	News Sentiment	0.0756	0.0087	0.1101	0.0193	0.0834	0.0177	0.0707	0.0093	0.0268	0.1364
	Fundamental Table	0.1751	0.0582	0.0796	0.0117	0.0758	0.0104	0.1819	0.0501	0.0691	0.2425
	Image Trend	0.0641	0.0082	0.0727	0.0104	0.0937	0.0202	0.0982	0.0201	0.0078	0.0721
48	Time Series	0.2154	0.1219	0.4306	0.5236	0.1997	0.0721	0.3228	0.1831	0.1171	0.3064
	News Sentiment	0.0770	0.0103	0.1955	0.0954	0.1130	0.0255	0.1149	0.0250	0.0736	0.2379
	Fundamental Table	0.1205	0.0269	0.1212	0.0225	0.1298	0.0235	0.1776	0.0446	0.0718	0.2508
	Image Trend	0.0996	0.0165	0.1686	0.0531	0.1723	0.0570	0.1655	0.0352	0.0885	0.2701
96	Time Series	0.2543	0.1481	0.6032	0.6982	0.1329	0.0311	0.1875	0.0633	0.1762	0.3917
	News Sentiment	0.0899	0.0123	0.1841	0.0882	0.1257	0.0277	0.1689	0.0514	0.0901	0.2404
	Fundamental Table	0.1262	0.0397	0.1305	0.0261	0.1628	0.0430	0.1810	0.0457	0.1567	0.3375
	Image Trend	0.1230	0.0399	0.1835	0.0537	0.1943	0.0881	0.1810	0.0626	0.0482	0.1751

Table 8: Model performance across modalities and prediction horizons (24, 48, 96) with fixed input length of 96 for S&P500 stocks.

Horizon	Modality	CNN		GRU		LSTM		RNN		TimeNet	
		MAE (↓)	MSE (↓)								
24	Time Series	0.0498	0.0045	0.0949	0.0234	0.1120	0.0230	0.2158	0.1215	0.1085	0.3199
	News Sentiment	0.0490	0.0039	0.0467	0.0041	0.0435	0.0036	0.0753	0.0102	0.0242	0.1329
	Fundamental Table	0.0463	0.0038	0.0931	0.0167	0.0888	0.0152	0.1664	0.0395	0.0894	0.2910
	Image Trend	0.0475	0.0041	0.0525	0.0055	0.0723	0.0122	0.1485	0.0301	0.0052	0.0594
48	Time Series	0.0820	0.0151	0.1433	0.0445	0.0895	0.0166	0.1754	0.1104	0.1646	0.3913
	News Sentiment	0.0559	0.0053	0.0569	0.0059	0.0880	0.0159	0.1148	0.0218	0.1089	0.3147
	Fundamental Table	0.0515	0.0051	0.1207	0.0280	0.0868	0.0125	0.0806	0.0099	0.1320	0.3587
	Image Trend	0.0714	0.0091	0.0708	0.0118	0.0799	0.0165	0.1452	0.0294	0.0970	0.2998
96	Time Series	0.0769	0.0095	0.2251	0.1129	0.1380	0.0322	0.1183	0.0209	0.2027	0.4381
	News Sentiment	0.0592	0.0055	0.0937	0.0178	0.1259	0.0266	0.1129	0.0213	0.0895	0.2643
	Fundamental Table	0.0775	0.0105	0.1163	0.0228	0.1465	0.0356	0.1158	0.0182	0.1846	0.4098
	Image Trend	0.0939	0.0201	0.1174	0.0340	0.1110	0.0312	0.1069	0.0149	0.0482	0.1937

Table 9: Performance comparison of three state-of-the-art bimodal models—Chat Time, CALF, and FTS-Text-MoE—using both text and time-series inputs. Evaluations are conducted on two datasets (S&P 500 and HS 300) across three stock subsets (5, 15, and 35 stocks). Prediction horizon is 24 with a fixed input length of 96. Best scores per row are highlighted.

