The 38th Canadian Conference on Artificial Intelligence
DOI: 0

Canadian AI 2025
26—29 May 2025, Calgary, Alberta

Financial Market Prediction using Contrastive Learning-based News Representation and Multi-Resolution Market Data

Saeede Anbaee Farimani[†], Majid Vafaei Jahan[†,*], Pouya Soltani[†] , Amin Milani Fard[‡]
[†]Islamic Azad University, Mashhad, Iran
[‡]New York Institute of Technology, Vancouver, Canada

Abstract

Decision support systems use LLM embeddings to convert market data into actionable trading insights. However, current financial prediction models often overlook valuable information within short-term intervals (e.g., 4 hours) within longer ones (e.g., a day). The dissemination of news within these shorter periods significantly impacts market movements even for multiple days. This study aims to determine the effectiveness of incorporating fine-grained information into market prediction models that traditionally rely on coarse-grained data. We design a neural network to simultaneously attend to important news and influential indicators present in short-term time slices as well as benefit from market data available in long-term timeframes. With the advancement of contrastive learning-based NLP, we utilize the Angle-optimized Embedding (AoE) sentence transformer for news representation, which generates discriminative embeddings leveraging angle-optimized loss. Besides, to tackle the problem of non-stationary series regression, we employed reversible instance normalization. Comparative results with baseline articles within Forex and cryptocurrencies demonstrate the superiority of the proposed method. Our ablation studies demonstrate that the simultaneous use of financial market data in both fine-grained hourly time slices and coarse-grained daily time slices improves prediction accuracy by up to 60%. Furthermore, utilizing the AoE method to generate informative vector representations for news documents outperformed other embeddings by up to 9.5%.

Keywords: Time series regression, News representation, AoE embeddings, Multi-resolution representation, ReVIN normalization

1. Introduction

Market participants in equities, foreign exchange, and cryptocurrencies closely monitor price movements across different timeframes, from intraday to monthly. News-driven volatility can disrupt price patterns at various scales, making it challenging for investors to make informed decisions aligned with their chosen timeframes. Behavioral economics focuses on studying the impact of information reaching financial market investors [1]. Scholars have developed decision support systems for financial investment that leverage multimode information [2–4]. Recent studies in behavioral finance have utilized deep learning networks with attention mechanisms to identify and extract key features from multimodal market time-series data and textual social media content [5–8]. A primary focus of these studies has been developing network architectures capable of assigning appropriate weights to different data modalities, such as news content, sentiment analysis, and market data, for accurate prediction [9–13]. Although these models have shown promise, they often rely solely on market data within specific time frames to forecast prices, overlooking the subtle impact of news and media on finer-grained price fluctuations. We design our approach to shed light on the importance of utilizing multi-resolution market data and studying the impact of news dissemination on price regression.

So far, scholars have used transformer models for financial text representation [14–16], especially some work that utilizes head-based fine-tuning of pre-trained models in the finance

[†]anbaee@mshdiau.ac.ir, VafaeiJahan@mshdiau.ac.ir

domain [17–19]. The contrastive learning-based NLP is a fast and accurate technique that learns meaningful text embeddings by maximizing similarity between semantically related text pairs while minimizing similarity between unrelated ones. Researchers in [20] used SBERT [21] trained based on cosine similarity of pair-sentences loss in a distillation learning-based method for financial domain text summarization. The authors in [22] introduced Angle-optimized Embeddings (AoE) for semantic textual similarity, in which an angle-based loss function is used to mitigate the vanishing gradient problem inherent to objectives based on cosine similarity, by optimizing the angle between embedding vectors in a complex space. AoE has shown to be very effective in sentence similarity tasks, however, it has not been explored in behavioral economics. This motivated us to use it for generating news document embeddings in our work.

Contributions. We introduce the AoE-MR method that leverages **AoE** LLM [22] for news representation and **M**ulti **R**esolution market data. AoE-MR incorporates two coarse- and fine-grained information to consider past correlations of market data and the recent effect of news on price fluctuation. The coarse-grained module captures long-term market trends by analyzing technical indicators through an RNN-based encoder layer. Conversely, the fine-grained module identifies influential news and their short-term market impact. This module comprises attention head layers to identify key news and temporal relationships within multiple short intervals between longer periods. Subsequently, sequences of short-term news and market features are processed by an encoder layer. Finally, a weighted fusion of short-term news and market based features and long-term market based features is performed for price regression. A significant challenge in time series normalization is the non-stationarity of data distributions, which occurs when the distribution of data in the training set differs from that in the testing set. To address that, we utilize RevIN normalization [23]. This technique introduces learnable weights to dynamically shift and scale time series data, enhancing the model's ability to predict price accurately. Our contribution includes:

- We analyze the importance of coarse-grained information between consecutive time intervals and finer-grained information in shorter time slices.
- We introduce a novel approach that leverages attention mechanisms to prioritize significant news within AoE embeddings, while also extracting features from multiple short-term intervals of market data buckets.
- We design an approach that simultaneously leverages contrastive learning for news representation and neural network-based normalization for financial time series price prediction. Our code is publicly available for download [1].
- The results of our experiments show the superiority of our approach compared to other works by up to 60%. Ablation studies indicate the strong ability of AoE in financial domain text representation.

