IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 30 May 2024, accepted 24 July 2024, date of publication 9 August 2024, date of current version 10 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3441029

RESEARCH ARTICLE

An Adaptive Multimodal Learning Model for Financial Market Price Prediction

SAEEDE ANBAEE FARIMANI[1], MAJID VAFAEI JAHAN[1], AND AMIN MILANI FARD[2]

[1]Department of Computer Engineering, Islamic Azad University, Mashhad Branch, Mashhad 9187147578, Iran
[2]Department of Computer Science, New York Institute of Technology, Vancouver, BC V5M 4X5, Canada

Corresponding author: Majid Vafaei Jahan (vafaeijahan@mshdiau.ac.ir)

This work was supported in part by Iran's Ministry of Science, Research and Technology under Grant 12-99-02-000049; and in part by the New York Institute of Technology - Vancouver Campus.

ABSTRACT Investors' trading behavior is influenced by a multimode of information sources such as technical analysis, news dissemination, and sentiment, which results in the non-stationary behavior of financial time series. With advancements in deep learning, studies considering temporal relationships in each data mode and applying heterogeneous data fusion techniques for market prediction are increasing. While net price change prediction is helpful for investors, most previous deep learning models only predict the up/down trend of price as the non-stationary behavior of price time series influences the regression performance. In this work, we present an adaptive model for price regression, which learns interdependencies between the distribution of multimode data and the amount of price change around an average price for snapshots of systems. We use news content, the mood in specialized newsgroups, and technical indicators for data representation. Different news topics, also known as modalities, can be absorbed by investors with different diffusion speeds; hence we use a concept-based news representation method that reflects news topics in a news vector. Also, our model considers the positive/negative mood in specialized newsgroups and technical indicators. To capture complex temporal characteristics in the distribution of economic concepts in the news sequence, we use a recurrent convolutional neural network and other recurrent layers to perceive changes in technical indicators and mood in specialized newsgroups. In the fusion layer, our model learns to normalize data points based on their estimated distribution and the importance weight of each data mode to handle multimodality challenges. To overcome the non-stationary behavior of price, we let the network learn how to drift the predicted values around the average price of that packet. Our experiments demonstrate a significant 40.11% error reduction compared to the baselines. We also discuss the adaptability, and price prediction capability of our proposed approach.

INDEX TERMS Adaptive fusion strategy, cryptocurrency market, financial sentiment analysis, market prediction, news document representation.

I. INTRODUCTION

In the past decades, technical and fundamental analyses have been used in behavioral economics [1], [2] to study the influence of psychological, social, and political factors on investors' decision making [3]. It has been shown that news related to economic, political, social, or emotional events posted on social media or newsgroups affect investors [2], [4], [5], [6], [7], [8]; hence, leveraging news and historical market information would be beneficial for financial decision making [9], [10], [11], [12], [13], [14].

The Mixed Distribution Hypothesis (MDH) [15] states that the net price change $\Delta P(\tau)$ over period τ, that has MIXED distribution, is hypothesized to be the sum of $\xi(\tau)$ incremental interperiod steps $\Delta P\tau_i$ as $\Delta P(\tau) = \sum_i^{\xi(\tau)} \Delta P\tau_i$ where randomized time $\xi(\tau)$ can be learned through the information events distribution that traders experienced during τ and $\Delta P(\xi(\tau))$ has a mixed distribution. Figure 1 depicts $\Delta P(\tau)$ hourly distribution and the distribution of *information flow* from specialized newsgroups in the form of hourly sentiment

The associate editor coordinating the review of this manuscript and approving it for publication was Ángel F. García-Fernández.

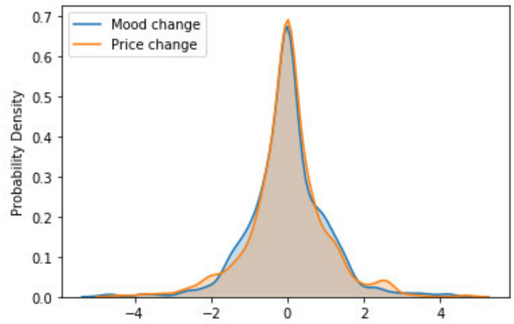


FIGURE 1. **Probability distribution of the net price change '$P(\tau)$ for BITCOIN/TETHERUS hourly close price from Jun 2020 to march 2022 and the distribution of changes in the sentiment time series of the bitcoin news.**

time series that bitcoin traders react to positive/negative moods. According to the MDH the variance of market return at a given time interval is proportional to the amount of information received by the market participants [9], [10], [12], [14]. Unstructured real-world news can inherently take many topics, also known as modalities, and be absorbed by investors with different diffusion speeds [16]. The qualitatively different types of *information flow* cause different innovations in returns [15]; therefore, a mixture of random variables causes changes in the mean and variance of price time series, also known as the non-stationary behavior of price.

So far, scholars have used deep neural networks to model *information flow* to the financial markets via embeddings of the textual data and technical indicators [17], [18], [19], [20], [21]. Most of these predictive models apply text mining on unstructured data to analyze news influence without considering latent relations among news with similar concepts. A few works [22], [23], [24] consider the joint effect of news, mood, and market-based information. However, the importance of such information sources and the lead-lag effect of an information source during the time that partially replicates the movements of the other in the price prediction problem requires more study.

To benefit from the information in the news, Farimani et al. in [20] propose a concept-based news representation learning method to create a low-dimensional space of latent economic concepts by clustering transformer-based word embeddings and then propose a news vectorization based on all information in the news title and body. This concept-based news representation approach calculates news vectors through latent concepts space, where a latent concept contains a group of semantically related words that occur in similar contexts, also known as a topic. Researchers in [25] and [26] analyze the efficacy of the clustering-based approach in enhancing the ability of contextualized embedding for sentence similarity tasks based on Bidirectional Encoder Representation (BERT) [27]. The results in [20], [25], and [26] indicate the capability of concept clustering based on BERT contextualized embedding in reflecting context proximity between relevant text. Also, the news representation method in [20] outperforms others, such as Word2Vec in [18], BERT CLS embedding in [28] and [29], and other concept-based approaches [30], [31] for price prediction. To tackle the problem of leveraging context proximity in the news, we use the concept modeling approach presented in [20].

Another information source for market analysis used by many researchers [21], [22], [32], [33] is the sentiment of social media posts and news feeds. To benefit from information of mood as the aggregated sentiment scores of specialized financial news during trading time intervals [34], Farimani et al. in [22] perform mood time series based on a BERT-based Transformer model (FinBERT) [35] fine-tuned for Financial Sentiment Analysis (FSA) and show FinBERT-based FSA outperforms lexicon-based [36], and machine learning-based sentiment analysis methods [37] for price forecasting problem. Selecting a proper set of market data and technical indicators as other data sources are addressed in [20] and [22] via a nonlinear information gain-based feature selection method.

The model in [20] can capture complex temporal patterns in the news based on news distribution through latent concepts and leverages the most informative market-based data. However, feature fusion is done by concatenating extracted features without considering their multimodal to move the market with respect to the price in a previous lag window. In time series prediction problems, models traditionally leverage a sequence of past values of features to predict the current time stamp, and lag is the length of previous steps [38]. The concatenation of such lag-based features extracted from various information sources in a compound vector makes it difficult for neural networks to distinguish the importance of such various lag-based features in price fluctuations.

In this work, we present our Adaptive Behaviour Model-Bag of Concepts and Sentiment and Informative Market Data (*ABM-BCSIM*) that is built on top of the BERT-based Bag of Economic Concept (BERT-BoEC) [20] and FinBERT-based Sentiment and Informative Market (FinBERT-SIMF) [22] methods. *ABM-BCSIM* extracts news-based features from BERT-BoEC via a recurrent temporal convolution layer (RCNN), and extracts market and mood-based features using FinBERT-SIMF with two separate recurrent layers. Our model leverages an adaptive fusion strategy that learns how to balance multimode of extracted features from various information sources to handle multimodalities of information sources, and predicts the amount of price change around a lag-based average price for price regression to overcome the challenge of non-stationary behavior of price time series.

Our model's primary objective is to predict the offset of price changes around the average price for each system snapshot τ. These price changes, observed in the time stamp $\tau + 1$, occur in response to the arrival of multimodal information to investors during the snapshot τ. Our proposed model extracts features from various information sources and learns the amount of price fluctuation from the average price based on market data distribution, news topics, and mood time series during τ. In the first step, following [39], we let the network learn how to adaptively normalize the input values according to their instance distribution during τ. Then, after feature extraction in the second step, we let the network learn the weights of each lag-based feature during end-to-end learning. Finally, in the third step, the network is forced to learn the fluctuation around the average price during τ and combine the predicted value with the average.

We implement *ABM-BCSIM* for exchange rate prediction in several important and large markets. We focus on major currency pairs in the Foreign Exchange Market (Forex), as well as Bitcoin, the most popular cryptocurrency, and gold in the commodities market. Our results demonstrate that considering both the multimodality of data sources and changes in the price distribution in each snapshot helps overcome the non-stationary behavior of price time series. This approach leads to a noteworthy reduction in errors within our regression model. The main contributions of this work are as follows:

1) Regarding the temporal dependencies between market movements and various news topics, we employ an RCNN layer to extract complex temporal patterns in the distribution of BERT-based latent concepts in the news during a delay window. Additionally, we utilize two RNN layers to extract complex temporal patterns from market and mood information simultaneously.
2) We introduce an adaptive fusion strategy for feature fusion by employing weighted concatenation of lag-based news, mood, and market-based extracted features. Regarding price regression, we utilize a normalization layer for adaptive data normalization based on the data distribution. Additionally, we incorporate a denormalization layer that learns how to adjust the predicted price fluctuation around the lag-based average price.
3) we investigate the lead-lag effects of multimode data and analyze the adaptability of our approach to diverse markets. Moreover, we assess the profitability of our proposed model.
4) We expand the *MarketNews* dataset presented in [22] *MarketPredict*[1] and available the source code of our RESTFul APIs for accessing and scraping specialized news.[2]

II. RELATED WORK

Extracting accurate features from various information sources, including news, trending hashtags, investor transactions, and social media moods, poses a challenge when applying machine learning models to market prediction tasks. In this study, we review the literature on diverse information sources that have been explored for market analysis. Additionally, we examine fusion strategies utilized for feature extraction from these heterogeneous data sources.

A. INFORMATION SOURCES

1) NEWS

Most of the fundamental market analysis methods are based on BoW, and feature selection using this technique ignores the semantics of the news content [2], [11]. A few works considered the influences of political, social, and cultural events on investors' behavior and used information retrieval techniques for identifying important events for stock return prediction [40], [41], [42].

Due to a linear space matrix decomposition [17], concept-based models such as Latent Semantic Analysis (LSA) [43] used by Feuerriegel et al. [44] fail to incorporate nonlinear semantics between the words. Latent Dirichlet Allocation (LDA) [45] has been used for one-day-ahead stock movement prediction [30] to generate latent concepts-based text vectors; however, it has time complexity limitations.

Huynh et al. [14] consider the semantic relationships among the words and use word embedding [46] within Bidirectional Gated Recurrent Unit (BGRU) to forecast S&P500 stock market. This method uses bidirectional relations between words in news titles. Lutz et al. [47] used Doc2vec embedding [31] for predicting the sentence sentiment in their financial decision support system. The generated low-dimensional vectors capture the proximity in the news title and description; however, the meaning of each feature is not explainable.

Transformer-based language models capture the deep semantic and syntactic information amongst words [27], [48], [49]. To extract explainable features from the news, [28] used BERT [27] to capture deep semantic information. Chen et al. [29] proposed a news aggregating method for Forex market prediction based on BERT. They use the *[CLS]* token as the vector representation of the news title while losing important information in the news content. Farimani et al. in [20] propose context-aware conceptual document representation to model the relevance between the news based on all information in financial news titles and bodies via clustering of contextualized BERT word embedding. The BERT-BoEC method in [20] leverages the conceptual relationship in the new that outperforms other concept-based approaches and other embedding techniques, such as BERT-based [cls] embedding.

