A FOREX Trading System Based on Semi-Supervised News Classification, Market Sentiment Analysis, and GRU-CNN Deep Learning Models

Nabil MABROUK, Marouane CHIHAB, Younes CHIHAB
Computer Sciences Laboratory, Ibn Toufail University, Kenitra, Morocco

***Abstract*—Investors access the foreign exchange market (FOREX) not only to preserve their wealth but also to generate profits and achieve specific financial goals. It is one of the largest financial markets that investors rely on, and it is based on fluctuations in currency exchange rates to make a profit in different time cycles: short-term, medium-term, and long-term. In this article, we propose an automated FOREX trading system that combines two artificial intelligence algorithms: the first to classify news by pertinence (semi-supervised) and then to analyze market sentiment. This algorithm plays a crucial role in replacing the traditional fundamental analysis, which is based on macroeconomic factors, political events, and news headlines. The use of GAN-BERT helped improve performance in classification tasks with limited labeled data and reduced execution time. This algorithm demonstrates impressive results, achieving a high accuracy of 97.5%, which makes its output data more reliable for use in the second algorithm, a combination of two deep learning models: the Gated Recurrent Unit (GRU) and the Convolutional Neural Network (CNN). We enrich the dataset used in this phase with additional technical indicators and features that may help explain market fluctuations. We evaluated our final algorithm over multiple time frames and several windows; the results were impressive and confirmed by back-testing its potential profitability and risk.**

***Keywords*—FOREX; trading; semi-supervised classification; sentiment analysis; machine learning; deep learning; RNN; CNN; GRU**

I. Introduction

The Forex market, which is known as the largest financial market in the world [1], [2], [3], presents a profitable opportunity for investors with solid risk management and a good understanding of macroeconomic and political factors, as well as geopolitical events and technical analysis. Those factors cause the creation of price variations and exchange rates. However, to grow and preserve their wealth, investors take risks and participate in the FOREX market to take advantage of currency fluctuations. Due to geographical dispersion and 24-hour availability, investors can trade currencies at any time from different time zones around the world, resulting in high liquidity and a large volume of trades [4], [5]. Many national and international companies, individuals, and institutional investors utilize exchange rate volatility to achieve specific financial objectives [6]. However, to mitigate risks and avoid emotional investment, traders have developed a range of strategies based on various types of analysis.

A. Traditional FOREX Analysis

Since the early twentieth century, the world has witnessed the emergence of what we now refer to as traditional trading strategies [7]. They were deployed and adapted by traders to suit their trading strategy styles, as well as a few other tools that traders have found helpful in previous trading. Over the years, traditional Machine Learning (ML) methods have shown a strong ability to predict the trend of stock prices [8], [9], [10]. FOREX investors can be divided into two categories. The first relies on analyzing historical data to calculate indicators that can help explain price variation, then using them to predict price movement; but unfortunately, many factors influence the FOREX market and make it highly volatile, such as political events, economic data, world news, and market sentiments [11].

1) Technical analysis: Technical analysis is generally based on various techniques, including the use of points and graphs, such as candlestick graphs. It generated significant excess returns, at least until the early 1990s [12], [13], [14], [15]. According to [16] by Rico Nur et al., this type of analysis utilizes historical data to generate new technical indicators based on mathematical calculations. These indicators can reduce the noise created by high fluctuations and can also help detect patterns in past prices. These patterns form the foundation for creating trading rules that are important for generating trading signals that are (buy, sell, and hold) [17]. In fact, technical analysis is widely used, particularly in the foreign exchange market [18]. However, nowadays there exists an exhaustive list of technical indicators, but traders do not use them all at once; they use a combination of them that go with their trading style, and there is evidence that, at least until the early 1990s, it generated significant excess returns.

2) Fundamental analysis: Another way to forecast financial markets is to analyze the causes of those fluctuations. Expert investors examine macroeconomic and political factors, geopolitical events, financial news, market sentiments, and other factors to predict stock prices or trends [11], [19]. Generally, fundamental analysis is used in long-term predictions; investors choose currencies and stocks with strong economic indicators and favorable political climates for investment.

3) Sentiment analysis: The primary drivers of currency market fluctuations are news, economic indicators, and market sentiment [20]. Analyzing the general attitude of traders toward a particular FOREX market helps to understand the market's behavior, making informed decisions, and minimizing the risk of loss. Optimistic emotions and expectations drive FOREX prices up, but pessimistic sentiment and news releases amplify the propensity to sell. Sentiment analysis increases the prediction accuracy of artificial intelligence algorithms [21].

In this study, we propose a comprehensive solution that leverages the three types of analysis, built on Algorithmic Trading techniques. Our system can automatically forecast and identify future price directions, trends, and cycles that characterize a FOREX dataset (time series dataset). In addition, it performs the optimal trade decision. Traditional trading strategies, data mining, and deep learning algorithms primarily rely on historical data that track price movements on the FOREX (OHLCV). These data are also enriched by adding more features based on technical indicators and trading rules for automatic or semi-automatic pattern recognition. In addition to that and to compensate for the fundamental analysis part that requires expert intervention, our proposed system in the first phase utilizes news articles and sentiment analysis to improve prediction accuracy. It utilizes a semi-supervised algorithm to resolve the issue of high-volume unlabeled data and also reduces time consumption. However, the primary goal of it is to classify and select the pertinent articles.

II. Related Works

In the past, the focus of many researchers has been concentrated on forecasting currency exchange rates, employing a variety of techniques and methodologies within this field. However, successful trading strategies have increasingly drawn attention to fundamental analysis. In this article, we will explore AI techniques, specifically deep learning and machine learning algorithms, while considering the three types of analysis. The first is technical analysis, grounded in mathematical calculations, whereas the second is fundamental analysis, which pertains to economic and political news, as well as other influencing factors [22]. The third is sentiment analysis, which can help improve forecasting accuracy [23].

Currency trading systems have been developed, allowing traders to execute all trading operations electronically through trading platforms [11]. This fact led to the availability of a vast volume of historical data, sparking an interest among researchers in resolving financial market problems using computational intelligence [24].

In 1967, Poole and, for the first time, utilized technical indicators to forecast Currency fluctuations [25], [26]. According to Poole and Dooley's viewpoint, currency exchange trading experiences are repeated over time, and historical data of past price trends can help shape them. They proved that, based on specific rules, traders could recognize past patterns and make profits; these rules enable traders to make informed decisions about when to buy or sell.

In addition to traditional trading strategies, another type of analysis has emerged, thanks to computer science, particularly artificial intelligence, and the high volume of historical data [27], resulting in numerous studies in this domain. The utilization of Data mining algorithms is for automatic or semi-automatic pattern recognition, establishing the connection and impact of different data [28], [29]. Machine learning (ML) algorithms have demonstrated significant efficacy in solving real-world categorization problems across multiple domains [30]. A variety of research teams have leveraged machine learning techniques to formulate trading strategies (Logistic regression, Random Forest, support vector machines, Neural Networks, etc.) [31].

In [22], using a linear regression model, they implemented a Forex forecasting system based on a sliding window of ten to 240 days of information on the seven major currency pairs; the model gets better results with the USD/CHF, USD/CAD, and EUR/USD, it was able to identify momentum opportunities to buy or sell and to decide if it is indifferent to buy or sell a currency. However, when it comes to classification problems, one of the classical machine learning methods is the Support Vector Machine (SVM), which still shows impressive results in solving significant classification problems, despite being mathematically complex and computationally expensive [32]. Many researchers have demonstrated the importance of this model in Forex trading. [33] Experimented with the advantages of the SVM model in forecasting Forex trends as a binary classification problem (uptrend or downtrend); The results showed that SVM models might help traders make automatic transaction decisions regarding Bid/Ask in the Forex Market. Additionally, Random Forest (RF) is widely used to resolve classification problems, especially in decision-making; it was introduced by Breiman in [34], who utilized bagging to enhance classification accuracy. In [35], the random forest algorithm is described as the top-performing algorithm for accurately predicting stock price directions. A small number of studies have been conducted on forecasting currency exchange rates using a random forest algorithm; one of these studies was carried out by [36]. In this article, the authors adopted the random forest algorithm to develop a trading strategy, and the proposed algorithm outperforms traditional technical indicators.