2. Related works

Normalization methods for time series. Traditional deep learning models mainly addressed binary prediction (up/down) of price or volatility trends [24, 25]. However, recent advances in neural network-based normalization techniques [4, 26, 27] have changed the focus to price regression. Using these methods, researchers aim to predict the absolute value of the price or volatility values rather than simply forecasting directionality. Time series normalization faces nonstationarity, where training and testing data distributions differ, prompting various normalization methods to address this issue.

Min-Max and z-score normalization scale data using fixed statistics, but for non-stationary time series, test data normalization must use test-derived values to avoid performance issues.

[1] https://github.com/anbaee/AoE-MR

LayerNorm [28], embedded in RNNs, normalizes hidden states using their own mean and variance, pioneering neural network normalization. Dynamic normalization methods like Dain [29] and RevIN [23] introduced learnable parameters, with RevIN offering a two-phase (normalization and denormalization) approach to handle non-stationary data, enhancing model robustness and adaptability over time. Research shows that integrating LayerNorm improves model convergence speed, while the Dain method [29] introduces a normalization layer in neural networks for classifying non-stationary time series, enabling the network to learn mean and variance for z-score normalization. Extending Dain, the RevIN method [23] processes data through a Dain layer for normalization and later restores the original distribution via inverse normalization using learnable parameters. Recent studies [30, 31] highlight the growing popularity of these normalization techniques, with RevIN being widely adopted in deep learning models for time series regression due to its effectiveness.

Sentence-transformer models. In recent years, researchers in the field of behavioral economics have utilized neural network-based text representations. Some methods have utilized the word embedding techniquefor representation [32]. However, this approach suffers from the problem of not considering the sentence context in generating word vectors. With the widespread use of large language models like BERT[33], some methods like [9] have employed this model in representing vectors of economic news documents. Researchers in [21] proposed the SBERT language model and demonstrated that the BERT language model is not capable of producing similar vectors for documents with similar topics.

Contrastive learning represents a significant advancement in refining sentence-transformers by optimizing the embeddings through a process of comparative analysis. This technique enhances the quality of sentence embeddings by encouraging the model to distinguish between similar and dissimilar text pairs, thus sharpening its ability to capture nuanced semantic distinctions. The researchers in [34] introduce a novel contrastive learning approach called SimCSE, which significantly improves sentence embeddings by using contrastive objectives for both supervised and unsupervised learning settings. Similarly, Zhang et al [35] explores the application of contrastive learning techniques in fine-tuning pre-trained language models, showcasing enhanced performance in sentence representation tasks. By embedding this approach into the training process, sentence-transformers are better equipped to capture the nuanced and often subtle semantic relationships present in textual data, leading to more accurate and insightful analyses in behavioral economics research.

With the recent advancements in GAN networks, researchers introduced a network based on BERT called Sentence-BERT [21], aiming to enhance the ability to generate similar vectors for similar documents. In this network, the loss function considers the cosine similarity of the generated vectors in the GAN network. Therefore, the network is capable of learning semantic similarities between documents and generating similar vectors for similar documents. Their results demonstrate that the Sentence-BERT model outperforms BERT in tasks involving semantic similarity of sentences in natural language processing. Subsequently, researchers in [22], by addressing challenges in the cosine function's saturation in certain areas, improved the performance of Sentence-BERT and introduced the AoE model. Given the superiority of this approach in tasks involving semantic sentence similarity, its use has become prominent across various fields as an open research area.

Given the advancements in Sentence-BERT and AoE, there is a prominent open research domain in utilizing these models in behavioral finance. The AoE model, with its enhanced ability to manage semantic similarity and address the cosine-based loss function challenges, presents opportunities for deeper exploration in financial contexts. Potential applications include improved sentiment analysis of financial news, more accurate clustering of market sentiments, and enhanced models to predict market trends. Investigating how AoE's advanced sentence embeddings can be applied to predicting finance behaviors could lead to more insights and refined predictive capabilities, ultimately advancing the field and offering new avenues for research and application.

In our study, we employ the AoE sentence transformer. The AoE sentence transformer introduces a novel optimization technique for sentence embeddings, addressing the limitations of traditional cosine similarity methods. Traditional models, such as BERT and RoBERTa, often use cosine similarity to measure the angle between vectors. However, cosine similarity can suffer from saturation issues, where the function becomes less sensitive to differences between vectors that are already close to each other. AoE addresses these issues by operating in complex space to compute the angle difference between sentence embeddings.