#	Model	S&P 500		HS 300	
		MAE (↓)	MSE (↓)	MAE (↓)	MSE (↓)
5	Chat Time	0.4753	0.4432	0.2647	0.1195
	CALF	0.3763	0.2688	0.5707	0.6290
	FTS-Text-MoE	0.0894	0.2263	0.2506	0.3789
15	Chat Time	0.6648	0.9103	0.4804	0.3975
	CALF	0.4996	0.5649	0.4921	0.5660
	FTS-Text-MoE	0.3299	0.4818	0.3232	0.4651
35	Chat Time	0.6431	0.6928	0.5643	0.5117
	CALF	0.5503	0.6011	0.5019	0.5778
	FTS-Text-MoE	0.3914	0.5344	0.3817	0.5010

TSQA[16] align stock news with prices; ACL18[41], CIKM18[37], SEP[15] use tweet sentiment. FinBEN [38] and FNSPID Nasdaq [9] add bilingual text yet remain text-price only, while a 2024 EMNLP Findings study[17] is first to fuse quarterly tables with text and prices, albeit at low frequency and small scale. Overall, these

datasets are modest in size and mostly single-market (chiefly U.S.), limiting their usefulness for pre-training and evaluating emerging large-scale financial LLMs and multimodal models.

6 Conclusion

In this work, we present FinMultiTime, the first large-scale, cross-market multimodal dataset for financial time-series forecasting. By temporally aligning stock prices, financial news, structured fundamentals, and K-line charts across the S&P 500 and HS300 universes, FinMultiTime provides a rich foundation for AI-driven financial prediction. Our experiments highlight three key findings. First, larger dataset scale consistently enhances model accuracy, confirming the importance of data volume for robust learning. Second, multimodal fusion significantly boosts predictive performance, as evidenced by substantial error reductions when incorporating sentiment, fundamentals, or image-based trends. Third, comparative analysis reveals that model choice and prediction horizon strongly influence outcomes. Finally, the reproducible pipeline ensures ongoing dataset expansion and adaptability, enabling the research community to build upon this resource. Overall, FinMultiTime advances financial forecasting by integrating heterogeneous signals at scale.