In recent years, the computing techniques of artificial intelligence, specifically Deep Neural Networks, have achieved significant advancements in natural language processing, image classification, voice translation, and other areas [37]. It is essential to highlight that specific Deep Neural Networks algorithms have been utilized for forecasting time series data and quantitative trading [38], [39], [40]. Deep learning algorithms have demonstrated improved performance and accuracy, yielding better returns in the finance sector [41], [42]. The broadening application of artificial intelligence has led to an increasing number of investors using deep learning models to predict and study stock and Forex prices. In [43], it has been demonstrated that fluctuations in stock and Forex prices can be predicted. Unlike traditional statistical and econometric models, deep learning can effectively describe complex influencing factors. The Forex, like any financial market, can be impacted by investor sentiment, especially fear, which leads to increased volatility or a shift towards safe-haven currencies, among other effects. However, numerous studies in the field of natural language processing (NLP) have studied sentiment analysis of textual content as a key to identifying public opinion and forecasting market trends. The majority of sentiment analysis techniques assess sentiment by calculating the polarity of words (positive, negative, neutral) or emotional words (such as happy, sad, like, dislike, etc.) that occur in textual documents [44]. In [45], they found an association between tweet sentiment and stock returns. They utilize a simple machine learning algorithm (Naive Bayesian) to transform tweets into signals indicating whether to buy, hold, or sell. In [46], they correlate Facebook's daily sentiment with trading behavior. They exploit Facebook's Gross National Happiness Index (FGNHI), which measures people's mood by examining the positive and negative terms used by Facebook participants. In [46], they employed Facebook's daily sentiment index to prove the existence of a correlation between trading behavior and Facebook participants' mood. They exploit Facebook's Gross National Happiness Index (GNHI), which relates to positive and negative terms used. In [47], they utilized sentiment analysis to extract opinions from user-generated reviews, merging sentiment with content-based recommendation techniques to provide personalized product recommendations. They demonstrate the advantages of their method in various categories of Amazon products.

Recently, text classification algorithms have received increased attention in the literature thanks to the growing number of available resources. Text classification is the process of automatically organizing a set of documents into one or more predetermined or undetermined categories based on their topics [48]. On the same subject, the primary objective of supervised learning is to develop methods for classifying natural language processing documents [49]. The general challenge of text categorization can be subdivided into various sub-challenges, such as sentiment classification, functional classification, subject classification, and other types of classification [50]. In [51], an algorithm is created by integrating natural language processing, statistical pattern identification, and sentiment analysis to develop a system that forecasts the directional trend of a specific currency pair in the FOREX market based on the text of breaking financial news headlines. Their system has shown impressive results with a high accuracy level. Another research project was conducted by [52] to forecast foreign exchange trends based on news articles, specifically financial news, which are rich in essential factors widely used in fundamental analysis. However, they developed their algorithms utilizing the FinBERT language model, which is a specifically pre-trained NLP model geared towards the financial sector, for this initiative. FinBERT is developed by enhancing the BERT language model. Google Researchers were the initial publishers of BERT in [53], which is based on Transformer networks [54]. Its purpose is to pre-train deep bidirectional models from unlabeled text by simultaneously considering both left and right contexts across all layers [53]. BERT impressively improved the state-of-the-art for large language models.

When dealing with text classification problems, two traditional approaches are adopted: supervised and unsupervised learning. Recently, researchers have combined these two approaches into one named semi-supervised learning, which is a machine learning algorithm concerned with using both labeled and unlabeled data to perform specific learning tasks [55]; This method is beneficial when there is a large amount of unlabelled data accessible, but collecting labeled data is costly or time-consuming. It requires less human effort and provides greater precision [56].

III. Our Proposed Solution

Our proposed solution consists of two major phases. The first phase is a news classification and sentiment analysis system based on semi-supervised algorithms. The second phase involves a deep learning system that takes as input the results of the first algorithm, along with other features, to forecast trades and predict future price directions.

A. Data Description

As our approach is a sequence of two subsystems, we have built our strategy utilizing two datasets:

1) Dataset of the first phase: We have collected the dataset from various credible sources, including BBC, CNN, Yahoo Finance, CNBC, DailyForex, and FxStreet. These sources are specifically designed for financial markets and advertise millisecond-level delivery, helping minimize news publication delays at the source. Each row describes an article, identified by date and time, title, content, source, author, category, and subcategory (if labeled). The data cover the period from January 01, 2006, until May 30, 2025. This data set will be the input for a semi-supervised algorithm, which combines a small amount of labeled data, often expensive and time-consuming, with a large amount of unlabeled data.

a) Time-zone handling: The FOREX market operates 24 hours a day, 5 days a week. However, trading activity in the market is not constant throughout the entire day. Instead, trading activity is divided into several distinct sessions, each with its own characteristics and level of activity [57]. The Forex market is divided into three major regions: Australia-Asia, Europe, and North America [58]. During news data collection, to prevent information leakage and avoid inadvertently using future data to train the model, we utilized Coordinated Universal Time (UTC) as a normalized timestamp.

The objective of this phase is to classify news articles. At this stage, chronological order is not essential. To carefully maintain the distribution of topics in each split, we prepared a balanced dataset using the same-sized set of each class. For training, we used 100 examples of labeled data and 900 unlabeled examples of each class. To evaluate the performance of our model, we used 200 examples of labeled data of each class; 100 examples for validation and 100 examples for testing. Resulting in a total of 34,800 examples for the general categories.

2) Dataset of the second phase: We collected data on the two most traded currencies in the world: the US Dollar (USD) and the Euro (EUR). The EUR/USD currency pair represents the exchange rate between the euro (the base currency) and the US dollar (the quote currency), where the euro is used to purchase the US dollar. This dataset is structured as a time series, obtained through repeated measurements over time, such as hourly, daily, or weekly data. Each row of the dataset contains observations for five variables: Open, High, Low, Close, and Volume (OHLCV). The variables in the data set (OHLCV) are continuous. To be clear, this dataset is noisy due to high-frequency trading and redundant information; in a typical scenario, feeding these data as is to an LSTM or GRU model may lead to an overfitting problem during the training phase; and the model will predict the same last value or a value very close to it; it may just memorize the previous data without learning. In our work, we tried to eliminate the noisy part of the dataset by first forecasting the next trend direction (classification problem) instead of the next price value (regression problem); the second step is training our model in different time cycles (short, medium, and long-term); the third is Minimizing the noisy parts of time series datasets; it can be achieved by utilizing a sliding window approach, which incorporates previous time steps as input. For example, a window size of four, which means that the model uses the current time step along with the three preceding time steps as input; see Fig. 1 for an illustrated example. In our article, we explore various window lengths, ranging from 1 to 15, to identify the most effective window size for our model. The fourth step involves generating additional features, specifically technical indicators, that are widely used to understand and analyze price movements in a specific market. However, technical indicators give an idea of where the price might go next in a given market at a specific time. We have used a large set of technical indicators generated over multiple periods with different parameters:

- The Weighted, Exponential, Simple, and Convergence Divergence Moving Average (WMA, EMA, SMA, and MACD).
- The Relatively Strength Index (RSI).
- The average directional index (ADX).
- The Commodity Channel Index (CCI).
- The Rate-of-Change (ROC).
- The Bollinger Band (BB), Moving Polynomial trending (MPT).

The goal of adding technical indicators is to improve the accuracy of our model; however, using a large set of features may lead to the opposite result. To avoid this, we reduced the high dimensionality of the dataset by retaining only the features that provide the most information and contribute the most to the prediction. According to Guyon [59], the objectives of feature selection are to reduce prediction time consumption, facilitate the understanding of predictors, and improve model accuracy. In [60], the authors have classified the majority of current feature selection algorithms into four primary categories: similarity-based methods, information-theoretic methods, sparse learning methods, and statistical methods.

The header of our final dataset is: DateTime, OHLCV variables, a set of selected technical indicators, sentiment_label, article_sentiment_score, article_source, Article_author.

3) Data splitting and prevention of information leakage: To prevent information leakage and ensure a realistic evaluation, we employed a strict chronological split for our data. The training set consisted of data from January 1, 2006, to December 31, 2019 (approximately 70% of the data). The validation set spanned January 1, 2020, to December 31, 2021 (approximately 10% of the data), and the test set covered January 1, 2022, to May 30, 2025 (approximately 20% of the data). Crucially, no data from the validation or test sets was used during training or feature engineering. All preprocessing steps were performed solely on the training data, and then the resulting transformations were applied to the validation and test sets. By maintaining a strict chronological order, we were able to more accurately assess the model's ability to generalize to unseen future data.

B. Phase 1: Semi-Supervised News Classification and Sentiment Analysis

Fundamental analysis is critical for FOREX traders. However, expert traders evaluate macroeconomic and political factors, fiscal policies, and other relevant factors to gain insight into future currency movements and develop strategies to capitalize on market trends. The high sensitivity to financial news makes the Forex market a risky environment for investors due to its high volatility and leverage, which can create fear in investors. This fear is an emotional driver that can amplify market volatility. In this phase, we have created two algorithms. The first algorithm utilizes news headlines to classify them based on their relevance, retaining only those that have an impact on FOREX trends. The second algorithm analyzes these news headlines and generates sentiment-based features.

1) GAN-BERT model for news classification: Let C be a set of categories, where n is the total number of categories in C, as shown in Eq. (1).

$$C = \{c_1, c_2, c_3, ..., c_n\} \tag{1}$$

The Article set belonging to each category forms the complete dataset where a_{ci} is the Article set attributed to the category a_i in the dataset; see Eq. (2).

$$A = \{a_{c1}, a_{c2}, a_{c3}...a_{cn}\} \tag{2}$$

Given a new text, "a_u" of an unknown category "c", the proposed model assigns the text to the most likely category from C. In [61], two models, a BERT-based model and a semi-supervised GAN [62], were combined to create a new model named GAN-BERT. Fig. 2 describes the architecture of the GAN-BERT model. As illustrated in the architecture of the GAN-BERT model (Fig. 2), the discriminator D classifies samples, while the generator G generates fake examples F. The discriminator utilizes the vector representations produced by BERT for unlabeled U and labeled L input texts. After training is complete, G is removed from the model, allowing the remaining components to be used for inference. The GAN-BERT model typically demonstrates better performance in classification tasks when there is a limited amount of labeled data. Another reason for employing GAN-BERT for news classification is that it takes into account not only news writing styles, but also potential fake writing styles generated by the fake data produced by the generator.

