Article

MM-iTransformer: A Multimodal Approach to Economic Time Series Forecasting with Textual Data

Shangyang Mou [1,*], Qiang Xue [1] , Jinhui Chen [2], Tetsuya Takiguchi [1] and Yasuo Ariki [1]

[1] Graduate School of System Informatics, Kobe University, Kobe 657-8501, Japan;
xueqiang199305@gmail.com (Q.X.); takigu@kobe-u.ac.jp (T.T.); ariki@kobe-u.ac.jp (Y.A.)
[2] Faculty of Systems Engineering, Wakayama University, Wakayama 640-8510, Japan; ckinki@wakayama-u.ac.jp
* Correspondence: phoebemou@outlook.com

Abstract: This paper introduces a novel multimodal framework for economic time series forecasting, integrating textual information with historical price data to enhance predictive accuracy. The proposed method employs a multi-head attention mechanism to dynamically align textual embeddings with temporal price data, capturing previously unrecognized cross-modal dependencies and enhancing the model's ability to interpret event-driven market dynamics. This enables the framework to model complex market behaviors in a unified and effective manner. Experimental results across multiple financial datasets, including the foreign exchange (Forex) and Gold-price datasets, demonstrate that incorporating textual information significantly enhances forecasting accuracy. Compared to models relying solely on historical price data, the proposed framework achieves a substantial reduction in Mean Squared Error (MSE) loss, with improvements of up to 26.79%. This highlights the effectiveness of leveraging textual inputs alongside structured time series data in capturing complex market dynamics and improving predictive performance.

Keywords: economic time series forecasting; feature engineering; multimodal data fusion; deep learning; data mining

1. Introduction

Time series forecasting is a critical challenge with significant practical implications across various domains, such as economics, energy, weather, traffic monitoring, and healthcare [1–4]. The field of economics, in particular, has demonstrated a longstanding interest in accurately predicting economic time series data. Financial markets, which inherently generate data in time series form, have been extensively studied over the past decades. The primary goal of these studies has been to uncover deeper insights into market trends. As researchers gain a better understanding of market behaviors, they are increasingly able to develop more effective investment strategies, thereby enhancing financial decision-making processes. However, the efficient market hypothesis (EMH) [5] posits that asset prices reflect all available information, theoretically rendering them unpredictable, as new data are instantly assimilated. Despite this, practical forecasting efforts continue to show that financial markets are indeed a combination of efficient and non-efficient markets, suggesting that they are, to some extent, predictable [6,7].

Time series forecasting primarily aims to project future values based on historical data, often exploiting repeating long-term or short-term patterns inherent in the data. Over the years, the field has explored various methodologies to capture these patterns, including linear and nonlinear time series models, and artificial intelligence techniques. However, in the economic field, the characteristics of data present unique challenges that complicate

the application of these models. Firstly, financial data are highly volatile and can exhibit extreme value changes, which undermines the assumption of stable patterns over time [8]. Secondly, compared to other types of time series, financial data typically exhibit more complex seasonality and more variable trends [9]. This variability is often exacerbated by sudden and unpredictable events, adding a layer of complexity to the forecasting. These factors make economic forecasting particularly challenging.

Different approaches for economic time series forecasting have been explored, ranging from econometric theory and statistical methodologies to artificial intelligence models. Statistical methods, such as the autoregressive integrated moving average mbox(ARIMA) model [10,11] and the semi-functional partial linear model [12], have also been employed in this field. These models benefit from a solid mathematical foundation and offer strong interpretability [13], but their predictions are often imprecise due to the nonlinear and irregular nature of economic data.

In response to the complexity of economic time series, neural networks and support vector machines (SVM) have demonstrated considerable success in modeling non-linear relationships, offering robust alternatives to traditional linear models [14]. Furthermore, recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have become particularly prevalent due to their ability to effectively extract time dependencies from time series data, thanks to the memory features obtained from their feedback mbox mechanisms [15,16]. More recently, transformer-based models have emerged for this task, employing attention mechanisms capable of learning global context and long-term dependencies, thus enhancing the depth of time series analysis [17,18].

As deep learning expands its application in time series forecasting, the integration of natural language processing (NLP) techniques is increasingly transforming economic predictions. Textual data, available in real-world economic time series forecasting applications, often contain nuanced information that may surpass the insights provided by numerical financial series alone [19]. Recent studies have begun using textual data in economic forecasting models to capture these additional insights. For instance, some models analyze large volumes of financial news to identify sentiments and factual information that may affect investor behavior and market trends [20].

In financial market forecasting, Farimani et al. [21] propose generating sentiment-based features over time and combining them with time series data to improve predictions. Similarly, Reis Filho et al. [22] focus on agricultural commodity prices, using a low-dimensional representation of domain-specific text data enriched with selected keywords, which addresses issues of high dimensionality and data sparsity, improving forecast accuracy. Baranowski et al. [23] take a different approach, developing a tone shock measure from European Central Bank communications to predict monetary policy decisions. In cryptocurrency forecasting, Erfanian et al. [24] apply machine learning methods such as SVR, OLS, ensemble learning, and MLP to examine how macroeconomic, microeconomic, technical, and blockchain indicators relate to Bitcoin prices over short and long terms. These studies show how NLP can enhance time series forecasting by integrating textual data with traditional economic indicators.

These studies demonstrate the potential of natural language processing (NLP) to enhance time series forecasting by incorporating textual data alongside traditional economic indicators. However, most existing approaches simplify the complexity of textual data by focusing on specific aspects, such as sentiment polarity, tone, or keyword-based representations, neglecting the rich, nuanced information embedded in the full text. This narrow scope limits their ability to capture the non-linear and qualitative impacts of news and other unstructured data sources. Incorporating comprehensive news content into financial forecasting can enhance datasets by introducing additional contextual information,

improving the adaptability and accuracy of predictive models. Such an approach allows models to better reflect real-world dynamics, capturing the multifaceted interactions that drive market behavior. Therefore, a more comprehensive framework is required to address these challenges, leveraging both unstructured textual data and structured time series to establish a robust and realistic forecasting methodology.

This paper introduces a novel approach to economic time series forecasting that integrates textual information directly into the predictive framework as an additional variable. Unlike sentiment-centric methods, this approach fully leverages the richness of textual data, such as economic news and reports, to extract predictive signals absent from numerical data alone. By treating textual data as an extension of time series variables, the proposed method provides a unified framework for modeling the interplay between textual and numerical inputs. The primary contributions of this study are as follows:

(1) Multimodal Attention-Based Integration: The proposed method employs an inverted attention mechanism to directly model interactions between textual data and multivariate time series. By treating text as an additional variable, the model captures intricate relationships between historical price movements and textual information, enabling more accurate and robust economic forecasts.

(2) Seamless Integration of Textual and Numerical Data: This study utilizes matched historical price and news datasets to construct multimodal datasets. By combining the contextual depth of textual data with the temporal patterns of price data, the proposed method enhances forecasting performance.

