Advancing Forex Prediction through Multimodal Text-Driven Model and Attention Mechanisms

Article *in* Intelligent Systems with Applications · April 2025

DOI: 10.1016/j.iswa.2025.200518

4 authors, including:

Fatima Dakalbab
University of Sharjah

Manar Abu Talib
University of Sharjah

Qassim MH Nasir
University of Sharjah

Intelligent Systems with Applications 26 (2025) 200518

Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.journals.elsevier.com/intelligent-systems-with-applications

Advancing Forex prediction through multimodal text-driven model and attention mechanisms

Fatima Dakalbab [a], Ayush Kumar [b], Manar Abu Talib [a,*], Qassim Nasir [c]

[a] Department of Computer Science, College of Computing and Informatics, University of Sharjah, Sharjah, United Arab Emirates
[b] School of Engineering and Information Technology, Manipal Academy of Higher Education, Dubai Campus, United Arab Emirates
[c] Department of Computer Engineering, College of Computing and Informatics, University of Sharjah, Sharjah, United Arab Emirates

ARTICLE INFO

Keywords:
Multimodality
Technical analysis
Sentiment analysis
Cross-modal
Attention-based fusion

ABSTRACT

The Forex market, characterized by high volatility and complexity, presents a significant challenge for accurate prediction of currency price movements. Traditional approaches often rely on either technical indicators or sentiment analysis, limiting their ability to capture the interplay between diverse data modalities. This research work introduces a novel multimodal deep learning framework that integrates technical analysis and sentiment analysis through a cross-modal attention mechanism, enabling a comprehensive understanding of market dynamics. The proposed model leverages innovative alignment techniques to synchronize sentiment from news articles with historical price trends, facilitating robust multiclass prediction of Forex price directions. To evaluate its effectiveness, the model was tested on three major currency pairs—EUR/USD, GBP/USD, and USD/JPY—using k-fold cross-validation. Multiple attention configurations, including no attention, self-attention, bi-cross attention, and a hybrid approach, were implemented to assess the impact of attention mechanisms on prediction performance. Experimental results highlight the superiority of the hybrid attention mechanism, which consistently outperformed single-modality models and other configurations across key metrics, such as Matthew's correlation coefficient, accuracy, directional accuracy, and F1-score. These findings underscore the importance of integrating sentiment and technical data for enhanced Forex prediction. This study contributes to the growing field of multimodal financial forecasting, offering a foundation for future research incorporating advanced risk metrics, real-time trading systems, and broader market applications.

1. Introduction

Financial markets can be classified as complex systems since they exhibit non-linear, non-stationary, and time-variant characteristics. Additionally, they are prone and exposed to various factors, including economic news, political events, and global impact (Leles et al., 2019). The use of Artificial Intelligence (AI) in financial markets has brought about significant advancements in the way financial transactions are carried out, greatly improving the effectiveness, safety, and customization of financial services. Financial technology, or FinTech, is a cutting-edge development in the field of finance that uses technology to address longstanding market issues. It encompasses automated trading, investments, insurance, and risk management (Gai et al., 2018).

The financial sector has recognized the growing usefulness of AI, especially in automating risk management, customizing investment strategies, and predicting market conditions (Ferreira et al., 2021). The utilization of AI in trading markets is wide-ranging and includes predicting future price trends or market movements, analyzing sentiment from social media and news, optimizing financial portfolios, and evaluating and reducing risks associated with financial assets (Ferreira et al., 2021; Dang, 2019). AI and machine learning techniques have become essential for predicting future pricing changes, underscoring their importance in algorithmic trading. Algorithmic trading, utilizing machine learning models, provides benefits such as improved transaction speed and elimination of emotional biases, hence enabling more logical and timely decision-making (Munkhdalai et al., 2019). However, despite the progress made, predicting future trends in financial markets is still difficult because of the unpredictable nature of market fluctuations, interference from irrelevant data, and vulnerability to sudden financial crises (Ferreira et al., 2021; Baek et al., 2020; Li et al., 2020).

Foreign Exchange, also referred to as Forex or FX, is a financial market where the buying and selling of currencies occur simultaneously.

* Corresponding author.
E-mail address: mtalib@sharjah.ac.ae (M.A. Talib).

https://doi.org/10.1016/j.iswa.2025.200518
Received 25 December 2024; Received in revised form 22 January 2025; Accepted 10 April 2025
Available online 11 April 2025

The Forex market is the largest financial market, holding a trading volume over $6 trillion (Chantarakasemchit et al., 2020). The market operates in a decentralized manner, functioning 24 h a day, conserve on weekends, which sets it apart from other marketplaces. Forex has distinct characteristics when compared to other markets. These variations provide benefits to Forex consumers for successful trading. These attributes include no commissions, absence of intermediaries, adjustable lot sizes, minimal transaction costs, significant liquidity, small margins and high leverage.

Financial time series in Forex are analyzed using three methods to anticipate future values: Sentiment Analysis (SA), Fundamental Analysis (FA), and Technical Analysis (TA). The SA utilizes economic news to examine the influence of sentiment on prices. The FA focuses on macroeconomic challenges, while the TA relies on previous data to predict future price movements. This research proposes an innovative method by investigating the development of a multimodal AI model specifically designed for the FOREX market. This model varies from conventional unidimensional methods by employing a multimodal attention approach that captures the dynamic interaction between technical analysis and sentiment analysis, thereby offering a comprehensive understanding of market movements.

The core of the multimodal attention technique resides in its capacity to simultaneously monitor market sentiments and technical indicators, identifying the complex connections between news sentiment and previous price movements. The thorough analysis allows the model to identify the optimal moments for purchasing or selling a certain currency pair, aiming to maximize profits and minimize risks. This research proposes that the use of multimodal attention technique represents a significant progress in the implementation of AI in financial markets. This research aims to examine the efficacy of this novel strategy in the FOREX market, an area that has not previously been investigated using this methodology. The results of this study aim to add to the broader discussion on AI in finance, showcasing the capability of multimodal attention models to improve and optimize trading techniques for maximum profitability.

Traders usually employ a variety of analytical techniques as the basis for their trading decisions. Nevertheless, the emotional strain and uncertainty that come with trading can make it hard to combine these different inputs. In order to tackle this issue, our proposed approach presents a novel multimodal model specifically built to combine and derive insights from different aspects of trading analysis.

The structure of this research project is as follows: after introducing AI in the FOREX Financial Market in Section 1, the background of Forex, trading analysis, neural networks, and multimodal and fusion techniques are explained in Section 2. The related work on the different FOREX-based prediction approaches is reviewed in Section 3. Section 4 discusses the research project methodology and the design of the attention-based model for directional price prediction. Section 5 provides the experiment and test results of implementing our proposed approach. Finally, the conclusion and future work are presented in Section 6.

2. Background

2.1. Forex market

The Foreign Exchange market, commonly referred to as FOREX or FX, ranks among the largest financial markets globally, with a daily average trading volume of $6.6 trillion (Chantarakasemchit et al., 2020). Forex is fundamentally decentralized, functioning across four primary time zones, which facilitates trading 24 h a day, five days a week, using multiple platforms (Li et al., 2015). Currency pairs, which denote the exchange rate between two currencies, are fundamental to Forex trading. These pairings are classified as major, minor, or exotic according to trade volumes and liquidity (Dautel et al., 2020). Major pairs include the US dollar and exhibit the highest liquidity, while minor pairs exclude the US dollar but remain actively traded. Exotic pairs involve emerging market currencies and often face lower liquidity and higher transaction costs (King et al., 2012).

Understanding essential Forex terminologies, such as base/quote currencies, bid/ask prices, spread, pip, and leverage, is crucial for operating this market. Table 1 summarizes these key terms. Additionally, candlestick charts, with their ability to visually represent price trends and market sentiment, are widely used for technical analysis. Bullish (green) and bearish (red) candles depict price movements, aiding traders in identifying patterns and forecasting trends (Faulkner, 2017).

2.2. Trading analysis techniques

Traders use a wide variety of trade analyses to study patterns and predict market movements. Depending on their trading strategy, traders may use one or more of these analysis techniques. Fig. 1 (Leles et al., 2019) shows that these types can be roughly grouped into three groups: sentiment analysis, fundamental analysis, and technical analysis.

- **Sentiment Analysis:** evaluates market sentiment using data from news, reports, and social media, helping traders anticipate market reactions (Yang et al., 2018).
- **Fundamental Analysis:** examines macroeconomic factors, including GDP, interest rates, and political events, to assess a currency's intrinsic value and its potential impact on exchange rates (Semiromi et al., 2020).
- **Technical Analysis:** focuses on identifying price patterns and trends using statistical indicators like Moving Average (MA), Relative Strength Index (RSI), and Bollinger Bands. These tools help traders

Table 1
FOREX fundamental terminologies (Yıldırım et al., 2021).

Terminology	Meaning
Base/Quote Currency	The first currency in a pair is the base currency, which is additionally referred to as the transaction currency. The second currency in the currency pair is the quote currency. For example, in the EUR/USD pair, the base is EUR, and the quote is USD.
Going Long/Short	The long action refers to purchasing the base currency or selling the quote currency. Going short involves selling the base currency or purchasing the quote currency.
Bid Price/Ask Price	The ask price is the price at which the base currency can be bought, and the bid price is the price at which it can be sold.
Spread	The difference between the ask and bid prices is the spread. A lowered spread enables the trader to generate profits from even the smallest fluctuations in price. The value of spreads is influenced by market volatility and liquidity.
Pip/pipette	The pip (percentage of point) represents the smallest change in currency. Pips are four decimal points (0.0001) of a currency. The pipette is a five-decimal point fraction (0.00001). In another way, one pip is equal to ten pipettes.
Leverage	It involves performing financial transactions with borrowed funds. A leverage of 1:200 means that if you open a position of volume 1, the actual transaction volume will be 200. Leverage can result in gains or losses of up to 200 times the original volume.
Margin	Consists of funds borrowed by the speculator from the broker in order to facilitate leveraged investments. This allows for greater gains or losses. A margin call is a rule established by a broker to limit a trader's losses in a leveraged position, protecting the company's funds.
Stop Loss	A stop loss is a trader's instruction to sell currency at a predetermined price. This term safeguards the trader against significant losses.
Take Profit	Take Profit instructs a trader to close an open position at a predetermined value to realize a profit. This term ensures that the trader will earn a profit irrespective of market price fluctuations.
Volume	The amount of an asset that were bought and sold in a market during a certain period.

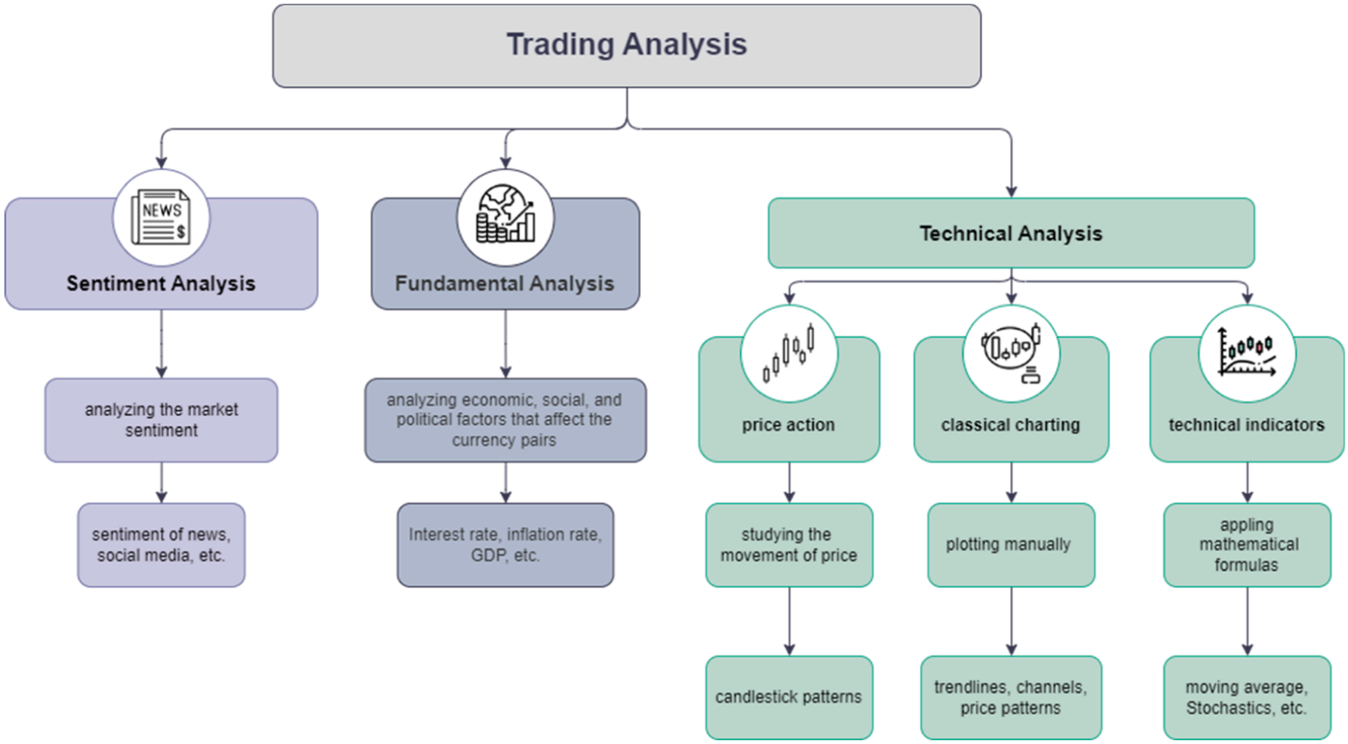

Fig. 1. Trading analysis types.

make data-driven decisions based on past market behavior (Dymova et al., 2016).

Table 2
Comparison between trading analysis techniques.

	Sentiment Analysis	Fundamental Analysis	Technical Analysis
Objective	Evaluating the market sentiment towards a currency pair to forecast its possible price fluctuations	Evaluating the asset's intrinsic value to determine if it is over- or undervalued	Determining optimal moments to enter, buy, or sell in the market
Concept	Conducting an analysis of opinions, news, reports, and social media to assess the general sentiment of the public towards a particular asset.	Focusing on market sentiment and economic factors	Using statistical tools to examine price patterns to anticipate movement
Dependency	News articles, investor opinions, social media, analyst reports, and market commentary	Based on economic data and financial reports	Involves charts (e. g., candlesticks) and technical indicators
Function	Trading, especially useful for predicting short-term market reactions and trends	Useful for trading and investment decisions	Mainly supports trading strategies
Usage	Analyzing massive amounts of qualitative data from diverse sources to extract sentiment.	Requires a thorough analysis of historical and current economic data	Utilizes past price data to form trading decisions
Time Frame	Depending on the source and type of sentiment analyzed, it can be used for long- and short-term trading.	Generally used for long-term trading strategies	Commonly applied for short-term trading

Table 2 contrasts the three main trading methods, sentiment analysis, fundamental analysis, and technical analysis. Sentiment analysis type analyzes qualitative data like news articles and social media to determine market sentiment toward an asset. Each analysis has various goals, concepts, dependencies, functions, usage, and time frames to suit different trading styles. Fundamental analysis analyzes economic indicators and financial documents to determine an asset's inherent worth to determine if it's undervalued or overvalued. Technical analysis uses charts and statistical indicators to anticipate price moves based on prior market activity.

2.3. *Artificial neural network*

Artificial Neural Networks (ANNs) simulate the human brain's ability to learn from data, making them effective for pattern recognition and predictive modeling. Modern advancements, like attention mechanisms, enhance neural networks by enabling them to focus on key data features. Self-attention captures dependencies within a single modality, while cross-attention aligns features across multiple modalities, such as price data and textual sentiment. These mechanisms have revolutionized tasks like sentiment analysis and market prediction by integrating diverse data sources (Vaswani et al., 2017).

