IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 18 January 2023, accepted 8 March 2023, date of publication 17 March 2023, date of current version 27 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3258695

RESEARCH ARTICLE

Hybrid Information Mixing Module for Stock Movement Prediction

JOOWEON CHOI[1], (Student Member, IEEE), SHIYONG YOO[2], XIAO ZHOU[3], AND YOUNGBIN KIM[4], (Member, IEEE)
[1]Department of Artificial Intelligence, Chung-Ang University, Dongjak, Seoul 06974, South Korea
[2]CAU Business School, Chung-Ang University, Dongjak, Seoul 06974, South Korea
[3]School of Journalism, Fudan University, Yangpu, Shanghai 200433, China
[4]Department of Imaging Science and Arts, Chung-Ang University, Dongjak, Seoul 06974, South Korea

Corresponding author: Youngbin Kim (ybkim85@cau.ac.kr)

This research was supported by the Chung-Ang University Research Grants in 2022 and Institute for Information & communications Technology Planning Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2021-0-01341, Artificial Intelligence Graduate School Program (Chung-Ang University)).

ABSTRACT With the continuing active research on deep learning, research on stock price prediction using deep learning has been actively conducted in the financial industry. This paper proposes a method for predicting stock price movement using stock and news data. The stock market is affected by many variables; thus, market volatility should be considered for predicting stock price movement. Because stock markets are efficient, all kinds of information are quickly reflected in stock prices. We create a new fusion mix by combining price and text data features and propose a hybrid information mixing module designed using two map blocks for effective interaction between the two features. We extract the multimodal interaction between the time-series features of the price data and the semantic features of the text data. In this paper, a multilayer perceptron-based model, the hybrid information mixing module, is applied to the stock price movement prediction to conduct a price fluctuation prediction experiment in a stock market with high volatility. In addition, the accuracy, Matthews correlation coefficient (MCC) and F1 score for the stock price movement prediction were used to verify the performance of the hybrid information mixing module.

INDEX TERMS Stock movement prediction, time-series forecasting, bidirectional encoder representations from transformer (BERT), gated recurrent units (GRU), multilayer perceptron (MLP).

I. INTRODUCTION

With the continuing deep learning research, deep learning technology has been introduced in the financial industry. As stock market volatility has expanded during the COVID-19 pandemic, the accuracy of stock price movement prediction has become a significant challenge for effective stock market forecasting research. The importance of studies on stock price prediction is increasing in natural language processing (NLP) and the financial industry. The stock market is a highly volatile market affected by company-related information and stock price indicators; thus, research on predicting stock price movement using various variables is constantly being conducted. First, time-series-based stock price movement prediction research has been conducted in two primary studies: one using stock price data and one using text data, such as stock-related news and Twitter [1], [8]. Research using stock price data generally predicts stock price movement by converting the opening, high, low, and closing prices and the trading volume into technical indicators. Methods for learning time-series characteristics using the convolutional neural network (CNN) or recurrent neural network (RNN) have been proposed to predict the variability of time-series data [1]. However, technical analyses using stock price data face a limitation in that they cannot reveal patterns that affect stock price fluctuations [2].

In addition to price and text data, the relationship between companies affects stock market volatility. By establishing an attention mechanism-based model and analyzing the influence on stocks using price, text, and company relationship data, stock price movement prediction studies have also been conducted [6], [8]. Because various types of information

The associate editor coordinating the review of this manuscript and approving it for publication was Hui Ma.

affect stock prices, a study is conducted to predict stock price movement by analyzing the relationship between financial data, social media, and stocks in a hierarchical fashion based on the hierarchical graph attention network [6].

We propose a new method to predict stock price movement. In this paper, we analyze market signals for stock market volatility using price data and text data. The patterns of stock market volatility are identified by analyzing stock data using RNN-based models: long short-term memory (LSTM) and gated recurrent units (GRU).

In addition, to reflect the stock market information contained in the text data for stock price movement prediction, the contextual information is identified through the contextual word embedding of the bidirectional encoder representations from transformers (BERT). The multimodal time-series market signals from the price and text data affect the stocks [7]. After extracting the time-series features of the price data and sematic features of the text data, the extracted features are combined to create a mixed feature containing multimodal information. The interaction between the features of the price and text data is strengthened by mixing the characteristics of the mixed feature via the hybrid information mixing module.

We devise a hybrid information mixing module consisting of two multilayer perceptron (MLP) blocks to improve the performance of stock price movement prediction by effectively mixing the information for two features. The hybrid information mixing module consists of the feature-mixing MLP and interaction-mixing MLP. The rows of the mixed feature contain channel information for each feature, and the columns of the mixed feature are embedding vectors.

First, feature-mixing MLP operates independently in each channel of the mixed feature. The channel information for each feature is learned by inputting a matrix transpose of the mixed feature, combining time-series and semantic features. This method allows communication between various tokens. The interaction-mixing MLP operates independently on each token of the mixed feature, and the mixed feature learns the embedding vector of each feature. This method allows communication between different channels.

This paper makes the following contributions. First, we extract time-series and semantic features by embedding the two data types using the GRU and BERT methods, respectively, to reflect the unique characteristics of the price and text data for predicting stock price movement. Afterward, the stock market information is strengthened by creating a mixed feature containing multimodal information combining the two features affecting stock price fluctuations. The multimodal interaction of stock market information possessed by two different characteristics is captured using the mixed feature, reflecting the unique characteristics of each data type.