(3) Comprehensive Experimental Validation: Extensive experiments on economic datasets demonstrate the effectiveness of the proposed approach. Results highlight the model's ability to outperform traditional forecasting techniques by leveraging both textual and numerical inputs, offering deeper insights into market dynamics. Figure 1 compares (b) our multimodal learning method with the existing (a) unimodal approach, demonstrating how this integration enhances the model's ability to utilize diverse data sources and extract valuable insights, thereby improving the accuracy of time series forecasting
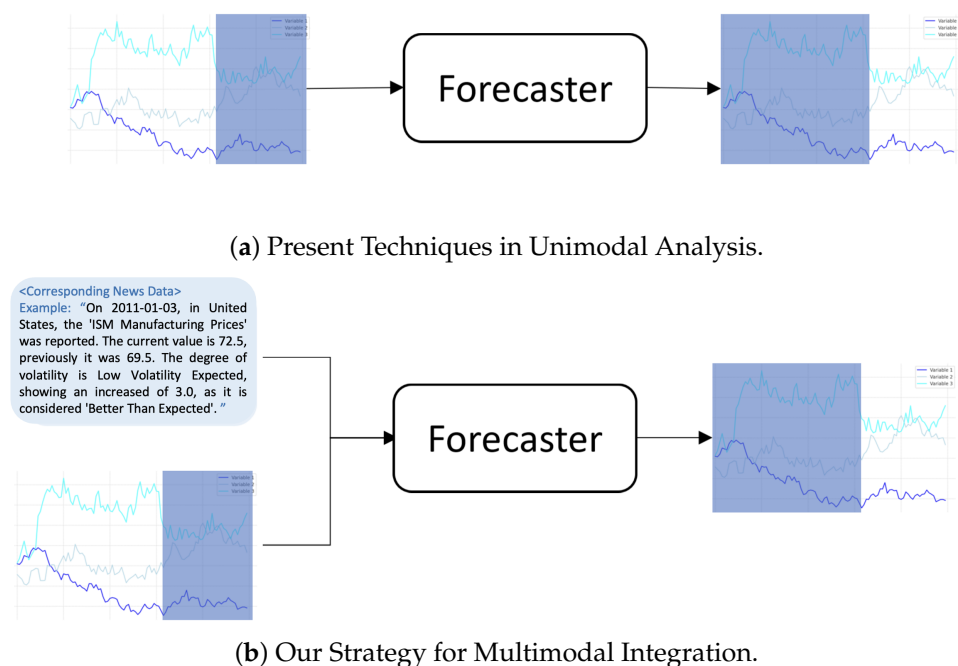

(**a**) Present Techniques in Unimodal Analysis.

(**b**) Our Strategy for Multimodal Integration.

Figure 1. Time series forecasting framework.

The remainder of this paper is organized as follows: Section 2 provides a detailed exploration of existing methods in time series forecasting. Section 3 describes the proposed method and outlines the framework in detail. Section 4 discusses the dataset, presents the experimental results, and explores some notable findings. Finally, Section 5 concludes the paper.

2. Related Work

2.1. Time Series Forecasting

Given that time series prediction tasks hold paramount significance in real-world applications, many classical approaches have been developed. The study [25] utilizes a hybrid ARIMA model that adopts the Markov process to construct the autoregressive model for iterative sequential forecasting. Nevertheless, the autoregressive process is incapable when handling complex sequences with non-linearity and non-stationarity. Machine learning (ML) methods have gained prominence in this field, an dconvolutional neural networks (CNNs) and long short-term memory (LSTM) neural networks, as two typical deep learning models, have attracted significant attention in recent years due to their unique advantages in modeling the long-term time dependency of data characteristics and local adjacent correlation features in time series [26,27]. For example, ref. [28] utilizes a deep long-short-term memory (DLSTM) architecture, enhanced by a genetic algorithm for optimal configuration, improving the forecasting accuracy of petroleum production.

Additionally, transformers have seen widespread adoption for time series analysis, demonstrating substantial advancements across various applications. Transformers leverage the self-attention mechanism to capture long-term dependencies in sequential data, enabling more efficient and accurate predictions [29,30]. For instance, PatchTST [31] introduces a patch-based input structure to reduce the quadratic complexity of attention calculations, improving transformer efficiency.

The iTransformer framework [32] introduces an inverted perspective for multivariate time series modeling, fundamentally redefining how Transformers process sequential data. Unlike traditional Transformer architectures, which treat all variables at a single time step as a single token (temporal token), iTransformer represents each variable's entire historical sequence as an independent entity called a variate token. Figure 2 illustrates this inverted structure, highlighting how variate tokens and the attention mechanism enable the iTransformer to model multivariate correlations more effectively. This inversion shifts the focus from time-centered embeddings to variable-centered embeddings, aligning the modeling perspective with the inherent structure of time series data, where each variable often reflects a distinct process. By embedding the full sequence of each variable, the attention mechanism naturally captures cross-variable correlations (multivariate correlations), which are essential for understanding complex time series. This design effectively addresses challenges such as misaligned timestamps, limited receptive fields, and semantically sparse temporal tokens, which often weaken the modeling capacity of traditional transformers for time series data.

While these advancements significantly improve time series forecasting, most methods rely solely on historical time series data, overlooking the potential contributions of external data sources. This limitation poses challenges in domains like economics, where external factors often play a pivotal role in influencing future trends.
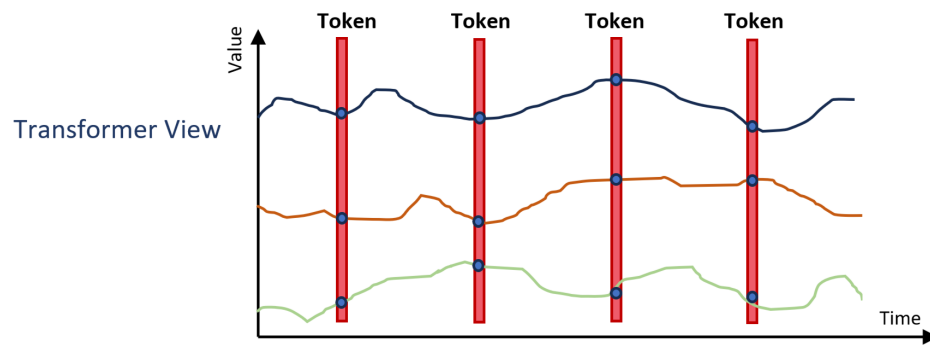
(**a**) Temporal tokens: multiple variables at each time step into a single token.
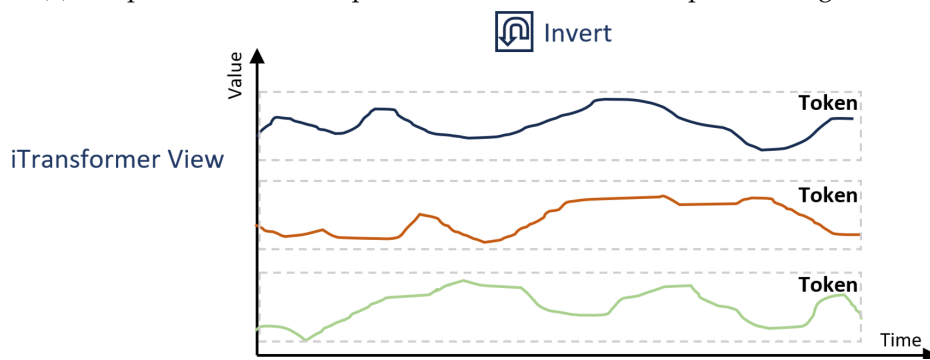
(**b**) Variate tokens: each series independently embedded into a single token for each variable.

Figure 2. Comparison of temporal and variate token embedding approaches in vanilla transformer and iTransformer.

2.2. Text-Enhanced Economic Time Series Forecasting

Multimodal learning, which incorporates both time series and textual corpora as input data sources, has generated significant interest for economic time series forecasting. Mostly, the focus has been primarily on sentiment analysis from financial news and social media to predict stock market trends [33]. In [34], a method that estimates social attention to stocks by sentiment analysis and influence modeling is proposed to predict the movement of the financial market. These methods treat the prediction as a classification problem, where the goal is to determine the direction of market movements based on sentiment polarity.

Moreover, when the objective shifts to forecasting specific future values of a time series, the task becomes a regression problem, which is inherently more challenging than simple directional predictions. Significant efforts have been made to enhance the accuracy of such forecasts. For instance, a study enhances stock forecasting by integrating external sentiment indicators derived from online text with traditional stock index series, showing improved accuracy on the Shanghai Stock Exchange Composite Index [35].

The Modality-aware Transformer, introduced in Ref. [36], is a novel multimodal transformer-based model designed to utilize both textual and numerical data to enhance time series forecasting. The model employs specialized attention mechanisms to focus on the most relevant features across different modalities, resulting in improved prediction accuracy demonstrated through extensive experiments on financial datasets. However, this study primarily relies on sentiment-based text processing, limiting the ability to fully leverage the nuanced information present in unstructured text.

While existing approaches have demonstrated the value of incorporating textual data, they often simplify textual inputs to sentiment polarity or other low-dimensional representations, potentially overlooking critical information. This paper addresses these limitations by proposing a novel framework that directly integrates textual and time

series data using language models to extract high-dimensional features from text. Unlike sentiment-based methods, our approach comprehensively models the interactions between text and time series, capturing deeper insights into the factors influencing predictions.