2.4. *Multimodal and fusion techniques*

Multimodal integration is the processing and correlation of knowledge from several sources of data or modalities. This method is vital for tasks when each modality offers complementing information necessary for task accomplishment. The integration of several modalities, including text, image, and audio, presents issues in alignment and representation due to the unique characteristics and semantic spaces of each data type. Multimodal learning models demonstrate enhanced performance compared to unimodal techniques across many applications, utilizing the depth of integrated data to attain more precise predictions and comprehensive representations (Ramachandram and Taylor, 2017). Fig. 2 displays several fusion strategies frequently utilized

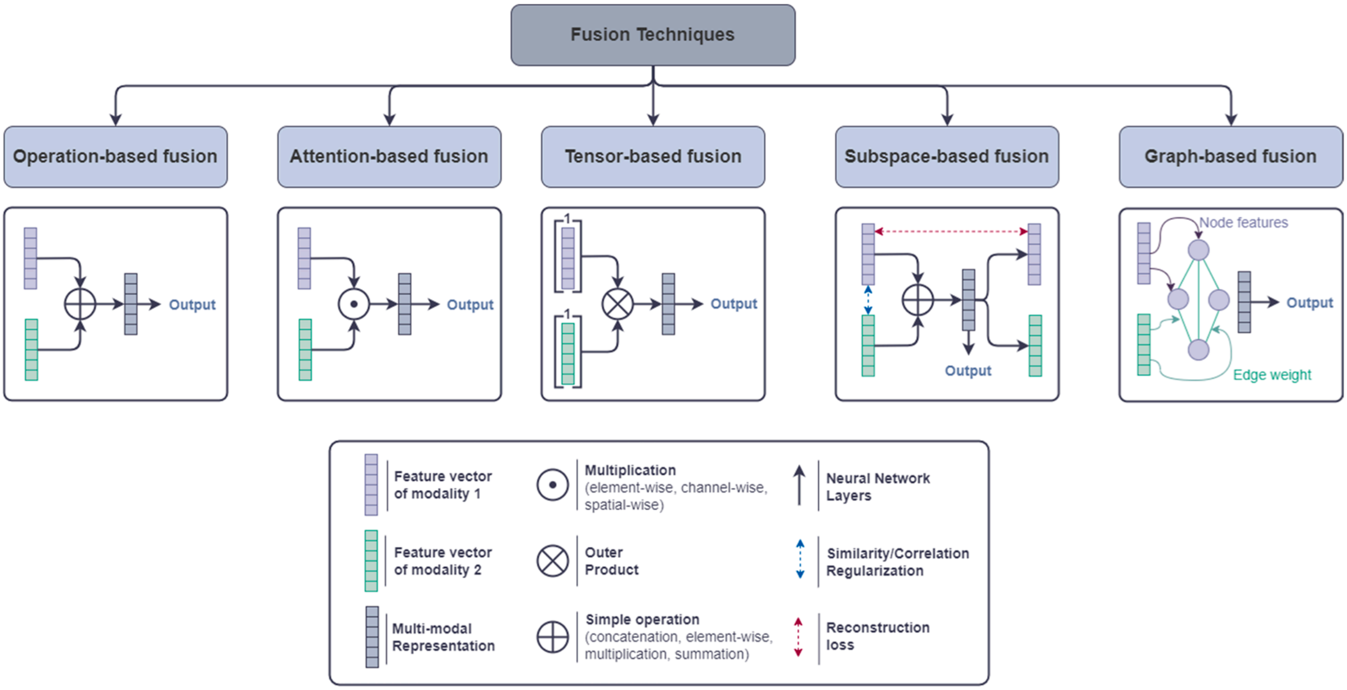

Fig. 2. Fusion techniques and approaches.

in multimodal models, encompassing operation-based fusion, attention-based fusion, tensor-based fusion, subspace-based fusion, and graph-based fusion (Cui et al., 2023). Each technique uniquely addresses the integration of modalities, seeking to capture significant interactions among data kinds. The following is a concise description of each method:

- **Operation-based Fusion**: This approach integrates feature vectors from many modalities using basic mathematical operations, such as concatenation, element-wise addition, or multiplication. The model incorporates information using fundamental operations, avoiding the introduction of extra learnable parameters, hence enhancing computing efficiency. Nevertheless, it may lack the sophistication required to clarify complex relationships between modalities, hence constraining its effectiveness in certain instances. Operation-based fusion has been extensively utilized in initial multimodal research due to its efficiency and rapidity (Chen et al., 2023).
- **Attention-based Fusion:** Attention mechanisms can prioritize features across modalities, enabling the model to concentrate on the most important data for a certain task. In attention-based fusion, weights are allocated to features from each modality according to their significance, dynamically modifying the emphasis as required. This method has demonstrated effectiveness in contexts where information relevance fluctuates, such as in multimodal sentiment analysis and image-text alignment tasks (Dammu et al., 2023).
- **Tensor-based Fusion:** Tensor-based fusion entails the creation of high-order interactions among features from several modalities through the computation of an outer product. This produces a more complex representation that encapsulates interactions across modalities at several levels. This method can augment the model's expressive ability, but it frequently experiences substantial computing expenses and might require dimensionality reduction strategies to address the increased complexity. Li et al. (2020) introduced the LSTM model with tensors representing events in finance to forecast stock movements utilizing fundamental and news inputs. Their methodology encompasses cross-modal interactions and tackles data heterogeneity challenges, including variations in sampling durations between continuous fundamental data and periodic news events.
- **Subspace-based Fusion:** This method maps features from each modality into a common subspace, facilitating closer alignment of analogous traits across multiple modalities. Subspace-based fusion seeks to reduce the dimensional differences among modalities and enhance alignment, hence promoting effective cross-modal learning. This method frequently employs similarity or correlation regularization to maintain significant links between modalities in the shared space, as demonstrated in multimodal learning applications for cross-modal retrieval and representation learning (Antelmi et al., 2019).
- **Graph-based Fusion:** Graph-based fusion denotes each modality as nodes inside a graph, with edges illustrating the relationships among features. This framework enables the model to acquire knowledge of both intra-modality interactions and inter-modal interdependence through the dynamic adjustment of edge weights. Graph-based fusion is especially advantageous for data characterized by relational structures, such as social network analysis and sensor data integration, where the interdependencies among modalities are essential for precise representation (Bintsi et al., 2023).

Each fusion technique presents distinct advantages, depending upon the characteristics of the task and data. Choosing a suitable fusion approach is essential for properly capturing complementary information between modalities, as evidenced by experiments employing the Fusilli library for multimodal applications (Townend et al., 2024).

3. Literature review

The integration of various trading analysis methodologies, including technical analysis, fundamental factors, and sentiment analysis, is emerging as a novel strategy for Forex prediction. This integration aims to utilize the advantages of each modality to enhance predictive precision and resilience in the complex Forex market. Numerous studies have sought to clarify this interaction by integrating sentiment data with traditional price patterns and technical indicators. For instance, Farimani et al. (2021, 2022, 2024) have made significant contributions to this field through three notable studies. These studies investigate diverse methodologies for integrating sentiment and technical indicators to improve forecasting abilities, highlighting the increasing recognition of sentiment's role as a complement to traditional trading metrics. Their initial work (Farimani et al., 2022), investigates a model that integrates sentiment data from financial news with technical indicators, including the RSI, BBs, and MAs. The authors employed an LSTM architecture to integrate these two modalities, illustrating how the connection between sentiment and technical metrics can enhance Forex price predictions, thereby underscoring the significance of sentiment as a crucial input alongside traditional trading indicator.

Expanding upon this foundation, Farimani et al.'s subsequent research (Farimani et al., 2021) presents a more advanced approach for integrating economic significance into sentiment data. The authors utilized the bag-of-economic-concepts methodology, aggregating BERT embeddings of news data into underlying economic concepts. A recurrent convolutional neural network subsequently analyzes these latent concepts and sentiment feature sequences. Their model additionally incorporates LSTM layers tailored for temporal patterns: one LSTM layer manages sequences related to sentiment and latent economic data, while another LSTM layer analyzes technical indicators, such as EMA, BB-Mean, OBV, Volume, and ATR. This dual-layered architecture allows the model to comprehend complex interdependencies between technical and sentiment data for forecasting currency pair fluctuations, including GBP/USD, BTC/USD, USD/JPY, and. EUR/USD. In their latest study (Farimani et al., 2024), Farimani et al. enhance their methodology by highlighting the integration and alignment of sentiment and technical modalities via an advanced, recurrent neural network model. This research presents a multimodal fusion method that adaptively learns from sentiment signals and technical indicators, thereby identifying complex patterns in Forex price movements. This study emphasizes that integrating these data sources significantly improves market prediction accuracy by combining sentiment from news sources with technical indicators, providing a comprehensive framework for effective financial forecasting.

In addition to Farimani's contributions, other research has further enhanced the integration of sentiment analysis and technical indicators in Forex forecasting. For instance, Semiromi et al. (2020) integrate news events from the economic calendar with various technical indicators, such as SMA, MACD, RSI, and Bollinger Bands. This model strengthens sentiment scores from news outlets, including Bloomberg and ForexFactory, and employs machine learning algorithms such as XGBoost, Random Forest, and SVM for classification purposes. The outcomes for currency pairs such as EUR/USD and GBP/USD exhibited commendable accuracy, with 63 % and 66 % classification rates.

In a study on EUR/USD Forex rate prediction utilizing the SMA technique and financial variables (Chantarakasemchit et al., 2020), the authors investigate a hybrid model that integrates SMA with fundamental economic indicators, including the dollar Index, inflation rates, and GDP. This methodology highlights the complementary connection between technical and fundamental data, employing models such as Multilayer Perceptron (MLP) and Linear Regression to forecast price movements. The findings demonstrate the efficiency of integrating technical indicators and economic factors into a unified dataset, resulting in insignificant error with linear regression (MSE: 0.000000003152 for combined data) and illustrating how the integration of diverse market data dimensions can enhance predictive precision.

Another integration between the technical indicators and fundamental data is presented by Yıldırım et al. (2021). The authors further examine the integration by utilizing Long Short-Term Memory (LSTM) networks to simulate technical indicators, including MACD, RSI, and CCI, as well as macroeconomic variables such as inflation and interest rates. The hybrid LSTM model, integrated with a rule-based decision framework, attains a notable profit accuracy of 73.09 % on currency pairs such as EUR/USD, illustrating the capability of LSTMs to effectively capture both technical patterns and economic fluctuations in a dynamic market.

In the field of deep learning, (Pornwattanavichai et al., 2022) utilizes an advanced cascading model that leverages BERT embeddings to analyze essential data (including FFD, GDP, DAX, PPI, CPI, Initial Unemployment Claim, Retail Sales, Unemployment rate, S&P 500, Inflation rate, interest rate) and technical indicators. This hybrid methodology integrates LSTM and ANN models to encapsulate economic sentiment and market trends. The model's efficiency underscores the increasing trend of employing BERT to analyze financial news sentiment and traditional technical analysis methods for enhanced Forex forecasting.

The Windsor and Cao (2022) investigates multimodal learning through a fusion-based LSTM model that analyzes sentiment from social media platforms, including Twitter and Sina microblogs, when combined with technical indicators like EMA and MACD. This study introduces a novel intermediate fusion methodology wherein distinct LSTM modules handle sentiment and technical data before integration. This parallel processing framework enables the model to identify the complex relationships between sentiment data and market trends, enhancing forecast accuracy for the USD/CNY exchange rate.

Nevertheless, several research has prioritized sentiment analysis while neglecting the importance of effective data alignment or fusion, which is essential for a comprehensive understanding of multimodal forecasting. For instance, Hajek and Novotny (2022) introduces a sentiment-focused model employing FinBERT to extract sentiment from internet news articles. This model incorporates certain technical indications, although it mainly emphasizes sentiment, with minimal consideration for fusion techniques. The lack of modality fusion constrains the model's capacity to leverage the relationship between sentiment and technical indicators.

An early-stage version of this work, including preliminary literature review findings, was presented at the International Conference on Innovations in Computing Research (ICR'23) (Dakalbab et al., 2023). However, the current paper represents a significant advancement beyond the conference proceedings. The literature review has been extensively revised and expanded to include recent studies and insights. As presented in Table 3 substantial efforts have been directed toward developing Forex prediction systems that incorporate multiple trading analysis techniques. However, to the best of our knowledge, no prior work has explored a multi cross-modal deep learning framework for Forex price prediction.

This study aims to advance the understanding and forecasting of financial market movements, with a focus on currency pairs including EUR/USD, GBP/USD, and EUR/JPY. In Forex trading, decisions are often influenced by a combination of historical price trends and the psychological or emotional impact of news events. Traders, however, face several critical challenges:

1. **Information Overload:** The vast volume of financial data and news can overwhelm traders, complicating the decision-making process.
2. **Isolated Analysis:** Existing systems often evaluate market data and news sentiment independently, limiting their capacity to capture the complex interdependencies between these modalities. While some methods attempt to merge these modalities into a single compound vector, they overlook critical interactions between the two.
3. **Data Alignment:** Integrating numerical market data with textual news data presents significant challenges. Market prices are consistently captured at specific time intervals, creating a structured timeline, whereas news is produced irregularly. This inconsistency complicates the development of analytical frameworks that accurately represent the timing and impact of news on market dynamics.

While fundamental data (e.g., GDP, inflation, interest rates) is important in long-term Forex forecasts, this study focuses on short-term forecasting powers, which are influenced more directly by market

Table 3
FOREX prediction using multiple trading analysis techniques.

Refs.	Trading Analysis	Fusion Analysis	Asset Pairs	Method	Problem Type
(Farimani et al., 2022)	TA ✓ FA × SA ✓	**Feature concatenation**: Sentiment features from FinBERT combined with market data into a compound matrix.	USD/JPY, BTC/USD, GBP/USD, EUR/USD	Fin BERT-based Sentiment and Informative Market Feature, LSTM	Predicting Price (Regression)
(Pornwattanavichai et al., 2022)	TA ✓ FA ✓ SA ×	**Hierarchical fusion:** Fundamental data processed with BERT (Stage 1) fused with technical indicators via autoencoder and further refined with BERT (Stage 2).	USD/EUR	LSTM, ANN	Predicting Price (Regression)
(Chantarakasemchit et al., 2020)	TA ✓ FA ✓ SA ×	**Unified dataset fusion:** Combines technical and fundamental data into a single dataset for joint analysis.	EUR/USD	MLP, Linear Regression	Predicting Price (Regression)
(Windsor and Cao, 2022)	TA ✓ FA × SA ✓	**Intermediate fusion**: Outputs of separate LSTM models (sentiment and market data) merged via an intermediate fusion layer.	USD/CNY	fusion-based LSTM (MF-LSTM)	Predicting Price (Regression)
(Farimani et al., 2021)	TA ✓ FA × SA ✓	**Linear activation fusion:** Combines outputs from sentiment-based and technical-data-based LSTMs through a linear activation layer.	USD/JPY, BTC/USD, GBP/USD, EUR/USD	BERT-BoEC RCNN, LSTMs	Predicting Price (Regression)
(Farimani et al., 2024)	TA ✓ FA × SA ✓	**Fusion layer integration**: Sentiment and technical indicators processed by recurrent layers and integrated through a dedicated fusion layer.	USD/JPY, BTC/USD, GBP/USD, EUR/USD XAU/USD	RCNN and other recurrent layers	Predicting Price (Regression)
(Semiromi et al., 2020)	TA ✓ FA × SA ✓	**Unified dataset fusion:** Combines sentiment and market data into a single dataset for analysis.	EUR/USD, USD/CHF, GBP/USD, USD/JPY	Random Forest, SVM, XGB	Price movement (Classification)
(Yıldırım et al., 2021)	TA ✓ FA ✓ SA ×	**Rule-based decision fusion:** Fundamental and technical indicators combined through rule-based mechanisms to guide predictions.	EUR/USD	LSTM	Price movement (Classification)

(continued on next page)

Table 3 (*continued*)

Refs.	Trading Analysis		Fusion Analysis	Asset Pairs	Method	Problem Type
(Nassirtoussi et al., 2015)	TA FA SA	✓ × ✓	**Rule-based decision fusion:** Fundamental and technical indicators combined through rule-based mechanisms to guide predictions.	EUR/USD	Multi-layer Dimension Reduction Algorithm	Price movement (Classification)
(Hajek and Novotny, 2022)	TA FA SA	✓ × ✓	**Rule-based decision fusion:** Fundamental and technical indicators combined through rule-based mechanisms to guide predictions.	EUR/USD	SVM, MLP, Random Forest	Price movement (Classification)
Ours	TA FA SA	✓ × ✓	**Cross-attention mechanism integration**: Aligns sentiment from news and technical data through cross-attention to capture dynamic relationships.	EUR/USD, GBP/USD. EUR/JPY	Deep Learning neural network	Price movement (multi-Classification)

sentiment and price changes. Incorporating fundamental analysis requires additional resources, such as organized datasets and processing overhead, which may impair the responsiveness of short-term predictive models. As a first stage, this study focuses on proving the effectiveness of the attention mechanism within a sentiment-technical analysis framework before expanding the model to include fundamental aspects. This research introduces a novel method for integrating and evaluating market data and news events, yielding a more thorough and precise comprehension of currency price movements. It aims to improve trading decisions by providing more insightful and less anxiety-inducing information.