2) MOOD

The application of sentiment analysis for performing mood signals as the aggregated sentiment score of news or posting during a specific granule of time in financial market prediction [50] promises good results in market prediction. Scholars in [4], [18], [22], and [19] show using mood together with market data results in better prediction compared

[1] https://github.com/MarketPredict-BoEC/Market-Predict
[2] https://github.com/FinBERT-SIMF/FinBERT-SIMF

to using solely technical indicators for market prediction. Researchers in [21], [28], [32], and [33] proposed sentiment analysis methods for the analysis of financial news headlines and examined social media mood influence on investors' decisions [4], [5]. Most of the previous works [21], [36] use lexicon-based sentiment analysis with domain-specific lexicons such as VADER [51] or Loughran-McDonald [52] prepared by human experts' intervention without considering the context of words. Other lines of works [13], [53] use a machine learning-based classifier model for FSA learned from an automatically labeled dataset indicated based on the market rise or fall after news publishing; however, ignore the asymmetric market response to good versus bad news [54]. Authors in [55] and [56] consider the trend prediction problem and show that transformer-based [27] sentiment analysis outperforms the other text representation methods. Scholars in [22] and [57] use a fine-tuning uncased version of the BERT model for FSA based on a manually labeled news dataset for the prediction problem. FinBERT [35] is a BERT-based [27] transformer language model fine-tuned for financial domain sentiment analysis. In this work, following [22], we use FinBERT for mood signal calculation.

3) MARKET DATA

Technical analysts employed trading disciplines based on analysis of basic market data such as price and volume information as well as technical indicators such as EMA, MACD, RSI, and On Balance Volume (OBV) [58], [59], [60]. So far, researchers have analyzed primary market data such as close, open, low, and high price and volume or $close - open$ for stock market prediction [61]. The author in [18] examined technical indicators for sentiment-based stock market prediction. Peng et al. in [62] propose a LASSO (Absolute Shrinkage and Selection Operator) based feature selection strategy for finding a profitable category of technical indicators for price direction prediction. Farimani et al. in [22] consider the interdependencies between mood signals and technical indicators and propose a feature selection based on information gain and nonlinear weights of the input gate of a recurrent neural network for finding the proper technical indicators in cryptocurrency and exchange rate prediction. This work aims at cryptocurrency, gold, and exchange rate prediction. As such, we use the feature selection procedure in [22] for price regression.

B. INFORMATION FUSION

First leading media-aware studies [11], [63] learn from the composition of textual data and market indicators in the form of a compound vector for market prediction. The authors in [13] propose feature selection methods to select the discriminative news-based features while concatenating prepared features from news, sentiment, or market data in a compound vector which makes it difficult to distinguish the effect of news from market data and often ignores the interdependencies between news. Another multimodal market decision support system works based on tensor decomposition solutions. Li et al. [23] use tensor decomposition to learn from news content representation based on Bag of Words (BoW), mood, and stock market data. Zhang [24] proposed a coupled matrix and tensor decomposition based on Coupled Stock Correlation (CSC) criteria. They used information from the title's verb embeddings as news events, the joint impacts of the users' sentiments from social media, and the stock price movements for stock trend prediction. Wang et al. [64] incorporate scalar data from the trading volume, functional data from intraday return series, and compositional data from investors' emotions analysis through the tensor decomposition-based framework to competently predict the open trend in the next trading day. Tensor-based approaches promise good results for handling heterogeneous data sources, while iterative estimation of parameters for each data mode is a time-consuming process and makes it difficult for models to learn the lead-lag effect of the news on market movements. These approaches extract features from a joint representation of multimodal data, while feature extraction from a sequence of news information makes it simpler to cover the lead-lag effect of the news on the market.

TABLE 1. Notations and symbols.

Notation	Description
$\mathcal{G}$	News Corpus including all news items relevant to target market
$\mathcal{M}$	Market information space including Open/Close/Low/High price, Volume and technical indicators time series.
k	Number of latent concepts in the market news corpus or document embedding dimension.
$\mathcal{C}^k$	Set of Embedded latent concepts space $\{c_1, \ldots, c_k\}$, where c_i represents the i-th economic latent concept cluster holding similar embedding vectors.
$X_N^{(t)}$	A sequence $\{x_1, \ldots, x_L\}$ of BERT-BoECE representation of News published during delay window d, where $x_i \in C^k$
L	Number of news in a delay window d.
$X_S^{(t)}$	Mood data items including aggregated positive, negative, and neutral scores of news published during delay window h, where $X_S^{(t)} \in \mathbb{R}^{h*3}$.
$X_M^{(t)}$	Market data items during delay window of h, that includes close price, Exponential Moving Average (EMA), Bollinger Bands (BBs) and On Balance Volume (OBV)
$a_{close}^{[t-h,t]}$	Average of close price P_{close} in $X_M^{(t-h,t)}$, $a_{close}^{[t-h,t]}$ $= \frac{1}{h} \sum_{i=0}^{h} P_{close}^{(t-i)}$,

Other works [7], [16], [18], [21], consider the interdependencies between each mode of data as well as the joint effect of chaotic data sources on each other via a deep neural network-based feature extraction. Xu et al. in [19] consider the lead-lag effects on news on market data and superior treatment of market stochasticity. Their generative model jointly exploits text and price signals for assessing continuous latent variables via a recurrent-based model; however, the Bi-directional Long Short Memory (Bi-LSTM) based text representation does not reflect the context proximity between news. Chen et al. in [29] presented a Bert-based Hierarchical Aggregation Model for extracting features from groups of news. They incorporated these

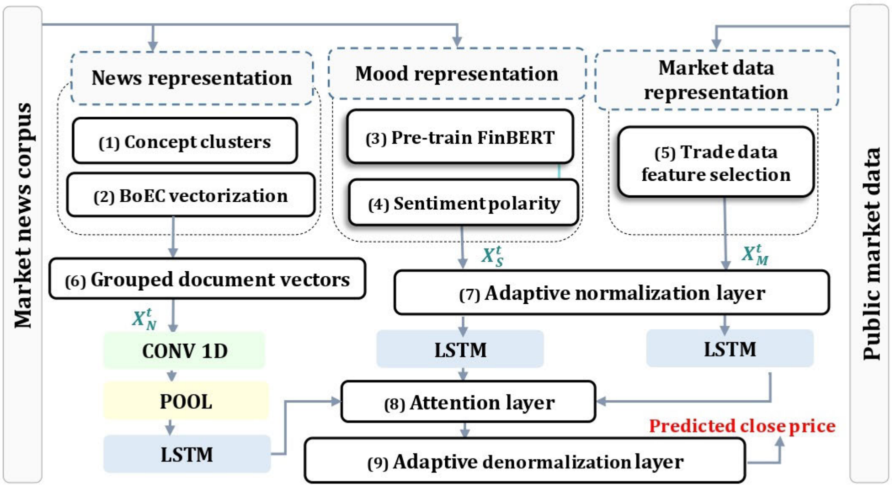


FIGURE 2. Overview of the system.

features with MLP-based extracted features from market data in the form of the compound vector into dens layer for decision making. Farimani et al. in [20] consider the temporal dependencies between the news and market data and propose temporal feature extraction from news and market data via two separate recurrent layers. However, decision-making is done based on the concatenated hidden states of these two recurrent layers making it difficult to adapt decision-making based on the contribution of news or market data in price fluctuations. Most existing methods predict the price trend as up/down binary classification, and price regression is not well explored. In this work, we consider the lead-lag effects of news on price fluctuations and propose an adaptive fusion strategy for decision-making based on the contribution of news, mood, and technical analysis on the price regression problem.

III. THE PROPOSED APPROACH

This section explains our problem statement and the adaptive decision-making process. Table 1 presents notations that we use throughout the paper.

Problem Statement: Our goal is to predict the absolute value of the exchange rate of an asset $p^{(t+1)}$ on target trading hour $t+1$. Similar to the MDH, let g be a function that assesses the price changes based on the composition of information from various data sources. Let $\hat{g}$ be the optimized loss function over all T instances in the training set, and $g \in \hat{g}$ be the autoregressive function that assesses the price changes due to information flow to the market around a lag-based average price based on different data modes, including news topic distributions, mood and technical indicators and the parameters of this function learned during end-to-end model training. We use three lag-based data from relevant news corpus $\mathcal{G}$ and market information space $\mathcal{M}$: (1) news content information $X_N^{(t)}$ during $[t-\Delta t, t]$, where $\Delta t = d$ is a fixed *news data lag size*, and (2) and (3) mood time series $X_S^{(t)}$ and market information $X_M^{(t)}$ during $[t-\Delta t, t]$, where $\Delta t = h$ is a fixed *market data lag size*. Therefore the objective is to find the optimum value of $\widehat{g}(X_N^{[T-d,T]}, X_S^{[T-h,T]}, X_M^{[T-h,T]})$ in Equation (1), as shown at the bottom of the next page. Formally, we define these data sources as follows:

Definition 1: (News Information) Let $X_N^{(t)}$ be a sequence of $\{x_1, \ldots, x_L\}$ of news representation arranged in temporal order based on publishing timestamps, where $x_l \in \mathcal{C}^K$ be a vector of the distributed representation of concepts in the news l-th based it's Term Frequencies (TF) of *title* and *content* words in each of concepts $\{c_1, \ldots, c_k\}$.

Definition 2: (Latent Economic Concepts) Let $\{c_1, \ldots, c_k\}$ be a group of concept clusters or topics, where c_k is a cluster of semantically related words in sentences of corpus $\mathcal{G}$. Let $E[w_{ij}]$ be the embedding vectors for the i-th token in the j-th sentence in corpus $\mathcal{G}$. We construct $\mathcal{C}^k$ by clustering all embeddings $E[w_{ij}]$ in the corpus into k clusters using the k-means algorithm, where u_l denotes the center of the l-th cluster, then, the token w_{ij} is assigned to the cluster $c = \arg\min_l (E[w_{ij}].u_l)$.

Definition 3: (Mood Time Series) Let $X_S^{(t)}$ be a 3-dimensional time series of aggregated sentiments scores of news that were published during consecutive trading time intervals, where the sentiment score of news comes from the probability distribution of news *title* embedding over three classes of positive, negative, and neutral in the classifier layer on top of a pre-train transformer model fine-tuned for financial sentiment analysis (FSA).

Definition 4: (Market Data) We define the market data time series $X_M^{(t)}$ as a 4-dimensional vector of observations $X_M^{(t)} = (m_1^{(t)}, \ldots, m_4^{(t)})'$ during consecutive trading time intervals, where $m_q \in \mathcal{M}$ is an informative market data including the close price, Exponential Moving Average (EMA), Bolinger Bands (BBs) and On Balance Volume (OBV) or Accumulation/Distribution Index (ADI) indicators. [22] suggest these indicators have the most information

Algorithm 1 Bag-of-Economic-Concept (BoEC) Slightly Different From [20] (Without Title Expansion)

Input : News document $N_l = \{title, content, timestamp\}$, embedded latent concepts space $\mathcal{C}^k = \{c_1, \ldots, c_k\}$
Output: Vector $x_l \in \mathcal{C}^k$ corresponding to the news document l

```
1 /* Step 1: Preprocessing */
2 titleKeywords ← preprocess(title)
3 descKeywords ← preprocess(content)
4 totalWords ← append(titleKeywords, descKeywords)
5 /* Step 2: Vectorization */
6 for i = 1 to k do
7     x_l[i] ← number of tokens in totalWords ∈ c_i
8 return x_l
```

gained from absolute values of the target close price time series, via considering inter-dependencies between sentiment information and various market data items.