2) FinBERT for news sentiment analysis: The specialized language and the scarcity of labeled data in the financial market domain hinder the effectiveness of general sentiment analysis models. The author in [63], proposed FinBERT, a BERT-based model designed to address NLP tasks specific to the financial sector. In our study, we fine-tune FinBERT using a small training set to create a new feature that enriches our datasets with sentiment labels for the FOREX market (positive, negative, and neutral). It is important to note that sentiments can be expressed by both experts and beginners. Therefore, we

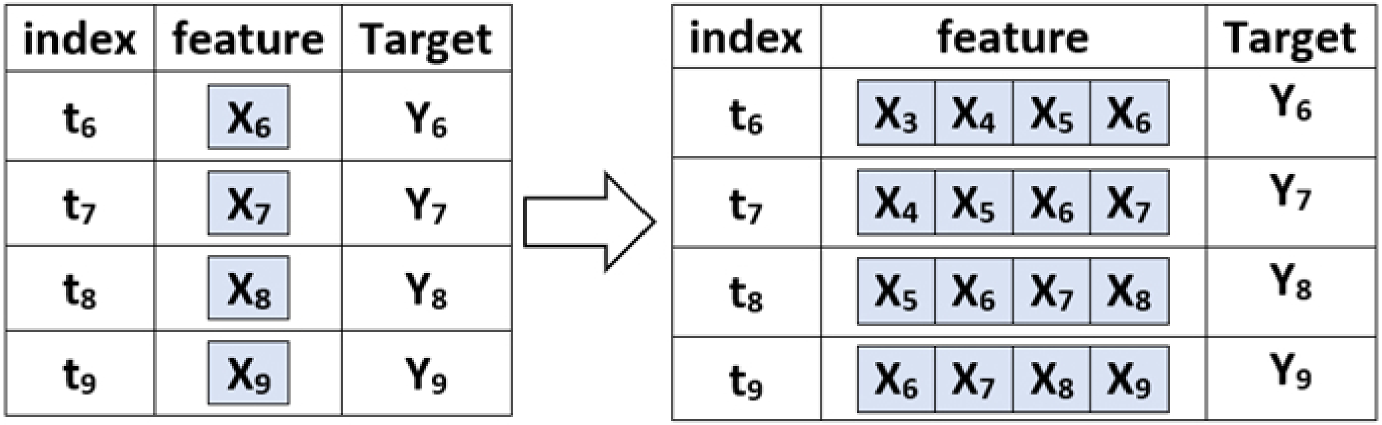

Fig. 1. Example of sliding-window approach.

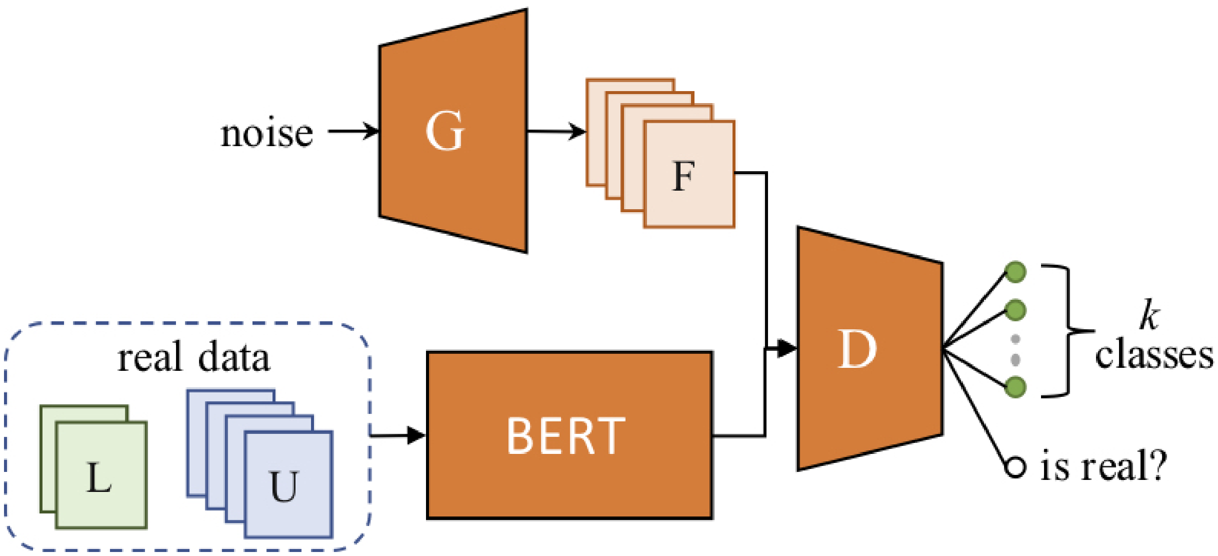

Fig. 2. GAN-BERT architecture.

will also consider the authors of the articles and assign a score to each based on their accuracy in making correctly expressed predictions.

3) GanBERT configuration and training hyperparameters:

- Base Transformer Model: BERT-base-cased.
- Discriminator Architecture. Consisted of the pre-trained BERT-base-cased model, fine-tuned for two tasks:
  - Binary Classification (Real vs. Fake): Distinguishing real news articles from generated articles.
  - Multi-Class Topic Classification: Classifying real news articles into one of the [Number of Topics] predefined topic categories.
- Output layer:
  - Real vs. Fake: Linear layer with sigmoid activation function.
  - Topic Classification: A separate linear layer with a Softmax activation function (for the multi-class classification task).
- Generator Architecture: A multi-layer perceptron (MLP) that takes random noise as input and generates synthetic feature vectors designed to resemble those of minority-class news articles. This helps to improve the discriminator's ability to classify minority topics, improving the accuracy of classification training.
- Layers: Two fully connected layers with 512 units, followed by a linear layer projected to the BERT embedding dimension (768).
- Activation Function: LeakyReLU activation (with a negative slope of 0.2) was used for the hidden layers of the generator.
- Training hyperparameters: Table I summarizes the key hyperparameters used in training the GanBERT

model for news classification. These parameters were carefully chosen to optimize the performance of the model and prevent overfitting.

TABLE I. GanBERT Training Hyperparameters

Hyperparameter	Value
Learning Rate (Disc.)	7e-7
Learning Rate (Gen.)	5e-4
Optimizer	Adam
Batch Size	64
Latent Dimension	100
Discriminator Dropout	0.2
Generator Dropout	0.2
Epochs	50
Early Stopping Patience	6
Loss Function (Disc.)	Binary Cross-Entropy
Loss Function (Class.)	Categorical Cross-Entropy

4) FinBERT Fine-Tuning Hyperparameters:

- Learning Rate: 2e-5
- Batch Size: 32
- Number of Epochs: 3
- Weight Decay: 0.01
- Optimizer: AdamW with a weight decay of 0.01

C. Phase 2: Forecasting Future Price Direction (CNN-GRU APPROACH)

1) Target variable: To get rid of the FOREX market disturbances and reduce high price variation, we reformulated our problem from regression into classification; In this article, we aim to forecast the next trend direction or next price direction; the price in the original dataset is a continuous variable, we had to generate a new categorical variable named Target based on specific rules related to our strategy; this new variable indicates our final decision, which will be a buy, sell, or hold signal.

Each operation is associated with transaction costs, which can be more expensive than returns. For that issue, we used a threshold called α; When the log return of the current day is inferior to -α, it indicates that it is an uptrend, so it is a SELL signal, and when the log return is superior to α, it indicates that it is a downtrend, so it is a BUY signal; otherwise, it is a HOLD signal. We encoded this feature using one-hot encoding; see Eq. (3). Another application of the threshold α is to address the issue of unbalanced classes within datasets, a challenge that we encounter when α is set too low or too high. However, artificially balancing the dataset can negatively impact sequential datasets. To explore this, we ran our algorithm using various thresholds, starting from zero (which created a binary classification problem) and increasing up to 0.01. Otherwise, when we increased the threshold to 0.01, we again encountered the issue of unbalanced datasets.

Before model training, we addressed the challenge of class imbalance by optimizing the log-returns threshold (α). Through experimentation, we identified a threshold value of 0.0014, which yielded a dataset with approximately equal representation across the three classes: 1598 BUY, 1585 SELL, and 1593 HOLD signals. This nearly balanced distribution was deemed crucial to ensure robust and unbiased model performance.

$$Target = \begin{cases} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}, & \text{if } Log(price_t/Price_{t-1}) > \alpha \quad "BUY" \\ \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}, & \text{if } Log(price_t/Price_{t-1}) < -\alpha \quad "SELL" \\ \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, & \text{Otherwise} \quad \text{"HOLD".} \end{cases} \tag{3}$$

2) Recurrent neural networks LSTM and GRU: To deal with sequential data problems, such as time series or text, researchers have developed the Recurrent Neural Network (RNN) [64], [65], [66], which is a specialized type of neural network that possesses a form of memory that can retain information from previous input. Thanks to the feedback connections in their architecture, which facilitate the retention of information across time steps. RNNs are particularly adept for sequence modeling tasks, where the temporal context of data points is crucial for accurate prediction and understanding.