This research's contributions include the development and implementation of a multimodal model that combines sentiment analysis and technical analysis. The suggested model uses a cross-attention mechanism for fusion, which is a novel way for predicting Forex prices that has not before been examined. The model carries out the following main functions:

1. **News Sentiment Analysis:** This research examines the sentiment in news articles and headlines associated with three major currency pairs: EUR/USD, EUR/JPY, and GBP/USD, providing valuable insights into how news sentiment can reflect market trends.
2. **Historical Price Data Analysis**: The model analyzes historical price data for the specified currency pairs by initially preprocessing the data, including technical indicators and log returns. Diverse labeling strategies are subsequently employed to categorize the market trend. This method aids in recognizing patterns and trends crucial for comprehending historical trading and establishes a robust basis for forecasting future price fluctuations.
3. **Data Alignment Algorithm:** The research proposes an algorithm to correlate news sentiment with historical price data. This procedure entails aligning both datasets' date formats and sentiment from news articles with the nearest relevant price data. In date mismatches, sentiment is handled specifically to ensure consistent alignment, facilitating the precise integration of both modalities for effective forecasting.
4. **Cross-Attention Integration:** The research employs cross-attention mechanisms to integrate sentiment and technical analyses. The model effectively comprehends and acquires knowledge about the dynamic relationship between sentiment obtained from news articles and technical indicators derived from historical price movements using a cross-attention mechanism. This integration improves the model's ability to accurately determine the price movement direction in the upcoming day. Furthermore, we compare the multimodal approach with singular module approaches and implement directional accuracy, a metric that enhances evaluation, to illustrate the superiority of the multimodal method.

4. Methodology

In this section, we detail the methodology for our approach. The process pipeline, from data collection to model evaluation and refining is illustrated Fig. 3. In the following sections, we will provide a comprehensive explanation of each phase, starting with data preprocessing and progressing to a detailed analysis of each modality and its alignment. Next, we examine the proposed multimodal model, which encompasses its individual modules, the fine-tuning process, and the evaluation metrics employed.

4.1. Dataset collection

During the data collection phase, this research expands on a corpus of financial news and market data associated with major currency pairs gathered from prior studies (Farimani et al., 2022). The major currency pairs consist of EUR/USD, GBP/USD, and USD/JPY. The researchers of that study gathered a significant dataset of news articles specifically related to these currency pairs. The two types of datasets are described as follows, and their statistics are presented in Fig. 4.

Market dataset: The authors obtained market data from Finhub spanning from September 2018 to December 2020. The Finhub REST API was utilized for scraping market data. Finhub extracts FOREX data from the FXCM broker using web scraping techniques. Every 60 min, a

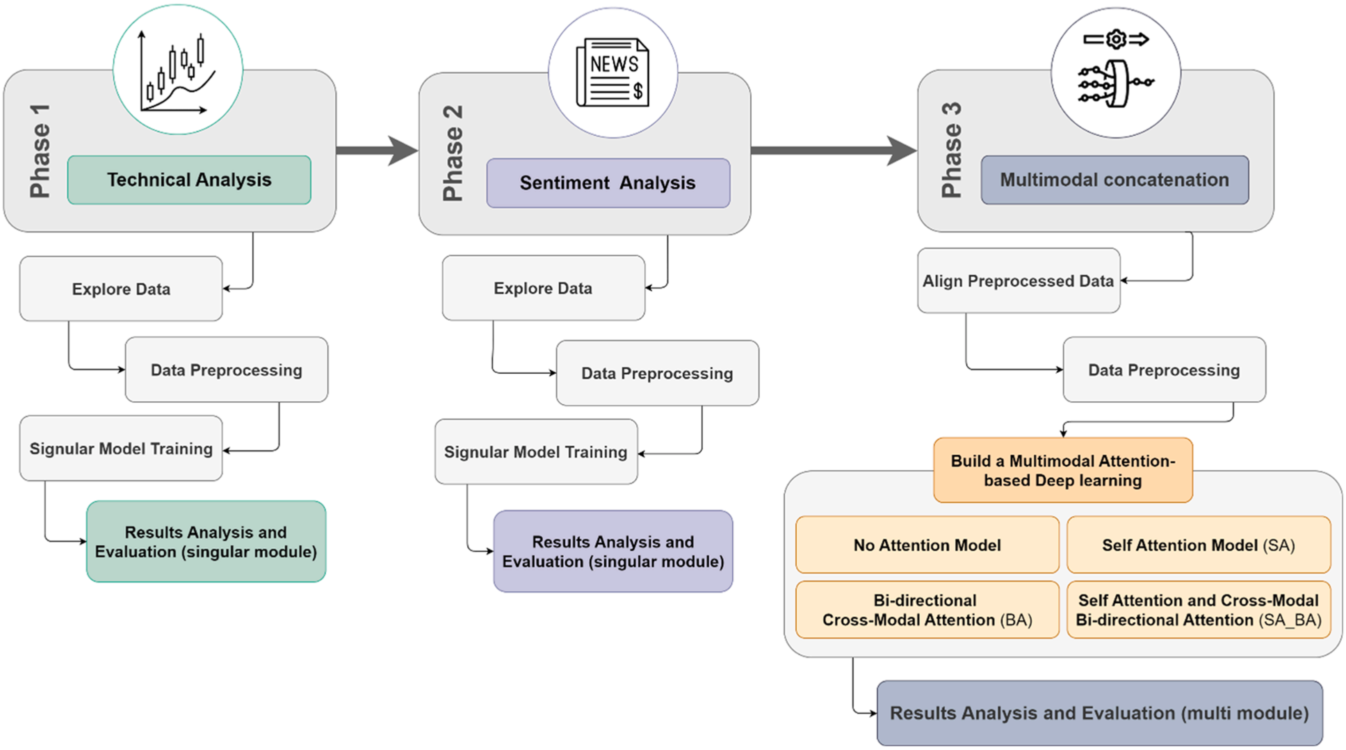

Fig. 3. Methodology framework applied.

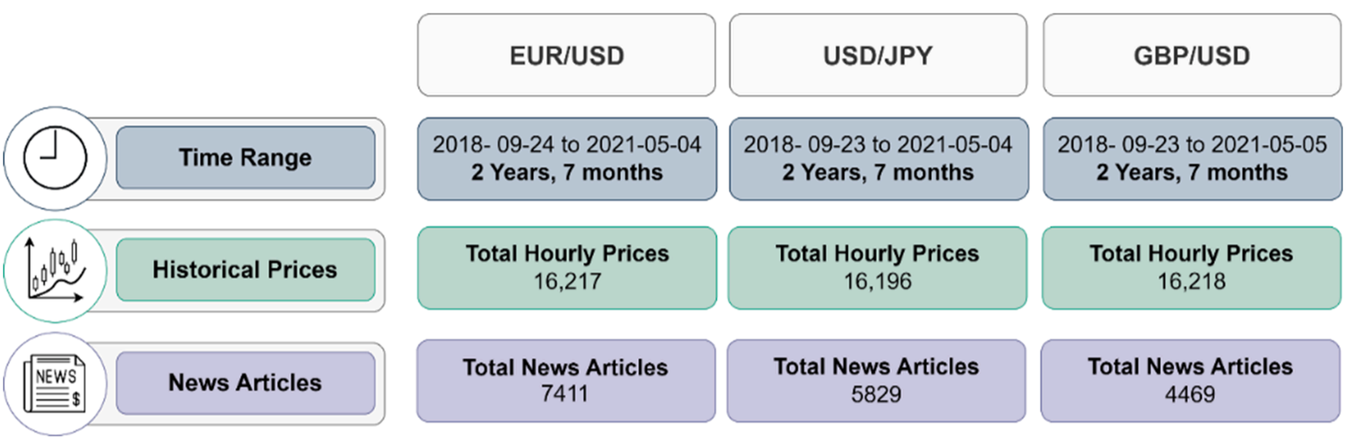

Fig. 4. Datasets description and statistics.

record is created with a timestamp in the GMT time zone. These records include information on the OHLC prices for three currency pairings, as well as the trading volume. Table 4 presents a sample of the market dataset.

News dataset: The authors employed news scraping techniques to obtain focused economic news relating to each distinct currency pair. Their news scraping method employs a variety of reliable financial sources, such as fxstreet, newsbtc, and investing websites. Their news dataset comprises financial news spanning 33 months, from September 2018 to May 2021. Table 5 presents a sample of the market dataset.

4.2. Dataset pre-processing

Data preprocessing is an important step that directly impacts the reliability and performance of any machine-learning model. This section provides an overview of the dataset preprocessing approaches used in both the technical and sentiment analysis modules.

Table 4
Sample of the market price datasets.

Currency Pair	Close	Open	Low	High	Volume	Time Stamp
EUR/USD	1.17357	1.17313	1.17233	1.17375	7613	2018-09-24T06:00
USD/JPY	112.519	112.598	112.472	112.611	2390	2018-09-24T06:00
GBP/USD	1.30825	1.30709	1.30618	1.30841	5810	2018-09-24T06:00

Table 5
Sample of the news datasets.

Currency Pair	Title	Article Body	Time Stamp
EUR/USD	EUR/USD consolidating at key levels ahead of FOMC, bulls beware of spinning formations	EUR/USD is trading between 1.1730 and 1.1793, piercing the July highs of 1.1790 and up modestly …	2018-10-11T12:31:05Z
USD/JPY	JPY futures: bullish momentum persists	Open interest in JPY futures markets increased by around 2.1K contracts ….	2018-09-24T13:28:32Z
GBP/USD	GBP/USD holds above 100-day MA ahead of the Fed	GBP/USD has found acceptance above the 100-day MA, having defended a key Fibonacci support on Friday….	2018-09-26T03:43:57Z

4.2.1. Technical module dataset pre-processing

The preprocessing phase of the technical analysis module is an essential step that prepares the dataset for thorough analysis. We begin by importing the EUR/USD Forex data. This data is then thoroughly examined to identify and correct any missing or incorrect values. The main step of the preprocessing flow involves plotting the high and low data for EUR/USD, USD/JPY, and GBP/USD, as shown in Fig. 5.

A key objective in our preprocessing involves incorporating technical indicators. These indicators are essential for interpreting market behavior and predicting price fluctuations. We incorporate several technical indicators into the dataset, including stochastic oscillator, momentum, Williams' R%, Average Directional Index (ADI), Exponential Moving Average (EMA), Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), On-Balance Volume (OBV), Bollinger Bands (including upper band, lower band, and middle band), and Average True Range (ATR). Additionally, we computed the dataset's log returns as part of our preprocessing to determine the percentage of change in the closing prices over time. When capturing the relative price changes, this method works better than absolute differences, which can have a wide range in scale. Following the computation of log returns, we evaluated multiple labeling strategies to identify the most effective class distribution for our predictive models. Table 6 provides specifics on the various labeling techniques and their goals.

Fig. 6 illustrates the correlation matrix of the several labeling approaches utilized on the dataset post-preprocessing. The figure highlights the relationships between each labeling method, showing how similar or different they are in classifying price movements. Each labeling technique explores specific features of price fluctuations, as indicated by the labels on the matrix axes. The correlation matrix shows that some labeling approaches, such as the Adjusted Threshold Method and the Price Change Direction Method, have a strong correlation because they use similar underlying logic to detect upward or downward trends. In contrast, approaches such as the Adjusted Standard Deviation Method and the Direction Method show weaker correlations, indicating differences in how they handle volatility and the size of price fluctuations. This research aids in understanding the alignment or divergence of multiple labeling procedures and facilitates the selection of complimentary labeling strategies for robust model training. The correlation matrix demonstrates the degree of similarity between the labeling strategies, providing insights into their relationships and helping justify the choice of a particular labeling approach.

Our predictive models are founded on a robust labeling scheme that comprises upward, downward, and hold classifications. Including both periods of stability and fluctuations enables a more accurate prediction of future price changes, thereby facilitating a comprehensive comprehension of market dynamics. Fig. 7 presents the distribution of labels—up, down, and hold—for each labeling approach across the EUR/USD, GBP/USD, and USD/JPY currency pairs. The pie charts provide a detailed overview of how frequently each classification appears in the dataset.

The Adjusted Standard Deviation Method dynamically modifies thresholds based on log return rolling windows, making it more sensitive to increasing market volatility during financial shocks or crises. However, this strategy may not capture all price rises that occur outside of usual market trends. Future iterations may add adaptive window sizes or volatility clustering methods to better manage irregular trends during such situations. This method computes the standard deviation of log returns using a rolling window of 20 time periods and dynamically modifies thresholds in response to market volatility.

- The movement is designated as "up" (signaling a significant upward trend) if the log return exceeds 0.25 times the rolling standard deviation.
- The movement is classified as "down" (signaling a substantial negative trend) if the log return is less than -0.25 times the rolling standard deviation.
- The log return is classified as "hold," denoting a steady market environment devoid of a discernible trend if it is within these limits.

The rolling window helps the model to remain responsive to current market conditions while not being excessively sensitive to short-term noise. The 0.25 multiplier ensures that the approach responds to substantial changes while minimizing noise from minor fluctuations. This approach gives an evenly distributed classification of market situations, resulting in more accurate and interpretable predictions.

To prepare our dataset for analysis, we used Z-scale normalization, which transformed each feature into a zero mean with a one-standard deviation. This normalization phase is critical for ensuring that each feature contributes equally to the model's performance, especially in algorithms that are sensitive to differences in feature magnitudes. Standardizing the input features improves value comparability and overall learning process stability.

The addition of technical indicators, well-defined labels, and normalized values considerably improves the dataset's suitability for future analysis. This thorough preparation provides uniformity at all data levels, laying the foundation for technical analysis and allowing for a more complete and precise understanding of market dynamics. Fig. 8 outlines the preparation techniques, outlining each step done to guarantee that the data is ready for successful modeling and analysis.

4.2.2. Sentiment module dataset pre-processing

The Forex news dataset, which consists of article bodies and titles for each news item, is imported to begin the preprocessing procedure. Combining these components results in a unified narrative that ensures that all relevant material is examined for subsequent research. The next step is to clean up the content by deleting any redundant characters, symbols, and formatting that might slow down the analysis.

After cleaning, the text is tokenized into individual words or sentences to prepare it for summarization. We used the FinBERT model to reduce longer articles into key instructive lines while keeping the necessary financial background. This stage ensures that sentiment analysis concentrates on the most important information.