We propose *ABM-BCSIM* for financial market decision support. Our approach consists of three phases: ① feature preparation, ② feature extraction, ③ fusion and market prediction, as depicted in Figure 2. This work contributes to adaptive decision-making based on news and market-based data importance. In the regression optimization problem, we consider the mean value of the target price and let the network learn how to balance the importance weight of news and market-based extracted features with the *different lags* to find the optimum value of price drift around the average close price. Our feature preparation phase is mostly based on [20], [22]. The main reason behind using concept clustering for news vectorization is to let the convolution kernels compress news vectors based on the most important concepts, where each concept frequency in the news vector reflects a topic distribution for a piece of news in the concept clustering space.

A. FEATURE PREPARATION

1) NEWS CONTENT EMBEDDING

In order to compute the news embedding vector x_l in $X_N^t = \{x_1, \ldots, x_L\}$, we first build latent economic concepts space $\mathcal{C}^k$ by clustering embedding vectors into k latent concept clusters of words in the sentences of news corpus $\mathcal{G}$ via BERT word embedding [27] (Block 1 of Fig 2). Then for BoEC news vectorization, we compute the term frequency of the news words in such transformer-based concepts (Block 2 of Fig 2). This distributed topic representation for news contributes to modeling different modalities of topics in the news through low-dimensional news vectors.

The BoEC vectorization procedure [20] is shown in algorithm 1 and figure 3. In the first step, the title and content of a news document are preprocessed by removing numbers, tags, URLs, and stop words (lines 2 and 3). Variable *totalWords* holds all keywords in content and title. Finally, in the vectorization step, keywords distribution through latent topics would be calculated (line 7). Hence, the dimension of the document vector x_l is identical to the number of latent concepts in embedding concepts space, i.e., k.

2) MOOD TIME SERIES

For performing $X_S^{(t)}$ as the 3-dimensional mood time series during $(t-h)$ and (t), for each consecutive trading time interval (an hour), we sum up the sentiment scores of news published between $t^{(i-1)}$ and t^i as $\{X_{S(pos)}^{(t_i)}, X_{S(neg)}^{(t_i)}, X_{S(neut)}^{(t_i)}\}$. Following [22], We use the predicted probabilities of the classifier layer of the FinBERT model [35] $P(title_l) =$ softmax $(W * title_l^{[CLS]} + b)$ for news $title_l^{[CLS]}$ embedding as the news sentiment scores in three classes of *positive*, *negative* and *neutral*, where b and W are the FinBERT classifier layer parameters.

3) MARKET DATA TIME SERIES

We fed each market data item $\{X_M^{(t)} \in \mathbb{R}^{4*h}, t = 1 \ldots h\}$ into the adaptive normalization layer [39] to let the network learn how to normalize current input instead of normalization of the whole market data time series based on global average and variances (Block 7). The adaptive normalization layer is responsible for learning how to normalize each $\{m_q^{(t)}, t = 1, \ldots, h\}$ in $X_M^{(t)}$ separately, in three steps: (1) linear shifting transformation, (2) linear scaling transformation, and (3) nonlinear gating layer, based on its variance and average. This normalization strategy contributes to making adaptable market price predictions around the moving average price under the presence of all market and news-based information, instead of static normalization that causes the predicted price to drift above or below the absolute price value, especially in the inference phase.

Let $\tilde{X}_M^{(t)}$ be the adaptive normalized form of $X_M^{(t)}$ based on the following equation:

$$\tilde{X}_M^{(t)} = (X_M^{(t)} - \alpha^{(t)}) \odot \beta^{(t)} \tag{2}$$

where $\odot$ is element-wise multiplication, and $\alpha^{(t)}$ and $\beta^{(t)}$ are learnable shifting and scaling factors of the distribution of current inputs that should be estimated during the end-to-end training of the model in figure 2. News dissemination stimulates investors to take rushed positions, leading to the non-stationary and multimodal distribution in market price time series [1]. Hence, learning parameters of the distribution and normalization of data points by leveraging news helps with a better estimation than using only market data for training.

$$\widehat{g}(X_N^{[T-d,T]}, X_S^{[T-h,T]}, X_M^{[T-h,T]}) = \underset{g \in \hat{g}}{argmin} \frac{1}{T} \sum_{t=1}^{T} \left| \frac{(g(X_N^{[t-d,t]}, X_S^{[t-h,t]}, X_M^{[t-h,t]}) + \frac{1}{h}\sum_{i=0}^{h}(P_{t-i})) - P_{t+1}}{P_{t+1}} \right| \tag{1}$$

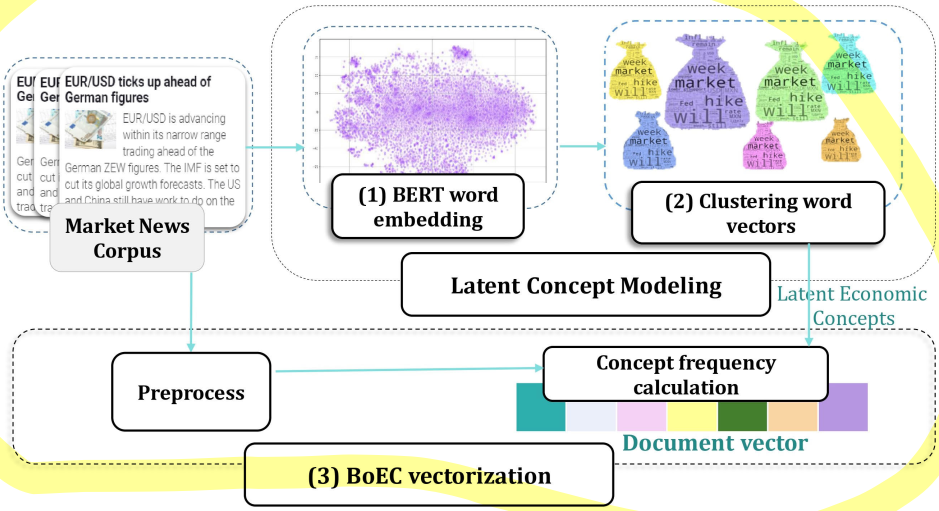


FIGURE 3. News document vectorization process.

In step (1), the linear shifting transformation is done by $\alpha^{(t)} = W_a a^{(t)} \in \mathbb{R}^4$, $a^{(t)} = \frac{1}{h}\sum_{j=1}^{h} m_q^{(j)}, q = 1, \ldots, 4$ where $W_a \in \mathbb{R}^{4*4}$ is the weight matrix, and $a^{(t)}$ is the averages of q features in $X_M^{(t)}$.

In step (2), the scaling transformation is performed to scale all shifted values, $\beta^{(t)} = W_b b^{(t)}$, where $W_b \in \mathbb{R}^{4*4}$ is the weight matrix, and $b^{(t)} = \sqrt{\frac{1}{h}\sum_{j=1}^{h} (m_q^{(j)} - \alpha_q)^2}, q = 1, \ldots, 4$ is the updated values of the standard deviation of data.

In step (3), adaptive gating transformation is done to suppress not relevant/useful features by $\tilde{\tilde{m}}^{(t)} = \tilde{m}^{(t)} \odot \gamma^{(t)}$ nonlinear transformation, where $\gamma^{(t)} = \sigma(W_g g^{(t)} + b)$, and σ is the sigmoid function, $W_g \in \mathbb{R}^{4*4}$, $b \in \mathbb{R}^4$ and $g^{(t)} = \frac{1}{h}\sum_{j=1}^{h} \tilde{m}_q^{(j)}, q = 1, \ldots, 4$ is the average of transformed data that is adaptively shifted and scaled in step (1) and step (2) based on Equation 2.

B. FEATURE EXTRACTION

1) NEWS FEATURES

News sequence $X_N^{(t)}$ is fed into a temporal convolutional layer followed by a pooling and recurrent layer. We use a temporal convolution layer to capture conceptual relationships in a sequence of news. Then, the temporal max pooling layer captures the most important information in the distributed representation of topics in the news sequence in the form of compressed feature vectors. In the recurrent layer, we use an LSTM unit after the convolutional layer, to learn time dependencies between the news sequence during delay window d and extract features from the compressed form of the conceptual representation of $X_N^{(t)}$.

2) MARKET AND MOOD DATA

Mood $X_S^{(t)}$ and market data $X_M^{(t)}$ are passed to the normalization layers (Block 7 of Fig. 2) and then passed to the recurrent layers to learn the time dependencies between technical indicators and aggregated scores of news sentiments during the delay window h. Note that we align news and market data based on the timestamp in chronological order and then perform $X_S^{(t)}$ and $X_M^{(t)}$ in a delay window length h.

C. FUSION AND PREDICTION

Regarding extracted feature from $X_M^{(t)}$ and $X_S^{(t)}$ via two RNN layers and $X_N^{(t)}$ via a recurrent convolution layer, we pass hidden states of three LSTM units to the attention layer (Block 8 of Fig. 2) and then pass the weighted concatenation of feature vectors to the denormalization layer (Block 9 of Fig. 2). Note that we align news and market data according to their timestamp in chronological order; therefore, extracted features are based on news and market data in temporal order.

Figure 4 shows the details of the attention layer of the model. The attention weight of each data modality is calculated with the temporal attention mechanism. The news attention weight $a_N^{(t)}$ is obtained by Equation 3 where $h_N^{[t-d,t]}$ is the hidden state vectors of corresponding news LSTM layer during lag d, and W_o^N and b_o^N are learnable weights. The context vector of news is obtained by $o_N^{(t)} = \sum_{i=0}^{d} a_N^{(t-i)} * h_N^{(t-i)}$. For calculating the context vector of sentiment $o_S^{(t)}$ and the context vector of market data $o_M^{(t)}$, we did the same way.

$$a_N^{(t)} = Softmax(tanh(W_o^N * h_N^{[t-d,t]} + b_o^N)) \tag{3}$$

The intuition behind this temporal attention mechanism is that the model dynamically chose whether to observe one of the sources more intensely during a period where it would be more important. The summation of $o_N^{(t)}$, $o_S^{(t)}$, and $o_N^{(t)}$ is passed to a dense layer with linear activation to obtain $\tilde{o}^{(t)}$ as the optimum value of the price changes function $g(X_N^{t-d,t}, X_S^{t-h,t}, X_M^{t-h,t})$ (Equation 2) in response to multimode *information flow* to the market. The predicted value $\tilde{o}^{(t)}$ in Equation 4 is an offset of the price change at

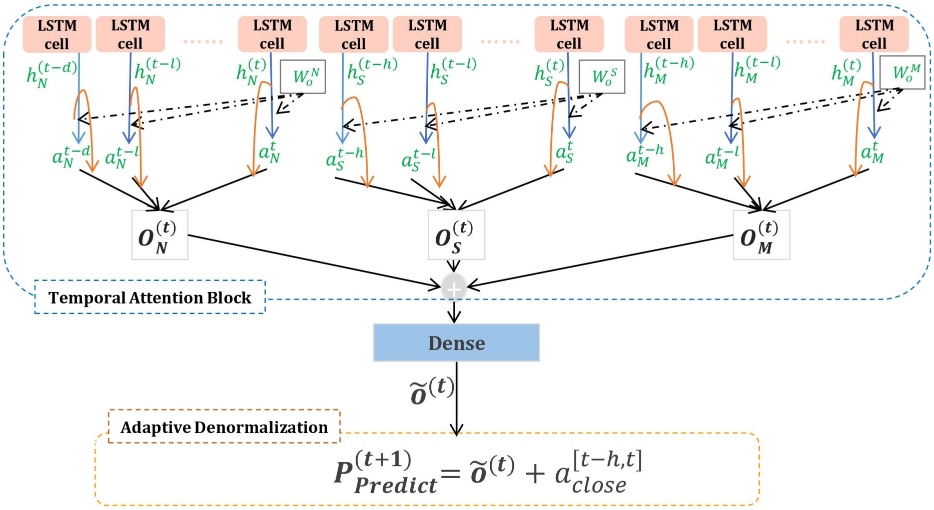


FIGURE 4. Network structure for fusion and prediction phases.

TABLE 2. Corpora statistics.