Long-Short-Term Memory (LSTM), published by [67], [68], is an architecture of recurrent neural networks (RNNs) that can help address the vanishing gradient problem faced by standard RNNs. LSTM can effectively bridge time lags exceeding 1,000 discrete time steps. This capability makes them suitable for tasks where long-range temporal dependencies are critical [67], [68], [69], [70].

A Gated Recurrent Unit (GRU) is a specific type of recurrent neural network unit within the broader category of deep neural networks. It was introduced by [71] to allow each recurrent unit to capture dependencies across multiple time scales adaptively. Recently, recurrent neural networks have shown promising results in various machine learning tasks, particularly when dealing with input and output of variable duration [72]. The GRU is recognized for its effectiveness in sequence learning tasks as it addresses the challenges of vanishing and exploding gradients that are often encountered in traditional recurrent neural networks (RNNs), making it well-suited for learning long-term dependencies.

3) Convolutional Neural Network: From a mathematical operation called convolution, the researchers derived the name of a new type of deep neural network, known as the convolutional neural network (CNN). It involves the linear interaction between matrices. The connectivity patterns of its neurons are modeled after the organization of the visual cortex in animals. This convolution technique enables the recognition of visual patterns in input data [73]. Like other artificial neural networks, CNNs consist of multiple layers, including the convolutional layer, which is the most critical. Each convolutional layer identifies features by sliding various filters over the input matrix and conducting piece-by-piece comparisons [74]. As indicated in [75], the two most significant advantages of convolutional neural networks are their ability to tackle complex tasks with a smaller number of parameters in the ANN and their ability to extract abstract features as data progress through deeper layers.

4) GRU-CNN model architecture: By combining the gated recurrent network and the convolutional neural network approaches, we developed our model. As shown in Fig. 3, it is composed of six layers:

- The input is a 3D tensor of shape (samples, timesteps, features). Each sample is a window of historical data. The time steps are the number of data points in a single window, and the features represent the values of all variables on the days to look back.
- The 1D convolutional layer: It takes the output of the previous layer as input; it automatically learns features by sliding multiple filters over the input matrix to extract local patterns.
  - Filters: 64
  - Kernel Size: 2
  - Activation: ReLU
- The third layer is the GRU layer; After the CNN, a GRU layer processes the output to capture the sequential nature of the patterns within the window. It is implemented in the many-to-one architecture, as shown in Fig. 4, because the goal is to predict the next price direction, not multiple prices. This layer is added to deal with the nature of sequential data.
  - Units: 64
  - Activation: tanh
  - Recurrent Activation: sigmoid
  - Dropout: 0.3
- The Dense Layer 1 (Fully Connected): The GRU output is flattened and fed into a dense layer.
  - Units: 32
  - Activation: ReLU
- The Dense Layer 2 (Fully Connected): The GRU output is flattened and fed into a dense layer.
  - Units: 8
  - Activation: ReLU
- The last layer is the output layer with three units; it provides a single decision (buy, sell, or hold).
  - Units: 3
  - Activation: Softmax

GRU-CNN Training Hyperparameters

- Window size: 1 to 10 days.
- Epochs: 50
- Batch Size: 64
- Learning Rate: 0.0005
- Optimizer: Adam
- Loss Function: Categorical Cross-Entropy
- Early Stopping Patience: 6

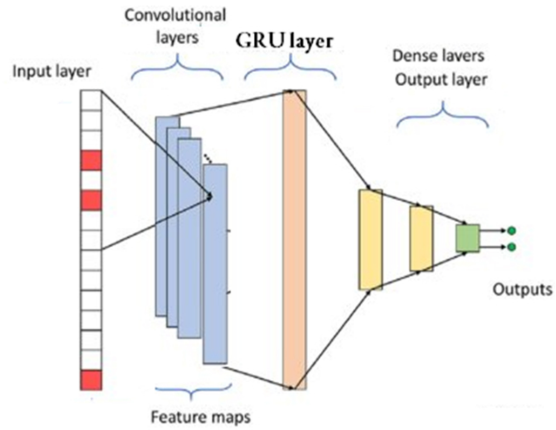

Fig. 3. GRU-CNN architecture.

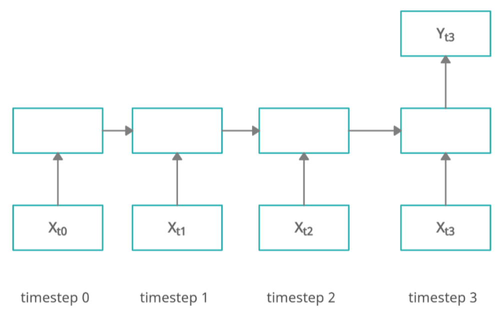

Fig. 4. Many to one GRU approach.

D. Phase 3: Our Investment Strategy and Decision Making

To develop an effective trading strategy, we must first determine the goal of our strategy: to generate short-term profits or build long-term wealth, and identify a personal risk profile or hedge against market volatility. Understanding whether we are more comfortable with conservative, moderate, or aggressive trading approaches will guide our decision-making process. Finally, we must evaluate the available time and resources.

In FOREX trading, the currency exchange rates are generally relatively low. This fact impacts the profitability of investments, particularly when the invested amounts are not significant; The returns may be less than the associated transaction costs. To resolve this issue, a third party, known as brokers in the FOREX market, offers a method called "leverage," which involves borrowing funds to acquire an asset. In our strategy, we employed a leverage of 1:100, allowing us to invest 1,000$ while trading up to 100,000$. Our investment strategy is detailed in Fig. 9. Our aim is to buy when prices are low and sell when they increase, which means that our machine learning model needs to predict the future direction of trades.

To facilitate this, we developed an algorithm that compares the logarithmic return of the current day against a threshold, α. If the return exceeds α, we proceed with a buy; if it falls below $-\alpha$, we opt to sell; and if it lies between $-\alpha$ and α, we hold our position. In this context, the variable we are considering is categorical, so we applied one-hot encoding to prepare the data for our algorithms and fit our model effectively.

In addition to the threshold α, risk management rules are necessary to avoid stacking in the hold position; Stop-loss and take-profit are calculated based on the entry point, unlike the threshold α, which is compared to the log return of the current day.

IV. Discussion and Results

For the news classification phase, which serves as a fundamental analysis, it is the most important and basic phase in our strategy; all other parts of the system rely on it. We trained our model for 50 epochs, with the option of early stopping. As shown in Fig. 5 and 6, the model was well trained. The goal is to classify news articles into general classes, including sports, technology, business, health, politics, and finance. Subsequently, we conducted a secondary categorization to identify subcategories, including stock market, banking, personal finance, investment, and economic indicators. The GAN-BERT model achieved an impressive accuracy of 97.5%, demonstrating both efficiency and performance; see Fig. 6. As illustrated in Fig. 5, the training loss line indicates that the model fits the training data well and can be generalized to new data, as confirmed by the validation loss line.

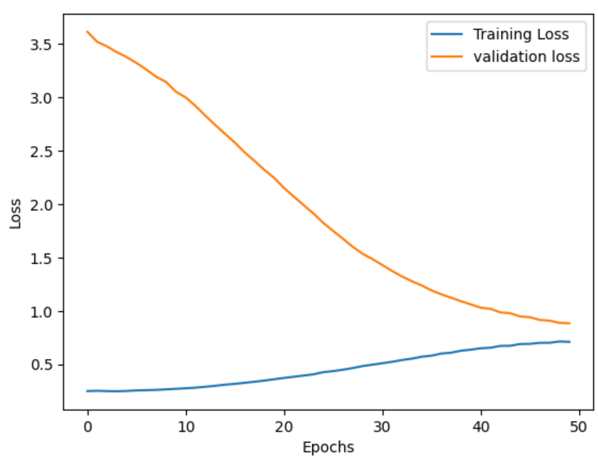

Fig. 5. GANBERT training validation loss.

The high efficiency of this model, as we have seen, makes the classified data a reliable source for the next stage. However, we have improved the persistence of these data by performing another NLP task, represented by fine-tuning the FinBERT model, which is specifically developed for the financial sector. We have executed it to label our dataset with the FOREX market sentiment, this variable is represented with three labels (Positive, Negative, and Neutral); by adding the sentiment feature, our dataset becomes able to describe the market mood and attitude of traders and investors towards the Forex market.

In the second phase, which focuses on forecasting future trends, we employed a combined CNN and RNN model, specifically a CNN and GRU architecture. We trained this model over 50 epochs, varying window lengths (1 day, 2 days, 3 days, 4 days, 5 days and 10 days), as shown in Fig. 10 to 19; by analyzing the training and validation loss graphics, we can see that our model fits the training data well and can be generalized for a larger time frame. Furthermore, the training validation accuracy graphics illustrated promising results in terms of both accuracy and precision.