The summarized articles are then classified using a sophisticated sentiment analysis framework, which increases the dataset's analytical depth. Sentiment is divided into five categories: very positive, positive, neutral, negative, and very negative. The summarization process is designed to focus on relevant information, ensuring that articles capturing major geopolitical or economic events are prioritized for sentiment extraction. The Twitter-RoBERTa model (Barbieri et al., 2020), a customized NLP model capable of collecting complex sentiment expressions, is employed to achieve this level of classification. The model effectively distinguishes between different levels of sentiment intensity using established criteria. Additionally, the use of the

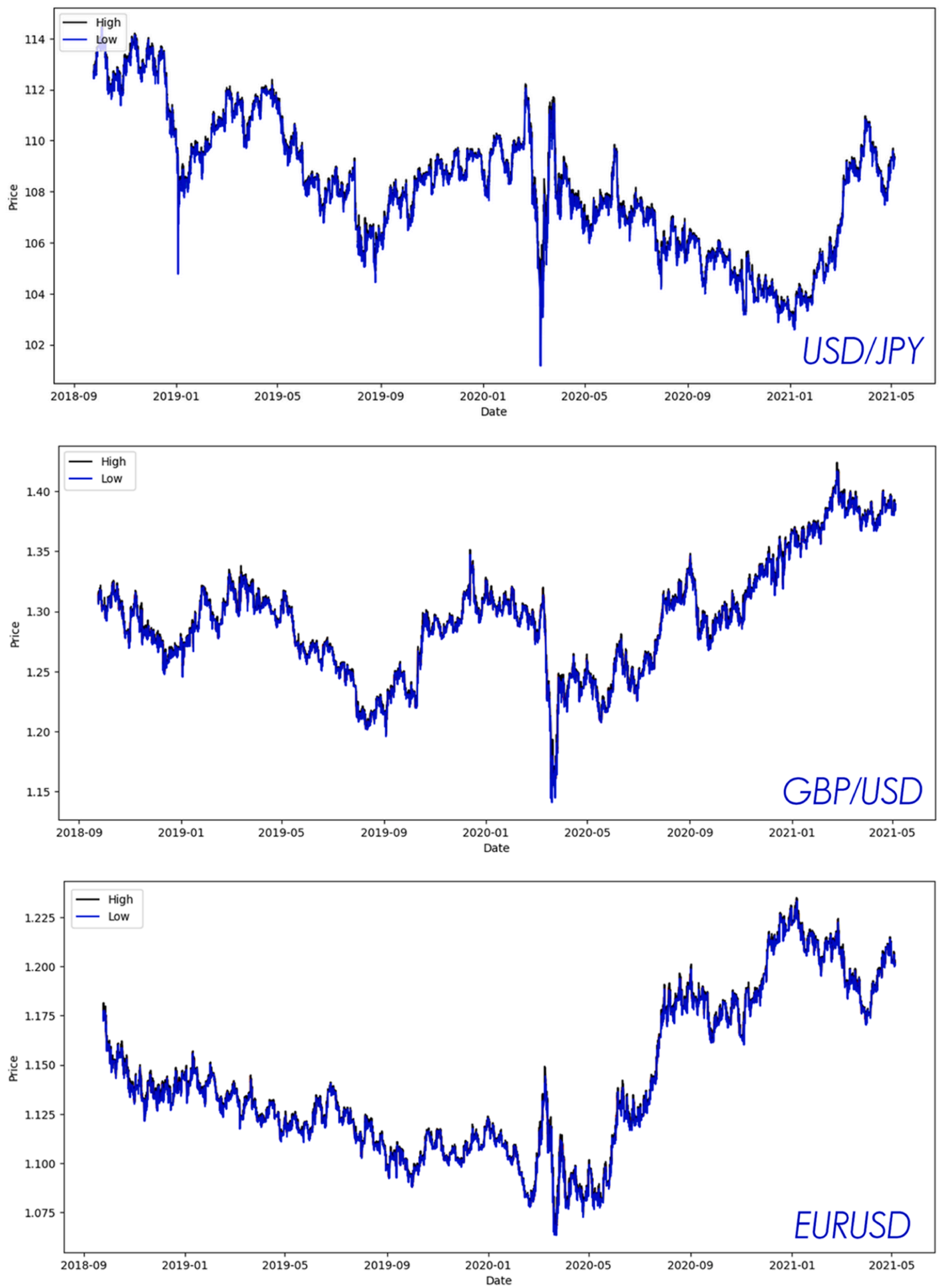

Fig. 5. EUR/USD main features plotting.

Table 6
Labeling techniques.

Labeling Technique	Description
Adjusted Threshold Method	This technique assigns a label to a data point: 'hold' if it falls inside a range near zero, 'down' if the log return is below a negative threshold, and 'up' if it surpasses a positive threshold. This approach offers a simple way to categorize trends by ignoring small variations and concentrating on noteworthy ones.
Adjusted Standard Deviation Method	By using the rolling window standard deviation of log returns, this technique creates dynamic thresholds that adjust to changes in market volatility. The standard deviation is multiplied to establish the cutoff points for labeling as "up," "down," or "hold". This approach offers a more balanced perspective on price movement and adjusts to the market, making it appropriate for varied levels of volatility.
Direction Method	A straightforward comparison of the log return value: 'hold' for zero, 'up' for positive, and 'down' for negative. This approach swiftly classifies trends without taking into account their magnitude, focusing only on the direction of change.
Adjusted Moving Average Method	Evaluate the price right now against a moving average that has been threshold-adjusted. For each data point, the label is 'up' if the price is much higher than the moving average, 'down' if it is significantly lower, and 'hold' otherwise. This method helps to capture changes in trends by identifying patterns based on how current prices differ from previous averages.
Price Change Direction Method	Labels are assigned by calculating the difference between the current closing price and the previous closing price. Like the Directional Method, but with a direct application of pricing disparities and an emphasis on absolute rather than relative changes
Volatility-Adjusted Method	By applying a variable multiplier to the rolling standard deviation, this technique adjusts thresholds in response to changes in the volatility of the market. Compared to a set standard deviation approach, it is more effective at recognizing subtle price movements since it adapts to real-time market swings.

Twitter-RoBERTa model for sentiment classification allows the model to detect nuanced emotional cues during crises.

Following identifying the sentiment categories, these five classes are mapped back into the three main categories of Positive, Neutral, and Negative to conform to the technical analysis framework. This makes integrating and comparing data consistently with other analytical elements possible. To maintain coherence with the current analytical structure, attitudes that are categorized as "Very Negative" and "Negative" are separated from sentiments that are categorized as "Very Positive" and "Positive." Ensuring compliance with the wider market analysis framework allows for a detailed initial sentiment capture and improves the dataset's usefulness for predictive modeling. This strategy provides a richer dataset for subsequent analysis by capturing small fluctuations in market sentiment using sophisticated NLP techniques such as the Twitter-RoBERTa models. In addition, this strategy minimizes the risk of sentiment bias from media sources, the summarized articles are aggregated across multiple sources to provide a balanced sentiment score for each day, reducing reliance on any single perspective.

After conducting sentiment analysis, we enhance our dataset by applying TF-IDF (Term Frequency-Inverse Document Frequency) analysis to the summarized articles. This statistical technique assesses the significance of a word in a document relative to its appearance in the entire dataset, adding an additional layer of informative features for machine learning models. The TF-IDF approach helps highlight terms that are common within a document but rare across the collection, emphasizing their importance. Incorporating TF-IDF features ensures that the textual characteristics capture meaningful distinctions across articles and sentiment contexts.

This TF-IDF transformation serves as a bridge between raw text analysis and advanced modeling, ensuring that the input features accurately represent variations in sentiment-laden language. The preprocessing steps, including cleaning, tokenization, summarization, and feature transformation, are summarized in Fig. 9.

4.3. Alignment and final data preparation

Developing a multimodal framework for financial research necessitated precise synchronization of currency pair-related news articles with corresponding price data. The alignment process posed a significant challenge due to the inherent inconsistencies in the timing of these data sources. Financial news articles are released irregularly, often in response to events of varying significance, whereas price data is consistently generated at fixed intervals. This discrepancy made it challenging to ensure that the sentiment extracted from news articles could be accurately linked to corresponding price changes, as illustrated in Fig. 10. To overcome this challenge, we developed an algorithm designed to align each news article with the most relevant price data point. The algorithm pairs each news story with the most recent price data before the article's publication to model short-term market reactions. However, we acknowledge that some news articles may explain past price movements or have an immediate market impact, reflecting the diverse roles that financial news can play in influencing market trends.

The first step involved standardizing the date formats in the sentiment and price movement datasets to ensure consistency and prevent data loss due to format incompatibility. This step allowed for precise synchronization between the two datasets, maintaining the integrity of the data alignment process. Once standardized, the sentiment analysis findings were matched with price movement data according to various scenarios:

- **Perfect Case:** When price data and news articles began on the same day (e.g., December 31, 2018), the datasets were directly matched without modification.
- **Case 1:** If the price data began earlier than the news articles (e.g., price data from December 31, 2018, and news articles starting January 7, 2019), there will be no prediction on that day.
- **Case 2:** If the price data extended beyond the availability of news articles, the sentiment from the last available news article was carried forward to fill the remaining period.
- **Case 3:** Additionally, when multiple articles were published on the same day, their sentiment scores were aggregated to produce a single sentiment score for that day. The aggregated sentiment was categorized as positive, neutral, or negative, based on the sum of the sentiment scores. This process ensured consistency between the datasets while preserving the nuances of the news sentiment data.

The aligned dataset comprised both sentiment labels and technical indicators such as volume, price movements (Open, Close, High, Low), and market indicators like momentum, stochastic, MACD, and RSI. To ensure equal contribution from all numerical features and to avoid dominance by features with larger scales, Z-score normalization was applied across all features. This rigorous alignment process produced a comprehensive dataset that accurately captured both market sentiment and technical trends. Fig. 11 illustrate the algorithm and sentiment aggregation process, respectively. The finalized dataset was ready for further analysis and served as the input for training the multimodal Forex prediction model.

The aligned dataset enabled the integration of sentiment analysis and technical indicators within the multimodal framework, providing a robust foundation for predicting market movements. By ensuring that each price data point was paired with a complete sentiment score, this approach supported more accurate and comprehensive market predictions. The combination of sentiment and technical data enhanced the model's ability to account for both psychological and quantitative

Fig. 6. Correlation between labeling methods.

market dynamics, setting the stage for improved trading decision-making.

4.4. Multimodal architecture and experimental design

Earlier approaches used attention mechanisms in the field of Forex trading (Zeng and Khushi, 2020; Grądzki and Wójcik, 2024). However, the majority of those approaches are founded on a single trading modality. In contrast, multimodal attention mechanisms have shown promising results in other financial domains, particularly for stock market prediction. Recent studies (Fataliyev and Liu, 2023; Ma et al., 2023; Anon, 2022; Yang, 2023) demonstrate the potential of multimodal techniques in financial analysis. However, none of these studies have focused on the Forex market. Therefore, we explore the field of Forex analysis by integrating multimodal attention mechanisms that have been previously effective outside of finance, such as in Alzheimer's disease using an attention-based deep learning framework that leverages several forms of data (Golovanevsky et al., 2022). Based on this innovative method, we modify its fundamental concepts to suit the realm of finance, particularly for investigating currency pairs. Our algorithm analyzes pre-processed news and market prices with the goal of forecasting market movements, namely whether they will go up, remain stable, or go down.

We employ neural networks for each modality in our multimodal system, which are founded on single-modality designs. This establishes the basis for our layers. We utilize three fully connected layers to process both sentiment (news) and technical (price) features, thereby capturing essential data characteristics. Subsequently, we investigate various attention mechanisms in order to optimize feature interactions and enhance prediction. Four primary mode configurations were evaluated:

1. ***No Attention Neural Network***: the initial model for our experiments, this baseline model does not include attention layers. It independently analyzes sentiment and technical data through fully connected layers, combining the outputs in a final dense layer for classification.
2. ***Self-Attention***: Self-attention prioritizes distinct parts of the input by computing attention weights based on the similarity of queries (Q), keys (K), and values (V). The technique determines how much emphasis to place on each element within the same modality by computing a weighted total of the values based on the dot product of queries and keys. This enables the algorithm to prioritize relevant patterns, such as key phrases in a news item or critical time periods in price data. However, because self-attention functions in a single modality, it is unable to capture the impact of sentiment on technical indicators or vice versa. Instead, it concentrates entirely on internal dependencies in news data (e.g., word- or sentence-level dependencies) or previous price movements (e.g., trend continuity). This limitation is especially obvious in volatile markets, where critical cross-modal signals may be disregarded.
3. ***Bi-Directional Cross-Modal Attention:*** This cross-modal technique addresses the limitation of self-attention by allowing for mutual interaction between modalities. It enables sentiment-related features to impact the perception of technical patterns and vice versa. During times of high market activity, such as geopolitical events or economic announcements, bi-directional attention aligns sentiment and technical elements to present a complete picture of market dynamics by capturing both short-term and long-term dependencies. The mechanism assigns attention weights to prioritize relevant features, enabling the model to highlight key cross-modal interactions. The technique consists of two unidirectional sub-layers, one of which aligns sentiment with technical data and the other with sentiment. This strategy enhances the model's response to external events and market fluctuations by linking features from both modalities, which helps capture non-linear dependencies. However, it is computationally expensive and may produce noise when the dependencies between modalities are minimal.
4. ***Combined Self-Attention and Bi-Cross Attention:*** In this configuration, self-attention is initiated to capture internal modality

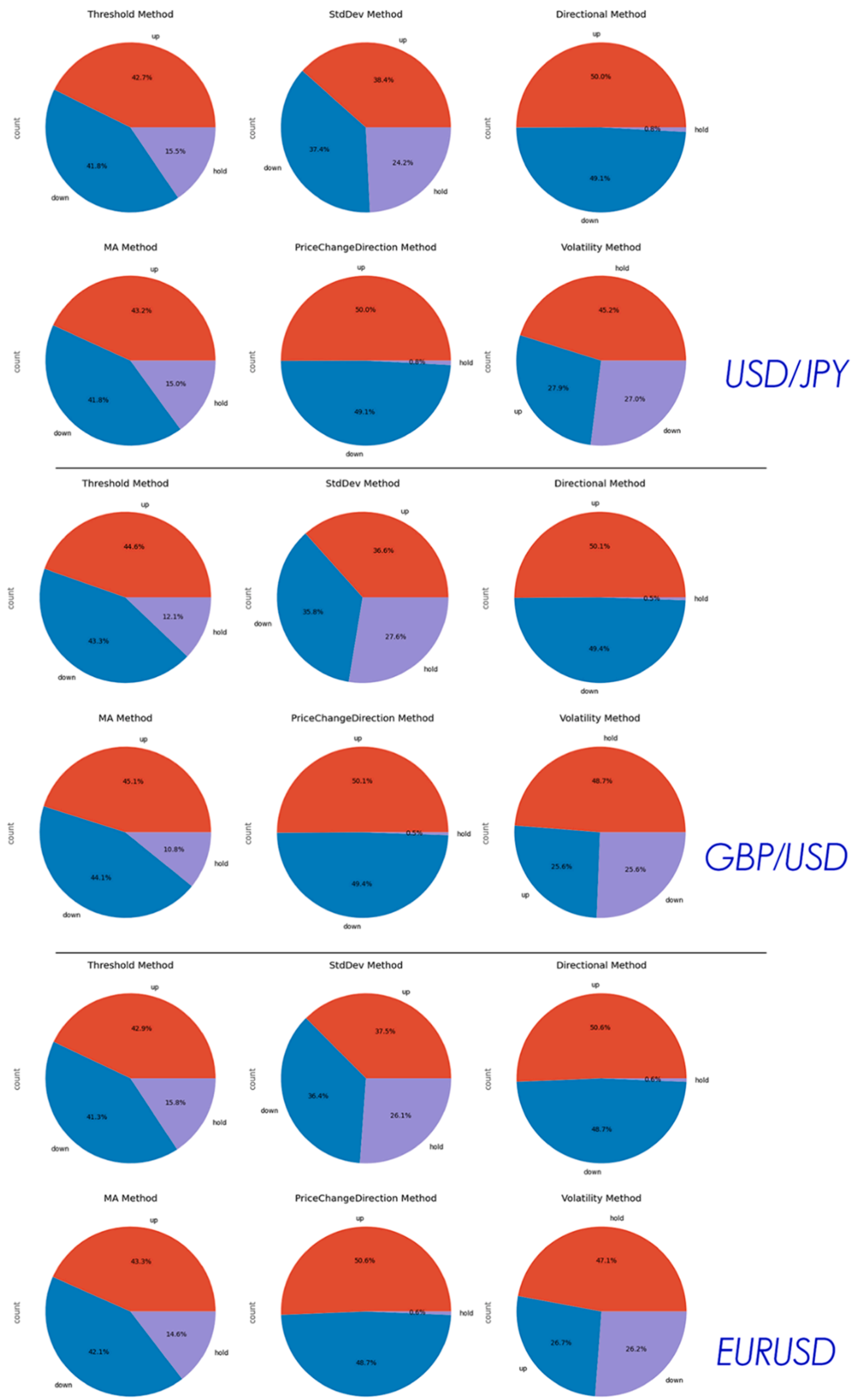

Fig. 7. Pie charts illustrating the distribution for each method in the selected currency pairs.