Corpus	News Count	keywords count	Sentences count	Sentence Length
EUR/USD	10,920	27,516	129,889	15/80
GBP/USD	7,600	23,705	83,748	17/49
USD/JPY	5,776	20,104	60,772	17/66
XAU/USD	5,070	26,338	81,875	15/16
BTC/USDT	11,147	46,905	159,686	15/12

time step $t+1$ which is added to the average of the close price during the lag window.

$$p_{predict}^{(t+1)} = \tilde{o}^{(t)} + a_{close}^{[t-h,t]} \quad (4)$$

Hence, the denormalization layer is responsible for adjusting this predicted price change around the lag-based average price $a_{close}^{[t-h,t]}$ for price regression. In the denormalization layer, we summed this obtained offset with the lag-based average price $a_{Close}^{[t-h,t]}$ to make sure that all the loss values during the backpropagation process are primarily dedicated to adjusting the predicted price change.

Since we formulate our task as a regression problem, we consider the mean squared error (MSE) loss function and use adaptive moment estimation (Adam) to minimize the loss value overall instance in the training set.

IV. EXPERIMENTS

A. DATASET

1) NEWS DATASET

In a review conducted by researchers [8], they discovered a noticeable imbalance in the attention given to studying the effects of news on FOREX and Cryptocurrencies compared to the extensive research focused on stock markets. We extended the *MarketNews* dataset presented in [22] to address this gap by increasing its duration and incorporating news related to XAU/USD. This dataset expansion will allow for a more comprehensive analysis of the impact of news on the FOREX market, specifically in relation to the XAU/USD currency pair. The news dataset includes over 150,000 unique records spanning 42 months, from September 2018 to March 2022, for the Forex and Gold markets. Additionally, it includes data from June 2020 to March 2022, specifically for the Bitcoin market. We explore subsets of news in our *MarketNews* dataset with keywords EUR/USD, GBP/USD, USD/JPY, XAU/USD as gold, and BTC/USD for training our predictive model, as shown in Tables 2 and 3. Note that the distribution

TABLE 3. Dataset statistics.

Dataset		Time Interval	# Instance	# News
EUR/USD	Training	2018-09-24 to 2020-12-11	13,804	5,956
	Validation	2020-12-11 to 2021-07-05	3,449	2,245
	Test	2021-07-05 to 2022-03-11	4,313	2,721
GBP/USD	Training	2018-09-23 to 2020-12-11	13,804	5,036
	Validation	2020-12-11 to 2021-07-05	3,450	1,240
	Test	2021-07-05 to 2022-03-11	4,312	1,322
USD/JPY	Training	2018-09-23 to 2020-12-11	13,811	3,885
	Validation	2020-12-11 to 2021-07-05	3,451	909
	Test	2021-07-05 to 2022-03-11	4,351	979
XAU/USD	Training	2018-09-24 to 2020-12-10	13,105	2,351
	Validation	2020-12-10 to 2021-07-04	3,275	1,435
	Test	2021-07-04 to 2022-03-11	4,095	1,192
BTC/USDT	Training	2020-06-24 to 2021-07-30	9,108	7,171
	Validation	2021-07-30 to 2021-11-07	2,276	2,267
	Test	2021-11-07 to 2022-03-11	2,846	1,709

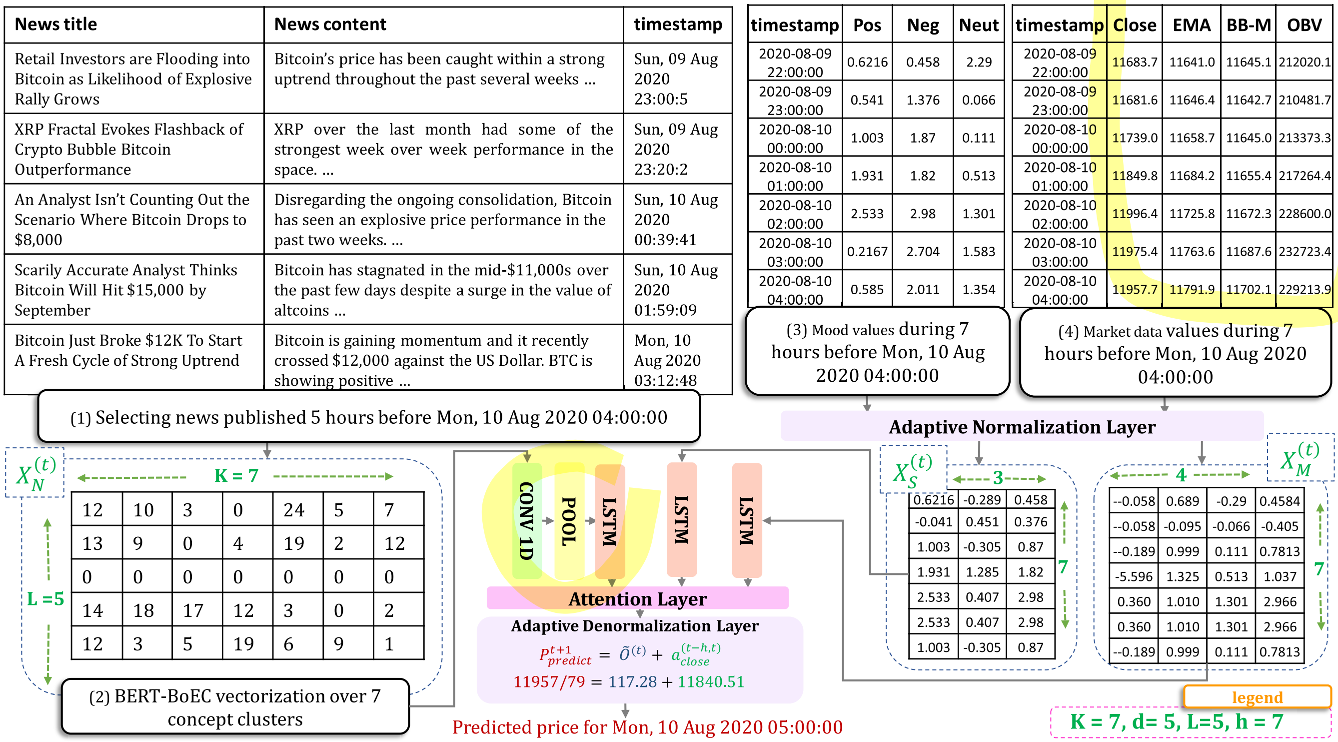


FIGURE 5. An example of jointly using news, mood, and market data for BTC/USDT price prediction on Aug 10, 2020 at 05:00:00.

of news publishing among days of the week is not uniform. For currency pairs in the Forex and gold market, a few news are released during holidays.

2) MARKET DATASET

Each record is time-stamped in the GMT zone every 60 minutes and contains Low, High, Open, and Close price as well as the trading Volume information. We use Binance:BTC/USDT for bitcoin and XAU/USD for gold exchange rate prediction.

B. EXPERIMENT SETTING

1) MODEL PARAMETERS SETTING

We evaluate our approach for price regression with two regression performance metrics, including Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). We set the kernel size in CNN to 3 and use 64 filter units with stride 1 and the same padding. We set the number of units in LSTMs to 128. For the optimizer, we used Adam with an initial learning rate of 0.001 and a decay rate of 1e-6. We set the batch size to 32. The minimum number of training epochs is 20. To eliminate timing dependencies and avoid NN from memorizing the pattern order and overfitting, we shuffle data samples before training and in each epoch to avoid NN falling in a local minimum. We use L1L2 regularization,[3] L2 bias regularization and L2 activation regularization for the CNN, LSTMs, and dense layers for model generalization.

We use Keras Model Subclassing APIs in Tensorflow 2 for the implementation.[4] Previous works [13] have studied the investors' reactions to the Forex market within an hour of the news release. We also consider the hourly time frames and work with *hourly* trading data and use 60% as training samples, 20% as validation, and 20% as test set.

2) NEWS SETTING

Table 2 shows news corpus statics. We build a corpus consisting of all news related to the target currency pair and then apply preprocessing steps on the corpus to remove numbers, URLs, tags, and stopwords. We then insert [*CLS*] token before each sentence and [*SEP*] token between adjacent sentences. Next, we apply embedding and clustering. For contextualized word embedding, we choose the pre-trained BERT implementation[5] and use the BERT-base-uncased[6] version with 768 hidden states and 12 attention. We sum the output of the top four layers of BERT and truncate each sentence in news documents to 128 tokens. For building embedded latent concepts space, we exploit k-means++ in the Sklearn package.[7] Following [20], we set the k as the number of concept clusters to 210.

[3] https://www.tensorflow.org/api_docs/python/tf/keras/regularizers/L1L2

[4] https://keras.io/guides/making_new_layers_and_models_via_subclassing/

[5] https://github.com/google-research/bert

[6] https://storage.googleapis.com/bert_models/2018_10_18/uncased_L-12_H-768_A-12.zip

[7] https://scikit-learn.org/

TABLE 4. Variations of our proposed ABM-BCSIM model.

Model Name	Training Data	News Vectorization	Sentiment Analysis	Fusion
ABM-IM	only trade data	–	–	Adaptive
ABM-SIM	Sentiment time series and trade data	–	FinBERT	Adaptive
ABM-BCIM	News and trade data	BoEC for news	–	Adaptive
ABM-BCS	News and sentiment time series	BoEC (Algorithm 1)	FinBERT	Adaptive
ABM-BCSIM	News, sentiment and Market data	BoEC (Algorithm 1)	FinBERT	Adaptive
ABM-BERTSIM	News, sentiment and trade data	BERT [cls] embedding	FinBERT	Adaptive
General-BCSIM	News, sentiment and trade data	BoEC (Algorithm 1)	FinBERT	z-score normalization and weighted fusion

3) MOOD SETTING

For sentiment analysis of news titles, we choose the FinBERT implementation[8] trained on TRC2 data set and fine-tuned on Financial PhraseBank[9] dataset for financial sentiment analysis.

4) MARKET DATA SETTING

We calculate technical indicators with TA (Technical Analysis) python package[10] and use the default values for an hourly time frame. As for the market-based features, authors in [22] suggest Close and EMA for all currency pairs shown in table 3. For EUR/USD, we also use Bolinger Bands High (BB-High) and Accumulation/Distribution Index (ADI). For USD/JPY, we use BB-Low and ADI as well. For GBP/USD, we also use BB-Mean and ADI. For BTC/USDT, we use BB-Low and OBV as well. Following [22], we investigate an information gain-based feature selection and nonlinear weight analysis of a recurrent network for XAU/USD and use Close, EMA, ADI, and BB-Low for Gold market decision supports. These indicators have multimodal close distributions to the target price for all assets we analyzed.

C. EXAMPLE

Figure 5 depicts an example of jointly using news, mood, and market data for price prediction on Mon, 10 Aug 2020 05:00:00. First, we align news based on the release time stamp. In this example, we set the market-based delay h window to 7 and the news-based delay d window to 5. We select news published between Sun, 9 Aug 2020 23:00:00 until Mon, 10 Aug 2020 04:00:00 (Block 1 at Figure 5). Then we calculate the BoEC representation for each news and group these representations and call them $X_N^{(t)}$ (Block 2 at Figure 5). Also, we use close price and technical indicators of market data from Sun, 09 Aug 2020 22:00:00 until Mon, 10 Aug 2020 04:00:00 and call them $X_M^{(t)}$ (Block 4 at Figure 5). For the mood time series, we sum up the sentiment scores of news that were published every hour on Sun, 09 Aug 2020 22:00:00 until Mon, 10 Aug 2020 04:00:00 and call them $X_S^{(t)}$. Regarding the training phase for price prediction as a regression task, we use the close price value of 05:00 as the target label. The price regression is formulated as market-based adaptive data normalization and feature extraction with a neural network for prediction of offset of price change $\tilde{o}^{(t)}$ around the average of the close price $a_{close}^{[t-h,t]}$ for this snapshot of the system. In the denormalization layer, the predicted offset is summed up with the average price of this snapshot to adjust the prediction for this period. This change prediction and shifting are adaptively made for each system snapshot based on the news, mood, and market data distributions during that period.