3. Proposed Method

3.1. Problem Definition

The multivariate economic time series forecasting problem can be formulated as follows. Given historical time series data denoted as $Y_{1:L} = \{y_1^t, y_2^t, ..., y_c^t\}_{t=1}^{L} \in \mathbb{R}^{L \times c}$, where L is the fixed historical window, c ($c > 1$) is the number of variates, and y_i^t denotes the value of the i_{th} variate at t_{th} time step. The forecasting task aiming to figure out the predicted series $\hat{Y}_{L+1:L+H} = \{\hat{y}_1^t, \hat{y}_2^t, ..., \hat{y}_c^t\}_{t=L+1}^{L+H} \in \mathbb{R}^{H \times c}$, where c ($c > 1$) denotes the number of variates, $\hat{y}_i^t$ is the predicted result of the i_{th} variate at the time step t, and H ($H > 1$) denotes the number of forecasting time steps. The ground truth for the time period from $L + 1$ to $L + H$ is denoted as $Y_{L+1:L+H} = \{y_1^t, y_2^t, ..., y_c^t\}_{t=L+1}^{L+H} \in \mathbb{R}^{H \times c}$.

3.2. Proposed MM-iTransformer Framework

We propose MM-iTransformer, where MM stands for multimodal, to overcome the limitations of existing time series forecasting models. This framework integrates textual embeddings derived from financial news and numerical embeddings from historical prices, providing a robust cross-modal approach to enhance predictive performance in economic applications (see Figure 3).

FinBERT for Textual Embeddings: To incorporate rich semantic information from financial news, we utilize FinBERT [37], a language model fine-tuned specifically for financial text. Unlike traditional sentiment analysis approaches, FinBERT captures nuanced financial terminology and contextual insights from complete news articles.
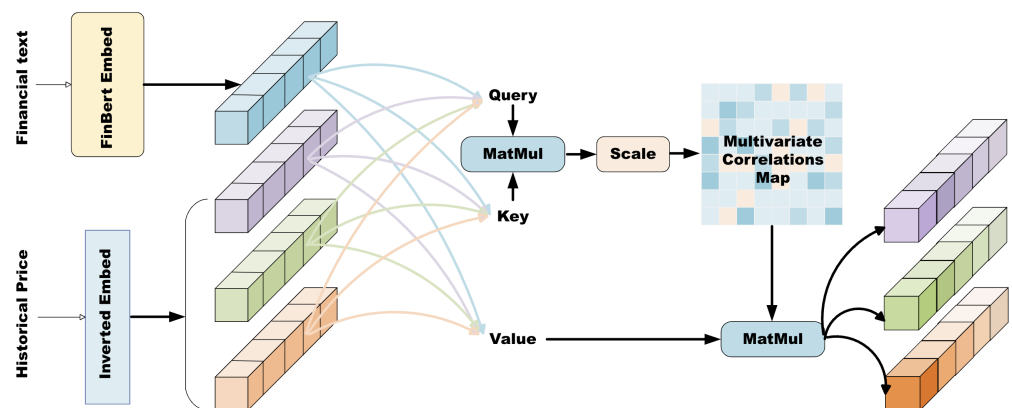

Figure 3. The workflow of MM-iTransformer: FinBERT Embed generates textual embeddings, Inverted Embed represents price data as variate tokens, and the Multivariate Correlations Map captures interactions for forecasting.

Given a set of historical news articles $x_1, x_2, ..., x_L$ within a historical window of length L, each article x_t, which is composed of a word sequence, is processed through FinBERT to obtain its contextualized embedding. Specifically, we extract the [CLS] token's last hidden state from FinBERT's output, resulting in a d-dimensional vector e_t that represents the overall context of the news at time t. This process yields a sequence of textual embeddings, as defined in Equation (1):

$$T_{1:L} = \{e_1, e_2, ..., e_L\} \in \mathbb{R}^{L \times d} \tag{1}$$

Each embedding e_t encapsulates the semantic and contextual information of the corresponding news article, effectively capturing critical financial insights that may influence market movements.

Integration with Time Series Data: The core of MM-iTransformer lies in effectively integrating the textual embeddings $T_{1:L}$ with the historical time series data. The historical price data $Y_{1:L}$ are embedded using Inverted Embed of iTransformer. Specifically, we describe these data as $Y_{1:L} = \{y_1, y_2, ..., y_L\} \in \mathbb{R}^{L \times c}$, where c is the number of numerical variables (e.g., different financial indicators). Our goal is to learn a function F that maps these inputs to future forecasts as shown in Equation (2):

$$\hat{Y}_{L+1:L+H} = F(Y_{1:L}, T_{1:L}) \tag{2}$$

Here, $\hat{Y}_{L+1:L+H} \in \mathbb{R}^{H \times c}$ represents the predicted future values over a forecast horizon *H*. To integrate textual and numerical data, we treat the textual embeddings as additional variate tokens, similar to how variables are represented in the inverted perspective of the iTransformer framework. This alignment allows us to leverage the attention mechanism to model cross-modal relationships effectively.

Cross-Attention Mechanism: The combined input consists of both numerical and textual embeddings concatenated along the variate (feature) dimension:

$$X = [Y_{1:L}; T_{1:L}] \tag{3}$$

In the cross-attention mechanism, we compute the query (*Q*), key (*K*), and value (*V*) matrices using learned weight matrices as shown in Equations (4), (5), and (6), respectively:

$$Q_h = XW_h^Q \tag{4}$$

$$K_h = XW_h^K \tag{5}$$

$$V_h = XW_h^V \tag{6}$$

Here, $W^Q, W^K \in \mathbb{R}^{(c+1) \times d_k}$ and $W^V \in \mathbb{R}^{(c+1) \times d_v}$ are learnable weight matrices for the queries, keys, and values, respectively. These matrices project the input features into appropriate representation spaces for attention computations. Note that d_k and d_v are distinct from the dimension d used for textual embeddings, as they specifically define the sizes of the query, key, and value projections within the attention mechanism:

- $c + 1$: The total number of variates, where c represents the numerical variables, and 1 represents the additional textual variable.
- d_k, d_v: The dimensionality of the projected query, key, and value vectors within the attention head h. These dimensions control the representation capacity of the attention mechanism, influencing how relationships between input features (both textual and numerical) are modeled and how information is aggregated into output representations.

The attention scores are computed using the scaled dot-product attention mechanism proposed in [38], which evaluates the relevance between queries and keys. This mechanism enables the model to weigh the importance of each feature (both textual and numerical) when making predictions, effectively capturing cross-modal interactions between financial news and historical price data.

Balanced Contribution of Modalities: To prevent richer textual embeddings from dominating the learning process and overshadowing numerical data, we implement a strategy to balance the contributions of both modalities.

- Selective Incorporation of Textual Data: We include only the most recent news articles in the textual embeddings. Given that the latest news tends to have the most significant impact on future market movements, this approach reduces the noise of older articles and focuses the model on the most relevant information.
- Dimension Alignment: A linear layer is applied to the textual embeddings, adjusting their feature dimension to align with that of the time series embeddings. This ensures that, although the number of variables may differ between modalities, they can be directly concatenated along the variate dimension without any dimensional mismatch.

The combined embeddings are then fed into the MM-iTransformer model, which processes them through multiple layers of attention and feed-forward networks to produce the final forecasts $\hat{Y}_{L+1:L+H}$. The details are outlined in Algorithm 1.

Algorithm 1 Modalities fusion under the MM-iTransformer framework.