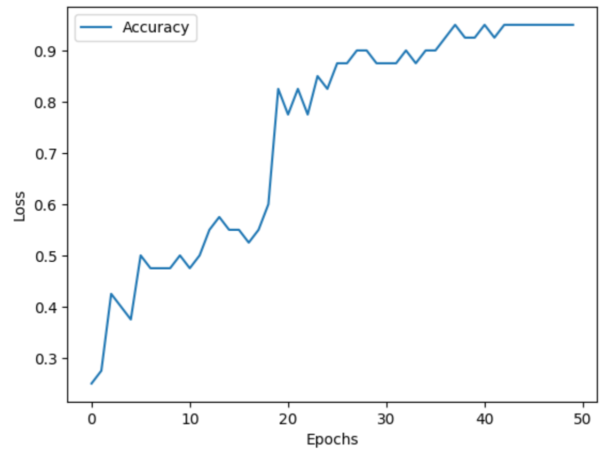

Fig. 6. GANBERT validation accuracy.

We conducted the experiment of our model over several time frames (1 day, 5 days, 10 days, and 15 days), as outlined in Table II. In general, the model achieved higher accuracy in forecasting the future trend direction.

In forecasting the price direction for the next day, our model was less accurate when using a window length of one day, achieving only 58% accuracy. This is due to the time it takes for the event to impact the FOREX market and the number of news articles that we used. However, when we used a window length of two days, we observed a slight improvement in the performance of the model, achieving an accuracy of 80%. Although using a 3-day window length, the results are more promising, achieving 83% accuracy, and continue to improve until a 10-day window length, when it decreases from 85% to 81%.

In the second experiment, we attempted to forecast the price for the next week (5 days). Although using a 1-day window length, unlike in the first experiment, the model's performance showed better results, with an accuracy of 81%. The wider we expanded the window length, the more accurate the model became, achieving 93% accuracy using a 5-day window length, and it starts to decrease with a 10-day window length.

In the third experiment, we attempted to forecast the direction of the next 10 days' prices (i.e. over the next two weeks, excluding non-business days). We obtained results similar to those of the last experiment. It achieved an accuracy between 83% and 92%. To conclude, the model was more accurate in forecasting the next week (5 days) trend direction, especially when the window length was 5 days. Fine-tuning our model is necessary to improve both the accuracy and the running time of our model simultaneously. However, Fig. 11, 13, 15, 17

TABLE II. MODEL ACCURACY BY WINDOW-SIZE AND TIME FRAME

Target	**Window lenght**					
	1 day	**2 days**	**3 days**	**4 days**	**5 days**	**10 days**
1 day	0.5876	0.8050	0.8387	0.8422	0.8511	0.8185
5 days	0.8092	0.8098	0.8321	0.8450	0.9291	0.9156
10 days	0.8341	0.8383	0.8667	0.8943	0.9165	0.9023

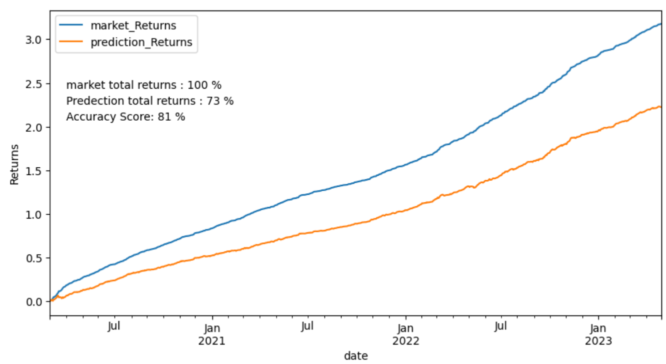

Fig. 7. The cumulative profits during the COVID-19 pandemic.

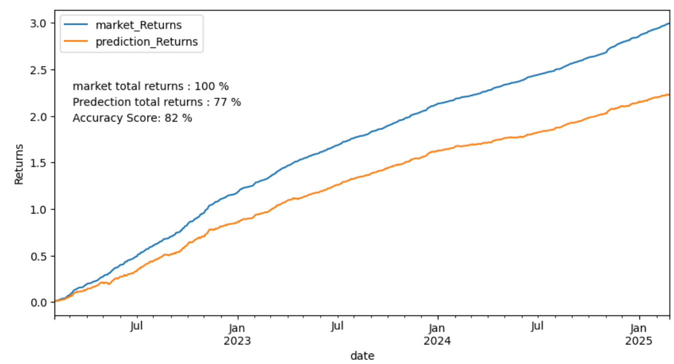

Fig. 8. The cumulative profits during the Russia-Ukraine war.

TABLE III. DIEBOLD-MARIANO TEST RESULTS (P-VALUES)

Model Pair	**DM Test Statistic**	**p-value**	**Conclusion** ($\alpha = 0.05$)
GRU-CNN vs Naïve Bayes	7.61	1.60×10^{-11}	Significant (GRU-CNN is better)
GRU-CNN vs Logistic Regression	9.84	2.44×10^{-16}	Significant (GRU-CNN is better)
GRU-CNN vs Random Forest	7.86	4.82×10^{-12}	Significant (GRU-CNN is better)
GRU-CNN vs XGBoost	3.19	1.91×10^{-03}	Significant (GRU-CNN is better)

and 19 show the training loss at each epoch; we notice that the training loss plot decreases to a point of stability.

To conduct a rigorous statistical analysis of our models' performance, we employed the Diebold-Mariano (DM) test to compare the predictive accuracy of our hybrid GRU-CNN model against the baseline models. The DM test evaluates the null hypothesis (H_0) that two models have the same forecast accuracy. A key component of the test is the zero-one loss function, which assigns a loss of 0 for a correct prediction and 1 for an incorrect one. A p-value below the chosen significance level (α=0.05) indicates that the difference in performance is statistically significant.

The results of the Diebold-Mariano test are summarized in Table III. The p-values for all comparisons are exceptionally low, all falling well below the 0.05 significance threshold. This finding provides strong statistical evidence that our GRU-CNN model is a better predictor of this task compared to all traditional machine learning benchmarks. The low p-values confirm that the performance gains of the hybrid model are not a result of random chance. Still, they are instead due to its inherent ability to capture more complex patterns in the financial time series data. This analysis validates our choice of a deep learning approach for this problem and confirms its effectiveness.

A. The Effectiveness of Our System During Times of Crisis

In times of crisis, as investors panic, the index of fear increases, causing adverse ripple effects throughout the economy, and the volatility of the market spikes dramatically. Each speculative attack on the Forex market often accompanies a crisis. Generally, we have tested our proposed solution over a long period that has witnessed several crises, and during this period, our system has consistently shown promising results. However, to confirm the effectiveness of our system during crisis times, we backtested our strategy on two crises: the COVID-19 pandemic from March 2020 to May 2023, and the Russia-Ukraine War, which began in February 2022 and is still ongoing. As shown in Fig. 7, which illustrates the cumulative profits during the COVID-19 pandemic, our system successfully predicted the future direction of the prices, achieving an accuracy of 81% with a total profit of 73% compared to the ideal. A similar result can be observed in Fig. 8, which illustrates cumulative profits during the Russia-Ukraine War; the accuracy was around 82% and the total profits were 77%; These results, while promising, should be interpreted with consideration for the complexities of real-world FOREX trading. Although our primary focus in this article is on demonstrating the effectiveness of integrating semi-supervised news classification, sentiment analysis, and time-series modeling in forecasting FOREX price directions, it is essential to acknowledge the potential impact of transaction costs (spreads), slippage, and inherent risks. To provide a more realistic perspective, we have applied a moderate estimate of a 10% reduction to the backtested profits achieved during both the COVID-19 pandemic and the Russia-Ukraine war. This reduction 10% accounts for the aggregate impact of transaction costs, primarily spreads on traded currency pairs, and