Predictive models. A broad variety of predictive models have during the past decades analyzed diverse portfolios. With the advent and widespread adoption of deep learning networks, particularly convolutional and recurrent networks, numerous methodologies have emerged [9, 10, 12, 13]. Recent advancements have incorporated attention mechanisms to enhance the extraction of critical features over time [5–8]. Recent studies have focused on integrating multiple data modalities to improve predictive performance. For example, Lee et al. [36] proposed a method that combines stock price data in various time frames of daily, monthly, and weekly with macroeconomic indicators, demonstrating improved prediction accuracy compared to unimodal approaches. Their approach highlights the effectiveness of the Multimodal fusion transformer for stock direction classification. Similarly, Gang Wang et al. [37] developed RCMA, an attention-based multimodal deep learning method that integrates financial and textual data to enhance fraud detection in finance, showcasing the potential benefits of multimodal data fusion in identifying suspicious financial activities. Another significant contribution in the realm of multimodal integration is presented by Farimani et al. [27], who developed a novel attention-based multimodal deep learning method, for financial market price regression. Their method integrates financial modality and the textual modality to enhance regression capabilities. This approach underscores the potential of multimodal data fusion as well as the normalization strategy for improving turbulent market prediction.

A comprehensive review in [2] emphasized the integration of diverse data sources, such as news, social media, and market data, with text mining and deep learning techniques for enhanced financial decision support. In sentiment analysis, the development of FinSoSent [38], a domain-specific large language model pre-trained on financial news and fine-tuned on financial social media corpora, is an example of the effective application of multimodal data sources financial news, technical indicators, and social media posts to improve sentiment analysis. The use of generative AI models, like GPT-3.5, for preprocessing text also demonstrates benefits in contextual understanding while addressing the challenges associated with domain-specific vocabulary [38]. The study underscores how combining different metrics and models, including those capable of handling multimodal inputs, can enhance forecasting accuracy and decision-making [39].

Many contemporary methods focus solely on information available within daily time intervals [10, 26] for predicting daily outcomes, often neglecting the potential value of information obtained by investors between these intervals. This paper aims to improve the prediction accuracy for price regression by incorporating information from hourly intervals between daily time slices. The proposed network structure is designed to integrate both granular hourly information and daily time slice data. This approach allows the model to prioritize significant news events and impactful price changes, thereby improving decision-making capabilities based on a more comprehensive dataset.

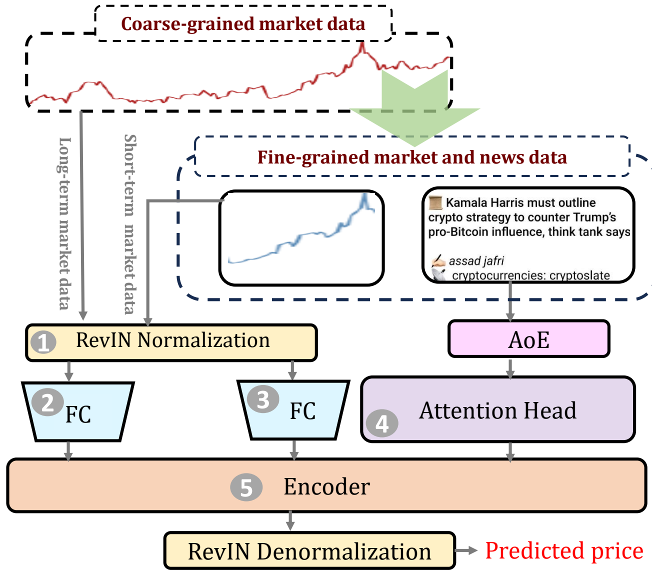

Figure 1. Our AoE-MR approach leverages the past history of daily market data and the last day coarse-grained news and hourly market data (green arrow) for price prediction.

3. The Proposed Model

Let $z_{i,t}$ be the i^{th} market-based feature at time t. Given the past values of market data $[z_{i,t_0}, z_{i,t_1}, ..., z_{i,t}] := z_{i,t_0:T}$ and two covariates of finer-grained market information $[\hat{z}_{i,\tau_0}, \hat{z}_{i,\tau_1}, ..., \hat{z}_{i,T}] := \hat{z}_{i,\tau_0:T}$ where $\tau = [t-1, T]$ and news sequence $[x_1, ..., x_n] := x_{t-1:T}$ published during τ, our goal is to predict the $z_{closeprice,t+1}$ as our target. Figure 1 depicts the data flow and the architecture of our AoE-MR method.

3.1. Coarse-grained Market Information Embeddings

Let $[z_{i,t_0}, z_{i,t_1}, ..., z_{i,t}] := z_{i,t_0:T}$ be the long-term market data encompassing OHLCV values and technical indicators. The time series $z_{i,t_0:T}$ is normalized based on its distribution using a RevIN normalization layer (Block 1 in Fig. 1). RevIN [23] is a wrapper on our base model to normalize each incoming series and de-normalize the output through learnable distribution parameters (Block 1 in Fig. 1). Subsequently, a market data embedding is generated through fully connected layers in a low-dimensional latent space (Block 2 in Fig. 1). Therefore, $Z^c_{i,t_0:T}$ would be the fully connected output for the coarse-grained component.

Our proposed model leverages both the historical market data (coarse-grained information) and recent news alongside market fluctuations (fine-grained information). The model attends to market data over the interval $[t_0 : T]$ as the past history of the system, while also focusing on the shorter period $\tau = [t-1 : T]$ within $[t_0 : T]$. This shorter interval captures the dissemination of news, which may induce significant fluctuations in the market on the upcoming day.