[8]https://huggingface.co/ProsusAI/finbert
[9]https://www.researchgate.net/publication/251231364_FinancialPhraseBank-v10
[10]https://technical-analysis-library-in-python.readthedocs.io/en/latest/

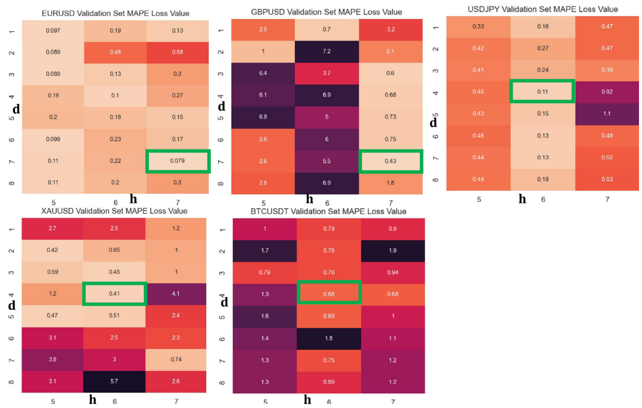


FIGURE 6. Sensitivity analysis for different values of d and h with respect to MAPE loss on the validation set for hourly market prediction. Lighter areas represent more accurate predictions. Better results surrounded with green boxes.

V. RESULTS AND ANALYSIS

We evaluate our method against the baselines and discuss the lead-lag effect of news, error analysis, and adaptability of the proposed *ABM-BCSIM* method.

A. LEAD-LAG SENSITIVITY ANALYSIS

Technical indicators are calculated based on the current and past price values and volumes of assets that change hands. At the same time, news authors publish specialized news based on their analysis of the market's future, considering the current and past economic, social, or political events. Regarding the lead-lag effects of various data sources, we use two hyper-parameters d as *news-based lag* and h as *market-based lag*. Figure 6 depicts the heat map plot of the MAPE loss in the validation set of our *ABM-BCSIM* model for all

TABLE 5. Comparison of the MSE and MAPE loss. Differences between *ABM-BCSIM* and others are statistically significant based on the Wilcoxon signed-rank test.

	EUR/USD		GBP/USD		USD/JPY		XAU/USD		BTC/USDT	
	MAPE LOSS									
Method	**Validation Loss**	**Test Loss**	**Validation Loss**	**Test Loss**	**Validation Loss**	**Test Loss**	**Validation Loss**	**Test Loss**	**Validation Loss**	**Test Loss**
ABM-IM	1.58 ± 2.09	1.14 ± 0.98	3.64 ± 0.80	0.80 ± 0.56	0.46 ± 0.15	0.50 ± 0.32^{*}	2.50 ± 1.01	7.64 ± 2.97	1.88 ± 0.70	4.36 ± 0.74
ABM-SIM	0.93 ± 1.09	0.98 ± 0.90	2.53 ± 0.80	2.62 ± 0.73	0.65 ± 0.30	0.47 ± 0.16^{*}	3.83 ± 2.81	1.94 ± 1.19^{**}	1.92 ± 0.45	13.62± 8.72
ABM-BCIM	0.67 ± 0.92	0.56 ± 0.30	0.60 ± 0.90	0.65 ± 0.60	0.46 ± 0.40	0.49 ± 0.80^{*}	1.94 ± 1.09	1.89 ± 1.02	2.29 ± 0.89	2.59 ± 0.89
ABM-BCS	1.79 ± 1.43	1.08 ± 0.94	3.08 ± 1.65	2.87 ± 1.46	0.58 ± 0.30	1.84 ± 0.89	2.76 ± 0.92	2.73 ± 1.02	10.63 ± 1.92	12.76± 1.07
FINBERT-SIMF [22]	0.16 ± 0.78	0.18 ± 0.32^{*}	**0.58 ± 0.67**	**0.44 ± 0.92**	0.44 ± 0.25	1.18 ± 1.42	1.50 ± 0.33	1.71 ± 0.54^{*}	75.84 ± 0.35	73.47± 0.30
BERT-BoEC [20]	6.03 ± 3.87	9.10 ± 3.93	6.72 ± 1.92	5.03 ± 1.98	1.56 ± 1.09	3.49 ± 1.34	5.03 ± 1.90	15.27± 4.09	48.82 ± 0.15	42.99± 0.30
BHAM [29]	6.06 ± 0.23^{*}	2.47 ± 0.23	6.29 ± 1.01^{*}	4.75 ± 1.53	49.69 ± 1.34	49.73± 1.89	45.13 ± 3.52	53.78± 3.91	2.44 ± 3.08	2.34 ± 3.04^{**}
SI-RCNN [18]	8.97 ± 3.53	9.87 ± 4.09	7.04 ± 2.03	6.45 ± 3.09	1.87 ± 0.89	3.94 ± 2.01	6.45 ± 3.05	16.21± 4.41	57.84 ± 3.43	58.14± 3.07
ABM-BCSIM	**0.12 ± 0.12**	**0.12 ± 0.10**	0.71 ± 0.50	0.65 ± 0.24	**0.46 ± 0.01**	**0.16 ± 0.06**	**1.06 ± 0.45**	**0.72 ± 0.36**	**0.74 ± 0.46**	**1.10 ± 0.22**
ABM-BERTSIM	1.52 ± 2.04	1.10 ± 0.89	0.72 ± 0.65	0.89 ± 0.80	0.56 ± 0.30	0.49 ± 0.35^{*}	1.68 ± 1.43	2.04 ± 1.23	2.55 ± 0.80	2.72 ± 0.21
General-BCSIM	5.92 ± 0.53	2.197± 0.42	5.02 ± 2.65	4.12 ± 2.15	1.34 ± 0.99	2.97 ± 1.11	4.12 ± 1.25	10.76± 2.09	39.65 ± 3.98	38.95± 3.70
	MSE LOSS(10^{-4})				**MSE LOSS**		**MSE LOSS**(10^{+4})			
ABM-IM	1.67 ± 4.15	1.21 ± 1.31	3.21 ± 2.34	1.23 ± 4.35	0.37 ± 0.22	0.63 ± 0.65	6.43 ± 7.32	2.81 ± 2.21	6.43 ± 4.26	21.2 ± 1.93
ABM-SIM	1.64 ± 0.98	1.63 ± 0.48	5.66 ± 9.84	4.45 ± 2.09	0.77 ± 0.68	0.49 ± 0.26	1.6 ± 0.60	1.11 ± 0.30	5.22 ± 0.97	26.5 ± 1.09
ABM-BCIM	6.78 ± 0.54	7.56 ± 4.87	1.20 ± 2.41	1.32 ± 1.98	3.39 ± 0.30	2.90 ± 0.67	10.32 ± 5.6	17.02± 0.51	24.31 ± 10.1	34.3 ± 19.8
ABM-BCS	4.65 ± 7.67	3.52 ± 9.45	5.43 ± 4.36	4.78 ± 2.17	0.40 ± 0.34	2.85 ± 1.07	27.02 ± 6.1	23.23± 2.01	21.09 ± 0.50	34.1 ± 1.00
FINBERT-SIMF [22]	0.11 ± 0.04	0.12 ± 0.03^{*}	**0.02 ± 0.05**	0.61 ± 0.23	1.0 ± 3.02	2.78 ± 0.90	9.92 ± 0.00	1.53 ± 0.23	15.44 ± 2.06	12.50± 6.35
BERT-BoEC [20]	6.72 ± 2.43	7.86 ± 2.35	7.09 ± 8.23	5.87 ± 3.23	3.09 ± 1.02	5.98 ± 2.83	3.04 ± 2.31	11.32± 3.95	54.96 ± 2.06	37.95± 4.84
BHAM [29]	0.35 ± 0.20^{*}	0.63 ± 0.26	0.92 ± 0.17^{*}	0.80 ± 0.49	30.05 ± 2.12	33.49± 3.45	5.49±2.37	1.18 ± 4.21	20.1 ± 4.21	16.4 ± 1.8
SI-RCNN [18]	9.80 ± 3.41	9.22 ± 2.04	7.91 ± 7.65	8.90 ± 8.94	2.57 ± 1.63	5.21 ± 2.78	4.02 ± 2.09	12.54± 4.13	59.8 ± 2.09	52.3 ± 2.03
ABM-BCSIM	**0.11 ± 0.03**	**0.11 ± 0.02**	0.40 ± 0.47	**0.50 ± 0.23**	**0.36 ± 0.03**	**0.05 ± 0.03**	**1.10 ± 0.02**	**1.01 ± 0.03**	**4.0 ± 2.6**	**5.1 ± 0.30**
ABM-BERTSIM	1.32 ± 2.35	3.21 ± 4.53	0.57 ± 0.83	0.89 ± 0.48	0.51 ± 0.39	0.51 ± 0.30^{**}	9.1 ± 0.01	9.83 ± 7.6	7.0 ± 2.12	29.8 ± 3.1
General-BCSIM	3.45 ± 4.52	4.53 ± 3.65	7.23 ± 3.25	9.25 ± 2.03	3.72 ± 1.32	5.30 ± 1.98	46.01 ± 25.7	88.9 ± 53.9	52.1 ± 3.01	54.42± 3.01

The confidence level for11 ± 0.00 unmarked cells is 99%. For cells marked with "**" and "*" the confidence level is 95% and 90% respectively.

currency pair predictions with respect to the different values of d and h.

The optimal results for EUR/USD and GBP/USD are observed when setting $d = 7$ and $h = 7$ hours ago, highlighting the importance of utilizing both news and market data for these markets. In addition, the best outcomes for BTC/USDT, XAU/USD, and USD/JPY are achieved by setting $d = 4$ and $h = 6$ hours ago. The two-hour delay between news and market data lag size for these currency pairs suggests that investors typically respond promptly to fresh news while considering a longer lag for tracking indicators changes. This indicates the leading impact of news on fluctuations in these markets.

B. COMPARISON WITH THE BASELINES

We consider six variations of our proposed model, as shown in Table 4 for ablation studies on the three folds of (1) the effect of data sources, (2) mainstream text representation methods, and (3) adaptive decision-making in our ABM-BCSIM approach. Moreover, we compare the effectiveness of our method in price prediction against time series-based financial prediction methods that directly use market price, mood time series, and news embedding in the training process. We repeat all experiments for 10 rounds with random seeds and report the average results in table 5.

1) ABLATION STUDY

We perform an ablation study on four variations of our ABM-BCSIM as shown in table 4. For ABM-IM, ABM-SIM, ABM-BCIM, and ABM-BCS, we switch off the input values of one or two modalities of data. For example, in ABM-IM, we only use market data for model training. For each variation, we find the best setting for h and/or d. The superiority of ABM-BCSIM against these four variations, as shown in table 5, implies the importance of simultaneous usage of news content, sentiment, and market data.

The better results of ABM-BCSIM against ABM-IM indicate the inefficiency of using only market-based data for financial decision support. The results of ABM-BCS against ABM-BCSIM show the weakness of market prediction based on only news-based information. The superiority of ABM-BCSIM against ABM-SIM (using mood and market time series) and ABM-BCIM (using concept-based news representation and market data) implies the importance of feature fusion based on news topics distribution in a sequence of news as well as perceiving mood in specialized newsgroups.

We assess the ability of the BoEC algorithm 1 for the financial document representation by comparing the results of ABM-BERTSIM against ABM-BCSIM as shown in table 5. In ABM-BERTSIM, we consider all parts of the experiment the same as our model and only replace BoEC text representation with BERT [*CLS*] representation for a sequence of news titles and body words with a length of 500 and 768 embedding dimension. The poor results of ABM-BERTSIM against our approach indicate the ability of the BoEC algorithm for financial text reorientation.