Input: Time Series Data: $Y_{1:L} = \{y_1^t, y_2^t, \ldots, y_c^t\}_{t=1}^L$,
$\quad TextData : \{x_1, x_2, \ldots, x_L\}$
Output: Future Forecasting H steps:
$\quad \hat{Y}_{L+1:L+H} = \{\hat{y}_1^t, \hat{y}_2^t, \ldots, \hat{y}_c^t\}_{t=L+1}^{L+H}$
1: **Process Text Data:**
2: $\quad TextEmb \leftarrow$ FinBert.Embed(x_L).*last_hidden_state*
3: $\quad TextEmb \leftarrow$ Linear(d, L)(TextEmb)
4: **Process Time Series Data:**
5: **for** $c = 1$ to C **do**
6: $\quad TimeSeriesEmb[c] \leftarrow$ Inverted Embed(y_c)
7: **end for**
8: **Combine Embeddings:**
9: $CombinedFeatures \leftarrow$ Concatenate($TimeSeriesEmb, TextEmb$)
$\quad$ along the variate dimension
10: **Apply iTransformer Model:**
11: Load iTransformer model with appropriate configurations
12: **for** $t = L + 1$ to $L + H$ **do**
13: $\quad \hat{y}^t \leftarrow$ iTransformer.Projection(CombinedFeatures)
14: $\quad \hat{Y}[t - L] \leftarrow \hat{y}^t$
15: **end for**
16:
17: **return** $\hat{Y}_{L+1:L+H}$

---

4. Experiment

4.1. Datasets

In the economic domain, despite the abundance of news text and historical pricing data, there is a notable lack of publicly available datasets that directly correlate these two types of information, especially in analyzing the impact of textual news on economic indicators. To validate the ideas presented, we develop two specialized datasets: one focusing on *Gold-price* and the other on *foreign exchange (Forex)*.

4.1.1. Gold Price Forecasting Dataset

For the first dataset, we utilize a publicly available Kaggle repository entitled "The Effect of Economic News on Gold Prices" (https://www.kaggle.com/datasets/fekihmea/the-effect-of-economic-news-on-gold-prices/data (accessed on 23 January 2025)). This dataset comprises two primary components: (i) historical gold-price data spanning from January 2019 to December 2023, capturing daily market dynamics and highlighting the evolving trends that shape gold's valuation; and (ii) an economic calendar detailing significant events over the same period from key global participants—including Canada, Japan, the United States, Russia, the European Union, and China. By integrating these two data streams, the dataset provides a comprehensive perspective on how economic news from these influential regions influences the highly volatile gold market.

Key Variables in the Dataset: The primary structure of the dataset utilized in this study is as follows:

- Historical Gold Prices:
  - Date: The specific date corresponding to the recorded gold price.
  - Price: The daily closing price of gold.
- Economic Calendar Data:
  - Date: The exact date of the anticipated or actual economic event.
  - Country: The country associated with the economic event.
  - Event: A concise label for the specific economic indicator or event (e.g., inflation rate release and unemployment rate).
  - Actual: The real reported value of the economic indicator for the current event.
  - Prior: The most recently released value of the same economic indicator for comparison.

Textual Description Generation: The economic calendar data, illustrated in Table 1, contain various events, numerical values, and dates. However, these raw data formats are not directly usable by language models, which are designed to process natural language text. To make these data compatible, we generate textual descriptions for each economic event using a standardized template. These descriptions include the event's date, name, the reported value, and its deviation from prior or consensus values, enabling clearer interpretation by language models. For example, "On 01/01/19, the Caixin Manufacturing PMI for December was reported at 49.70, down from the previous value of 50.20, indicating a decrease of 0.50". This approach transforms tabular numerical data into concise, natural-language summaries, enhancing the dataset's utility for language-based analyses.

Table 1. Examples of economic calendar data.

Day	Country	Event	Current	Previous
01/01/19	CN	Caixin Manufacturing PMI DEC	49.7	50.2
01/04/19	US	Non Farm Payrolls DEC	312 K	176 K
01/07/19	CA	Ivey PMI s.a DEC	59.7	57.2

Data Alignment and "No News" Filling: To integrate textual descriptions with historical gold price data, we align each event-based description to the corresponding trading date in the price dataset. For dates without economic events, we fill the gap with the placeholder phrase "no news" to ensure temporal consistency. This approach guarantees that every date in the Gold-price dataset has a corresponding textual entry, maintaining completeness and coherence in the resulting data structure.

Resulting Samples: The integration of historical gold prices with textual economic event descriptions produces a dataset of 1352 samples. Each sample consists of the following:

- Date: The trading day associated with the gold price and any economic news.
- Gold Price: The recorded closing price of gold on the specified date.
- Economic News Description: A one-sentence summary of the economic event(s) released that day, or the placeholder phrase "no news" for days without scheduled announcements.

This curated dataset encapsulates the intricate interplay between global economic events and gold price fluctuations, offering a granular perspective on market behavior. By synthesizing structured numerical data with natural language descriptions, the dataset provides an invaluable resource for examining temporal patterns, identifying key correlations, and understanding the role of economic news in shaping financial market dynamics.

4.1.2. Foreign Exchange Forecasting Dataset

For the second dataset, we consolidate data from two distinct Kaggle sources. The first, titled "Economic Calendar Invest Forex" (https://www.kaggle.com/datasets/devorvant/economic-calendar (accessed on 23 January 2025)) serves as an extensive archive of significant economic events, news, and volatility assessments, focusing on the economies of developed nations, especially the United States, Britain, and Japan, due to their profound influence on global forex markets. This dataset includes critical economic indicators and historical data publications. The second source, "Foreign Exchange Rates 2000–2019", (https://www.kaggle.com/datasets/brunotly/foreign-exchange-rates-per-dollar-20002019 (accessed on 23 January 2025)) produced by the Federal Reserve's Data Download Program, features a comprehensive time series of foreign exchange rates per dollar for multiple countries.

Key Variables in the Dataset: The primary structure of the dataset utilized in this study is as follows:

- Foreign Exchange Data:

  - Date: The specific date when the foreign exchange rate was recorded.
  - Japan exchange: The exchange rate of the US dollar against the Japanese yen.
  - Euro exchange: The exchange rate of the US dollar against the Euro.
  - UK exchange: The exchange rate of the US dollar against the British pound.

- Economic Calendar Data:

  - Date: The exact date of the anticipated or actual economic event.
  - Country: The country associated with the economic event.
  - Volatility: The expected impact of the event on currency or indexes.
  - Event: A concise summary of the economic event.
  - Assessment: Evaluation of the event's outcome compared to forecasts—better, worse, or as expected.
  - Actual Data: The actual data reported for the event.
  - Prior Data: Previous data related to the event, with notes on any interim changes.

Textual Description Generation: Similar to the Gold-price dataset, the economic calendar data for the Forex dataset are converted into natural language summaries to make them suitable for language models. For example, "On 2011-01-03, in United States, the 'ISM Manufacturing Prices' was reported. The current value is 72.5, previously it was 69.5. The degree of volatility is Low Volatility Expected, showing a increased of 3.0, as it is considered 'Better Than Expected'". This method transforms raw numerical data into concise, contextual summaries that enhance the dataset's utility for language-based analyses.

Data Alignment and Focused Selection: To ensure consistency across datasets, the Forex data are aligned with the dates of economic events, using the same "no news" filling approach as in the Gold-price dataset to address dates without significant economic events. Given that the majority of economic events in the calendar dataset are pertinent to the United States, Britain, and Japan, our dataset exclusively incorporates exchange rate data for the euro, yen, and pound to enhance the dataset's relevance and ensure a high correlation with the included economic events. This targeted selection strategy helps to maintain a focused and relevant dataset for analyzing forex fluctuations.

Resulting Samples: The combination of foreign exchange rates with textual descriptions of economic events resulted in a dataset of 2347 samples. Each sample contains the following:

- Date: The trading day to which the forex rates and any economic news pertain.
- Forex Rates: The recorded exchange rates for the day.

- Economic News Description: A concise summary of the economic event(s) on that day, or the phrase "no news" for days without scheduled announcements.

This curated dataset facilitates an in-depth analysis of the relationship between major economic announcements and their impact on forex rates, offering insights into the dynamics of the foreign exchange market. By combining structured numerical data with descriptive text, the dataset supports a comprehensive approach to understanding the economic indicators that influence forex fluctuations, thereby enhancing our understanding of market behavior.