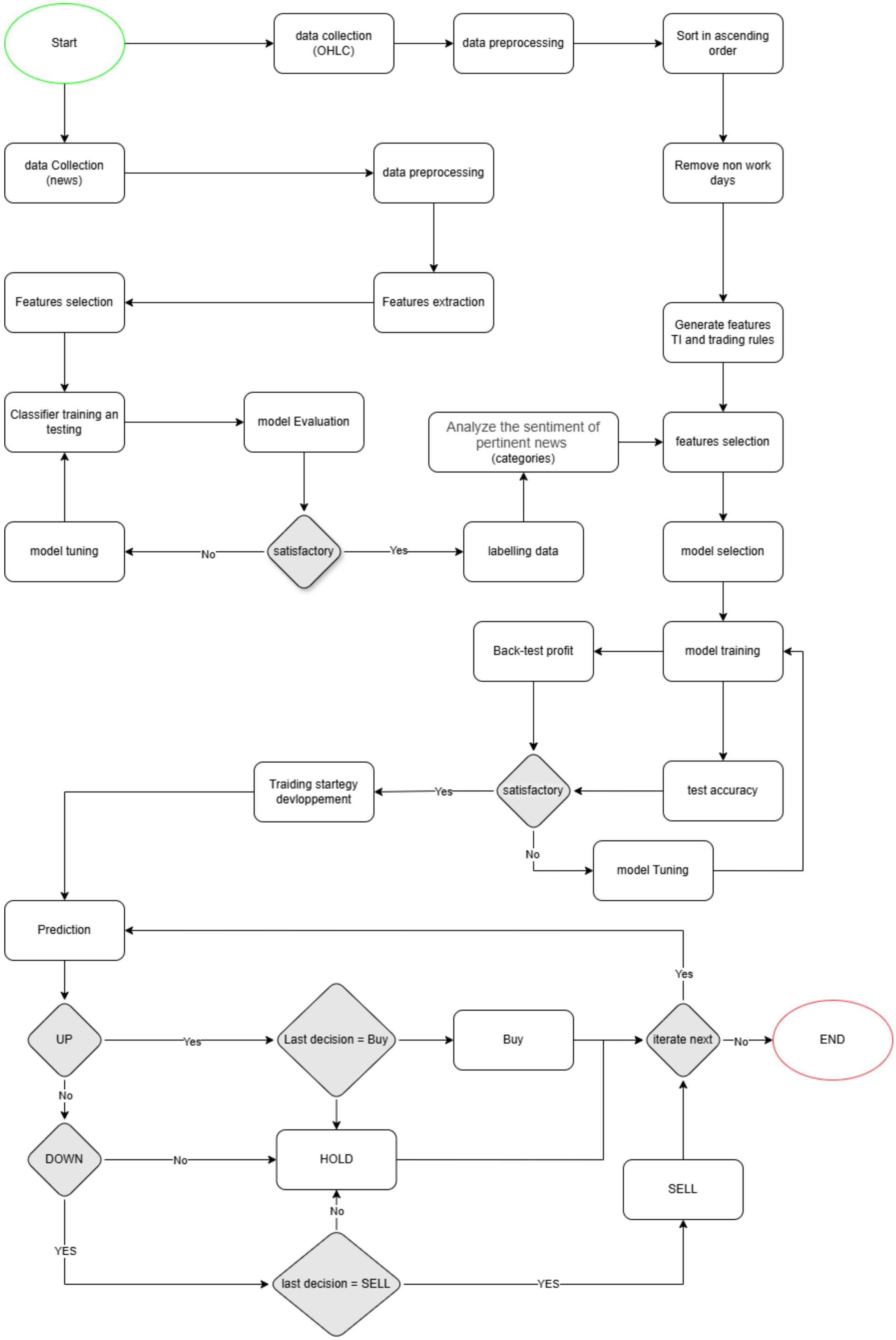

Fig. 9. Our investment strategy for trading foreign exchange.

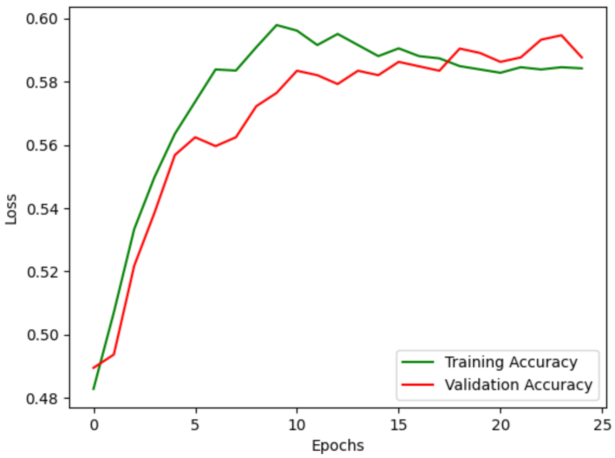

Fig. 10. CNN-GRU Training validation Accuracy 1 day window length.

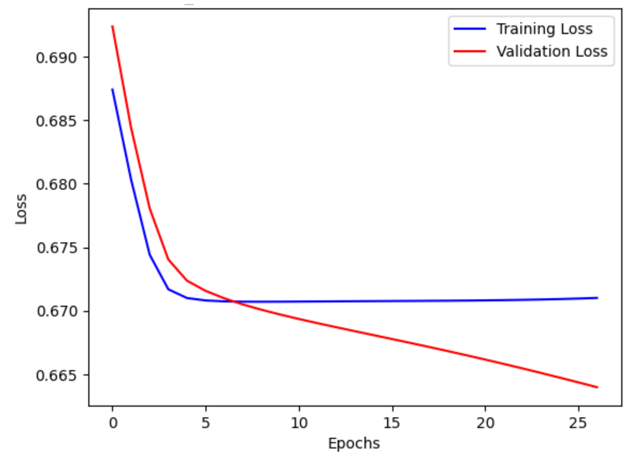

Fig. 11. CNN-GRU Training validation loss 1 day window length.

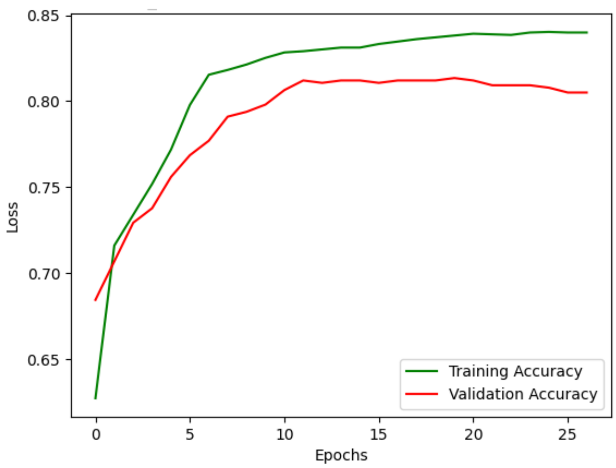

Fig. 12. CNN-GRU Training validation Accuracy 2 days window length.

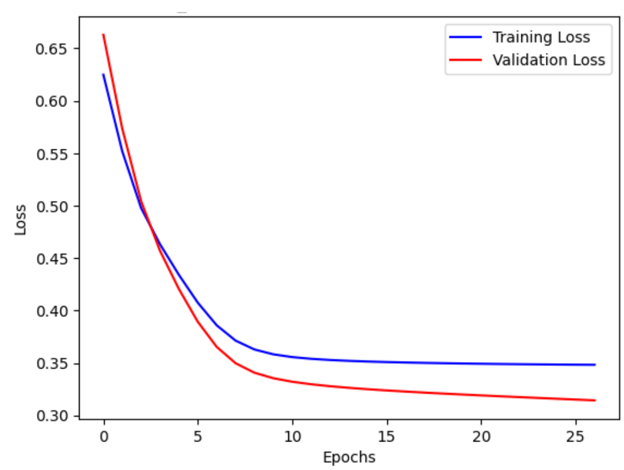

Fig. 13. CNN-GRU Training validation loss 2 days window length.

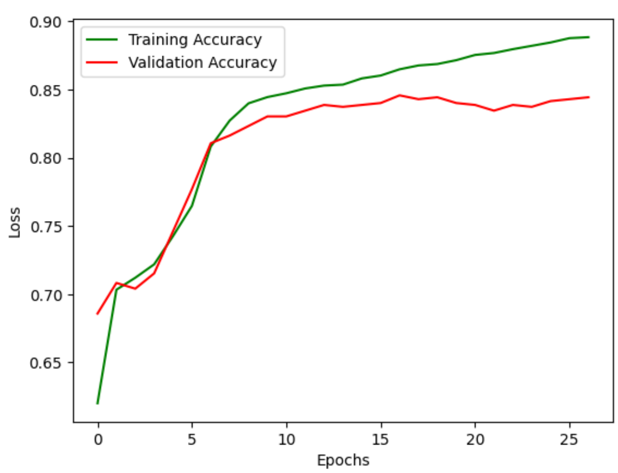

Fig. 14. CNN-GRU Training validation Accuracy 3 days window length.

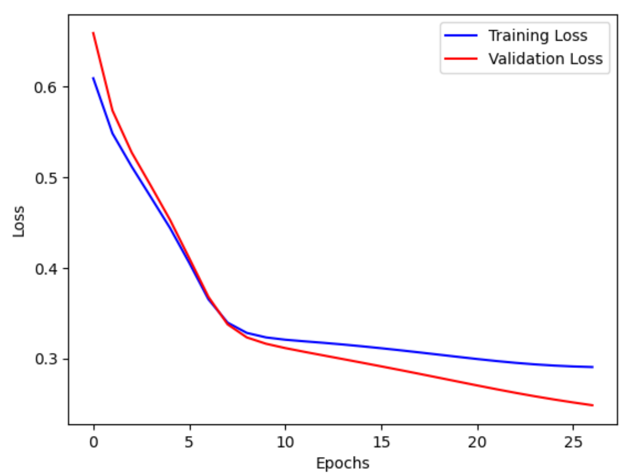

Fig. 15. CNN-GRU Training validation loss 3 days window length.

estimated slippage derived from historical volatility patterns. We maintain that even with this moderate adjustment, the system shows a strong potential for profitability and resilience in forecasting FOREX price directions during turbulent market conditions. By analyzing these two results, we can conclude that our proposed strategy can overcome obstacles during crisis

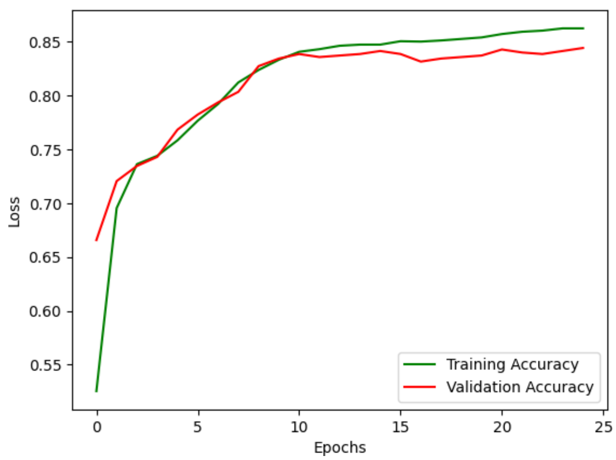

Fig. 16. CNN-GRU Train validation Accuracy 5 days window length.