In our ABM-BCSIM we extract features from an adaptive normalized form of the market and mood time series data with two separate normalization layers (Block 7 in Figure 2) and a sequence of news topic distribution. Then, after feature extraction from the multimode of data, in the fusion layer, the network learns the deviation around the average price at that window (Block 9 in Figure 2) based on the corresponding extracted knowledge and the estimated time series distribution. The normalization and denormalization layers are the key parts of our adaptive behavior model. The normalization layer estimates the distribution of mood or market data at each window, and the denormalization layer learns the deviation from the average price based on the received information at that window of data.

In General-BCSIM, we replace the adaptive normalization layer with the general z-score normalization and remove the denormalization layer from our model architecture shown in figure 2. In general-BCSIM, we train the network while normalizing all training samples based on the mean and variance of the whole training samples, and prediction is made based on normalized data via the test set mean and variances. While both ABM-BCSIM and General-BCSIM use the same data input and we set their best configuration setting, results in table 5 show significantly lower errors in ABM-BCSIM against General-ABM. This indicates the strong ability of our adaptive decision-making model for price regression. The non-stationary behavior of price time series and neglecting the changes of data distribution at each window of data in the General-BCSIM training process causes the predicted values to drift below the real price value, leading to a large loss value. In section V-D, we discuss the ability of ABM-BCSIM to overcome the non-stationary behavior of price.

2) DISCUSSION ON BASELINES

We compare ABM-BCSIM against the following methods:

- FinBERT-SIMF: Scholars in [22] use FinBERT pre-train model for sentiment analysis of news titles. Their proposed method extracts features from the sentiment of news titles and technical indicators via two layers of LSTMs and passes the extracted features to a dense layer for price regression. The authors use z-score normalization and set the lag window size to 7 hours ago and predicted the price for the following hours. To implement this work, we use technical indicators the same as ABM-BCSIM. Other settings are done the same as the settings reported in [22].
- BERT-BoEC: In the proposed approach by Farimani et al. [20], feature extraction is done from news title sentiment, BERT-BoEC-based news embedding together that pass to a recurrent CNN, and market data that normalized via z-score normalization and pass to an LSTM layer. In their proposed approach, the concatenation of extracted features from two LSTM layers passes to a dense layer with a linear activation function for price regression. To implement this work, we use technical indicators the same as ABM-BCSIM. The window size for feature extraction is set to 7 hours ago. Other settings are done the same as reported in [20].
- SI-RCNN: As presented in [18], after performing an average Word2vec embedding of all the words in the news title, a hybrid model is composed of an RCNN for the financial news and an LSTM for technical indicators. We adopted this model for hourly market price prediction and used z-score normalization. Other parameters are set according to [18]. We extract word embedding through the Google News pre-train model with an embedding dimension of 300 and set the delay window to 7 hours ago same as our method.
- BHAM: The BERT-based Hierarchical Aggregation Model [29] first performs [CLS] token through BERT pre-trained model for news title, then groups news based on timestamp and applies summarization. They used a SOTA-based summarization method. We used the top 5 news for group aggregation. This model uses a multi-layer perceptron for feature extraction from trading data and then predicts the market trend based on jointly using text information and trading data. We adopted this model for hourly market prediction and for labeling, we perceive hourly changes in close price. Following [29], we use min-max normalization for trade data preprocessing. We use three dense layers for feature extraction from trade data. Market information is also the same as our model. For the MLP layer and the attention mechanism, we implemented the model architecture using Keras in Tensorflow 2 and used the same loss function and optimizer as our model.

The results in Table 5 show the superiority of our ABM-BCSIM technique compared to the baselines of up to 54.95% for EUR/USD, 63.61% for USD/JPY, 66.69% for XAU/USD, and 52.90% for BTC/USDT in MAPE error reduction. To perform the comparison between our regression method and other methods, we utilized the non-parametric Wilcoxon signed-rank test from the Scipy Python package. This statistical test tests the null hypothesis that two related paired samples originate from the same distribution. By applying this test, the obtained p-values indicate we can reject the null hypothesis, which indicates the significance of the difference between the accuracy of our method compared to the other methods for all the cases outlined in table 5. The obtained p-values reject the null hypothesis and indicate the difference (improvement) is statistically significant. This considerable error reduction is essentially related to the adaptive decision-making regarding the information received

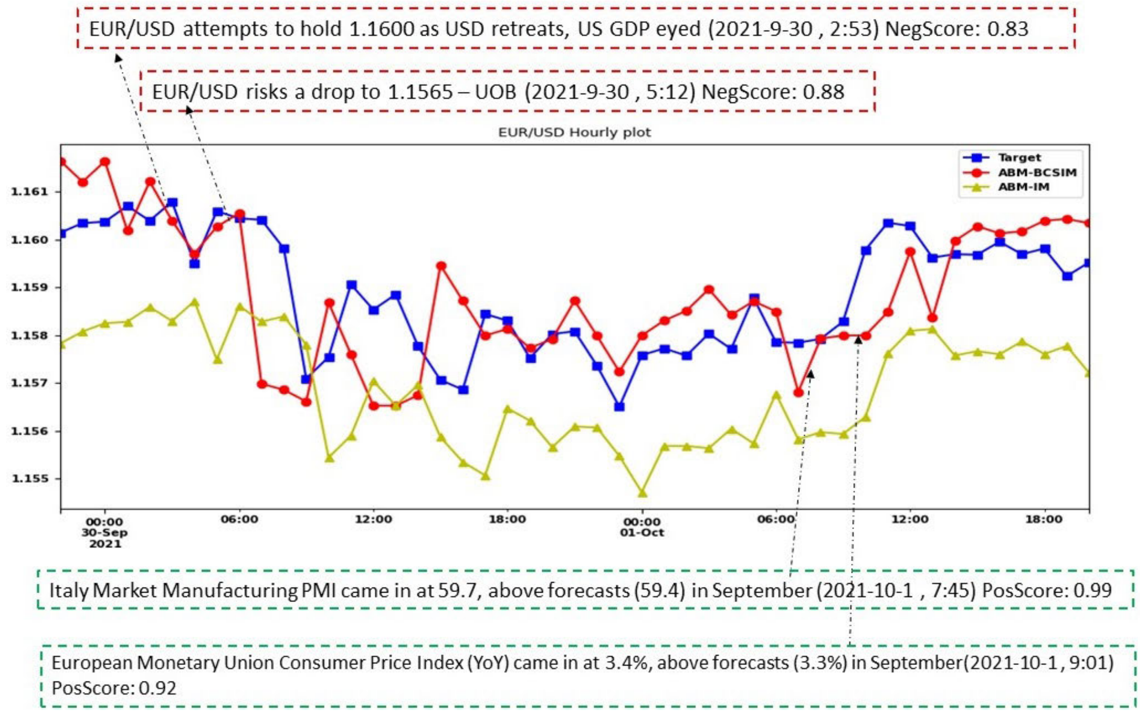


FIGURE 7. EUR/USD prediction plot from 2021-9-30 to 2021-10-1 in the test set.

by investors from a multimode of data in a snapshot of the system and forcing the network to learn the amount of price change around the average price of that snapshot regarding the input information. It is important to note that the test set data may have a different average and standard deviation compared to the training set, however, to avoid prediction drift below the real close price values, we forced the model to learn the change price around the average price of input data instance.

Both ABM-BCSIM and BERT-BoEC [20] use BoEC algorithm for text representation and use the same network structure for feature extraction from market data and news vectors. As presented in SI-RCNN [18], the news vectorization is based on Word2vec average embedding, and news-based feature extraction is done via RCNN layer, the same as our model. The main difference between our work and these two baselines is the adaptive normalization and denormalization steps, which are presented in ABM-BCSIM. Therefore, the main reason behind the significant error reduction of our ABM-BCSIM, especially for turbulent markets of BTC/USDT, USD/JPY, and XAU/USD, is due to decision-making for a snapshot of the system as τ regarding both the information events that investors experienced during τ and change in the data distribution during τ in the model training phase. The BERT-BoEC model [20] leverages news-group information; however, it does not consider the change in market data distribution and its prediction significantly deviates from the actual close price. The FinBERT-SIMF model [22] predicts EUR/USD and GBP/USD markets with little loss values; however, it does not perform well for turbulent BTC/USD, XAU/USD, and USD/JPY currency pairs due to the high frequency of price changes in these markets.

We formulate our work as price regression problem. Given the mean value of 1.02 for EUR/USD and the small (less than one) values of typical error for EUR/USD prediction, the MSE loss will tend to be very low, on the order of 10^{-4}. This is because MSE squares the errors, and smaller errors have a much smaller impact on the loss when squared. Conversely, for BTC/USDT with a much higher mean value (around 24,000), even small errors are magnified significantly due to squaring, leading to larger MSE loss values (around 10^{+4}).

C. ERROR ANALYSIS AND LIMITATION

We analyze the performance of our adaptive behavior model on major currency pairs in the Forex, Commodities, and Cryptocurrencies markets, as shown in table 3, while maintaining the same basic set of information sources. Figure 7 depicts the prediction plot from 2021-30-9 to 2021-10-01, where the blue line shows the target close price (the close price of the next hour with respect to the timestamp on the x-axis), the red line with dot markers presents ABM-BCSIM, and the yellow line with triangle markers is the prediction of ABM-IM that does not use the news information. Positive and negative news, based on FinBERT judgments, published in this period are shown in green and red boxes, respectively. This indicates that ABM-BCSIM predicted the market fall in 09-30, 03:00 and the market rise in 10-01, 10:00 sooner than ABM-IM, since ABM-BCSIM is

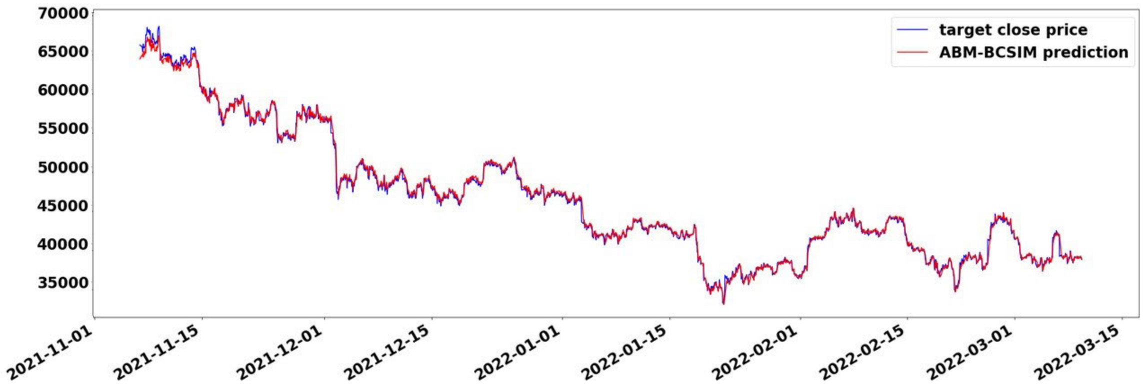


FIGURE 8. BTC/USDT prediction plot during 9 months of the test set in hourly timeframe.

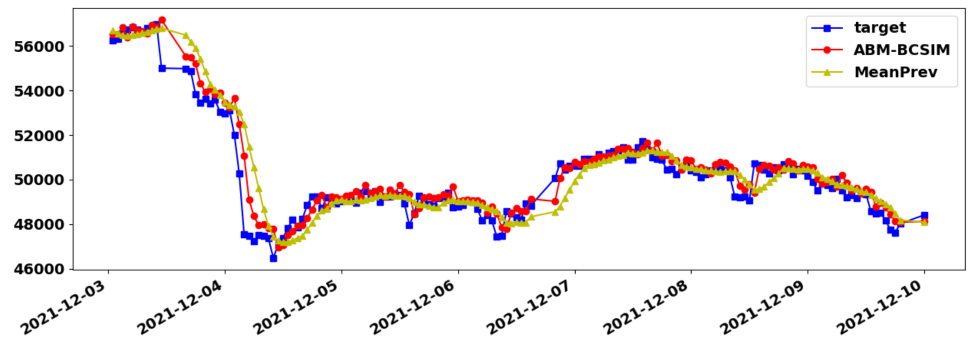


FIGURE 9. BTC/USDT prediction plot from 2021-12-03 to 2021-12-10 in the test set. The yellow line, known as MeanPrev, represents the average closing price ($a_{close}^{[t-h,t]}$) observed by the model. It is used as a bias to predict the price at time t+1.