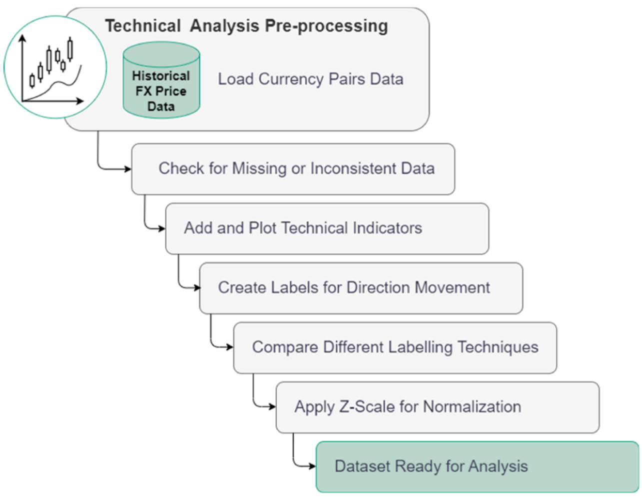

Fig. 8. Technical market price module dataset pre-processing flow.

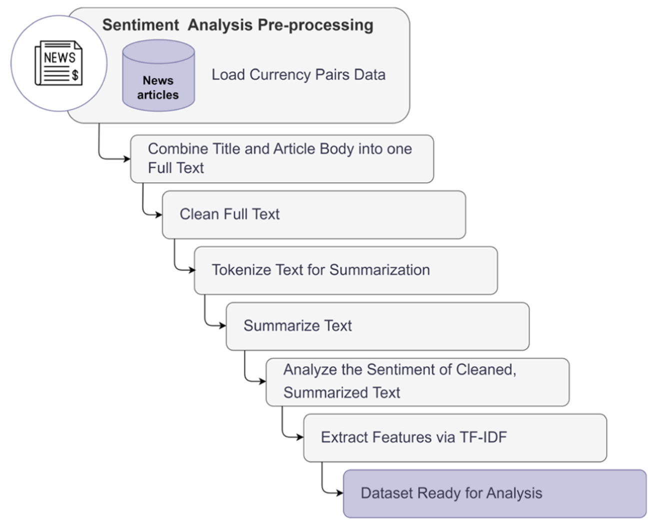

Fig. 9. Sentiment analysis module dataset pre-processing flow.

relationships, and cross-modal attention is subsequently applied to align sentiment and technical features. This method aims to optimize insights by capturing both modality-specific features and cross-modal interactions.

Table 7 compares the two primary attention mechanisms deployed: self-attention and bidirectional cross-modal attention. Self-attention focuses on internal relationships within a single modality, such as sentiment or technical data, emphasizing interdependence within the input. This technique effectively catches trends within specific modalities, making it especially useful for currency pairs with persistent patterns and minimal volatility. However, in highly volatile markets, self-attention may struggle to catch interactions between modalities, such as the impact of mood on price changes. In contrast, bi-directional cross-modal attention allows for mutual interaction between modalities by synchronizing sentiment and technical modules. This bidirectional method improves the model's responsiveness to external events and market shifts by allowing emotion data to help interpret technical signals and vice versa. While this strategy improves the model's ability to capture complicated cross-modal interactions, it is computationally demanding and may add noise when there are weak dependencies between modalities.

Fig. 12 illustrate the configurations that were implemented to capture distinctive feature interactions, outlining the design structure of each model. These figures offer a visual representation of the process in which the features of each modality are transmitted through the attention layers and affect the final predictions.

Generalized attention: Attention is a cognitive mechanism that prioritizes influential features by highlighting the relationships between states. In machine learning, an attention layer takes in queries (Q), keys (K), and values (V). Queries and keys have dimensions d_k, while values have a dimension d_v. Keys identify features, and queries evaluate them to find the best match. The process involves calculating dot products between the query and each key, normalizing by $\sqrt{d_k}$, and applying a *Softmax* function to convert the results into weights:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

Self-Attention: Self-attention mechanisms allow the model to evaluate relationships within a single modality, enhancing the understanding of its internal dynamics. In this work, self-attention operates on two latent feature matrices: the sentiment news matrix (S) and the technical price matrix (T). The mechanism applies as follows:

$$Self\_attention(S \rightarrow S) \tag{2}$$

$$Self\_attention(T \rightarrow T) \tag{3}$$

Cross-modal attention: Cross-modal attention facilitates interaction between modalities, enabling a deeper understanding of their relationships. This mechanism extracts outputs from self-attention layers and establishes bidirectional attention between sentiment (S_A) and technical (T_A) data. It concatenates the outputs of two unidirectional cross-modal attention layers:

$$concatenation(cross\_modal\ attention(S_A \rightarrow T_A),\ cross\_modal\ attention(T_A \rightarrow S_A)) \tag{4}$$

A benchmark model without attention mechanisms directly maps neural network outputs to the final prediction layer. This serves as a baseline for comparison with multimodal architecture.

4.5. Hyperparameter optimization approach

To optimize hyperparameters, each modality was tuned independently while ensuring no test set overlap. The dataset was split into 70 % training and 30 % testing, and K-Fold Cross-Validation (five folds) was applied to mitigate overfitting and improve generalization. Consistent dataset alignment was maintained by resetting indices before splitting. Each fold used a unique random seed, ensuring robust performance evaluation. During training, key hyperparameters such as learning rate, batch size, and dropout ratio were systematically adjusted to identify configurations yielding the highest validation accuracy.

4.6. Performance evaluation metrics

The model's performance was evaluated using accuracy, precision, recall, F1-score, Directional Accuracy (DA), and the Matthews

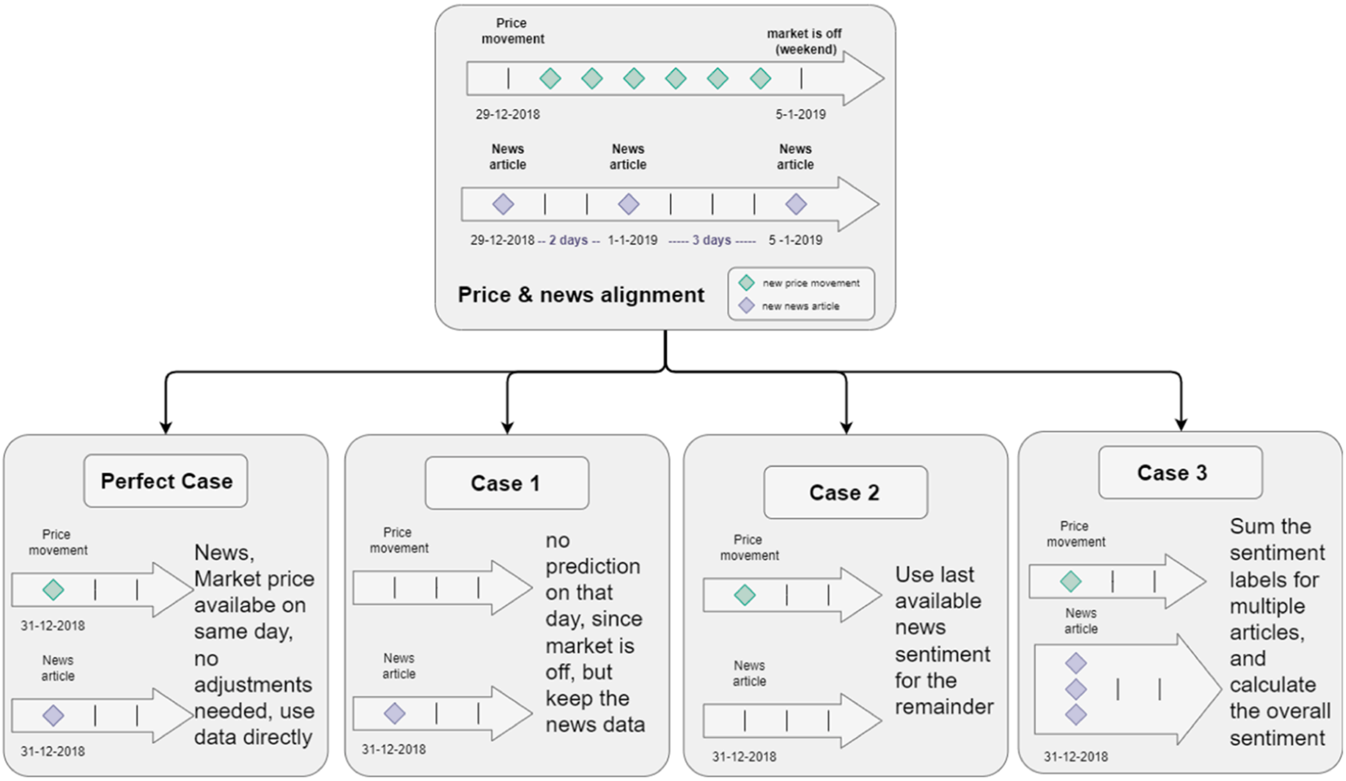

Fig. 10. Alignment cases.

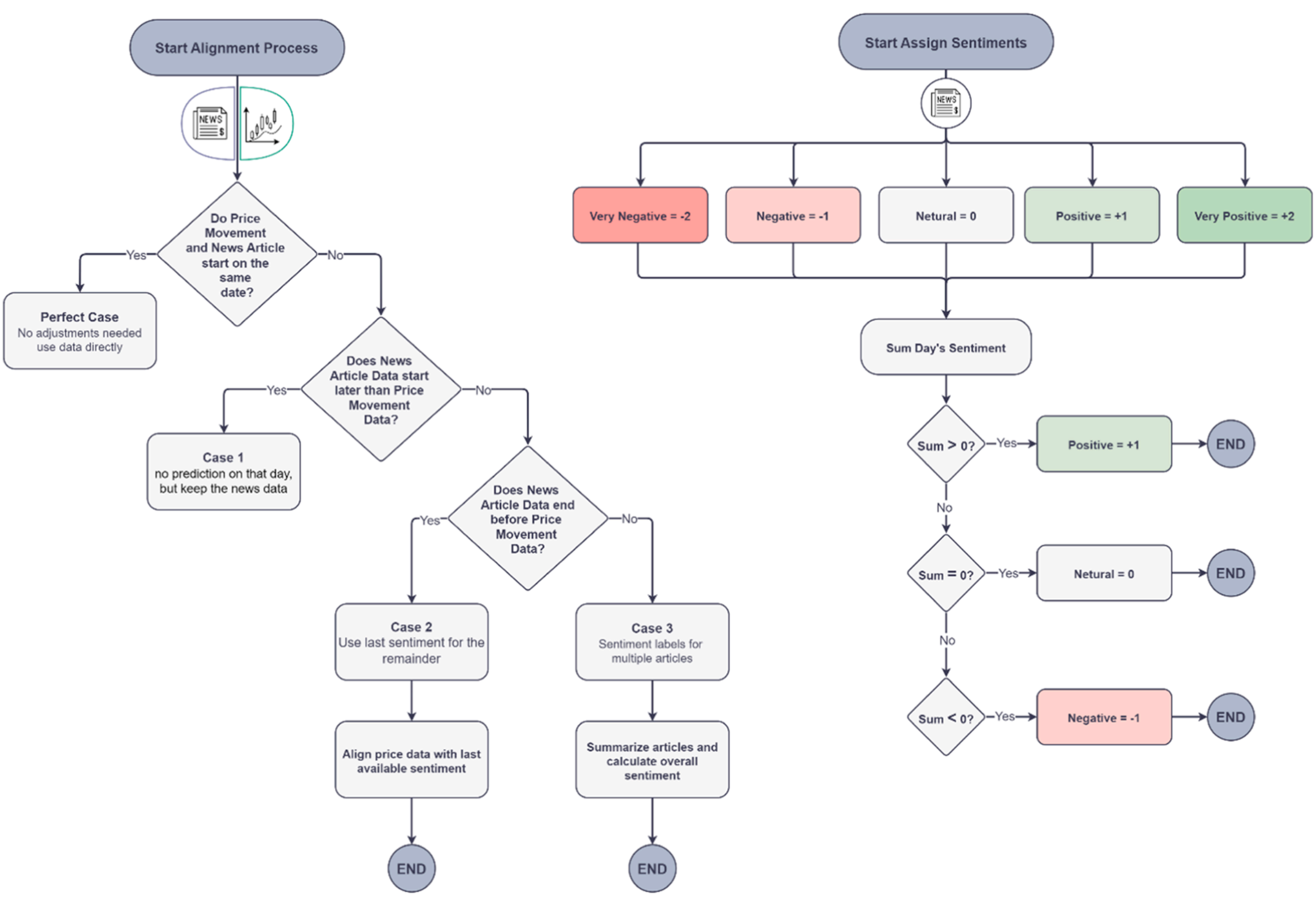

Fig. 11. Aggregation of daily news sentiment and the alignment process.

Table 7
Comparison between self-attention and bi-directional cross-modal attention.

Feature	Self-Attention	Bi-Directional Cross-Modal Attention
Description	Focuses on **internal relationships** within a particular modality (such as sentiment or technical data).	Enables a **mutual interaction between modalities** by coordinating sentiment and technical features.
Strengths	Captures **internal dependencies** efficiently, highlighting key trends within a single modality.	Captures **complicated interactions between modalities,** increasing responsiveness to external events.
Limitations	In volatile markets, it may be difficult to capture cross-modal interdependence.	Computationally intensive; may generate noise if interdependence is low.
Directionality	**Single-directional** (within the same input)	**Bidirectional** (each modality attends to the other)
Example	Analyzing word dependencies in a news article.	Linking a news sentiment to a technical indicator.

Correlation Coefficient (MCC). While accuracy, precision, and recall assess general classification tasks, DA and MCC are particularly critical for financial forecasting. Directional Accuracy (DA) evaluates the model's ability to predict directional price movements. Matthews Correlation Coefficient (MCC) provides a robust measure of classification accuracy, accounting for imbalanced datasets.

$$\textit{Directional Accuracy} \;=\frac{1}{N}\sum_{i=1}^{N-1}\left[sign(a_{i+1}-a_i)=sign(f_{i+1}-f_i)\right] \tag{5}$$

$$MCC=\frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{6}$$

Here, a denotes actual values, f represents forecasts, N is the number of instances, and *sign()* extracts the price movement's direction. MCC values range from -1 (disagreement) to +1 (perfect prediction), with 0 indicating random guessing. These metrics, particularly DA and MCC, are well-suited for financial domains, where false predictions carry significant costs. Please note that fig.

5. Results and discussion

This section outlines the results of the proposed multimodal approach for Forex price prediction, integrating sentiment analysis (news data) and technical analysis (price data). The methodology was evaluated on three currency pairs (EUR/USD, GBP/USD, JPY/USD), ensuring effective alignment of sentiment data with price movements. The multimodal architecture was tested on these pairs to assess its applicability and effectiveness in the Forex market.

The experiments were conducted on a 13th Gen Intel® Core™ i7-13700HX processor with 32 GB RAM in a Jupyter Lab environment, which supported data preparation, model training, and evaluation efficiently. This setup facilitated systematic evaluation of diverse attention mechanisms incorporated into the model.

5.1. Unified models performance

This section assesses the outcomes of models trained individually on technical indicators and news sentiment for forecasting Forex price patterns. We compare these independent models with a unified, multimodal approach that uses both technical indicators and news sentiment for trend forecasting.