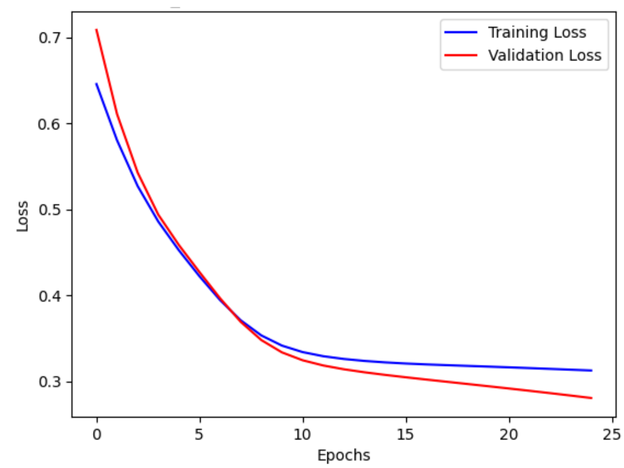

Fig. 17. CNN-GRU Training validation loss 5 days window length.

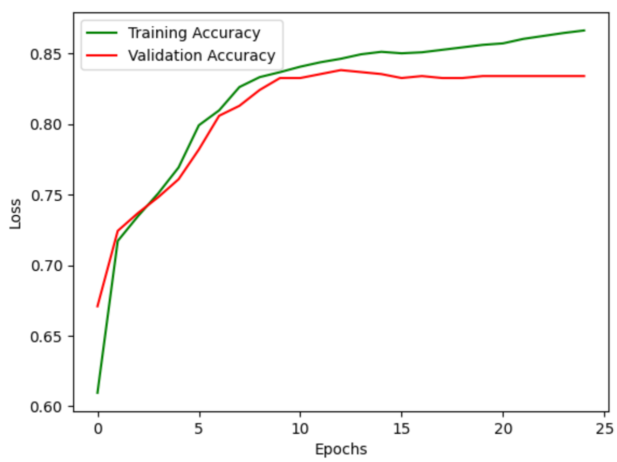

Fig. 18. CNN-GRU Training validation Accuracy 10 days window length.

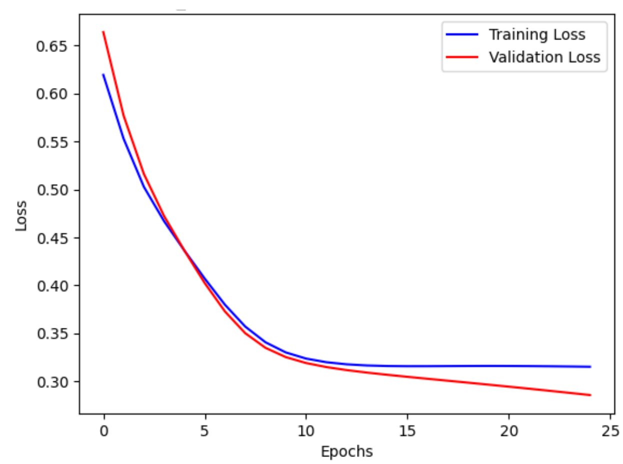

Fig. 19. CNN-GRU Training validation loss 10 days window length.

times by imitating fundamental analysis through the analysis of news headlines and social media posts. Future research will incorporate a more detailed, high-fidelity simulation incorporating granular transaction cost and slippage data, as well as a thorough risk assessment using industry-standard metrics, such as Sharpe and Sortino ratios, to validate these findings further and provide a more complete picture of the system's performance across diverse market environments.

V. Conclusion

In the FOREX market, numerous factors can affect the state of the market in various ways, making it too complex to develop an optimal trading strategy. This study presents a comprehensive system for automated Forex trading, utilizing an NLP-based strategy to analyze sentiment and forecast the expected impact on the Forex market through news reports. The results demonstrate that the system exhibits high profitability and can be effectively deployed for real-life trading purposes, overcoming external events such as regime changes, crisis times, epidemics, and natural disasters. This solution was developed and back-tested for a specific period; it may not remain helpful and profitable in the future.

Data Availability Statement

Due to privacy and ethical restrictions. The data that support the findings of this study cannot be made publicly available. However, a simplified version of the codebase is available on request from the corresponding author.

References

[1] T. Keeris, "the world's largest financial market: forex," *Social Science Research Network*, p. 2, 2018.

[2] T. Mahmud, T. Akter, S. Anwar, M. T. Aziz, M. S. Hossain, and K. Andersson, "Predictive modeling in forex trading: A time series analysis approach," in *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*. 2024 Second International Conference on Inventive Computing and Informatics (ICICI), 2024, p. 390–397.

[3] A. Chaboud, D. Rime, and V. Sushko, *The foreign exchange market*. Edward Elgar Publishing, 2023.

[4] H. Law, S. Aghabozorgi, S. Lim, Y. Teh, and T. Herawan, "Processing of kuala lumpur stock exchange resident on hadoop mapreduce," in *Proceedings of the International Conference on Data Science (ICDATA)*. Proceedings of the International Conference on Data Science (ICDATA), 2014, p. 1.

[5] G. McLeod, "Forex market size: A traders advantage," *DailyFX (January 23).*, 2014.

[6] L. Puspitawati, “Information system for forex investment and their effects on investment growth in foreign currencies,” *International Journal of Research and Applied Technology (INJURATECH)*, 2021.

[7] Y. Pei, J. Cartlidge, A. Mandal, D. Gold, E. Marcilio, and R. Mazzon, “Cross-modal temporal fusion for financial market forecasting,” *arXiv preprint arXiv:2504.13522*, 2025.

[8] L. Malagrino, N. Roman, and A. Monteiro, “Forecasting stock market index daily direction: A bayesian network approach,” *Expert Systems with Applications*, vol. 105, pp. 11–22, Sep. 2018.

[9] M. M. Kumbure, C. Lohrmann, P. Luukka, and J. Porras, “Machine learning techniques and data for stock market forecasting: A literature review,” *Expert Systems with Applications*, vol. 197, p. 116659, 2022.

[10] P. K. Sarangi, M. Chawla, P. Ghosh, S. Singh, and P. K. Singh, “Forex trend analysis using machine learning techniques: Inr vs usd currency exchange rate using ann-ga hybrid approach,” *Materials Today: Proceedings*, vol. 49, p. 3170–3176, 2022.

[11] L. Abednego, “Forex trading robot with technical and fundamental analysis,” *Journal of Computers*, p. 1089–1097, 2018.

[12] C. Neely, J. Ulrich, and P. Weller, “The adaptive markets hypothesis: Evidence from the foreign exchange market,,” *J. Financ. Quant. Anal*, vol. 44, p. 467–488, 2009.

[13] I. K. Nti, A. F. Adekoya, and B. A. Weyori, “A systematic review of fundamental and technical analysis of stock market predictions,” *Artificial Intelligence Review*, vol. 53, p. 3007–3057, 2020.

[14] R. Pramudya and S. Ichsani, “Efficiency of technical analysis for the stock trading,” *International Journal of Finance & Banking Studies*, vol. 9, p. 58–67, 2020.

[15] R. Schabacker, *Technical analysis and stock market profits*. Harriman House Limited, 2021.

[16] R. N. Ilham, I. Sinta, and M. Sinurat, “The effect of technical analysis on cryptocurrency investment returns with the 5 (five) highest market capitalizations in indonesia,” *Jurnal Ekonomi*, vol. 11, p. 1022–1035, 2022.

[17] S. Galeshchuk and S. Mukherjee, *FOREX Trading Strategy Optimization*. Springer International Publishing, 2018.

[18] M. Fisichella and F. Garolla, “Can deep learning improve technical analysis of forex data to predict future price movements?” *Ieee Access*, vol. 9, p. 153083–153101, 2021.

[19] K. Cao and H. You, “Fundamental analysis via machine learning,” *Financial Analysts Journal*, vol. 80, p. 74–98, 2024.

[20] L. Abednego, C. E. Nugraheni, and A. Fedora, “Forex sentiment analysis with python,” *International Journal of Advanced Research in Economics and Finance*, vol. 4, p. 46–55, 2022.