3.2. Fine-grained Market Information Embeddings

We divide $\tau = [t-1 : T]$ to h time steps and consider covariate finer-grained market data, denoted as $\hat{z}_{\tau_0:\tau}$, is passed through the RevIN normalization layer (Block 1) followed by an encoder module to generate its embeddings (Block 3). Finally a sequence of finer-grained market data $\hat{Z}^f_{\tau_0:T}$ with length h is ready to be fed to the encoder for price regression.

3.3. Fine-grained News Information Embeddings

For covariate x as a news title, let $[x_1, ..., x_n] := x_{t-1,T}$ be a sequence of recently published news within the last market state interval $[t-1:T]$. We leverage AoE LLM [22] to create news embeddings for each interval during τ. Considering h fine-grained intervals within $[t-1:T]$, we granularly align news embeddings with their corresponding fine-grained market data in temporal order $[(\hat{z}_{i,\tau_0}, X_{\tau_0}), ..., (\hat{z}_{i,\tau}, X_{\tau})]$, where $\mathbf{x}_{\tau_i} = [AoE(x_j)|S.T \tau_{i-1} < timestamp(\mathbf{x}_j) < \tau_i]$. AoE is a sentence-transformer model trained contrastively based on following objective function:

$$\mathcal{L} = w_1 * log\Big[1 + \sum_{s(\mathbf{x}_i,\mathbf{x}_j) > s(\mathbf{x}_m,\mathbf{x}_n)} e^{\frac{cos(\mathbf{x}_m,\mathbf{x}_n) - cos(\mathbf{x}_i,\mathbf{x}_j)}{\gamma}}\Big] + w_2 * log\Big[1 + \sum_{s(\mathbf{x}_i,\mathbf{x}_j) > s(\mathbf{x}_m,\mathbf{x}_n)} e^{\frac{\Delta\theta_{ij} - \Delta\theta_{mn}}{\gamma}}\Big]$$

where, the first term is a cosine loss function designed to generate highly similar embeddings for the high-similarity text pair $\mathbf{x}_i, \mathbf{x}_j$, while ensuring the similarity value $s(\mathbf{x}_i, \mathbf{x}_j)$ exceeds that of the low-similarity pair $\mathbf{x}_m, \mathbf{x}_n$. However, the cosine function's saturation zone can impede the optimization process. To address this, the second term is introduced to minimize the normalized angle difference $\Delta\theta_{i,j} - \Delta\theta_{m,n}$ in complex space for pairs with high similarity $\mathbf{x}_i, \mathbf{x}_j$ compared to those with low similarity $\mathbf{x}_m, \mathbf{x}_n$. w_1 and w_2 are constants. γ is the temperature parameter.

Given h fine-grained intervals within τ, we employ h separate attention head blocks (Block 4) to enforce network assign larger weights to important news published during each interval τ_i. The output of each attention block is computed as $X^f_{\tau_i} = Softmax(tanh(W_n * X_{\tau_i} + b_n))$. This process results in a sequence of weighted news embeddings with length h $X^f_{\tau_0:\tau}$ which is ready to be fed to the encoder module for price regression.

3.4. Encoder Module

The encoder in our model is a compromise of three RNN blocks followed by temporal attention. Each RNN block in the encoder module, fed by a temporal sequence of prepared features $Z^c_{i,t_0:T}$, $X^f_{\tau_0:\tau}$, $\hat{Z}^f_{\tau_0:T}$. The temporal attention placed at the output of each RNN block is responsible for assigning heavy weights to important time steps of fine-grained news and market data as well as coarse-grained market indicators. Finally, the output features from the attention modules are concatenated together and sent to a linear regression layer. Subsequently, they are passed through an inverse normalization layer (Block 5), to shift and scale the predicted price around target distribution.

4. Experiments

4.1. Dataset

Following [2, 4], we analyzed major currency pairs in the Forex market such as EUR/USD, GBP/USD, USD/JPY, and the Bitcoin / TetherUS at the cryptocurrency market. The MarketPredict Dataset presented in [4] includes economic news published from 2021 to 2022 and market price indices in hourly time slices. To assess the impact of fine-grained market and news data within shorter timeframes, we analyze daily and hourly market data. In this regard, we collected daily OHLC data from Finhub [2] and used hourly market data

[2] https://finnhub.io/

Currency Pair	Time Span	Number of Samples(Hour)	Number of Samples(Day)
EUR/USD	14/1/2021 - 13/3/2022	7,872	326
USD/JPY	13/1/2021 - 14/3/2022	7,872	328
GBP/USD	14/1/2021 - 14/3/2022	7,872	327
BTC/USDT	1/3/2020 - 13/3/2022	7,872	901

Table 1. Market Price Indices Dataset.

available in the marketpredict dataset. Table 1 shows the period and the number of samples of market price indices for both daily and hourly time frames.

4.2. Settings

We aim to predict the daily close price. Regarding the multi-horizon market data analysis in our AoE-MR model, we employ daily market data as coarse-grained information and hourly time frame data as fine-grained information. The daily lag is set to six days prior, and the hourly lag is set to 24 hours prior. We set $h = 6$ as the number of short-term intervals, dividing the data into 6 fine-grained buckets, each spanning 4-hour intervals (24/4=6), along with the corresponding news published within those periods. Technical indicators remain crucial resources for investors' decision-making.