5.1.1. Technical analysis model performance discussion

The technical analysis model utilized historical price data and technical indicators to predict trends. Despite using dense layers, batch normalization, and dropout to prevent overfitting, performance varied across currency pairs. This variation is influenced by the unique characteristics of each currency pair, such as the level of market volatility and the presence of discernible trends. Some pairs, like GBP/USD, exhibit more stable market patterns, making them more predictable, while pairs like USD/JPY tend to be more volatile and challenging to model. It is also important to note that the singular models are intentionally designed as comparative baselines rather than highly sophisticated predictors. Both the technical and sentiment analysis models

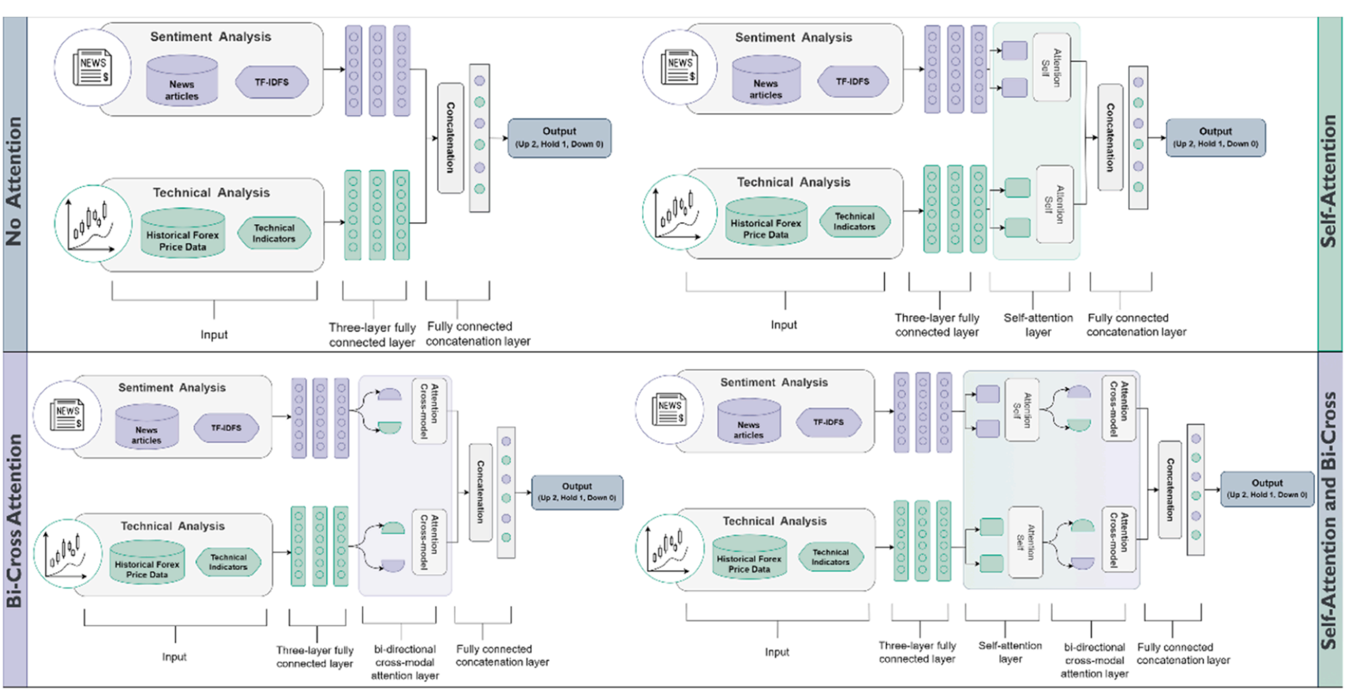

Fig. 12. Baseline neural network architecture without attention mechanisms.

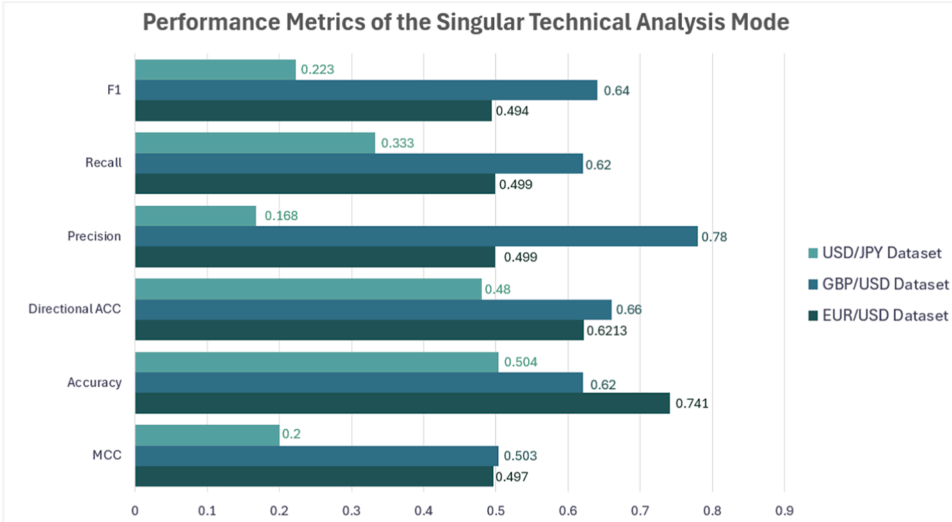

Fig. 13. Bar chart illustrating performance metrics of the singular technical analysis model.

Table 8
Performance metrics of the singular technical analysis model.

Dataset\Metric	MCC	Accuracy	DA	Precision	Recall	F1-Score
EUR/USD Dataset	0.497	**0.741**	0.621	0.499	0.499	0.494
GBP/USD Dataset	**0.503**	0.62	**0.66**	**0.78**	**0.62**	**0.64**
USD/JPY Dataset	0.2	0.504	0.48	0.168	0.333	0.223

utilize neural network architectures but are kept relatively simple to provide a fair comparison with the multimodal approach. The main performance indicators, encompassing the MCC, Accuracy, DA, Precision, Recall, and F1-score, are all presented in Fig. 13 and Table 8:

- **EUR/USD:** The model achieved an accuracy of 0.741, along with an MCC of 0.497, signifying a degree of predictive capability. Nevertheless, the F1-score was comparatively low at 0.494, indicating that the model had difficulty consistently identifying trend shifts.
- **GBP/USD:** The GBP/USD dataset yielded reasonable outcomes, exhibiting an MCC of 0.503 and an accuracy of 0.62. The directional accuracy was comparatively elevated (0.66), indicating superior performance in forecasting overall trend direction.
- **USD/JPY:** The USD/JPY dataset exhibited the poorest performance, recording an MCC of merely 0.2 and an accuracy of 0.504. The poor F1-score (0.223) signifies that the model encountered considerable difficulties with this dataset.

Overall, although the technical analysis of the Singular Model showed considerable success on specific parameters for EUR/USD and GBP/USD, its performance was inadequate for dependable trend prediction. These drawbacks underscore the necessity of a multimodal strategy to combine technical indicators and sentiment analysis, as the next sections will demonstrate.

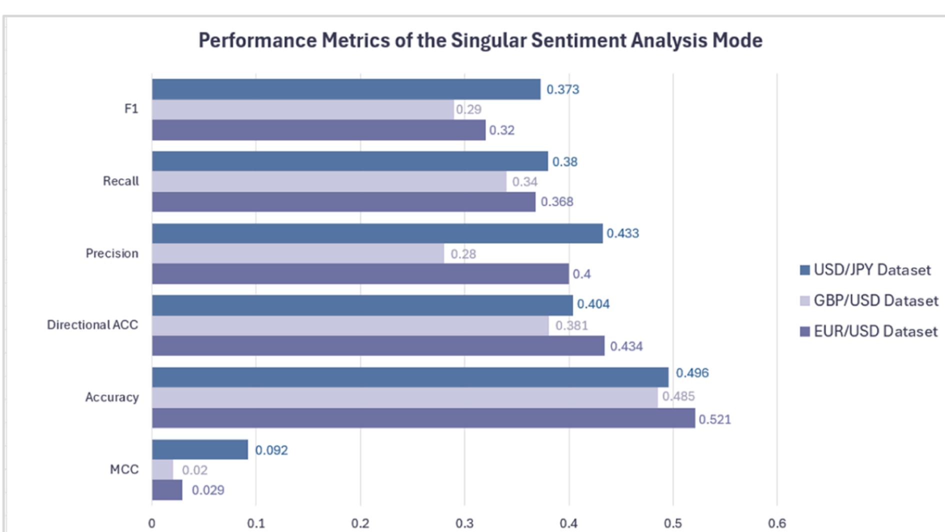

Fig. 14. Bar chart illustrating performance metrics of the singular sentiment analysis model.

Table 9
Performance metrics of the singular sentiment analysis model.

Dataset\Metric	MCC	Accuracy	DA	Precision	Recall	F1-Score
EUR/USD Dataset	0.029	**0.521**	**0.434**	0.4	0.368	0.32
GBP/USD Dataset	0.02	0.485	0.381	0.28	0.34	0.29
USD/JPY Dataset	**0.092**	0.496	0.404	**0.433**	**0.38**	**0.373**

5.1.2. Sentiment analysis model performance discussion

The sentiment analysis model's performance varied across currency pairs. This can be attributed to the different levels of influence that sentiment has on the respective markets. Some currency pairs may be more sensitive to news-driven sentiment changes, whereas others, particularly those with high-frequency fluctuations, may be less influenced by qualitative factors. The performance differences, coupled with the simplicity of the singular sentiment model, highlight the need for a multimodal approach to better capture the complex interdependencies between market sentiment and historical trends. In Fig. 14 and Table 9 the primary performance indicators are presented.

- **EUR/USD:** The model demonstrated minimal predictive capability, as evidenced by an MCC of 0.029 and an accuracy of 0.521. The F1-score was low at 0.32, suggesting that it is difficult to consistently identify sentiment-driven trend shifts.
- **GBP/USD:** The GBP/USD dataset also produced insignificant results, with an MCC of 0.02 and an accuracy of 0.485. The DA was 0.381, suggesting that the forecasting of the overall trend direction was limited.
- **USD/JPY:** The USD/JPY dataset yielded marginally superior results, with an MCC of 0.092 and an accuracy of 0.496. Nevertheless, the F1-score of 0.373 continues to indicate the model's challenges with this dataset.

In general, the Sentiment Analysis Singular Model demonstrated some success in specific parameters; however, its performance was insufficient for the accurate prediction of trends. These constraints underscore the significance of a multimodal approach that integrates sentiment analysis with technical indicators, as explained in the following sections. The singular models serve as a comparative baseline, emphasizing the enhancements provided by the multimodal approach.

5.2. Multimodal performance

The multimodal Forex trading model is intended to capture a more comprehensive understanding of market trends by combining technical analysis from price data with sentiment analysis from financial news. This model builds on complementary insights by integrating both data types: qualitative narratives from news articles and quantitative signals from historical prices. The architecture's adaptability enables it to function in a variety of modes, each of which employs distinct attention mechanisms to investigate the impact of each data source on the final prediction. The model can operate in four modes depending on the sort of attention that is applied. In the following subsections, we will dive deeper into each mode of the multimodal approach and present their performance.

5.2.1. No attention model

The multimodal architecture in the No Attention model is entirely dependent on the integration of technical and sentiment data without any form of attention mechanism. This method functions as a benchmark for evaluating the influence of incorporating attention layers on predictive performance. The result of this mode is presented in Fig. 15 and Table 10. The No Attention module demonstrated moderate performance levels on each currency pair in the absence of the additional complexity of attention. Although the module captured some general patterns from the input data, it did not benefit from the refined intermodal relationships that attention mechanisms contribute.

The key performance metrics that were observed for the "No Attention" module across the currency pairs are listed below.

Table 10
Performance evaluation of the no attention mode.

Evaluation Metric\Dataset	EUR/USD	JPY/USD	GBP/USD
MCC	0.637	**0.69**	0.632
Accuracy	0.753	**0.792**	0.75
DA	0.765	**0.805**	0.764
Precision	0.76	**0.782**	0.758
Recall	0.738	**0.774**	0.74
F1-score	0.736	**0.774**	0.736

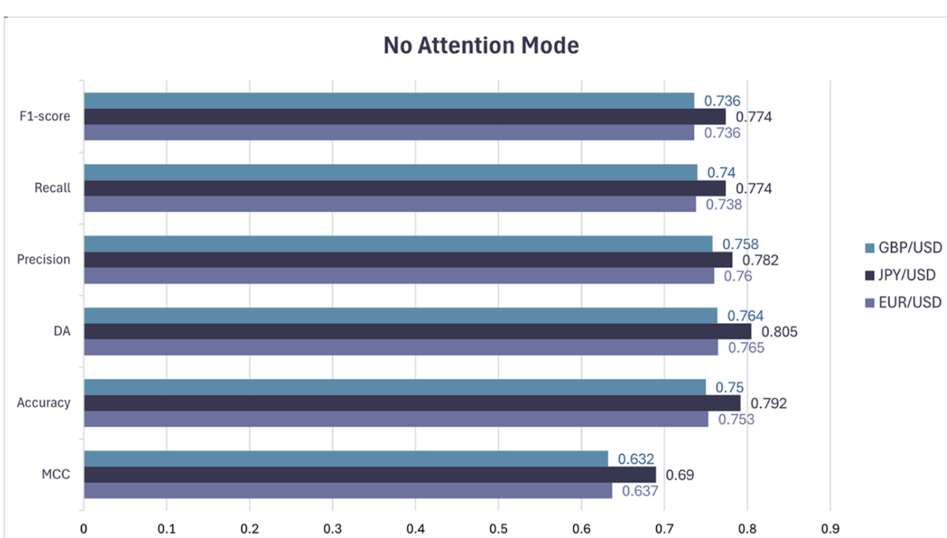

Fig. 15. Performance metrics for the no attention mode.

- **EUR/USD:** An accuracy of 0.753 and an MCC of 0.637 suggest that this pair has a relatively strong predictive ability. The module demonstrated a directional accuracy of approximately 76.5 %, indicating a reasonable level of reliability in identifying trend directions. However, its ability to recognize nuanced patterns is restricted by the absence of attention mechanisms.
- **GBP/USD:** An MCC of 0.632 and an accuracy of 0.75 indicate moderate performance. This module outperforms a random baseline; however, its recall and F1-scores, while consistent, indicate that the model may struggle to comprehend the nuanced movements of this currency pair without additional attention. The depth of predictive insights that can be achieved with more complex relationships is limited by the lack of attention.
- **USD/JPY:** The "No Attention" module exhibited its peak performance on this pair, with an MCC of 0.69 and an accuracy of 0.792. This outcome emphasizes a degree of ability to identify more extensive patterns, as evidenced by an 80.5 % directional accuracy. Nevertheless, the lower precision and recall values suggest that, despite its ability to classify trends accurately, its sensitivity to smaller, complex market movements is restricted in the absence of attention layers to focus on specific signals.

The significance of integrating attention mechanisms to improve performance was illustrated by the "No Attention" module. The module's inability to profit efficiently on the interdependencies between news sentiment and price trends is indicated by the limited accuracy and MCC scores across currency pairs. This comparison demonstrates the value that attention mechanisms contribute to the capture of the complex, interconnected nature of financial market data.

5.2.2. Self-attention model

The Self-Attention Mode introduces an improved architectural approach by incorporating self-attention layers into the model, which allows it to concentrate independently on critical components of technical and sentiment data. This method, in contrast to the No Attention model, offers a focused perspective on each modality, thereby improving the model's ability to identify essential features within individual data streams and thereby improving interpretability. The self-attention mechanism significantly improves the model's ability to capture and interpret isolated trends within each modality, offering a more intricate understanding of price and sentiment features compared to the No Attention module. Fig. 16 illustrates the evaluation metrics for each currency pair, showcasing the model's effectiveness across different metrics, such as MCC, accuracy, and directional accuracy. Each metric highlights the model's ability to capture and interpret trends uniquely, given the Self-Attention's capacity to focus on crucial data points. The results of the self-attention mode are presented in Table 11.

Table 11
Performance evaluation of the self-attention mode.

Evaluation Metric\Dataset	EUR/USD	JPY/USD	GBP/USD
MCC	0.714	**0.774**	0.74
Accuracy	0.808	**0.848**	0.826
DA	0.801	**0.833**	0.813
Precision	0.826	**0.862**	0.834
Recall	0.808	**0.848**	0.824
F1-score	0.814	**0.85**	0.826

The following are the primary observations regarding the performance of the self-attention Mode across the currency pairs

- **EUR/USD:** The model has a significant predictive potential, with an accuracy of 0.808 and an MCC of 0.714. The self-attention mechanism emphasizes critical data points, enhancing the model's understanding of trend behaviors with a directional accuracy of 80.1 %. This mode is significantly more capable in concentrating on critical data elements within each modality, resulting in a comprehensive comprehension of trend behavior.
- **GBP/USD:** The model's accuracy of 0.826 and MCC of 0.74 for the GBP/USD pair demonstrates the self-attention mechanism's enhancement of sensitivity to fluctuations in this currency pair. The increased F1-scores and recall emphasize the model's attention to minor yet influential patterns, improving on areas that the No Attention module struggled with.
- **USD/JPY:** The model achieves its highest performance on the USD/JPY pair, with an accuracy of 0.848 and MCC of 0.774. The high precision and recall further emphasize the model's strength in interpreting complex signals for USD/JPY.