4.2. Evaluation Metrics

To assess the performance of the forecasting models, we adopt four evaluation metrics: Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). MSE focuses on larger prediction errors, making it sensitive to outliers. RMSE provides an interpretable error magnitude by aligning the scale of errors with the original data. MAE offers a balanced view by equally weighting errors of different magnitudes, while MAPE expresses errors as a percentage of the actual values, indicating the average deviation in percentage terms. These metrics are defined as follows:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{7}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{10}$$

where y_i and $\hat{y}_i$ denote the actual and predicted values, respectively, and n is the total number of samples.

4.3. Experimental Settings

The proposed framework is trained using the MSE loss function and optimized with ADAM [39], initialized with a learning rate of 1×10^{-4}. Considering the relatively small size of our datasets, we evaluate various forecast horizons—$\{4, 8, 12, 24, 36, 48, 60, 96\}$ days—to identify the most practical forecast horizon for economic forecasting. As illustrated in Figure 4, the 12-day forecast horizon is selected based on its balance between predictive accuracy and practical applicability. The time series historical window is set at 96 days, providing a comprehensive temporal context for generating accurate and actionable forecasts.

As described in Section 3.2, we incorporate only the most recent day's news as textual input. Aligning textual inputs with the full historical window would produce text embeddings with dimensions far exceeding those of the time series features, creating an imbalance between the two modalities and hindering effective fusion.
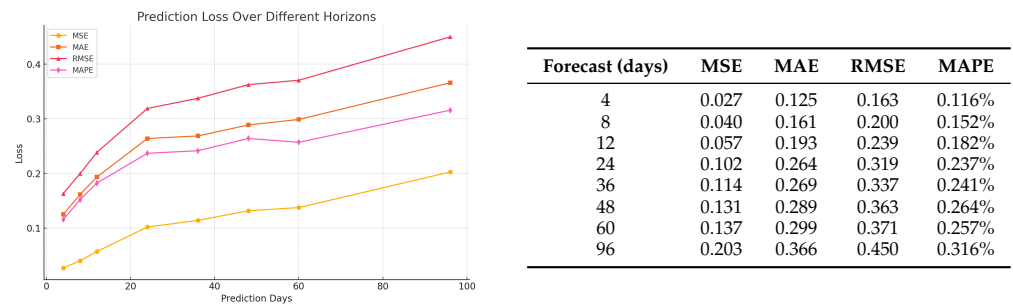
Prediction Loss Over Different Horizons

Forecast (days)	MSE	MAE	RMSE	MAPE
4	0.027	0.125	0.163	0.116%
8	0.040	0.161	0.200	0.152%
12	0.057	0.193	0.239	0.182%
24	0.102	0.264	0.319	0.237%
36	0.114	0.269	0.337	0.241%
48	0.131	0.289	0.363	0.264%
60	0.137	0.299	0.371	0.257%
96	0.203	0.366	0.450	0.316%

Figure 4. (**Left**): Line graph showing performance for the MM-iTransformer model using a historical window of $L = 96$ days across various forecast horizons ($T = 4, 8, 12, 24, 36, 48, 60, 96$ days). (**Right**): Detailed tabular representation of predictive performance metrics (MSE, MAE, RMSE, MAPE).

4.4. Baselines

To evaluate the effectiveness of the proposed MM-iTransformer, we compare its performance with the baseline iTransformer model. This analysis highlights the added value of incorporating textual information into economic forecasting through our multimodal approach.

In addition to the iTransformer and MM-iTransformer comparison, we employ several classical and baseline methods for further evaluation. First, we use the ARIMA model [40], a well-established approach for time series forecasting, to assess the advantages of transformer-based architectures over traditional statistical models. Additionally, we include the Random Walk Model [41], a widely used benchmark in financial forecasting, to establish a baseline and quantify the predictive improvements achieved by our proposed MM-iTransformer framework.

4.5. Experimental Results

This study compares the MM-iTransformer framework with the iTransformer, ARIMA, and the Random Walk Model using the Forex and Gold-price datasets. We primarily focus on the Mean Squared Error (MSE), as the models are trained using this metric. Additionally, we evaluate their predictive capabilities using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The results, presented in Table 2, show the following:

1. Comparison between MM-iTransformer and iTransformer: The MM-iTransformer integrates textual data with historical price data and demonstrates superior performance over the iTransformer. On the Forex dataset, it reduces the MSE from 0.022 to 0.019, showing a 13.64% reduction. For the Gold-price dataset, the MSE is reduced from 0.065 to 0.057, a reduction of 12.31%. These results highlight the significant benefits of incorporating textual data into forecasting models, enhancing the MM-iTransformer's accuracy and robustness.

2. Single-variable and Multi-variable Prediction Performance: MM-iTransformer excels in both single-variable (Gold-price) and multi-variable (Forex) predictions, affirming its capability to utilize textual data effectively. This inclusion provides additional predictive signals, enabling the model to capture market dynamics more accurately than using historical price data alone.

3. Comparison with Classical Models: Compared to classical models, both the MM-iTransformer and iTransformer exhibit superior performance, with MM-iTransformer consistently achieving the lowest MSE across datasets. However, the iTransformer does not consistently outperform the Random Walk Model on the Forex dataset, possibly due to difficulties in adapting to abrupt market changes—a scenario where the Random Walk Model, which assumes price continuity, may have an advantage.

Table 2. Performance comparison of different models on Forex and Gold-price datasets. The underlined values indicate the best performance for each metric.

Models	Forex				Gold-Price			
	MSE	MAE	RMSE	MAPE	MSE	MAE	RMSE	MAPE
MM-iTransformer (Ours)	0.019	0.094	0.137	0.160%	0.057	0.185	0.239	0.182%
iTransformer [32]	0.022	0.102	0.147	0.189%	0.065	0.207	0.254	0.196%
ARIMA [40]	0.030	0.092	0.174	0.859%	0.105	0.279	0.325	2.786%
Random Walk Model [41]	0.020	0.094	0.142	9.495%	0.110	0.288	0.331	19.727%

In summary, the MM-iTransformer framework marks a significant advancement in economic forecasting by leveraging multimodal data, particularly through the integration of textual information. This study confirms the potential of advanced deep learning techniques to revolutionize financial market predictions by incorporating diverse data types.

Ablation study: To assess the impact of temporally aligned news on prediction accuracy, we conduct an ablation study by randomizing the order of news texts within the time series window. This change disrupts the contextual relevance of the news data, simulating scenarios where the financial news is not aligned with historical market prices. By removing this alignment, we could isolate and examine the specific effect of the temporal relevance of textual inputs on forecasting performance.

As shown in Table 3, using randomized news texts shows a clear decline in performance compared to temporally aligned news. In the Gold-price dataset, the MSE increases to 0.069, and for the Forex dataset, it increases to 0.021 when random news is used. This indicates that non-contextual, misaligned news does not contribute meaningfully to prediction accuracy and may even reduce it. The observed rise in error rates underscores the critical role of temporal alignment, affirming that relevant, timely news data significantly enhance forecasting accuracy.

Table 3. Ablation study on the impact of textual data temporal alignment on forecasting accuracy.

	Forex				Gold-Price			
	MSE	MAE	RMSE	MAPE	MSE	MAE	RMSE	MAPE
With random text	0.021	0.099	0.146	0.201%	0.069	0.215	0.263	0.205%

Statistical Significance Analysis of Textual Data Integration: To rigorously ascertain the impact of integrating textual data into economic forecasting models, we utilize the Diebold–Mariano (DM) test to statistically evaluate the enhancements observed with the MM-iTransformer compared to the traditional iTransformer. The DM test results are summarized in Table 4, where the DM statistics for both the Gold-price and Forex datasets are presented.

Table 4. Diebold–Mariano (DM) test results comparing the forecasting performance of MM-iTransformer vs. iTransformer. Significance levels: * $p < 0.05$, ** $p < 0.01$.

Dataset	DM Test Statistic (*p*-Value)
Gold-price	2.030 (0.043 *)
Forex	3.010 (0.003 **)

The DM test results show statistically significant improvements in forecasting accuracy when textual data are integrated. Specifically, the Gold-price dataset yields a DM statistic of 2.030 with a *p*-value of 0.043, and the Forex dataset exhibits a DM statistic of 3.010 with a *p*-value of 0.003. These results indicate that the improvements in predictive performance

with the MM-iTransformer, which incorporates textual data, are statistically significant compared to the iTransformer, which relies solely on the historical price data.