References

- [1] Dogu Araci. 2019. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063* (2019).
- [2] Adebiyi A Ariyo, Adewumi O Adewumi, and Charles K Ayo. 2014. Stock price prediction using the ARIMA model. In *2014 UKSim-AMSS 16th international conference on computer modelling and simulation*. IEEE, 106–112.
- [3] Luc Bauwens, Sébastien Laurent, and Jeroen VK Rombouts. 2006. Multivariate GARCH models: a survey. *Journal of applied econometrics* 21, 1 (2006), 79–109.
- [4] Lei Chai, Hongfeng Xu, Zhiming Luo, and Shaozi Li. 2020. A multi-source heterogeneous data analytic method for future price fluctuation prediction. *Neurocomputing* 418 (2020), 11–20.
- [5] Clayton Leroy Chapman, Lars Hillebrand, Marc Robin Stenzel, Tobias Deußer, David Biesner, Christian Bauckhage, and Rafet Sifa. 2022. Towards generating financial reports from tabular data using transformers. In *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer, 221–232.
- [6] Jou-Fan Chen, Wei-Lun Chen, Chun-Ping Huang, Szu-Hao Huang, and An-Pin Chen. 2016. Financial time-series data analysis using deep convolutional neural networks. In *2016 7th International conference on cloud computing and big data (CCBD)*. IEEE, 87–92.
- [7] Zihan Chen, Lei Nico Zheng, Cheng Lu, Jialu Yuan, and Di Zhu. 2023. Chatgpt informed graph neural network for stock movement prediction. *arXiv preprint arXiv:2306.03763* (2023).
- [8] Junyan Cheng and Peter Chin. 2024. Sociodojo: Building lifelong analytical agents with real-world text and time series. In *The Twelfth International Conference on Learning Representations*.
- [9] Zihan Dong, Xinyu Fan, and Zhiyuan Peng. 2024. Fnspid: A comprehensive financial news dataset in time series. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4918–4927.
- [10] Kui Fu and Yanbin Zhang. 2024. Incorporating Multi-Source Market Sentiment and Price Data for Stock Price Prediction. *Mathematics* 12, 10 (2024), 1572.
- [11] Udit Gupta. 2023. GPT-InvestAR: Enhancing stock investment strategies through annual report analysis with large language models. *arXiv preprint arXiv:2309.03079* (2023).
- [12] Yu Huang, Chenzhuang Du, Zihui Xue, Xuanyao Chen, Hang Zhao, and Longbo Huang. 2021. What makes multi-modal learning better than single (provably). *Advances in Neural Information Processing Systems* 34 (2021), 10944–10956.
- [13] Bryan Kelly, Dacheng Xiu, et al. 2023. Financial machine learning. *Foundations and Trends® in Finance* 13, 3-4 (2023), 205–363.
- [14] Kyoung-jae Kim. 2003. Financial time series forecasting using support vector machines. *Neurocomputing* 55, 1-2 (2003), 307–319.
- [15] Kelvin JL Koa, Yunshan Ma, Ritchie Ng, and Tat-Seng Chua. 2024. Learning to generate explainable stock predictions using self-reflective large language models. In *Proceedings of the ACM Web Conference 2024*. 4304–4315.
- [16] Yaxuan Kong, Yiyuan Yang, Yoontae Hwang, Wenjie Du, Stefan Zohren, Zhangyang Wang, Ming Jin, and Qingsong Wen. 2025. Time-MQA: Time Series Multi-Task Question Answering with Context Enhancement. *arXiv preprint arXiv:2503.01875* (2025).
- [17] Ross Koval, Nicholas Andrews, and Xifeng Yan. 2024. Financial Forecasting from Textual and Tabular Time Series. In *Findings of the Association for Computational Linguistics: EMNLP 2024*. 8289–8300.
- [18] Bjoern Krollner, Bruce Vanstone, and Gavin Finnie. 2010. Financial time series forecasting with machine learning techniques: A survey. In *European Symposium on Artificial Neural Networks: Computational Intelligence and Machine Learning*. 25–30.
- [19] Moreno La Quatra and Luca Cagliero. 2020. End-to-end training for financial report summarization. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*. 118–123.
- [20] Geon Lee, Wenchao Yu, Kijung Shin, Wei Cheng, and Haifeng Chen. 2025. Time-cap: Learning to contextualize, augment, and predict time series events with large language model agents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 18082–18090.
- [21] Haoxin Liu, Shangqing Xu, Zhiyuan Zhao, Lingkai Kong, Harshavardhan Prabhakar Kamarthi, Aditya Sasanur, Megha Sharma, Jiaming Cui, Qingsong Wen, Chao Zhang, et al. 2024. Time-mmld: Multi-domain multimodal dataset for time series analysis. *Advances in Neural Information Processing Systems* 37 (2024), 77888–77933.
- [22] Peiyuan Liu, Hang Guo, Tao Dai, Naiqi Li, Jigang Bao, Xudong Ren, Yong Jiang, and Shu-Tao Xia. 2025. Calf: Aligning llms for time series forecasting via cross-modal fine-tuning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 18915–18923.
- [23] Xiao-Yang Liu, Guoxuan Wang, Hongyang Yang, and Daochen Zha. [n. d.]. Fin-GPT: Democratizing Internet-scale Data for Financial Large Language Models. In *NeurIPS 2023 Workshop on Instruction Tuning and Instruction Following*.