The self-attention mechanism enables the model to selectively prioritize individual features within each data stream, thereby enabling a comprehensive understanding of currency movements that surpasses the limitations observed in the No Attention module. These findings emphasize the potential of multimodal approaches with self-attention, which improves the model's ability to capture the complex patterns that are inherent in financial market data.

5.2.3. Bi-directional cross-attention model

By employing bi-directional attention layers to facilitate and intensify the interaction between sentiment and technical data, the Bi-directional Cross-Attention (BA) Mode is a sophisticated improvement.

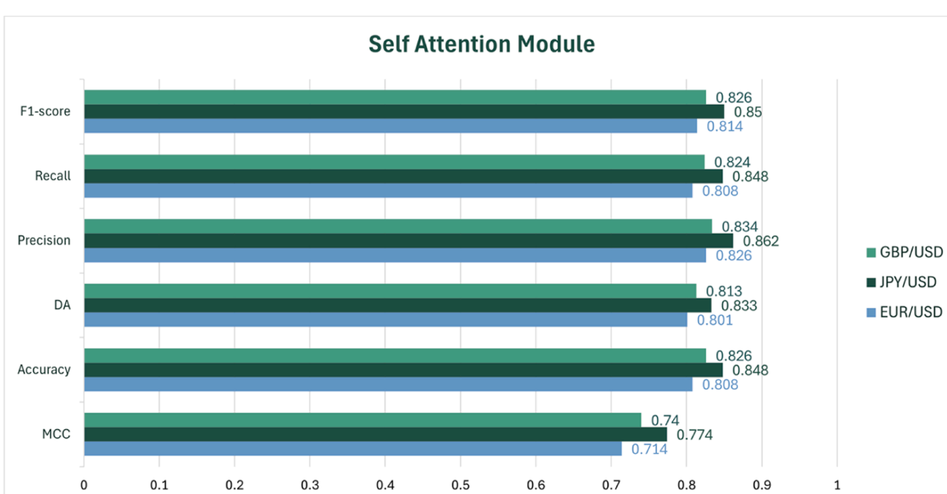

Fig. 16. Performance metrics for the self attention mode.

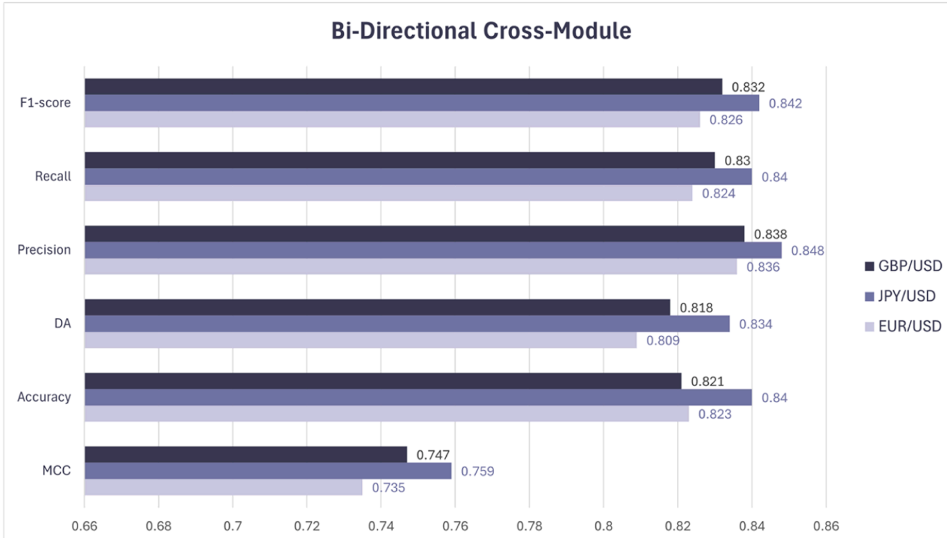

Fig. 17. Performance metrics for the bi-directional cross-attention mode.

Table 12
Performance evaluation of the bi-directional cross-attention mode.

Evaluation Metric\Dataset	EUR/USD	JPY/USD	GBP/USD
MCC	0.735	**0.759**	0.747
Accuracy	0.823	**0.84**	0.821
DA	0.809	**0.834**	0.818
Precision	0.836	**0.848**	0.838
Recall	0.824	**0.84**	0.83
F1-score	0.826	**0.842**	0.832

This method enables the model to dynamically capture and exchange critical information across both data sources, in contrast to the self-attention mode, which applies attention within each modality independently. The interpretive context is enhanced by the bi-directional design, which allows each modality to concentrate on its inherent features and complementary insights from the other. Fig. 17 illustrates the outcomes of the BA Mode, which are further elaborated upon in Table 12. This mode considerably enhances the model's ability to capture the interplay between sentiment and price, thereby illustrating its superiority over single-modality approaches.

The following are the primary findings for the BA Mode across currency pairs:

- **EUR/USD:** The model demonstrates robust predictive potential for this pair, achieving an accuracy of 0.823 and an MCC of 0.735. Its capacity to consistently identify directional trends is underscored by its directional accuracy of 80.9 %. This mode's trend prediction performance surpasses that of the previous models due to the bidirectional exchange of information, which enables the identification and integration of impactful cross-modal patterns.
- **GBP/USD:** This mode continues to exhibit robust predictive insights, with an MCC of 0.747 and an accuracy of 0.821. Its capacity to precisely record slight price fluctuations is emphasized by its precision and recall values. The bi-directional cross-attention mechanism allows the model to more effectively synthesize information from both modalities, resulting in enhanced trend recognition and generalization for this currency pair.
- **USD/JPY:** This pair results in the most benefit from the bi-directional attention mechanism, as evidenced by its 0.84 accuracies and an MCC of 0.759, which is the highest among all tested pairs. An advanced capability to adaptively capture patterns that

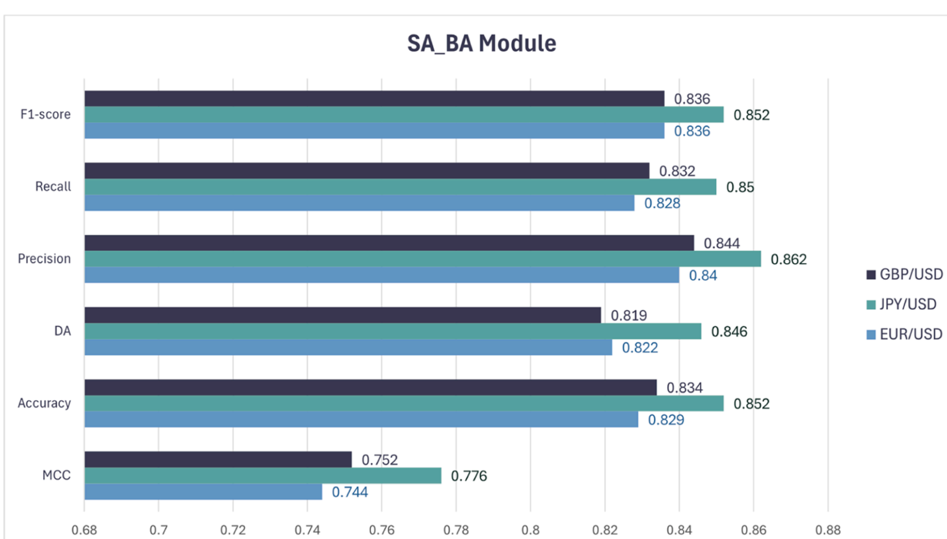

Fig. 18. Performance metrics for the self-attention and bi-directional cross-attention mode.

Table 13
Performance evaluation of the self-attention and bi-directional cross-attention mode.

Evaluation Metric\Dataset	EUR/USD	JPY/USD	GBP/USD
MCC	0.744	**0.776**	0.752
Accuracy	0.829	**0.852**	0.834
DA	0.822	**0.846**	0.819
Precision	0.84	**0.862**	0.844
Recall	0.828	**0.85**	0.832
F1-score	0.836	**0.852**	0.836

would otherwise remain latent is indicated by robust F1-scores and directional accuracy of 83.4 %. The cross-modal attentional focus enables the model to identify complex shifts that impact this currency pair, underscoring the cross-attention mode's superior capacity to manage complicated financial signals.

The BA Mode's performance illustrates the influence of cross-modal information exchange in the improvement of predictions. This mode enhances its predictive reliability by capturing complex relationships between sentiment and price data. These results emphasize the benefits of sophisticated attention mechanisms in multimodal architectures, demonstrating the model's ability to leverage multi-source data to achieve superior performance in Forex market predictions.

5.2.4. Self-attention and bi-directional cross-attention model

The Self-Attention with Bi-directional Cross-Attention Mode (SA_BA) optimizes interpretative depth within and across modalities by integrating self-attention and bi-directional cross-attention mechanisms. This mode represents the most comprehensive approach in the model group by layering self-attention to capture internal modality patterns and bi-directional cross-attention to synchronize and enrich information exchange between technical and sentiment data. Fig. 18 illustrates the outcomes of the SA_BA Mode, which are summarized in Table 13. This dual-attention approach demonstrates the most exceptional performance among the tested modes, thereby confirming its effectiveness in capturing complex relationships between modalities.

The following are the primary findings for the SA_BA Mode across currency pairs:

- **EUR/USD:** This mode reveals exceptional predictive accuracy, obtaining a directional accuracy of 82.2 %, with an accuracy of 0.829 and an MCC of 0.744. The model's prediction capability is robust and well-rounded due to the effective utilization of both internal trends within each modality and complex inter-modal relationships, which is achieved through the combination of self-attention and bi-directional cross-attention.
- **GBP/USD:** The SA_BA mode exhibits a high degree of adaptability to this currency pair, with an MCC of 0.752 and an accuracy of 0.834. It effectively captures nuanced shifts in GBP/USD trends, with precision and recall scores of 0.844 and 0.832, respectively. The model's enhanced sensitivity to intricate market movements is reflected in the model's ability to integrate information from both modalities at a deeper level in this dual-attention setup.
- **USD/JPY:** This pair obtained the highest metrics among all tested modes, recording an accuracy of 0.852 and an MCC of 0.776. The SA_BA mode's dual-attention mechanism enables 84.6 % directional accuracy, suggesting a more comprehensive understanding of the complex interactions within the JPY/USD market. The model can identify and capitalize on both direct and interrelated dependencies between technical indicators and sentiment data by effectively integrating both self-attention and cross-modal attention.

In conclusion, the SA_BA Mode is superior to previous models, resulting in a level of predictive complexity that exceeds them. The advantages of a multi-attention framework for Forex market predictions are underscored by the comprehensive and sophisticated interaction of data streams that is facilitated by the integration of SA with bi-directional cross-attention. The significance of sophisticated attention mechanisms in improving the interpretability and accuracy of multimodal forecasting frameworks is underscored by the success of this model.

5.2.5. Final comparison

In comparison to the Singular Approach (which consists solely of the technical analysis and sentiment analysis models), the Multimodal Approach exhibits a distinct performance advantage across all currency pairs. This advantage is especially visible when attention mechanisms are implemented.

- **EUR/USD Pair:** The SA_BA model in the Multimodal Approach demonstrated the greatest accuracy of 0.829 and an MCC of 0.744, which is a significant improvement over the Technical Analysis (Accuracy: 0.741, MCC: 0.497) and Sentiment Analysis (Accuracy: 0.521, MCC: 0.029) models in the Singular Approach. The SA_BA model also exhibited a considerably higher level of directional accuracy, with a score of 82.2 %, as opposed to 62.1 % for Technical Analysis and 43.4 % for Sentiment Analysis. These findings underscore the effectiveness of integrating both data modalities with attention mechanisms, which leads to improved pattern recognition and trend prediction.
- **JPY/USD Pair:** The SA_BA model demonstrated superior performance, with an accuracy of 0.852 and an MCC of 0.776, surpassing the Singular Approach models (Technical Analysis: Accuracy 0.504, MCC: 0.04; Sentiment Analysis: Accuracy 0.496, MCC: 0.092). The model demonstrates a strong comprehension of the trend directions for JPY/USD, with a directional accuracy of 84.6 %. This is a significant advance from the 48 % achieved in Technical Analysis and the 40.4 % achieved in Sentiment Analysis. The SA_BA model exhibited the highest precision, recall, and F1-scores, demonstrating its ability to manage complex, nuanced relationships within the multimodal data.
- **GBP/USD Pair:** The SA_BA model, like the other pairs, obtained the highest performance metrics, with an accuracy of 0.834 and an MCC of 0.752. This model outperformed both the technical analysis (Accuracy: 0.62, MCC: 0.503) and sentiment analysis (Accuracy: 0.485, MCC: 0.02) models. The SA_BA model achieved a directional accuracy of 81.9 %, which is higher than the Technical and Sentiment models' accuracy of 66 % and 38.1 %, respectively. The SA_BA model's enhanced recall, precision, and F1-scores indicate that it is more sensitive and accurate in capturing the dynamics of the GBP/USD pair, thereby providing a more reliable forecast than either singular modality solely.

Fig. 19 demonstrates the importance of the multimodal approach over the singular approach. The Multimodal Approach, particularly when combined with the SA_BA attention mechanism, consistently outperforms the Singular Approach across all tested metrics. This suggests that the combination of technical and sentiment data with sophisticated attention mechanisms captures more profound market insights. The SA_BA model's capacity to balance SA and cross-modal attention leads to improved accuracy, reliability, and directional awareness for all currency pairs. The value of a multimodal, attention-integrated model in capturing complex, interdependent market patterns for enhanced predictive accuracy is supported by these results.

To validate the effectiveness of our proposed multimodal Forex trading model, we compare its performance with state-of-the-art approaches presented in the literature. Table 14 highlights key distinctions between our methodology and related studies. It is important to note that while the datasets used in the comparisons focus on the same main currency pairs, the time periods analyzed differ across studies, which may impact performance metrics. Our approach stands out due to its

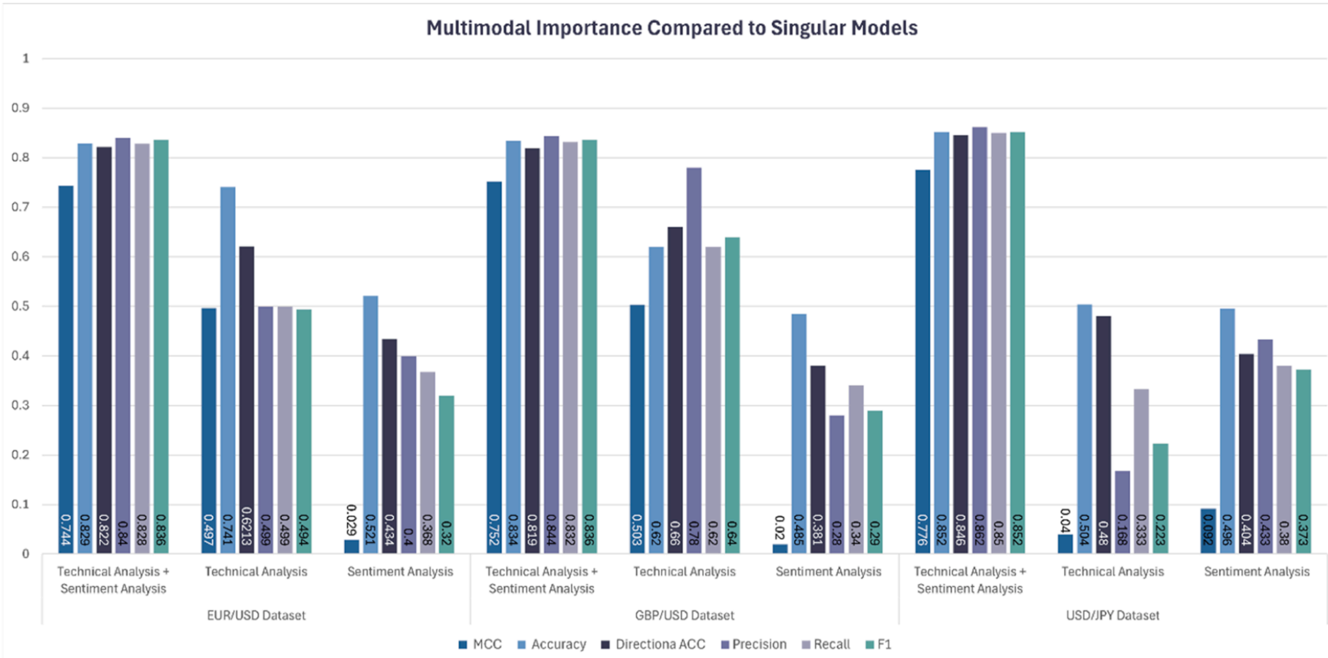

Fig. 19. Multimodal importance compared to singular models.