[21] Y. Dave, “Predicting forex pair movements: Integrating sentiment analysis, technical, and fundamental indicators using machine learning and deep learning models,” 2024.

[22] C. Maté, “Forecasting in forex the spot price interval of tomorrow with the same information of today. an analysis of the seven majors using a linear regression model based on interval arithmetic,” *Knowledge-Based Systems*, vol. 258, p. 109923, Dec. 2022.

[23] J. Saettele, *Sentiment in the forex market: indicators and strategies to profit from crowd behavior and market extremes*. Wiley Online Library, 2021.

[24] E. Gerlein, M. Mcginnity, A. Belatreche, and S. Coleman, “Evaluating machine learning classification for financial trading: An empirical approach,” *Expert Systems with Applications*, vol. 54, pp. 193–207, Jul. 2016.

[25] W. Poole, “Speculative prices as random walks: An analysis of ten time series of flexible exchange rates,” *Southern Economic Journal*, vol. 33, pp. 468–478, 1967.

[26] R. Hudson and A. Urquhart, “Technical trading and cryptocurrencies,” *Annals of Operations Research*, vol. 297, p. 191–220, 2021.

[27] R. Moscinski and D. Zakrzewska, *Building an Efficient Evolutionary Algorithm for Forex Market Predictions*. Poland: Springer, 2015.

[28] I. Witten, E. Frank, and M. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier, 2011.

[29] W. Haoxiang, S. Smys *et al.*, “Big data analysis and perturbation using data mining algorithm,” *Journal of Soft Computing Paradigm (JSCP)*, vol. 3, p. 19–28, 2021.

[30] I. H. Sarker, “Machine learning: Algorithms, real-world applications and research directions,” *SN computer science*, vol. 2, p. 160, 2021.

[31] N. MABROUK, M. CHIHAB, Z. HACHKAR, and Y. CHIHAB, “Intraday trading strategy based on gated recurrent unit and convolutional neural network: forecasting daily price direction,” *International Journal of Advanced Computer Science and Applications*, vol. 13, 2022.

[32] S. Suthaharan, *Support Vector Machine*. Boston, MA: Springer US, 2016, vol. 36.

[33] T. T. Thu and V. Xuan, “Supervised support vector machine in predicting foreign exchange trading,” *International Journal of Intelligent Systems and Applications*, vol. 10, pp. 48–56, Sep. 2018.

[34] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001.

[35] M. Ballings, D. Van den Poel, N. Hespeels, and R. Gryp, “Evaluating multiple classifiers for stock price direction prediction,” *Expert Systems with Applications*, vol. 42, pp. 7046–7056, 2015.

[36] A. Santuci, E. Sbruzzi, L. Araújo-Filho, and M. Leles, “Evaluation of forex trading strategies based in random forest and support vector machines,” *IEEE Latin America Transactions*, vol. 20, p. 2146–2152, 2022.

[37] L. Dongdong, S. Yuan, M. Li, and Y. Xiang, “An empirical study of machine learning algorithms for stock daily trading strategy,” *Mathematical Problems in Engineering*, vol. 2019, pp. 1–30, Apr. 2019.

[38] D. Lv, Z. Huang, M. Li, and Y. Xiang, “Selection of the optimal trading model for stock investment in different industries,” *PLOS ONE*, vol. 14, p. e0212137, Feb. 2019.

[39] B. Yue, J. Rao, and Y, “A deep learning framework for financial time series using stacked autoencoders and long short term memory,” *PLoS ONE*, vol. 12, p. 1–24, 2017.

[40] R. Akita, A. Yoshihara, T. Matsubara, and K. Uehara, *Deep learning for stock prediction using numerical and textual information*. Okayama, Japan: IEEE, Jun. 2016.

[41] Z. Hu, Y. Zhao, and M. Khushi, “A survey of forex and stock price prediction using deep learning,” *Applied System Innovation*, vol. 4, p. 9, Feb. 2021.

[42] H. Jia, “Deep learning algorithm-based financial prediction models,” *Complexity*, vol. 2021, p. 5560886, 2021.

[43] J. Sirignano and R. Cont, “Universal features of price formation in financial markets: Perspectives from deep learning,” *SSRN Electronic Journal*, vol. 19, p. 1449–1459, 2019.

[44] S. Seifollahi and M. Shajari, “Word sense disambiguation application in sentiment analysis of news headlines: an applied approach to forex market prediction,” *Journal of Intelligent Information Systems*, vol. 52, pp. 57–83, 2019.

[45] T. Sprenger, A. Tumasjan, P. Sandner, and I. Welpe, “Tweets and trades: the information content of stock microblogs,” *European Financial Management*, vol. 20, pp. 926–957, 2014.

[46] A. Siganos, E. Vagenas-Nanos, and P. Verwijmeren, “Facebook’s daily sentiment and international stock markets,” *Journal of Economic Behavior &; Organization*, vol. 107, pp. 730–743, Nov. 2014.

[47] R. Dong, M. O’mahony, M. Schaal, K. Mccarthy, and B. Smyth, “Combining similarity and sentiment in opinion mining for product recommendation,” *Journal of Intelligent Information Systems*, vol. 46, pp. 285–312, 2016.

[48] P. Joseph and S. Christy, “Improved precision in detecting the early symptoms of covid using k-nearest neighbor and support vector machine algorithm,” in *AIP Conference Proceedings*, vol. 2821, AIP Conference Proceedings. AIP Publishing, 2015, p. 080005.

[49] K. Horecki and J. Mazurkiewicz, *Natural Language Processing Methods Used for Automatic Prediction Mechanism of Related Phenomenon*. Cham: Springer International Publishing, 2015.

[50] V. Bijalwan, V. Kumar, P. Kumari, and J. Pascual, “Knn based machine learning approach for text and document mining,” *International Journal of Database Theory and Application*, vol. 7, pp. 61–70, Feb. 2014.

[51] A. K. Nassirtoussi, S. Aghabozorgi, T. Y. Wah, and D. Ngo, “Text mining of news-headlines for forex market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment,” *Expert Systems with Applications*, vol. 42, pp. 306–324, Jan. 2015.

[52] S. Senarathna and R. Rajapaksha, *Forex Market Prediction Through Text Mining of Financial News*. IEEE, Dec. 2024.

[53] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*. Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), 2019, p. 4171–4186.

[54] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, A. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[55] J. E. V. Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine learning*, vol. 109, p. 373–440, 2020.

[56] J. M. Duarte and L. Berton, "A review of semi-supervised learning for text classification," *Artificial intelligence review*, vol. 56, p. 9401–9469, 2023.

[57] J. Hao, F. He, F. Ma, and T. Fu, "Trading around the clock: Revisit volatility spillover between crude oil and equity markets in different trading sessions," *Journal of Futures Markets*, vol. 43, p. 771–791, Apr. 2023.

[58] I. Ghosh and P. Dragan, "Can financial stress be anticipated and explained? uncovering the hidden pattern using eemd-lstm, eemd-prophet, and xai methodologies," *Complex & Intelligent Systems*, vol. 9, p. 4169–4193, Dec. 2022.

[59] I. Guyon, "Feature selection: A data perspective," *ACM Computing Surveys*, vol. 50, 2017.

[60] C. J. Neely, D. E. Rapach, J. Tu, and G. Zhou, "Forecasting the equity risk premium: The role of technical indicators," *Institute for Operations Research and the Management Sciences*, 2014.

[61] D. Croce, G. Castellucci, and R. Basili, "Gan-bert: Generative adversarial learning for robust text classification with a bunch of labeled examples," in *Proceedings of the 58th annual meeting of the association for computational linguistics*. Proceedings of the 58th annual meeting of the association for computational linguistics, 2020, p. 2114–2119.

[62] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *Advances in neural information processing systems*, vol. 29, 2016.

[63] D. Araci, "Finbert: Financial sentiment analysis with pre-trained language models," *arXiv preprint arXiv:1908.10063*, 2019.

[64] K. K. L. Wong, *The Structure of Neural Network*. Wiley-IEEE Press, 2024, pp. 69–81.

[65] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent neural networks: A comprehensive review of architectures, variants, and applications," *Information*, vol. 15, p. 517, 2024.

[66] F. M. Salem, *Recurrent neural networks*. Springer, 2022.

[67] hochreiter Sepp and schmidhuber jürgen, "Memory, long short-term," *Neural Computation*, vol. 9, p. 1735, 1997.

[68] S. Hochreiter and J. Schmidhuber, "Lstm can solve hard long time lag problems," *Advances in neural information processing systems*, vol. 9, 1996.

[69] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," *Neural computation*, vol. 12, p. 2451–2471, 2000.

[70] F. A. Gers and E. Schmidhuber, "Lstm recurrent networks learn simple context-free and context-sensitive languages," *IEEE transactions on neural networks*, vol. 12, p. 1333–1340, 2001.

[71] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*. Springer, 2012.

[72] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint*, p. 1409–1259, 2014.

[73] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, p. 2278–2324, 1998.

[74] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 1–9, 2015.

[75] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," *International Conference on Engineering and Technology (ICET)*, p. 1–6, 2017.