- [24] Xiao-Yang Liu, Hongyang Yang, Qian Chen, Runjia Zhang, Liqing Yang, Bowen Xiao, and Christina Dan Wang. 2020. FinRL: A deep reinforcement learning library for automated stock trading in quantitative finance. *arXiv preprint arXiv:2011.09607* (2020).
- [25] Alejandro Lopez-Lira and Yuehua Tang. 2023. Can chatgpt forecast stock price movements? return predictability and large language models. *arXiv preprint arXiv:2304.07619* (2023).
- [26] Mantas Lukauskas, Vaida Pilinkienė, Jurgita Bruneckienė, Alina Stundžienė, Andrius Grybauskas, and Tomas Ruzgas. 2022. Economic activity forecasting based on the sentiment analysis of news. *Mathematics* 10, 19 (2022), 3461.
- [27] Burton G. Malkiel and Eugene F. Fama. 1970. EFFICIENT CAPITAL MARKETS: A REVIEW OF THEORY AND EMPIRICAL WORK. *The Journal of Finance* 25, 2 (1970), 383–417. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1970.tb00518.x>
- [28] Eliza Mik. 2017. Smart contracts: terminology, technical limitations and real world complexity. *Law, innovation and technology* 9, 2 (2017), 269–300.
- [29] Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu. 2020. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied soft computing* 90 (2020), 106181.
- [30] Guizhu Shen, Qingping Tan, Haoyu Zhang, Ping Zeng, and Jianjun Xu. 2018. Deep learning with gated recurrent unit networks for financial sequence predictions. *Procedia computer science* 131 (2018), 895–903.
- [31] Wataru Souma, Irena Vodenska, and Hideaki Aoyama. 2019. Enhanced news sentiment analysis using deep learning methods. *Journal of Computational Social Science* 2, 1 (2019), 33–46.
- [32] Chengsen Wang, Qi Qi, Jingyu Wang, Haifeng Sun, Zirui Zhuang, Jinming Wu, Lei Zhang, and Jianxin Liao. 2025. Chattime: A unified multimodal time series foundation model bridging numerical and textual data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 12694–12702.
- [33] Xinlei Wang, Maike Feng, Jing Qiu, Jinjin Gu, and Junhua Zhao. 2024. From news to forecast: Integrating event analysis in llm-based time series forecasting with reflection. *Advances in Neural Information Processing Systems* 37 (2024), 58118–58153.
- [34] Sanford Weisberg. 2005. *Applied linear regression*. Vol. 528. John Wiley & Sons.
- [35] Andrew Robert Williams, Arjun Ashok, Étienne Marcotte, Valentina Zantedeschi, Jithendaraa Subramanian, Roland Riachi, James Requeima, Alexandre Lacoste, Irina Rish, Nicolas Chapados, et al. 2024. Context is key: A benchmark for forecasting with essential textual information. *arXiv preprint arXiv:2410.18959* (2024).
- [36] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. 2022. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186* (2022).
- [37] Huiwu Hu, Wei Zhang, Weiwei Shen, and Jun Wang. 2018. Hybrid deep sequential modeling for social text-driven stock prediction. In *Proceedings of the 27th ACM international conference on information and knowledge management*. 1627–1630.
- [38] Qianqian Xie, Weiguang Han, Zhengyu Chen, Ruoyu Xiang, Xiao Zhang, Yueru He, Mengxi Xiao, Dong Li, Yongfu Dai, Duanyu Feng, et al. 2024. Finben: A holistic financial benchmark for large language models. *Advances in Neural Information Processing Systems* 37 (2024), 95716–95743.
- [39] Frank Z Xing, Erik Cambria, and Roy E Welsch. 2018. Natural language based financial forecasting: a survey. *Artificial Intelligence Review* 50, 1 (2018), 49–73.
- [40] Wenyuan Xu, Dawei Xiang, Rundong Wang, Yonghong Hu, Liang Zhang, Jiayu Chen, and Zhonghua Lu. 2025. Learning Explainable Stock Predictions with Tweets Using Mixture of Experts. *arXiv preprint arXiv:2507.20535* (2025).
- [41] Yumo Xu and Shay B Cohen. 2018. Stock movement prediction from tweets and historical prices. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1970–1979.
- [42] Hang Yang, Yubo Chen, Kang Liu, Yang Xiao, and Jun Zhao. 2018. Dcfcc: A document-level chinese financial event extraction system based on automatically labeled training data. In *Proceedings of ACL 2018, System Demonstrations*. 50–55.
- [43] Runze Yang, Longbing Cao, Xin You, Kun Fang, Jianxun Li, and Jie Yang. 2025. Fourier Basis Mapping: A Time-Frequency Learning Framework for Time Series Forecasting. *arXiv preprint arXiv:2507.09445* (2025).
- [44] Yi Yang, Yixuan Tang, and Kai Yan Tam. 2023. Investlm: A large language model for investment using financial domain instruction tuning. *arXiv preprint arXiv:2309.13064* (2023).
- [45] Wentao Zhang, Lingxuan Zhao, Haochong Xia, Shuo Sun, Jiaze Sun, Molei Qin, Xinyi Li, Yuqing Zhao, Yilei Zhao, Xinyu Cai, et al. 2024. A multimodal foundation agent for financial trading: Tool-augmented, diversified, and generalist. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4314–4325.
- [46] Yanzhao Zou and Dorien Herremans. 2023. PreBit—A multimodal model with Twitter FinBERT embeddings for extreme price movement prediction of Bitcoin. *Expert Systems with Applications* 233 (2023), 120838.