Table 14
Comparison of our proposed methodology with related work.

Refs.	Method Key Features	Fusion	Proposed Model's Advancements	Reference Performance	Proposed Model's Performance
(Semiromi et al., 2020)	Unified dataset fusion using Gradient Boosting, Random Forest, and SVMs; no explicit fusion method described.	Feature Concatenation: News events, sentiment scores, and technical indicators combined.	Employs SA and bi-cross attention mechanisms to capture deeper interdependencies between news sentiment and technical indicators.	Accuracy: ~66 %	Accuracy **85.2** % MCC: **0.776** Directional Accuracy: **84.6** % AUC **0.955**
(Yıldırım et al., 2021)	Separate LSTMs for macroeconomic indicators and technical indicators; outputs combined via rule-based fusion.	Rule-Based Decision Fusion: Outputs from ME-LSTM and TI-LSTM combined using confidence-based rules.	Integrates modalities in a single architecture via cross-attention, enabling joint learning and interpretability.	Accuracy: ~79.42 %	
(Pornwattanavichai et al., 2022)	Cascading model: Fundamental data processed by BERT1, fused with technical indicators via an autoencoder, refined through BERT2.	Hierarchical fusion: Fundamental data processed with BERT (Stage 1) fused with technical indicators via autoencoder and further refined with BERT (Stage 2).	Utilizes SA and bi-cross attention mechanisms for direct integration of sentiment and technical indicators.	Accuracy: 84.38 %	
(Hajek and Novotny, 2022)	FinBERT sentiment scores combined with technical indicators and external factors for classification.	Not clearly stated: Combines FinBERT sentiment scores with technical indicators, but fusion methodology is unspecified.	Incorporates cross-attention mechanisms for nuanced interaction between modalities, offering improved directional accuracy and robustness.	AUC: ~0.707	
(Nassirtoussi et al., 2015)	Multi-layer algorithm: Integrates semantic abstraction, sentiment analysis, and targeted feature reduction.	Hierarchical Fusion: a multi-layer algorithm that sequentially integrates semantic abstraction, sentiment analysis, and feature reduction to enhance prediction accuracy.	Combines SA and bi-cross attention mechanisms for deeper alignment of modalities, resulting in a more cohesive integration of sentiment and technical analysis.	Accuracy: ~83.33 %	

integration of self-attention and bi-cross attention mechanisms, enabling the model to capture interdependencies between news sentiment and technical indicators effectively. These enhancements contribute to improved accuracy, MCC, and DA when forecasting Forex price movements, demonstrating the model's robustness and predictive capabilities.

6. Conclusion

This research presents a comprehensive method for forecasting Forex price fluctuations by combining technical indicators with sentiment analysis derived from news, employing a multi cross-modal deep learning model. To the best of our knowledge, this is the first thorough use of cross-modal attention mechanisms for Forex prediction. It combines qualitative sentiment data with quantitative data from technical analysis to produce a novel combination of market dynamics. This multimodal Forex trading approach utilizes textual narratives from news sources alongside structured financial data to increase forecasting capabilities in the complicated and unstable Forex market.

Our method employs a cross-attention framework that indicates the complex links between historical price data and sentiment derived from news. We confirmed the model's generalizability and performance by testing it across multiple currency pairs: EUR/USD, GBP/USD, and JPY/USD. The findings indicated that our multimodal model, especially with the self-attention and Bi-Cross Attention module, markedly surpassed single-modality models depending exclusively on technical indicators or sentiment, validating the benefits of a comprehensive approach to Forex prediction. This research underscores the alignment issues between pricing and sentiment, a novel finding that enhances its methodology's reliability. This research primarily contributes a multimodal Forex prediction model that enhances directional accuracy by identifying latent interactions between market sentiment and price trends. This model architecture demonstrates the effectiveness of cross-modal attention in financial forecasting and offers a scalable solution adaptable to various currency pairs and market situations.

Although this study focuses on Forex markets, the proposed approach can be applied to other financial prediction domains, such as stocks and commodities, with minor changes. By adding appropriate sentiment sources (e.g., earnings reports, social media) and changing labeling criteria to suit domain-specific characteristics, this multimodal technique has the potential to improve predictions in broader financial applications. Future research could investigate cross-domain validation to analyze performance and the necessary modifications across different financial markets.

Future iterations could also focus on experimenting with alternative preprocessing approaches to enhance model robustness and performance. For example, subword-level tokenization and pre-trained contextual embeddings. In addition, feature-specific scaling methods could be explored to ensure that technical indicators with different ranges, such as RSI and MACD, are normalized in a way that preserves their unique interpretations. Another direction could involve dynamic window adaptation for rolling standard deviation-based labeling, adjusting the window size according to changing market conditions to better capture both long-term and short-term trends. By refining these preprocessing steps, future research could provide more interpretable, robust, and accurate predictions, enhancing the practical utility of the proposed multimodal model. Additionally, future directions for this work could involve the development of a front-end system to enable online, real-time testing and assessment. Moreover, using financial indicators such as profit, Sharpe ratio, and other risk considerations would improve the practical evaluation. Implementing these features could further test the model's stability, rendering it a viable instrument for real-world Forex trading applications.

CRediT authorship contribution statement

Fatima Dakalbab: Methodology, Data curation, Resources, Formal analysis, Writing – original draft, Validation, Visualization, Investigation, Software. **Ayush Kumar:** Methodology, Resources, Visualization, Data curation, Investigation, Writing – original draft, Software. **Manar Abu Talib:** Conceptualization, Funding acquisition, Writing – review & editing, Validation, Supervision, Project administration. **Qassim Nasir:** Conceptualization, Investigation, Validation, Supervision, Resources, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Dr. Manar Abu Talib, the corresponding author, along with her co-authors, would like to express their appreciation to the University of Sharjah and the OpenUAE Research and Development Group for their support of this research. Our research assistants also contributed to conducting the tests, reporting, and analyzing the results.

Data availability

Data will be made available on request.

References

Anon. (2022). MMDL: A novel multi-modal deep learning model for stock market prediction. In *Proceedings of the IEEE 9th international conference on data science and advanced analytics (DSAA)* (pp. 1–2). https://doi.org/10.1109/DSAA54385.2022.10032436

Antelmi, L., Ayache, N., Robert, P., & Lorenzi, M. (2019). Sparse multi-channel variational autoencoder for the joint analysis of heterogeneous data. In *Proceedings of the international conference on machine learning* (pp. 302–311).

Baek, S., Glambosky, M., Oh, S. H., & Lee, J. (2020). Machine learning and algorithmic pairs trading in futures markets. *Sustainability (Switzerland), 12*(17). https://doi.org/10.3390/SU12176791

Barbieri, F., Camacho-Collados, J., Neves, L., & Espinosa-Anke, L. (2020). TWEETEVAL: Unified benchmark and comparative evaluation for tweet classification. *Findings of the association for computational linguistics findings of ACL: EMNLP 2020* (pp. 1644–1650). https://doi.org/10.18653/V1/2020.FINDINGS-EMNLP.148

Bintsi, K.-M., Baltatzis, V., Potamias, R. A., Hammers, A., & Rueckert, D. (2023). Multimodal brain age estimation using interpretable adaptive population-graph learning. In *Proceedings of the international conference on medical image computing and computer-assisted intervention* (pp. 195–204).

Chantarakasemchit, O., Nuchitprasitchai, S., & Nilsiam, Y. (2020). Forex rates prediction on EUR/USD with simple moving average technique and financial factors. In *Proceedings of the 17th international conference on electrical engineering/electronics, computer, telecommunications and information technology, ECTI-CON 2020* (pp. 771–774). Institute of Electrical and Electronics Engineers Inc.. https://doi.org/10.1109/ECTI-CON49241.2020.9157907

Chen, Q., Li, M., Chen, C., Zhou, P., Lv, X., & Chen, C. (2023). MDFNet: Application of multimodal fusion method based on skin image and clinical data to skin cancer classification. *Journal of Cancer Research and Clinical Oncology, 149*(7), 3287–3299. https://doi.org/10.1007/S00432-022-04180-1/METRICS

Cui, C., et al. (2023). Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: A review. *Progress in Biomedical Engineering, 5*(2), Article 022001. https://doi.org/10.1088/2516-1091/ACC2FE

Dakalbab, F. M., Talib, M. A., & Nasir, Q. (2023). Machine learning-based trading robot for foreign exchange (FOREX). *Lecture Notes in Networks and Systems, 721*, 196–210. https://doi.org/10.1007/978-3-031-35308-6_17/COVER

Dammu, H., Ren, T., & Duong, T. Q. (2023). Deep learning prediction of pathological complete response, residual cancer burden, and progression-free survival in breast cancer patients. *PLOS ONE, 18*(1), Article e0280148. https://doi.org/10.1371/JOURNAL.PONE.0280148

Dang, Q. V. (2019). Reinforcement learning in stock trading. *Advances in Intelligent Systems and Computing, 1121*, 311–322. https://doi.org/10.1007/978-3-030-38364-0_28

Dautel, A. J., Härdle, W. K., Lessmann, S., & Seow, H.-V. (2020). Forex exchange rate forecasting using deep recurrent neural networks. *Digital Finance, 2*(1–2), 69–96. https://doi.org/10.1007/s42521-020-00019-x

Dymova, L., Sevastjanov, P., & Kaczmarek, K. (2016). A FOREX trading expert system based on a new approach to the rule-base evidential reasoning. *Expert Systems with Applications, 51*, 1–13. https://doi.org/10.1016/j.eswa.2015.12.028

Farimani, S. A., Jahan, M. V., Fard, A. M., & Haffari, G. (2021). Leveraging latent economic concepts and sentiments in the news for market prediction. In *Proceedings of the IEEE 8th international conference on data science and advanced analytics (DSAA)* (pp. 1–10).

Farimani, S. A., Vafaei Jahan, M., Milani Fard, A., & Tabbakh, S. R. K. (2022). Investigating the informativeness of technical indicators and news sentiment in financial market price prediction. *Knowledge-Based Systems, 247*, Article 108742. https://doi.org/10.1016/j.knosys.2022.108742

Farimani, S. A., Jahan, M. V., & Fard, A. M. (2024). An adaptive multimodal learning model for financial market price prediction. *IEEE Access, 12*, 121846–121863. https://doi.org/10.1109/ACCESS.2024.3441029

Fataliyev, K., & Liu, W. (2023). MCASP: Multi-modal cross attention network for stock market prediction. In *Proceedings of the 21st annual workshop of the Australasian language technology association* (pp. 67–77).

Faulkner, A. E. (2017). Global financial data. *Journal of Business & Finance Librarianship, 22*(1), 61–67. https://doi.org/10.1080/08963568.2017.1264051

Ferreira, F. G. D. C., Gandomi, A. H., & Cardoso, R. T. N. (2021). Artificial intelligence applied to stock market trading: A review. *IEEE Access, 9*, 30898–30917. https://doi.org/10.1109/ACCESS.2021.3058133

Gai, K., Qiu, M., & Sun, X. (2018). A survey on FinTech. *Journal of Network and Computer Applications, 103*, 262–273. https://doi.org/10.1016/J.JNCA.2017.10.011

Golovanevsky, M., Eickhoff, C., & Singh, R. (2022). Multimodal attention-based deep learning for Alzheimer's disease diagnosis. *Journal of the American Medical Informatics Association, 29*(12), 2014–2022. https://doi.org/10.1093/jamia/ocac168

Grądzki, P., & Wójcik, P. (2024). Is attention all you need for intraday FOREX trading? *Expert Systems, 41*(2). https://doi.org/10.1111/EXSY.13317

P. Hajek, J. Novotny, J. K. the 2022 6th I. C., and undefined 2022, "Predicting exchange rate with FinBERT-based sentiment analysis of online news," dl.acm.org, pp. 133–138, Oct. 2022, doi: 10.1145/3572647.3572667.

King, M. R., Osler, C., & Rime, D. (2012). Foreign exchange market structure, players, and evolution. *Handbook of exchange rates* (pp. 2–44). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118445785.ch1

Leles, M. C. R., Sbruzzi, E. F., De Oliveira, J. M. P., & Nascimento, C. L. (2019). Trading switching setup based on reinforcement learning applied to a multiagent system simulation of financial markets. In *Proceedings of the SysCon 2019 - 13th annual IEEE international systems conference*. Institute of Electrical and Electronics Engineers Inc.. https://doi.org/10.1109/SYSCON.2019.8836887

Li, W., Wong, M. C. S., & Cenev, J. (2015). High frequency analysis of macro news releases on the foreign exchange market: A survey of literature. *Big Data Research, 2*(1), 33–48. https://doi.org/10.1016/j.bdr.2015.02.003

Li, Y., Ni, P., & Chang, V. (2020). Application of deep reinforcement learning in stock trading strategies and stock forecasting. *Computing, 102*(6), 1305–1322. https://doi.org/10.1007/s00607-019-00773-w

Li, Q., Tan, J., Wang, J., & Chen, H. (2020). A multimodal event-driven LSTM model for stock prediction using online news. *IEEE Transactions on Knowledge and Data Engineering, 33*(10), 3323–3337.

Ma, N., et al. (2023). Comprehensive investigating of MMR gene in hepatocellular carcinoma with chronic hepatitis B virus infection in Han Chinese population. *Frontiers in Oncology, 13*. https://doi.org/10.3389/fonc.2023.1124459

Munkhdalai, L., Munkhdalai, T., Park, K. H., Lee, H. G., Li, M., & Ryu, K. H. (2019). Mixture of activation functions with extended min-max normalization for FOREX market prediction. *IEEE Access, 7*, 183680–183691. https://doi.org/10.1109/ACCESS.2019.2959789

Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2015). Text mining of news-headlines for FOREX market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment. *Expert Systems with Applications, 42*(1), 306–324.

Pornwattanavichai, A., Maneeroj, S., & Boonsiri, S. (2022). BERTFOREX: Cascading model for FOREX market forecasting using fundamental and technical indicator data based on BERT. *IEEE Access, 10*, 23425–23437. https://doi.org/10.1109/ACCESS.2022.3152152

Ramachandram, D., & Taylor, G. W. (2017). Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Processing Magazine, 34*(6), 96–108. https://doi.org/10.1109/MSP.2017.2738401

Semiromi, H. N, Lessmann, S., & Peters, W. (2020). News will tell: Forecasting foreign exchange rates based on news story events in the economy calendar. *North American Journal of Economics and Finance, 52*, Article 101181. https://doi.org/10.1016/j.najef.2020.101181

Townend, F., Roddy, P. J., & Goebl, P. (2024). *florencejt/fusilli: Fusilli v1.1.0 (v1.1.0)*. Zenodo. https://doi.org/10.5281/zenodo.10463697.

Vaswani, A., Noam, S., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.)*, 30. Advances in neural information processing systems*. Curran Associates, Inc.

Windsor, E., & Cao, W. (2022). Improving exchange rate forecasting via a new deep multimodal fusion model. *Applied Intelligence, 52*(14), 16701–16717.

Yang, S. Y., Yu, Y., & Almahdi, S. (2018). An investor sentiment reward-based trading system using Gaussian inverse reinforcement learning algorithm. *Expert Systems with Applications, 114*, 388–401. https://doi.org/10.1016/J.ESWA.2018.07.056

Yang, H. (2023). Multimodal stock price forecasting using attention mechanism based on multi-task learning. In *Proceedings of the Asia-Pacific web (APWeb) and web-age information management (WAIM) joint international conference on web and big data* (pp. 454–468).

Yıldırım, D. C., Toroslu, I. H., & Fiore, U. (2021). Forecasting directional movement of FOREX data using LSTM with technical and macroeconomic indicators. *Financial Innovation, 7*(1), 1–36. https://doi.org/10.1186/s40854-020-00220-2

Zeng, Z., & Khushi, M. (2020). Wavelet denoising and attention-based RNN- ARIMA model to predict FOREX price. In *Proceedings of the international joint conference on neural networks*. Institute of Electrical and Electronics Engineers Inc.. https://doi.org/10.1109/IJCNN48605.2020.9206832