These findings provide statistical evidence that supports the hypothesis that textual integration enhances the ability of forecasting models to capture complex market dynamics more effectively than models relying on time series data alone. This validation confirms the relevance and efficacy of incorporating textual data into economic forecasting models, reinforcing the value of multimodal approaches in enhancing prediction accuracy.

Analysis of Historical Window Lengths: To assess model performance stability, we experimented with various historical window lengths ranging from 36 to 384 days on the Gold-price and Forex datasets. The results indicate that shorter windows typically result in lower loss, with a 96-day window yielding optimal performance as shown in Figures 5 and 6. Extending the window beyond 96 days led to increased errors from outdated and less relevant data, which is detrimental for short-term forecasting.
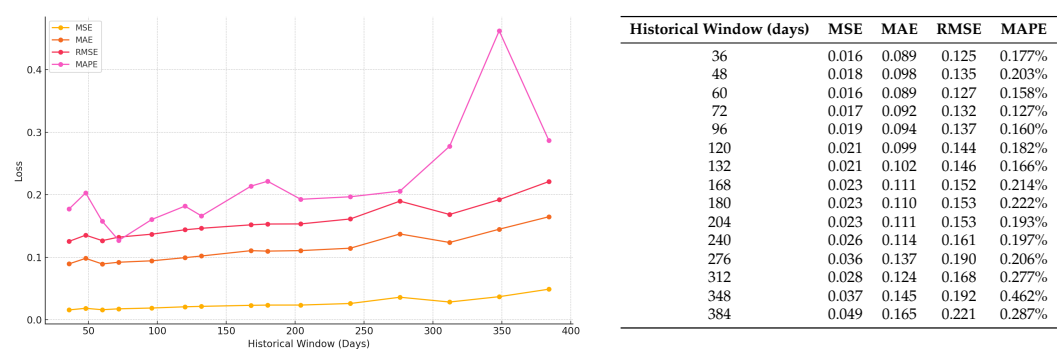

Historical Window (days)	MSE	MAE	RMSE	MAPE
36	0.016	0.089	0.125	0.177%
48	0.018	0.098	0.135	0.203%
60	0.016	0.089	0.127	0.158%
72	0.017	0.092	0.132	0.127%
96	0.019	0.094	0.137	0.160%
120	0.021	0.099	0.144	0.182%
132	0.021	0.102	0.146	0.166%
168	0.023	0.111	0.152	0.214%
180	0.023	0.110	0.153	0.222%
204	0.023	0.111	0.153	0.193%
240	0.026	0.114	0.161	0.197%
276	0.036	0.137	0.190	0.206%
312	0.028	0.124	0.168	0.277%
348	0.037	0.145	0.192	0.462%
384	0.049	0.165	0.221	0.287%

Figure 5. Comparison of model performance with different historical window lengths ($L = 36, 48, 60, 72, 96, \ldots, 384$) on the Forex for a 12-day forecast horizon ($H = 12$).
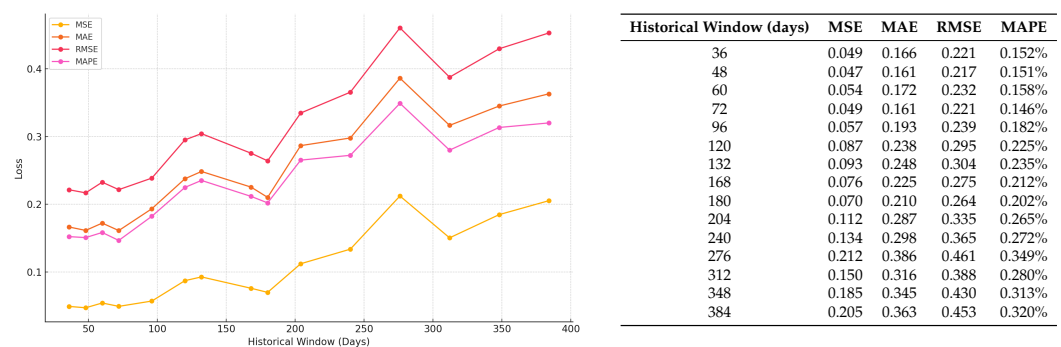

Historical Window (days)	MSE	MAE	RMSE	MAPE
36	0.049	0.166	0.221	0.152%
48	0.047	0.161	0.217	0.151%
60	0.054	0.172	0.232	0.158%
72	0.049	0.161	0.221	0.146%
96	0.057	0.193	0.239	0.182%
120	0.087	0.238	0.295	0.225%
132	0.093	0.248	0.304	0.235%
168	0.076	0.225	0.275	0.212%
180	0.070	0.210	0.264	0.202%
204	0.112	0.287	0.335	0.265%
240	0.134	0.298	0.365	0.272%
276	0.212	0.386	0.461	0.349%
312	0.150	0.316	0.388	0.280%
348	0.185	0.345	0.430	0.313%
384	0.205	0.363	0.453	0.320%

Figure 6. Comparison of model performance with different historical window lengths ($L = 36, 48, 60, 72, 96, \ldots, 384$) on the gold price for a 12-day forecast horizon ($H = 12$).

The sizes of our datasets—1352 samples for Gold-price and 2347 for Forex—also guided our choice of window length to avoid overfitting. Larger windows risk models memorizing specific data features instead of generalizing from underlying patterns, reducing their effectiveness on unseen data. Furthermore, we did not opt for smaller windows, as they could miss significant economic cycles and trends crucial for robust predictions. The chosen 96-day window provides a balanced approach, capturing essential market dynamics without the noise and overfitting associated with longer windows. This balance facilitates efficient learning and reliable forecasting, making it well suited for our dataset constraints.

Example Forecasting Results: Figure 7 presents a detailed comparison of forecasting outcomes from the MM-iTransformer and iTransformer models applied to Forex datasets. The graphs illustrate a marked improvement in prediction accuracy through the integration of textual data. Specifically, the MM-iTransformer, which incorporates textual

inputs alongside historical price data, demonstrates superior alignment with actual market trends and captures future price movements with significantly higher precision compared to iTransformer, which relies solely on historical price data. This visual representation compellingly demonstrates how integrating relevant textual information can substantially enhance a model's ability to comprehend and predict market dynamics, leading to more precise forecasts.
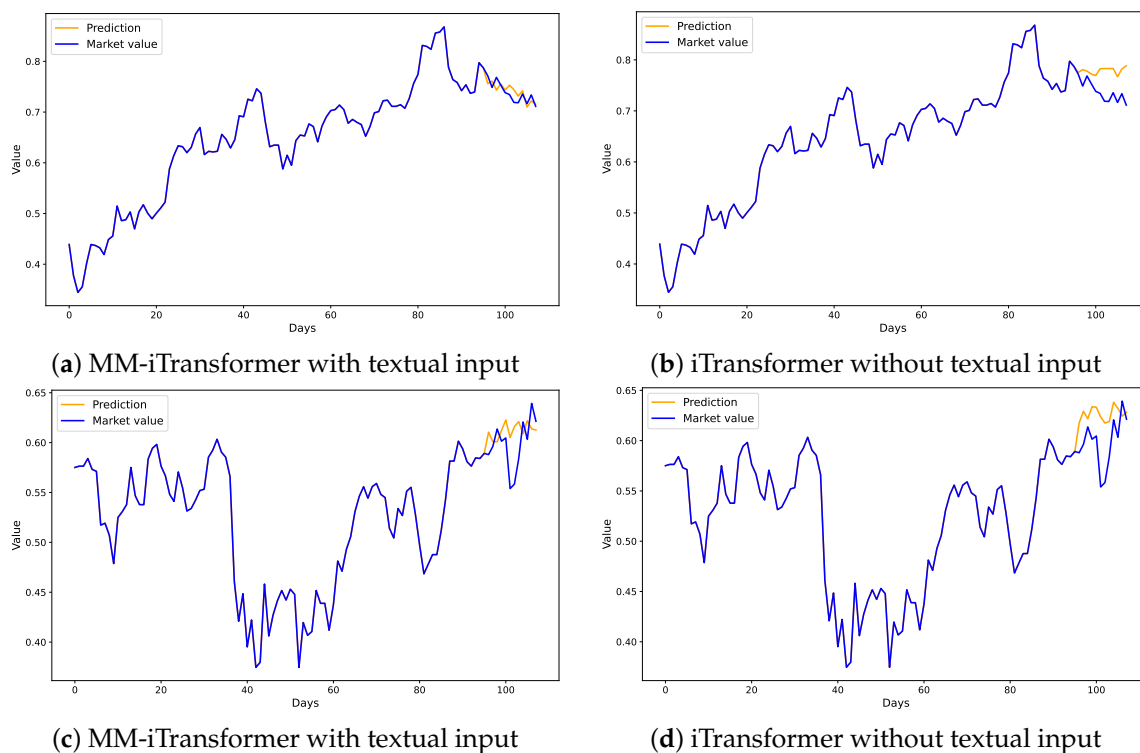