A Bilingual News Summarize Algorithm

In reference to FNNSPID [9], we introduce a weight model W_z to enhance summarization and emphasize relevant stocks. In the sumy

package, all terms are included in the summary. Exclusiveness involves rephrasing sentences rather than extracting terms. We parse the graph G into sentences and assign a weight W_p based on relevance to the stock symbol, setting $k = 1$ for sentences containing the symbol. For summarized sentences S_{sum} , a score of $t = 1$ is given if the sentence is longer. In Equation (6), we combine W_p and W_q to calculate the final weight W_z , with irrelevant sentences receiving a weight of 0. The sentences are sorted by weight to form the final summary.

$$W_p(S, s) = \begin{cases} k & \text{if } S \in G \\ 0 & \text{otherwise} \end{cases}$$

$$W_q(S_{sum}, S_{long}) = \begin{cases} t & \text{if } S_{sum} \in S_{long} \\ 0 & \text{otherwise} \end{cases}$$

$$W_z = W_p + W_q$$

B FinMultiTime Applications

This section critically examines the potential uses of the FinMultiTime dataset in financial-market research, the technical hurdles encountered during its construction, and the attendant ethical challenges, while outlining avenues for future work.

B.1 Construction Challenges

Bilingual news extraction and sentiment labelling. We experimented with lightweight extractive algorithms (Luhn, LexRank, TextRank) and generative models (distilbert-cnn-12-6). Although both approaches handle simple sentiments (e.g., “sharp price rise” or “steep decline”) reasonably well, extractive methods often miss key context in longer passages, whereas generative models suffer from summary repetition, unstable scores, and attention drift on lengthy documents.

Modal imbalance. Relying on a small set of tabular variables or on trend labels derived solely from candlestick images fails to unlock the complementary value of FinMultiTime’s four modalities. These limitations underscore the need for more efficient architectures that can exploit mutual information among text, images, and structured data to reveal genuine predictive power.

B.2 Prospective Use Cases

Multimodal model training. The temporally aligned text, numeric, image, and table streams enable the development of joint-learning models for stock prediction. Such models can bolster robustness to short-horizon noise and improve reinforcement-learning agents in sequential decision-making—especially for trend forecasting and strategy design.

Sentiment and trend-signal analysis. Combining news-sentiment scores with long-horizon trend labels allows researchers to assess the incremental explanatory power of non-price signals within a modern portfolio-theory framework. Batch processing of sentiment and trend indicators across many tickers further refines market forecasts and portfolio allocation.

Correlation and anomaly detection. The four aligned modalities facilitate granular studies of how sentiment, image-based trends,

and fundamentals correlate with price dynamics, potentially revealing latent market drivers. Pattern matching on historical data can surface precursors of systemic risk, offering fresh tools for volatility warnings and risk management.

Financial generative-AI applications. With its large, heterogeneous corpus, FinMultiTime serves as prime fine-tuning material for large language models, powering next-generation robo-advisers, automated report writers, and other finance-oriented AI services.

B.3 Ethical Considerations

Privacy and data security. Financial records often contain sensitive personal or institutional information. We employ state-of-the-art anonymisation and de-identification techniques and adhere strictly to GDPR, CCPA, and related regulations to safeguard privacy throughout data collection and processing.

Misuse risks. Predictive models built on FinMultiTime could be misappropriated, leading to market manipulation or systemic risk. We therefore conduct bias and fairness audits and publish explicit usage guidelines to curb discriminatory or misleading outcomes.

Transparency and traceability. Every record is source-tagged, and detailed processing documentation is released publicly, ensuring reproducibility, auditability, and responsible research practice.

By addressing construction bottlenecks, enriching multimodal use cases, and enforcing rigorous ethical safeguards, FinMultiTime not only provides a solid empirical foundation for financial-market analysis but also sets a high academic and ethical benchmark for future industry and scholarly endeavours.

C Future Work

Expanding the FinMultiTime Dataset: Although our coverage of stock-related data is extensive, financial data remain inherently time-sensitive. We plan to develop an automated pipeline to continuously ingest and update news feeds, thereby substantially enlarging the dataset’s scope and currency.

Unlocking FinMultiTime’s Full Potential: As the most comprehensive resource aligning price series, sentiment annotations, long-term trend signals, and corporate fundamental data, FinMultiTime can support several frontier research directions:

Multimodal Modeling: Multimodal modeling will integrate heterogeneous sources—text, images, tables, and time series—to construct more robust market-prediction models; sentiment-impact analysis will quantitatively assess how news sentiment drives stock-price volatility, thereby advancing sentiment-analysis algorithms; trend-signal evaluation will investigate the contribution of long-term trend indicators to forecasting accuracy; and fundamental-data integration will examine the auxiliary role of financial-statement features in investment decision-making to enhance real-world applicability. Although our news coverage is already extensive, the synergistic exploitation of chart images, textual summaries, and tabular data remains underexploited. In future work, we will explore pre-training language models within a reinforcement-learning framework to improve multimodal feature extraction and its downstream applications.

By identifying these limitations and outlining targeted research avenues, we aim to inspire subsequent studies and further enhance the value and impact of the FinMultiTime dataset.