TABLE 6. KPSS unit root test results for drift & trend of close price time series as well as skewness and kurtosis of price return. All of Close price time series are non-stationary at 1% significant level with p-value of 0.01 < 0.05.

EUR/USD	Intercept	KPSS	8.45
		C(1%)	0.73
	Intercept & Trend	KPSS	2.41
		C(1%)	0.21
	$Skew(r^{(t)})$		-0.13
	$Kurtosis(r^{(t)}))$		12.20

USD/JPY	Intercept	KPSS	4.4
		C(1%)	0.13
	Intercept & Trend	KPSS	4.29
		C(1%)	0.21
	$Skew(r^{(t)})$		-2.91
	$Kurtosis(r^{(t)}))$		204.2

GBP/USD	Intercept	KPSS	12.75
		C(1%)	0.73
	Intercept & Trend	KPSS	2.17
		C(1%)	0.21
	$Skew(r^{(t)})$		0.29
	$Kurtosis(r^{(t)}))$		20.43

XAU/USD	Intercept	KPSS	19.18
		C(1%)	0.73
	Intercept & Trend	KPSS	4.01
		C(1%)	0.21
	$Skew(r^{(t)})$		-0.96
	$Kurtosis(r^{(t)}))$		30.57

BTC/USDT	Intercept	KPSS	12.06
		C(1%)	0.73
	Intercept & Trend	KPSS	4.01
		C(1%)	0.21
	$Skew(r^{(t)})$		0.02
	$Kurtosis(r^{(t)}))$		14.34

informed about the negative/positive news dissemination via BoEC text embedding and sentiment scores together, while ABM-IM, which only uses market data, learns to replicate market movements after they have already occurred.

Specifically, considering the up/down spike of the target in Fig. 7, ABM-IM achieves a high accuracy of 84% in predicting the target of the previous step. However, its accuracy drops to 35% when attempting to predict the target at time t. On the other hand, ABM-BCSIM demonstrates improvements in prediction accuracy compared to ABM-IM. It achieves 65% accuracy for predicting $target^t$. This represents a 29% improvement in prediction accuracy for $target^t$ with the incorporation of news-based data modalities. ABM-BCSIM predicted $target^{t-1}$ with 53% higher accuracy than random chance. However, to further enhance its performance, it needs to incorporate a mood time series as a Markovian process. This addition enables the consideration of current and past values of news sentiments, taking into account time attenuation factors to reduce the dependency of ABM-BCSIM on highly correlated market data.

To analyze the performance of ABM-BCSIM in predicting the bitcoin market, we plot the prediction for 4 months of the test set (hourly timeframe) in figure 8. This figure clearly shows the strong adaptability of our behavior model in the turbulent bitcoin market thanks to the adaptive decision-making in the fusion layer. To better analyze our model performance, we plot a snapshot of BTC/USDT data from 2021-13-03 to 2021-12-10 in figure 9. The yellow line as denoted "MeanPrev" in this figure is the average close price, i.e., $a_{close}^{[t-h,t]}$ of $h = 6$ hours ago. This figure shows the ability of our approach for trend behavior modeling during the \$10,000 fall of BTC/USDT price at only 24 hours. In many cases, the predicted values differ from the average price $a_{close}^{[t-h,t]}$ that was passed to the network at the denormalization layer (Block 9 in figure 2), which indicates the importance of forcing the model to learn the deviation around average price based on information events.

D. ADAPTABILITY AND GENERALIZABILITY

Table 6 shows the kurtosis and skewness of return distribution $r^{(t)} = \frac{p^{(t)}}{p^{(t-1)}} - 1$ based on total training, validation, and test sets for hourly data. The small skewness values and large kurtosis values for the market return rate indicate that we applied our approach to some turbulent market with high probabilities of extreme price movement (kurtosis 204 for USDJPY). Also, these statics indicates the variance of the price is not constant over time and depends on the amount of information that investors have experienced during that time [15]. We also test the non-stationary behavior of the price time series with the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) unit root test [65] in table 6. KPSS test is done to indicate both random walk with drift and random walk with drift and deterministic trend state of the non-stationary process, and the results indicate that all of the price time series are non-stationary. We reported the Kpss value and the value of the 99% quantile under H0 as C(1%) at table 6.

In such turbulent markets where the price time series is non-stationary and return is leptokurtic with kurtosis greater than a normal distribution, the non-stationary behavior of price and turbulent price change is a challenging task for applying deep learning models. Results of our experiments reveal that other predictive models perform worse than ours, especially for currency pairs with large values of kurtosis for price return.

We analyzed at least three years of hourly market data for various assets across major Forex currency pairs, Gold, and Bitcoin markets. This data encompassed periods of both low and high volatility regimes, with assets experiencing both high and low volume orders. To ensure the generalizability of our approach, we employed volume-based indicators as market features. We use ADI and OBV as volume based indicators, which are selected based on weight analysis of input gate of a recurrent neural network.

E. COMPUTATIONAL OVERHEAD

The most time-consuming part of our method is Algorithm III-A, which computes news vectors in a latent space. Considering K as the latent space dimension, for n news documents each having at least l tokens, the computational complexity of BoEC is $O(K.n.l)$. We compare our model with BHAM [29] as it also incorporates news content and technical indicators for decision making. For our ABM-BCSIM, the news vectorization, which can be done offline, took 406.79 seconds, and the training process lasted 316.93 seconds. As BHAM incorporates BERT-based news vectorization utilizing a SOTA summarizer module and market-based feature extraction through a multilayer perceptron, which in total took 119.79 seconds. To analyze computational overhead, we compared the execution time of our model training approach with BHAM for BTC/USDT prediction (BHAM predicts BTC/USDT better than other baselines). Running this experiment on the same machine, we observed that training phase for ABM-BCSIM takes 62% longer than BHAM. However, for BTC/USD, our approach exhibited significant improvement, achieving up to 52.90% more accurate price prediction.

For more analysis, we developed a trading strategy for BTC/USDT to compare the efficacy of ABM-BCSIM and BHAM. Our simple strategy starts with \$100,000 assets and considers 0.1% for trading commission. When we observed an upward trend, buy, and when we observed a downward trend sell and close the position. We did this experiment for both ABM-BCSIM and BHAM. The results indicated that the total profit of ABM-BCSIM was \$233,075.32, and BHAM's total profit was \$147,444.04. The computation complexity of ABM-BCSIM is much larger than BHAM; however, it achieved more profit.

VI. CONCLUSION AND FUTURE WORK

In this work, we presented an adaptive data-driven model for price regression in financial markets, considering the temporal dependencies between market movements and different topics in news stories. We focused on predicting the exchange rate for five major currency pairs in the Forex, Gold, and Cryptocurrency markets. To overcome the challenge of the non-stationary behavior of price time series, we model information events that investors experience during a period that is proportional to the amount of price change. Our main contributions are (1) using relevant news information to the target market, (2) leveraging important economic concepts in the news, tracking mood in specialized newsgroups, and informative technical indicators as three modes of data, and (3) considering the change in the market data distribution and proposing adaptive fusion strategy for system snapshots. In the fusion layer, we let the network learn how to balance the multimode of extracted features to handle data multimodalities and the amount of fluctuation of the predicted values around an average price.

Results of our experiments show a significant error reduction of at least 40.11% for XAU/USD, USD/JPY, and BTC/USDT. We perform an ablation study to investigate the effect of various data sources as well as the effectiveness of our adaptive fusion strategy. Also, we analyze the error of prediction and adaptability of our approach to the financial market. Although our proposed approach was presented and evaluated in the financial domain, it can be applied in other time series prediction use cases, such as Business Development, where behavior modeling based on unstructured data becomes handy. Besides, we can use our scraped news about various assets, such as oil, gas, minor currency pairs in the Forex market, and altcoins, to perform other market prediction studies. As future work, we plan to improve mood time series estimation as a Markovian process, and use a deep-Q network for training a news-based trading agent based on an adaptive behavior model.

REFERENCES

[1] E. F. Fama, "The behavior of stock-market prices," *J. Bus.*, vol. 38, no. 1, pp. 34–105, Jan. 1965. [Online]. Available: https://www.jstor.org/stable/2350752

[2] R. P. Schumaker, Y. Zhang, C.-N. Huang, and H. Chen, "Evaluating sentiment in financial news articles," *Decis. Support Syst.*, vol. 53, no. 3, pp. 458–464, Jun. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167923612000875

[3] B. Wuthrich, V. Cho, S. Leung, D. Permunetilleke, K. Sankaran, and J. Zhang, "Daily stock market forecast from textual web data," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, vol. 3, Oct. 1998, pp. 2720–2725.

[4] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, Mar. 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S187775031100007X

[5] X. Zhang, H. Fuehres, and P. A. Gloor, "Predicting stock market indicators through Twitter 'I hope it is not as bad as I fear,'" in *Proc. Social Behav. Sci.*, vol. 26, 2011, pp. 55–62. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877042811023895

[6] P. C. Tetlock, "Does public financial news resolve asymmetric information?" *Rev. Financial Stud.*, vol. 23, no. 9, pp. 3520–3557, Sep. 2010.

[7] Q. Li, J. Tan, J. Wang, and H. Chen, "A multimodal event-driven LSTM model for stock prediction using online news," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 10, pp. 3323–3337, Oct. 2021.

[8] S. A. Farimani, M. V. Jahan, and A. M. Fard, "From text representation to financial market prediction: A literature review," *Information*, vol. 13, no. 10, p. 466, Sep. 2022. [Online]. Available: https://www.mdpi.com/2078-2489/13/10/466

[9] Y. Shi, K.-Y. Ho, and W.-M. Liu, "Public information arrival and stock return volatility: Evidence from news sentiment and Markov regime-switching approach," *Int. Rev. Econ. Finance*, vol. 42, pp. 291–312, Mar. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1059056015002294

[10] A. C. MacKinlay, "Event studies in economics and finance," *J. Econ. Literature*, vol. 35, no. 1, pp. 13–39, 1997. [Online]. Available: https://www.jstor.org/stable/2729691

[11] M. Hagenau, M. Liebmann, and D. Neumann, "Automated news reading: Stock price prediction based on financial news using context-capturing features," *Decis. Support Syst.*, vol. 55, no. 3, pp. 685–697, Jun. 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167923613000651

[12] W. Long, L. Song, and Y. Tian, "A new graphic kernel method of stock price trend prediction based on financial news semantic and structural similarity," *Expert Syst. Appl.*, vol. 118, pp. 411–424, Mar. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095741741830650X

[13] A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, "Text mining of news-headlines for FOREX market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment," *Expert Syst. Appl.*, vol. 42, no. 1, pp. 306–324, Jan. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417414004801

[14] H. D. Huynh, L. M. Dang, and D. Duong, "A new model for stock price movements prediction using deep neural network," in *Proc. 8th Int. Symp. Inf. Commun. Technol.*, Dec. 2017, pp. 57–62.

[15] P. K. Clark, "A subordinated stochastic process model with finite variance for speculative prices," *Econometrica*, vol. 41, no. 1, pp. 135–155, Jan. 1973.

[16] D. Cheng, F. Yang, S. Xiang, and J. Liu, "Financial time series forecasting with multi-modality graph neural network," *Pattern Recognit.*, vol. 121, Jan. 2022, Art. no. 108218. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S003132032100399X

[17] H. K. Kim, H. Kim, and S. Cho, "Bag-of-concepts: Comprehending document representation through clustering words in distributed representation," *Neurocomputing*, vol. 266, pp. 336–352, Nov. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231217308962

[18] M. R. Vargas, C. E. M. dos Anjos, G. L. G. Bichara, and A. G. Evsukoff, "Deep leaming for stock market prediction using technical indicators and financial news articles," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

[19] Y. Xu and S. B. Cohen, "Stock movement prediction from tweets and historical prices," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*. Melbourne, VIC, Australia: Association for Computational Linguistics, 2018, pp. 1970–1979. [Online]. Available: https://www.aclweb.org/anthology/P18-1183

[20] 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 *Proc. IEEE 8th Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2021, pp. 1–10.