(**a**) MM-iTransformer with textual input   (**b**) iTransformer without textual input

(**c**) MM-iTransformer with textual input   (**d**) iTransformer without textual input

Figure 7. Comparison of predictions on the Forex dataset with a historical window of 96 and a forecast horizon of 12; the vertical axis represents series values, (**a**,**c**) showcase predictions from the MM-iTransformer, integrating textual data alongside historical price inputs, while (**b**,**d**) illustrate results from the iTransformer, which utilizes only historical price data.

4.6. Computational Complexity Analysis

In our research, the MM-iTransformer model integrates textual data with historical price data to improve the accuracy of the economic time series forecast. This integration introduces additional computational complexity primarily due to the use of pre-trained neural networks to generate text embeddings. Given the minimal computational demands of traditional models like ARIMA and the Random Walk Model, which do not use deep learning frameworks, our analysis focuses exclusively on comparing the MM-iTransformer with the original iTransformer.

Table 5 illustrates that MM-iTransformer exhibits slower training speeds and higher memory requirements than iTransformer. Specifically, on the Forex dataset, the MM-iTransformer processes at 232.89 milliseconds per iteration compared to 49.91 milliseconds for the iTransformer. Furthermore, the memory usage for the MM-iTransformer is 1.02 GB versus 0.59 GB for the iTransformer.

These findings underscore the increased computational costs associated with the MM-iTransformer. However, the significant improvement in prediction accuracy by effectively leveraging multimodal data justifies these costs. This capability to utilize diverse data types is crucial for accurately capturing complex market dynamics. Future efforts will aim to optimize the model's architecture to reduce computational demands without compromising

performance. Enhancing the efficiency of the model will broaden its applicability, making it more suitable for real-time trading systems and large-scale financial applications.

Table 5. Model efficiency comparison under input-96-predict-96 of weather and traffic.

Dataset	Forex		Gold-Price	
	Training Speed (ms/iter)	Memory Footprint	Training Speed (ms/iter)	Memory Footprint
MM-iTransformer	232.89 ms	1.02 GB	235.23 ms	1.02 GB
iTransformer	49.91 ms	0.59 GB	47.72 ms	0.44 GB

4.7. Cross-Modality Correlation Interpretation

We conduct detailed case studies on both the Gold-price dataset and the Forex dataset. These studies illustrate how our model, through its attention mechanism, adeptly captures and quantifies the interdependencies between textual embeddings and time series data.

As visualized in the attention heatmaps (Figure 8), the model utilizes the attention mechanism to pinpoint and quantify the interdependencies that exist between diverse data modalities. In the Gold-price dataset, the heat map reveals substantial attention weights on interactions between gold-related economic news and the price movements of gold itself. A notable correlation of 0.45 underscores the substantial influence of economic news on market prices as shown in Figure 8a. This high correlation coefficient indicates that our model not only recognizes relevant economic news but also assesses its impact on market fluctuations significantly. Similarly, the heat map for the Forex dataset shows focused attention on economic narratives associated with specific currency pairs. For instance, the attention weight of 0.31 for the GBP/USD exchange rate highlights the significant impact of economic narratives on currency valuation, as demonstrated in Figure 8b. This focused attention reflects the model's ability to discern and emphasize the economic indicators that are the most predictive of currency movements.
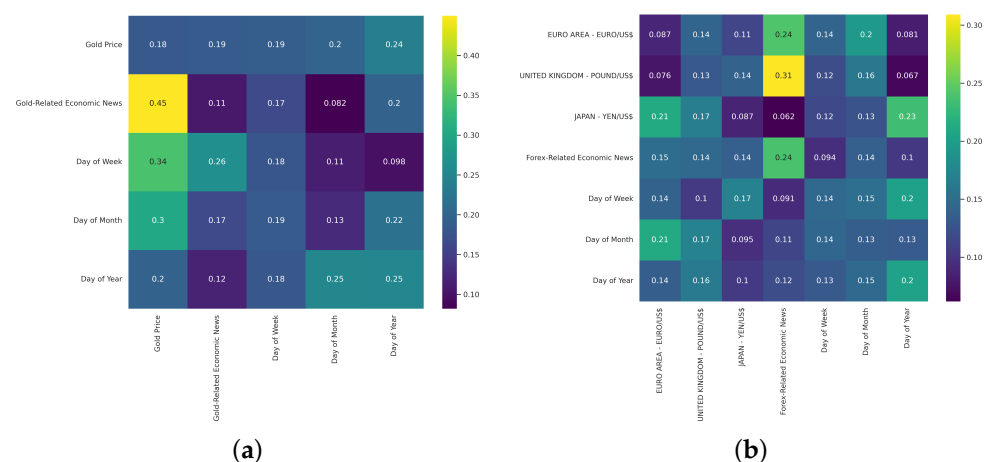

(a)   (b)

Figure 8. Comparison of attention heatmaps illustrating the interactions between textual and time series data. Values indicate the similarity between the high-dimensional representations of time series and textual data, where higher values suggest a stronger correlation. (**a**) Attention Heatmap of Head 0, Layer 1 on the Gold-price Dataset. Each row and column corresponds to features including gold price, gold-related economic news, and temporal features (Day of Week, Day of Month, Day of Year). (**b**) Attention Heatmap of Head 0, Layer 1 on the Forex Dataset. Features represented include three currency exchange rates (EURO/USD, GBP/USD, JPY/USD), forex-related economic news, and temporal features (Day of Week, Day of Month, Day of Year).

These findings provide quantitative evidence of meaningful patterns and relationships in different types of data. The ability of our model to discern these complex relationships not only demonstrates advanced multimodal comprehension capabilities but also significantly enhances the model's forecast performance. By effectively integrating and analyzing multimodal data, the model offers a robust tool to capture complex market dynamics that are often missed by more traditional models.

5. Conclusions

This study has successfully demonstrated the significant potential of integrating textual data into economic time series forecasting with the proposed MM-iTransformer. We have shown that incorporating textual information markedly enhances forecasting accuracy. Our findings underscore that the inclusion of news text not only aligns with the inherently multimodal nature of economic forecasting but also significantly improves predictive accuracy as evidenced by the notable reduction in MSE for both the Gold-price and Forex datasets.

However, we acknowledge certain limitations in this study, particularly concerning the scope and synchronization of the datasets used. While our results validate the utility of textual information for forecasting, the relatively limited scope of our datasets may not fully encompass the diverse and region-specific characteristics of global financial markets. Furthermore, the challenge of aligning textual data with time series data is exacerbated by the scarcity of large, well-structured, and temporally synchronized datasets in the financial sector.

To overcome these limitations, we plan to develop a larger, more comprehensive dataset in future work. This enhanced dataset will aim to better synchronize region-specific and real-time news data with financial time series, providing a more robust basis for evaluating and validating the proposed method. Through these efforts, we hope to further advance the field of multimodal economic forecasting by leveraging a broader spectrum of synchronized data, thereby enriching our understanding and predictive capabilities within this complex domain.