Researchers in [4] proposed an information gain analysis approach for selecting technical indicators. Following their work, we also used Close price, Exponential Moving Average (EMA), and Bollinger bands low (BBL) for all analyzed currency pairs. Additionally, for EUR/sUSD, USD/JPY and GBP/USD we use Accumulation/ Distribution Index (ADI) and for BTC/USDT, we use On Balance Volume (OBV) as volume-based indicator. For indicator calculation, TA [3] package in Python was used, and all indicators were calculated based on the default values in this package.

The network structure was implemented using TensorFlow. The learning rate for all currencies was set to 0.001, and the decay coefficient was set to 1×10^{-6}. The number of epochs was set to 300, and the ADAM optimizer was used. The error function used was the Mean Absolute Percentage Error (MAPE). To compute vector representations of news documents, we employed the *angle-bert-base-uncased-nli-en-v1* model[4]. For each news item, the average of the CLS token embeddings from the news title was used as the document vector, resulting in a 784 dimensional representation. The same seed was used in all experiments. We considered the first 60% part of dataset for training set, the next 20% for validation set, and the last 20% for test set, while preserving the temporal order of data.

5. Results and Analysis

Our experiments involved benchmarking against existing methods, conducting ablation studies to assess the contribution of the language model to vector quality, and examining the influence of network topology and data granularity (hourly vs. news-based).

5.1. Multi-resolution Data Analysis

We investigate the impact of fine-grained hourly information on daily market price prediction by switching off hourly news and hourly market data in turn and assess performance.

Fine-grained news Analysis. We analyzed the effect of covariate fine-grained news in price regression accuracy. Our findings underscore the critical role of fine-grained, recently published news in predicting long-term daily trends. While coarse-grained information, such

[3] https://pypi.org/project/ta/

[4] https://huggingface.co/SeanLee97/angle-bert-base-uncased-nli-en-v1

Model name	Coarse grained market data	Fine-grained news	Fine-grained market data
NoNews-MR	✓	×	✓
AoE-SR	✓	AoE $[CLS]$ representation	×
FinBERT-MR	✓	FinBERT $[cls]$ embedding	✓
BERTBoEC-MR [9]	✓	Bag of Economic Concept clusters based on BERT representation	✓
AoE-MR [Proposed method]	✓	AoE $[CLS]$ representation	✓

Table 2. Variations of our proposed AoE-MR method.

as technical indicators from the past six days, proved valuable, incorporating fine-grained news from the current day led to a significant reduction in prediction error for all currency pairs. This highlights the importance of forcing the network to attend to recent news for accurate long-term forecasting.

Fine-grained market data Analysis. We examined the influence of fine-grained hourly market data on daily price forecasting performance. A key innovation of this research is the novel integration of price indicators from both daily and hourly timeframes to forecast daily price movements. Specifically, hourly price indicators lagged by one day, and daily price indicators lagged by six days were employed as predictors for the subsequent day's price. To quantify the influence of hourly price indicators, we conducted an ablation study by excluding fine-grained market data from our proposed AoE-MR network and meticulously monitored the resulting model error.

The results of this experiment are shown in Table 3. The AoE-SR row in Table 3 presents MAPE loss values obtained when excluding hourly price indicators. These results underscore the critical role of fine-grained market data as covariates in daily price regression. AoE-MR demonstrated superior performance, achieving a 44% reduction in MAPE for EUR/USD, 55% for GBP/USD, 60% for USD/JPY, and 53% for BTC/USDT compared to AoE-SR.

5.2. **Ablation Studies**

To study the effect of news embedding representation, we conducted ablation studies. We assessed the contribution of news representation by comparing AoE embeddings to alternative text representation methods. Subsequently, Table 2 depicts various modifications that we analyzed for ablation studies. We compared two adaptations of our model to assess the influence of fine-grained news document vector representation.

We employed FinBERT [40] as a head-based fin-tuning model for the financial domain. Scholars in [12] analyze the efficacy of FinBERT for text representation and financial domain sentiment evaluation. We used $[cls]$ representation of the FinBERT model. The second modification for assessing the importance of news representation is BERT-BoEC [9] method. The BoEC method has two phases of clustering and frequency calculation within each cluster. In the clustering phase, the vector representation of words in each news item is obtained from the BERT model, then these words are clustered into several clusters. The vector corresponding to each news item is then represented based on the frequency of its words in each cluster, such that the length of the resulting vector equals the number of clusters, and each entry corresponds to the frequency of the news item's words in that cluster. This method's advantage is producing low-dimensional vectors while using all the information available in each news item.