[21] Y.-F. Chen and S.-H. Huang, "Sentiment-influenced trading system based on multimodal deep reinforcement learning," *Appl. Soft Comput.*, vol. 112, Nov. 2021, Art. no. 107788. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494621007092

[22] 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," *Knowl.-Based Syst.*, vol. 247, Jul. 2022, Art. no. 108742. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095070512200346X

[23] Q. Li, Y. Chen, L. L. Jiang, P. Li, and H. Chen, "A tensor-based information framework for predicting the stock market," *ACM Trans. Inf. Syst.*, vol. 34, no. 2, pp. 1–30, Feb. 2016, doi: 10.1145/2838731.

[24] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang, and P. S. Yu, "Improving stock market prediction via heterogeneous information fusion," *Knowl.-Based Syst.*, vol. 143, pp. 236–247, Mar. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705117306032

[25] S. Rajaee and M. T. Pilehvar, "An isotropy analysis in the multilingual BERT embedding space," 2021, *arXiv:2110.04504*.

[26] S. Rajaee and M. T. Pilehvar, "A cluster-based approach for improving isotropy in contextual embedding space," in *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics*, 2021, pp. 575–584. [Online]. Available: https://aclanthology.org/2021.acl-short.73

[27] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Computational Linguistics*, vol. 1, Jun. 2019, pp. 4171–4186. [Online]. Available: https://www.aclweb.org/anthology/N19-1423

[28] J. Z. G. Hiew, X. Huang, H. Mou, D. Li, Q. Wu, and Y. Xu, "BERT-based financial sentiment index and LSTM-based stock return predictability," 2019, *arXiv:1906.09024*.

[29] 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," in *Proc. 2nd Workshop Econ. Natural Lang. Process.* Hong Kong: Association for Computational Linguistics, 2019, pp. 41–50. [Online]. Available: https://www.aclweb.org/anthology/D19-5106

[30] P. Hajek and A. Barushka, "Integrating sentiment analysis and topic detection in financial news for stock movement prediction," in *Proc. 2nd Int. Conf. Bus. Inf. Manage.*, Sep. 2018, pp. 158–162.

[31] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proc. 31st Int. Conf. Int. Conf. Mach. Learn.*, vol. 32, 2014, pp. 1188–1196.

[32] B. Liang, H. Su, L. Gui, E. Cambria, and R. Xu, "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks," *Knowl.-Based Syst.*, vol. 235, Jan. 2022, Art. no. 107643. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705121009059

[33] T. Daudert, "Exploiting textual and relationship information for fine-grained financial sentiment analysis," *Knowl.-Based Syst.*, vol. 230, Oct. 2021, Art. no. 107389. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705121006511
[34] Q. Li, T. Wang, P. Li, L. Liu, Q. Gong, and Y. Chen, "The effect of news and public mood on stock movements," *Inf. Sci.*, vol. 278, pp. 826–840, Sep. 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025514003879
[35] D. Araci, "FinBERT: Financial sentiment analysis with pre-trained language models," 2019, *arXiv:1908.10063*.
[36] M. Atzeni, A. Dridi, and D. Reforgiato Recupero, "Using frame-based resources for sentiment analysis within the financial domain," *Prog. Artif. Intell.*, vol. 7, no. 4, pp. 273–294, Dec. 2018.
[37] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," 2019, *arXiv:1910.01108*.
[38] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Econometrics*, vol. 31, no. 3, pp. 307–327, Apr. 1986. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0304407686900631
[39] N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosifidis, "Deep adaptive input normalization for time series forecasting," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3760–3765, Sep. 2020.
[40] M. Daniel, R. F. Neves, and N. Horta, "Company event popularity for financial markets using Twitter and sentiment analysis," *Expert Syst. Appl.*, vol. 71, pp. 111–124, Apr. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417416306571
[41] Y.-C. Wei, Y.-C. Lu, J.-N. Chen, and Y.-J. Hsu, "Informativeness of the market news sentiment in the Taiwan stock market," *North Amer. J. Econ. Finance*, vol. 39, pp. 158–181, Jan. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S106294081630122X
[42] C. Wong and I.-Y. Ko, "Predictive power of public emotions as extracted from daily news articles on the movements of stock market indices," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Oct. 2016, pp. 705–708.
[43] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discourse Process.*, vol. 25, nos. 2–3, pp. 259–284, Jan. 1998.
[44] S. Feuerriegel and J. Gordon, "News-based forecasts of macroeconomic indicators: A semantic path model for interpretable predictions," *Eur. J. Oper. Res.*, vol. 272, no. 1, pp. 162–175, Jan. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0377221718304879
[45] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
[46] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*.
[47] B. Lutz, N. Pröllochs, and D. Neumann, "Predicting sentence-level polarity labels of financial news using abnormal stock returns," *Expert Syst. Appl.*, vol. 148, Jun. 2020, Art. no. 113223. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095741742030049X
[48] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, *arXiv:1907.11692*.
[49] J. Zhang, Y. Zhao, M. Saleh, and P. J. Liu, "PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization," 2019, *arXiv:1912.08777*.
[50] F. Z. Xing, E. Cambria, and Y. Zhang, "Sentiment-aware volatility forecasting," *Knowl.-Based Syst.*, vol. 176, pp. 68–76, Jul. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705119301546
[51] C. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 8, May 2014, pp. 216–225.
[52] T. Loughran and B. Mcdonald, "Measuring readability in financial disclosures," *J. Finance*, vol. 69, no. 4, pp. 1643–1671, Aug. 2014.
[53] T. Ito, H. Sakaji, K. Izumi, K. Tsubouchi, and T. Yamashita, "GINN: Gradient interpretable neural networks for visualizing financial texts," *Int. J. Data Sci. Anal.*, vol. 9, no. 4, pp. 431–445, May 2020, doi: 10.1007/s41060-018-0160-8.
[54] I. Medovikov, "When does the stock market listen to economic news? New evidence from copulas and news wires," *J. Banking Finance*, vol. 65, pp. 27–40, Apr. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378426616000145
[55] D. Othan, Z. H. Kilimci, and M. Uysal, "Financial sentiment analysis for predicting direction of stocks using bidirectional encoder representations from transformers (BERT) and deep learning models," in *Proc. Int. Conf. Innov. Intell. Technol.*, 2019, pp. 30–35.
[56] W. Cheng and S. Chen, "Sentiment analysis of financial texts based on attention mechanism of FinBERT and BiLSTM," in *Proc. Int. Conf. Comput. Eng. Appl. (ICCEA)*, Jun. 2021, pp. 73–78.
[57] M. G. Sousa, K. Sakiyama, L. d. S. Rodrigues, P. H. Moraes, E. R. Fernandes, and E. T. Matsubara, "BERT for stock market sentiment analysis," in *Proc. IEEE 31st Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2019, pp. 1597–1601.
[58] D. T. Tran, A. Iosifidis, J. Kanniainen, and M. Gabbouj, "Temporal attention-augmented bilinear network for financial time-series data analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 5, pp. 1407–1418, May 2019.
[59] Z.-R. Lai, D.-Q. Dai, C.-X. Ren, and K.-K. Huang, "Radial basis functions with adaptive input and composite trend representation for portfolio selection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6214–6226, Dec. 2018.
[60] M. M. Kumbure, C. Lohrmann, P. Luukka, and J. Porras, "Machine learning techniques and data for stock market forecasting: A literature review," *Expert Syst. Appl.*, vol. 197, Jul. 2022, Art. no. 116659. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417422001452
[61] C. Huang, D. Yang, and Y. Chuang, "Application of wrapper approach and composite classifier to the stock trend prediction," *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2870–2878, May 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417407001819
[62] Y. Peng, P. H. M. Albuquerque, H. Kimura, and C. A. P. B. Saavedra, "Feature selection and deep neural networks for stock price direction forecasting using technical analysis indicators," *Mach. Learn. Appl.*, vol. 5, Sep. 2021, Art. no. 100060. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S266682702100030X
[63] P. C. Tetlock, M. Saar-Tsechansky, and S. Macskassy, "More than words: Quantifying language to measure firms' fundamentals," *J. Finance*, vol. 63, no. 3, pp. 1437–1467, Jun. 2008.
[64] H. Wang, S. Lu, and J. Zhao, "Aggregating multiple types of complex data in stock market prediction: A model-independent framework," *Knowl.-Based Syst.*, vol. 164, pp. 193–204, Jan. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705118305276
[65] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *J. Econometrics*, vol. 54, no. 1, pp. 159–178, 1992. [Online]. Available: https://www.sciencedirect.com/science/article/pii/030440769290104Y

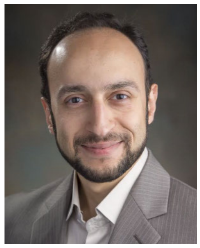
SAEEDE ANBAEE FARIMANI received the B.Sc. degree in computer software engineering from the Ferdowsi University of Mashhad, Iran, and the M.Sc. and Ph.D. degrees in computer software engineering from Islamic Azad University, Mashhad Branch, Iran.

She is currently a Faculty Member of computer engineering with Islamic Azad University. Her research and industry interests include NLP and business data analysis.

Dr. Farimani was a recipient of an applied research grant from Iran's Ministry of Science, Research and Technology. She has served as a Reviewer for journals, such as IEEE Transactions on Neural Networks and Learning Systems, *Expert Systems with Application*, and *Electronic Commerce Research and Applications*.

MAJID VAFAEI JAHAN received the B.Sc. degree from the Ferdowsi University of Mashhad, the M.Sc. degree from the Sharif University of Technology, and the Ph.D. degree from Islamic Azad University, Science and Research Branch, Iran.

He is currently an Associate Professor in computer engineering with Islamic Azad University, Mashhad Branch, Iran. He is also a Distinguished Academic and a Researcher in the field of computer software engineering. Throughout his career, he has been recognized for his outstanding contributions to engineering and computer science. His research and industry interests include stochastic modeling, big data analysis, and data science. His work in these areas has made significant contributions to both academic knowledge and practical applications, earning him a reputation as a leading expert in his field. He has received numerous awards, including the Top Researcher Award in the engineering field, the Best Published Book Writer Award, the Best Published Paper Award, and the prestigious Top Researcher Award in Iran.



AMIN MILANI FARD received the B.Sc. degree in computer software engineering from the Ferdowsi University of Mashhad, Iran, the M.Sc. degree in computer science from Simon Fraser University, Canada, and the Ph.D. degree in computer software engineering from The University of British Columbia, Canada.

He is currently an Associate Professor in computer science with New York Institute of Technology, Vancouver Campus. His research and industry interests include software engineering, data analysis, and cybersecurity. He was a recipient of a number of awards and grants, such as the Most Influential Paper Award (IEEE SCAM'23), the Best Paper Award (ECIR'19), the NYIT Research Grants, the National Research Council Canada (NRC) Industrial Research Assistance Program, the Natural Sciences and Engineering Research Council of Canada (NSERC) Alexander Graham Bell Canada Graduate Scholarship, the ACM SIGSOFT CAPS Grant Awards, the Research Medal from the National Research Institute for Science Policy, Iran Ministry of Science, and the First Rank Khwarizmi Award. He has served as a PC Member for major conferences, such as IEEE/ACM ASE, ICPC, and DSAA, as well as a Reviewer for journals, such as IEEE Transactions on Software Engineering, *ACM Transactions on Knowledge Discovery from Data*, IEEE Transactions on Neural Networks and Learning Systems, *Empirical Software Engineering* (Springer), *Nature* (Springer), *Automated Software Engineering*, *Financial Innovation*, *Journal of Systems and Software*, *Information Systems*, and *Machine Learning with Applications*.