Abbreviations

The following abbreviations are used in this manuscript:

EMH	Efficient Market Hypothesis
ARIMA	Autoregressive Integrated Moving Average
SVM	Support Vector Machines
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
NLP	Natural Language Processing
ML	Machine Learning
CNN	Convolutional Neural Networks
DLSTM	Deep Long Short-Term Memory
SVR	Support Vector Regression
OLS	Ordinary Least Squares
MLP	Multi-Layer Perceptron
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
MAPE	Mean Absolute Percentage Error

References

1. Kelany, O.; Aly, S.; Ismail, M.A. Deep learning model for financial time series prediction. In Proceedings of the International Conference on Innovations in Information Technology (IIT), Al Ain, United Arab Emirates, 17–18 November 2020; pp. 120–125.
2. Alghamdi, T.; Mostafi, S.; Abdelkader, G.; Elgazzar, K. A comparative study on traffic modeling techniques for predicting and simulating traffic behavior. *Future Internet* **2022**, *14*, 294. [CrossRef]
3. Ma, K.; Pang, X.; Pang, H.; Lv, C.; Gao, T.; Chen, J.; Huo, X.; Cong, Q.; Jiang, M. A novel method for favorable zone prediction of conventional hydrocarbon accumulations based on RUSBoosted tree machine learning algorithm. *Appl. Energy* **2022**, *314*, 119983. [CrossRef]
4. Volkova, S.; Ayton, E.; Porterfield, K.; Corley, C.D. Forecasting influenza-like illness dynamics for military populations using neural networks and social media. *PLoS ONE* **2017**, *12*, e0188941. [CrossRef] [PubMed]
5. Fama, E.F. Efficient capital markets: A review of theory and empirical work. *J. Financ.* **1970**, *25*, 383–417. [CrossRef]
6. Phan, D.H.B.; Sharma, S.S.; Narayan, P.K. Stock return forecasting: Some new evidence. *Int. Rev. Financ. Anal.* **2015**, *40*, 38–51. [CrossRef]
7. Wang, J.J.; Wang, J.Z.; Zhang, Z.G.; Guo, S.P. Stock index forecasting based on a hybrid model. *Omega* **2012**, *40*, 758–766. [CrossRef]
8. Nava, N.; Matteo, T.D.; Aste, T. Financial time series forecasting using empirical mode decomposition and support vector regression. *Risks* **2018**, *6*, 7. [CrossRef]
9. He, H.; Gao, S.; Jin, T.; Sato, S.; Zhang, X. A seasonal-trend decomposition-based dendritic neuron model for financial time series prediction. *Appl. Soft Comput.* **2021**, *103*, 107488. [CrossRef]
10. Mondal, P.; Shit, L.; Goswami, S. Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. *Int. J. Comput. Sci. Eng. Appl.* **2014**, *4*, 13. [CrossRef]
11. Khan, S.; Alghulaiakh, H. ARIMA model for accurate time series stocks forecasting. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 524–528. [CrossRef]
12. Aneiros-Perez, G.; Vieu, P. Nonparametric time series prediction: A semi-functional partial linear modeling. *J. Multivar. Anal.* **2008**, *99*, 834–857. [CrossRef]
13. Bhardwaj, G.; Swanson, N.R. An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. *J. Econom.* **2006**, *131*, 539–578. [CrossRef]
14. Pai, P.F.; Lin, C.S. A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega* **2005**, *33*, 497–505. [CrossRef]
15. Widiputra, H.; Mailangkay, A.; Gautama, E. Multivariate CNN-LSTM Model for Multiple Parallel Financial Time-Series Prediction. *Complexity* **2021**, *2021*, 9903518. [CrossRef]
16. Vidal, A.; Kristjanpoller, W. Gold volatility prediction using a CNN-LSTM approach. *Expert Syst. Appl.* **2020**, *157*, 113481. [CrossRef]

17. Zeng, Z.; Kaur, R.; Siddagangappa, S.; Rahimi, S.; Balch, T.; Veloso, M. Financial time series forecasting using cnn and transformer. *arXiv* **2023**, arXiv:2304.04912.

18. Lezmi, E.; Xu, J. Time Series Forecasting with Transformer Models and Application to Asset Management. *SSRN Electron. J.* **2023**. Available online: https://ssrn.com/abstract=4375798 (accessed on 23 January 2025). [CrossRef]

19. Kloptchenko, A.; Eklund, T.; Karlsson, J.; Back, B.; Vanharanta, H.; Visa, A. Combining data and text mining techniques for analysing financial reports. *Intell. Syst. Account. Financ. Manag. Int. J.* **2004**, *12*, 29–41. [CrossRef]

20. Liapis, C.M.; Karanikola, A.; Kotsiantis, S. Investigating deep stock market forecasting with sentiment analysis. *Entropy* **2023**, *25*, 219. [CrossRef]

21. Farimani, S.A.; Jahan, M.V.; Fard, A.M.; Tabbakh, S.R.K. Investigating the informativeness of technical indicators and news sentiment in financial market price prediction. *Knowl.-Based Syst.* **2022**, *250*, 108742. [CrossRef]

22. Reis Filho, I.J.; Marcacini, R.M.; Rezende, S.O. On the enrichment of time series with textual data for forecasting agricultural commodity prices. *MethodsX* **2022**, *9*, 101758. [CrossRef] [PubMed]

23. Baranowski, P.; Bennani, H.; Doryń, W. Do the ECB's introductory statements help predict monetary policy? Evidence from a tone analysis. *Eur. J. Political Econ.* **2021**, *68*, 101964. [CrossRef]

24. Erfanian, S.; Zhou, Y.; Razzaq, A.; Abbas, A.; Safeer, A.A.; Li, T. Predicting bitcoin (BTC) price in the context of economic theories: A machine learning approach. *Entropy* **2022**, *24*, 1487. [CrossRef]

25. Alizadeh, M.; Rahimi, S.; Ma, J. A hybrid ARIMA–WNN approach to model vehicle operating behavior and detect unhealthy states. *Expert Syst. Appl.* **2022**, *188*, 116515. [CrossRef]

26. Munir, M.; Siddiqui, S.A.; Dengel, A.; Ahmed, S. DeepAnT: A deep learning approach for unsupervised anomaly detection in time series. *IEEE Access* **2018**, *6*, 1991–2005. [CrossRef]

27. Bandara, K.; Bergmeir, C.; Hewamalage, H. LSTM-MSNet: Leveraging forecasts on sets of related time series with multiple seasonal patterns. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 1586–1599. [CrossRef]

28. Sagheer, A.; Kotb, M. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* **2019**, *323*, 203–213. [CrossRef]

29. Kitaev, N.; Kaiser, Ł.; Levskaya, A. Reformer: The efficient transformer. *arXiv* **2020**, arXiv:2001.04451.

30. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *203*, 22419–22430.

31. Nie, Y.; Nguyen, N.H.; Sinthong, P.; Kalagnanam, J. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv* **2022**, arXiv:2211.14730.

32. Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; Long, M. itransformer: Inverted transformers are effective for time series forecasting. *arXiv* **2023**, arXiv:2310.06625.

33. Akita, R.; Yoshihara, A.; Matsubara, T.; Uehara, K. Deep learning for stock prediction using numerical and textual information. In Proceedings of the 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), Okayama, Japan, 6–10 June 2016; pp. 1–6.

34. Zhang, L.; Zhang, L.; Xiao, K.; Liu, Q. Forecasting price shocks with social attention and sentiment analysis. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, 18–21 August 2016; pp. 559–566.

35. Wang, Y. Stock market forecasting with financial micro-blog based on sentiment and time series analysis. *Shanghai Jiaotong Univ. (Sci.)* **2017**, *32*, 173–179. [CrossRef]

36. Emami, H.; Dang, X.H.; Shah, Y.; Zerfos, P. Modality-aware Transformer for Time series Forecasting. *arXiv* **2023**, arXiv:2310.01232.

37. Liu, Z.; Huang, D.; Huang, K.; Li, Z.; Zhao, J. Finbert: A pre-trained financial language representation model for financial text mining. In Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, Montreal, QC, Canada, 19–25 July 2021; pp. 4513–4519.

38. Vaswani, A. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.

39. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

40. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.

41. Hull, J.C.; Basu, S. *Options, Futures, and Other Derivatives*; Pearson Education: New Delhi, India, 2016.