For the AoE representation of news titles, we consider the vector representation of each news title as the average representation of the $[CLS]$ token. The AoE-MR model outperformed the FinBERT and BoEC-based news representation, which emphasizes the efficacy of contrastive learning over transfer learning for financial text representation. A crucial factor contributing to AoE's superior embeddings within our AoE-MR model is its targeted optimization to address the vanishing gradient problem arising from the cosine similarity

Method	EUR/USD Validation Loss	Test Loss	GBP/USD Validation Loss	Test Loss	USD/JPY Validation Loss	Test Loss	BTC/USDT Validation Loss	Test Loss
NoNews-MR	1.858	2.657	1.708	1.997	1.025	1.392	6.346	5.354
FinBERT-MR [40]	0.740	1.145	0.834	1.161	0.785	0.866	4.137	3.448
BERTBoEC-MR	1.905	1.436	2.546	1.909	0.785	0.952	4.401	3.367
AoE-SR	3.745	1.892	2.360	2.404	2.272	2.174	13.213	7.097
AoE-MR	**0.651**	**1.058**	**0.834**	**1.072**	**0.786**	**0.866**	**4.120**	**3.273**
Self Attention and FC layers [12]	2.466	3.091	1.271	3.373	0.767	0.938	3.525	3.612
Multi layer LSTMs [4]	3.273	6.172	1.621	3.148	3.113	3.704	8.413	9.684
Parallel CNNs [41]	4.877	7.452	3.024	4.030	5.999	6.380	5.395	5.553

Table 3. Comparison of Mean Absolute Percentage Error (MAPE).

function's saturation regions. The AoE Loss component of the model is instrumental in enhancing AoE's capacity to generate discriminative embeddings.

5.3. Comparison with Baselines

Recent research in deep learning time series analysis has primarily focused on regression [27, 30, 31] rather than binary classification. We compare the effectiveness of our method in price regression to other time series-based financial prediction methods that directly incorporate news embeddings and market prices during training. To ensure a fair comparison, we adapted all baselines to the price regression problem by embedding the RevIN normalization method as a wrapper layer in their network architectures. This decision was motivated by the significant impact of normalization on financial time series regression performance. Furthermore, we employed the same price indicators used in our proposed AoE-MR method across all baseline models. The baseline network architectures are as follows:

Self Attention and FC layers. In [12] news was represented using the BERT language model, and a self-attention mechanism was used for news summarization. Feature extraction from daily price indicators was performed using a multilayer fully connected network.

Multi-layer LSTMs modules. In [4], feature extraction from news embeddings is done by multi-layer stacked LSTM layers. Also, daily price indicators were fed into other LSTM layers. The news was vectorized using the FinBERT model.

Parallel CNN modules. In [41], a parallel temporal CNN network was used for feature extraction from news and daily price data. The news was vectorized using the BERT.

Our experimental results unequivocally demonstrate the superiority of the proposed method over established baselines. This enhanced performance can be attributed to three key innovations. (1) *Multi-horizon price indicator integration* By incorporating price indicators from multiple time horizons, specifically hourly and daily data, we effectively capture the intricate dynamics of price movements. Hourly indicators, closely aligned with investor behavior for day trading, provide granular insights into short-term trends, while daily indicators offer a broader market perspective. (2) *Attention-based news summarization*: The integration of an attention mechanism within our network architecture enables the model to dynamically focus on the most relevant news segments within a 24-hour window. This selective attention mechanism significantly improves the model's ability to extract meaningful information from news data. (3) *Advanced news representation*: Our novel approach leverages the AoE language model for robust news representation. Coupled with the attention mechanism, this enables the model to effectively weigh the contributions of news, price, and time-series data in prediction, resulting in more accurate and informative output.

The results of the experiments show the importance of contrastive learning in document representation and the application of news semantic similarity. On the other hand, the use of price data in shorter time slices that reflect price changes more quickly, alongside long-term daily price indicators, significantly improves the accuracy of predictive models. Additionally, attention to the normalization method of price indicators to overcome the non-stationary behavior of time series is crucial for price regression.

6. Conclusion

Prediction of long-term time frame of the market is a challenging task because of the delayed update of long-term indices according to the recent finer-grained price fluctuations that occurred in response to recent news dissemination. To address this issue, our proposed model was designed to use past long-term market data as well as two fine-grained covariates for forex and cryptocurrency pair exchange rate forecasting. Our model incorporates multi-mode data including past days price indicators, fine-grained hourly market information, and recent financial news. Ablation studies indicate the significant effects of hourly time frame indicators and recently published news in the daily price regression improvements by up to 60% error reduction. To overcome the challenge of non-stationary time series regression, we embedded a learnable normalization layer. Additionally, we employ the AoE language model for news representation. Our AoE-MR model significantly outperformed other text representation benchmarks, demonstrating the efficacy of contrastive learning in generating informative news embeddings for financial forecasting. With the expansion of the use of transformers in time series prediction, and their application in financial time series prediction we plan to predict multi-horizon prices as a future work.

Acknowledgments

This work was supported in part by Iran's Ministry of Science, Research and Technology grant; and in part by the New York Institute of Technology - Vancouver Campus.

References

[1] E. F. Fama. "The Behavior of Stock-Market Prices". In: *The Journal of Business* 38.1 (1965), pp. 34–105. ISSN: 00219398, 15375374.

[2] S. A. Farimani, M. V. Jahan, and A. Milani Fard. "From Text Representation to Financial Market Prediction: A Literature Review". In: *Information* 13.10 (2022). ISSN: 2078-2489. DOI: 10.3390/info13100466.

[3] L. Avramelou, P. Nousi, N. Passalis, and A. Tefas. "Deep reinforcement learning for financial trading using multi-modal features". In: *Expert Systems with Applications* 238 (2024), p. 121849. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2023.121849.

[4] S. Anbaee Farimani, M. Vafaei Jahan, A. Milani Fard, and S. R. K. Tabbakh. "Investigating the informativeness of technical indicators and news sentiment in financial market price prediction". In: *Knowledge-Based Systems* 247 (2022), p. 108742. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2022.108742.

[5] X. Chen, X. Ma, H. Wang, X. Li, and C. Zhang. "A hierarchical attention network for stock prediction based on attentive multi-view news learning". In: *Neurocomputing* 504 (2022), pp. 1–15. ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2022.06.106.

[6] Y. Ma, R. Mao, Q. Lin, P. Wu, and E. Cambria. "Multi-source aggregated classification for stock price movement prediction". In: *Information Fusion* 91 (2023), pp. 515–528. ISSN: 1566-2535. DOI: https://doi.org/10.1016/j.inffus.2022.10.025.

[7] H. Xu, D. Cao, and S. Li. "A self-regulated generative adversarial network for stock price movement prediction based on the historical price and tweets". In: *Knowledge-Based Systems* 247 (2022), p. 108712. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2022.108712.

[8] P. Zeng, G. Hu, X. Zhou, S. Li, P. Liu, and S. Liu. "Muformer: A long sequence time-series forecasting model based on modified multi-head attention". In: *Knowledge-Based Systems* 254 (2022), p. 109584. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2022.109584.

[9] S. A. Farimani, M. V. Jahan, A. M. Fard, and G. Haffari. "Leveraging Latent Economic Concepts and Sentiments in the News for Market Prediction". In: *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*. 2021, pp. 1–10. DOI: 10.1109/DSAA53316.2021.9564122.

[10] Y. Zou and D. Herremans. "PreBit — A multimodal model with Twitter FinBERT embeddings for extreme price movement prediction of Bitcoin". In: *Expert Systems with Applications* 233 (2023), p. 120838. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2023.120838.

[11] H. Wang, S. Lu, and J. Zhao. "Aggregating multiple types of complex data in stock market prediction: A model-independent framework". In: *Knowledge-Based Systems* 164 (2019), pp. 193–204. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2018.10.035.

[12] D. Chen, S. ma, K. Harimoto, R. Bao, Q. Su, and X. Sun. *Group, Extract and Aggregate: Summarizing a Large Amount of Finance News for Forex Movement Prediction*. 2019. arXiv: 1910.05032 [cs.CL]. URL: https://arxiv.org/abs/1910.05032.

[13] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang, and P. S. Yu. "Improving stock market prediction via heterogeneous information fusion". In: *Knowledge-Based Systems* 143 (2018), pp. 236–247. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2017.12.025.

[14] K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev, and D. Trajanov. "Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers". In: *IEEE Access* 8 (2020), pp. 131662–131682. DOI: 10.1109/ACCESS.2020.3009626.

[15] M. Suzuki, H. Sakaji, M. Hirano, and K. Izumi. "Constructing and analyzing domain-specific language model for financial text mining". In: *Information Processing & Management* 60.2 (2023), p. 103194. ISSN: 0306-4573. DOI: https://doi.org/10.1016/j.ipm.2022.103194.

[16] A. H. Huang, H. Wang, and Y. Yang. "FinBERT: A Large Language Model for Extracting Information from Financial Text". In: *Contemporary Accounting Research* 40.2 (2023), pp. 806–841. DOI: https://doi.org/10.1111/1911-3846.12832.

[17] K. Mo, W. Liu, X. Xu, C. Yu, Y. Zou, and F. Xia. *Fine-Tuning Gemma-7B for Enhanced Sentiment Analysis of Financial News Headlines*. 2024. arXiv: 2406.13626 [cs.CL]. URL: https://arxiv.org/abs/2406.13626.

[18] C. Jeong. "Domain-specialized LLM: Financial fine-tuning and utilization method using Mistral 7B". In: *Journal of Intelligence and Information Systems* 30.1 (Mar. 2024), 93–120. ISSN: 2288-4882. DOI: 10.13088/jiis.2024.30.1.093.

[19] W. Luo and D. Gong. *Pre-trained Large Language Models for Financial Sentiment Analysis*. 2024. arXiv: 2401.05215 [cs.CL]. URL: https://arxiv.org/abs/2401.05215.

[20] T. Gokhan, P. Smith, and M. Lee. "Extractive financial narrative summarisation using sentencebert based clustering". In: *Proceedings of the 3rd Financial Narrative Processing Workshop*. 2021, pp. 94–98.

[21] N. Reimers and I. Gurevych. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Ed. by K. Inui, J. Jiang, V. Ng, and X. Wan. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 3982–3992. DOI: 10.18653/v1/D19-1410.

[22] X. Li and J. Li. "AoE: Angle-optimized Embeddings for Semantic Textual Similarity". In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 1825–1839. DOI: 10.18653/v1/2024.acl-long.101.

[23] T. Kim, J. Kim, Y. Tae, C. Park, J.-H. Choi, and J. Choo. "Reversible Instance Normalization for Accurate Time-Series Forecasting against Distribution Shift". In: *International Conference on Learning Representations (ICLR)*. 2022.

[24] H.-H. Tung, C.-C. Cheng, Y.-Y. Chen, Y.-F. Chen, S.-H. Huang, and A.-P. Chen. "Binary Classification and Data Analysis for Modeling Calendar Anomalies in Financial Markets".

In: *2016 7th International Conference on Cloud Computing and Big Data (CCBD)*. 2016, pp. 116–121. DOI: 10.1109/CCBD.2016.032.

[25] T.-T. Nguyen and S. Yoon. "A Novel Approach to Short-Term Stock Price Movement Prediction using Transfer Learning". In: *Applied Sciences* 9.22 (2019). ISSN: 2076-3417. DOI: 10.3390/app9224745.

[26] J. Chen. "Analysis of Bitcoin Price Prediction Using Machine Learning". In: *Journal of Risk and Financial Management* 16.1 (2023). ISSN: 1911-8074. DOI: 10.3390/jrfm16010051.

[27] S. A. Farimani, M. V. Jahan, and A. M. Fard. "An Adaptive Multimodal Learning Model for Financial Market Price Prediction". In: *IEEE Access* (2024), pp. 1–1. DOI: 10.1109/ACCESS.2024.3441029.

[28] J. L. Ba, J. R. Kiros, and G. E. Hinton. *Layer Normalization*. 2016. arXiv: 1607.06450 [stat.ML]. URL: https://arxiv.org/abs/1607.06450.

[29] N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosifidis. "Deep Adaptive Input Normalization for Time Series Forecasting". In: *IEEE Transactions on Neural Networks and Learning Systems* 31.9 (2020), pp. 3760–3765. DOI: 10.1109/TNNLS.2019.2944933.

[30] D. Cao, F. Jia, S. O. Arik, T. Pfister, Y. Zheng, W. Ye, and Y. Liu. "TEMPO: Prompt-based Generative Pre-trained Transformer for Time Series Forecasting". In: *The 12th International Conference on Learning Representations (ICLR'24)*. 2024.

[31] Y. Liu, H. Wu, J. Wang, and M. Long. "Non-stationary Transformers: Exploring the Stationarity in Time Series Forecasting". In: *Advances in Neural Information Processing Systems (NeurIPS)*. Ed. by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh. Vol. 35. Curran Associates, Inc., 2022, pp. 9881–9893.

[32] Q. Li, Y. Chen, L. L. Jiang, P. Li, and H. Chen. "A Tensor-Based Information Framework for Predicting the Stock Market". In: *ACM Trans. Inf. Syst.* 34.2 (2016). ISSN: 1046-8188. DOI: 10.1145/2838731.

[33] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *Proceedings of The 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*. Vol. 1. 2. Minneapolis, Minnesota. 2019.

[34] T. Gao, X. Yao, and D. Chen. "SimCSE: Simple Contrastive Learning of Sentence Embeddings". In: *Empirical Methods in Natural Language Processing (EMNLP)*. 2021.

[35] H. Shi, J. Gao, H. Xu, X. Liang, Z. Li, L. Kong, S. M. S. Lee, and J. T. Kwok. "Revisiting Over-smoothing in BERT from the Perspective of Graph". In: *International Conference on Learning Representations (ICLR'22)*. 2022.

[36] T.-W. Lee, P. Teisseyre, and J. Lee. "Effective Exploitation of Macroeconomic Indicators for Stock Direction Classification Using the Multimodal Fusion Transformer". In: *IEEE Access* 11 (2023), pp. 10275–10287. DOI: 10.1109/ACCESS.2023.3240422.

[37] G. Wang, J. Ma, and G. Chen. "Attentive statement fraud detection: Distinguishing multimodal financial data with fine-grained attention". In: *Decision Support Systems* 167 (2023), p. 113913. ISSN: 0167-9236. DOI: https://doi.org/10.1016/j.dss.2022.113913.

[38] J. Delgadillo, J. Kinyua, and C. Mutigwe. "FinSoSent: Advancing Financial Market Sentiment Analysis through Pretrained Large Language Models". In: *Big Data and Cognitive Computing* 8.8 (2024). ISSN: 2504-2289. DOI: 10.3390/bdcc8080087.

[39] K. Chahuán-Jiménez. "Neural Network-Based Predictive Models for Stock Market Index Forecasting". In: *Journal of Risk and Financial Management* 17.6 (2024). ISSN: 1911-8074. DOI: 10.3390/jrfm17060242.

[40] D. Araci. "Finbert: Financial sentiment analysis with pre-trained language models". In: *arXiv preprint arXiv:1908.10063* (2019). URL: https://arxiv.org/abs/1908.10063.

[41] S. Gangopadhyay and P. Majumder. "Text representation for direction prediction of share market". In: *Expert Systems with Applications* 211 (2023), p. 118472. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2022.118472.