Second, this paper proposes the hybrid information mixing module for stock price movement prediction. Through the hybrid information mixing module comprising two MLP blocks, the mixed feature takes the hybrid information mixing module as input. In addition, the multimodal information contained in the mixed feature is double-learned for each row and column for the mixed feature. In the double-learning process, this module attempts to predict stock price movement by capturing market signals that affect stock price fluctuations.

Like other MLP-based models [8], [9], [10], [11], [13], [14], the hybrid information mixing module in this paper is much simpler structurally than the transformer-based model because the hybrid information mixing module also consists of MLP blocks. In this paper, an MLP-based model is applied to the stock price movement prediction in the hybrid information mixing module to improve the accuracy of stock price fluctuation prediction. The StockNet [1] dataset is used in this study.

In this paper, we focus on the performance of a hybrid information mixing module which is designed with two map blocks for effective interaction between the two features. The rest of the paper is organized as follows. Section II presents related work about research on technical analysis and fundamental analysis. In Section III and IV, we present the proposed Hybrid information mixing module. Section V reports experimental results and ablation study. The final section provides conclusions and future research directions.

II. RELATED WORK

Stock price movement prediction is an interesting task in NLP and finance. As interest in stock market prediction has increased, many studies have been conducted using various factors to predict stock prices. Regarding data utilization, models use historical price data in the stock market analysis [13] and text data from the news or social media [14], [15]. Furthermore, studies [1], [16], [17] have used price and text data together. Regarding algorithm utilization, some studies [4], [20] have analyzed the historical price data using the RNN and RNN-based variants, such as GRU and LSTM. Some studies [19], [20] have predicted the stock price by understanding investor emotions through sentiment analysis using text data. Research [6], [21] has been conducted to predict stock price movement by identifying correlations between companies using the graph network. Many event-driven methods [16], [27], [38] predict recent stock price movement trends using events extracted from news and social media. In addition, several studies [39], [40] have analyzed temporal factors of the stock market and extracted the characteristics of these temporal factors to reflect their effects on stock market fluctuations in predicting stock movement.

A. RESEARCH ON TECHNICAL ANALYSIS

Technical analyses depend on numerical features, such as the historical price [6]. Time-series data analyses are central to elements in stock price prediction. For an effective time-series analysis of the historical price of the stock market, the RNN variants LSTM and GRU are used to consider the sequential change in data. The RNN and its variants constitute the most commonly used class of tools for time-series forecasting tasks [22].

Zhang et al. [18] proposed a state-frequency memory regression network based on the RNN model to capture multifrequency trading patterns from historical price data for stock price prediction trends. In addition, Li et al. [23] built a multitask RNN framework to extract informative features from raw market data of individual stocks and predicted the stock price movement direction by capturing the consistency within the same stock portfolio. Li et al. [4] proposed a sentiment-autoregressive moving average model to represent stock prices by integrating the variables (price and news) that influence stock price prediction. Li et al. also suggested a differential privacy LSTM deep neural network that predicts stock prices with an LSTM-based model that combines the sentiment scores of news articles derived using the valence-aware dictionary and sentiment reasoner model and different news sources through differential privacy methods. Wu et al. [41] proposed a cross-modal attention-based hybrid RNN that selects trend-related social texts through cross-modal attention mechanisms and further incorporates representations of text sequences and trend series. Additionally, Nelson et al. [24] studied the LSTM network to predict future stock price trends based on price history and technical analysis indicators. Moreover, Xu and Cohen. [1] and Sawhney et al. [6] proposed a method to encode the temporal trend in historical prices using GRU to capture the price movement of stocks.

In stock market movement, price and text data and the relationship between companies affect stock price fluctuations. A study on stock price prediction is conducted by building a graph-based model to reflect the relationship between companies in the stock price movement. Sawhney et al. [6] proposed a novel architecture called multipronged attention network for stock forecasting that robustly blends temporal signals from social media, price, and interstock relationship data through a hierarchical graph neural network. In addition, Kim et al. [13] introduced a hierarchical attention network for stock prediction (HATS) that uses relational data for stock price movement prediction. The HATS method is designed to learn useful node representations by selectively aggregating information on different relationship types.

To learn momentum spillover signals, Zhao et al. [7] designed dual-attention network stock movement prediction (DANSMP). The DANSMP approach is based on a more comprehensive market knowledge graph that includes bi-typed heterogeneous entities, including listed companies and their associated executives, and hybrid relations, including explicit and implicit relations. Yoo et al. [42] proposed the data-axis transformer with multilevel contexts for stock price movement prediction. This model uses transformer encoders to learn temporal correlations within each stock and inter-stock correlations in an end-to-end manner. Moreover, the analysis of correlating stock data with its relevant factors involves data uncertainty from the data and modeling perspectives [43]. Wang et al. [43] addressed this uncertainty problem by designing a Copula-based contrast predictive coding (CoCPC) method to learn better stock representations with less uncertainty, using hierarchical coupling from the macro to the micro level.

To predict stock prices, Mukherjee et al. [44] used artificial neural networks or deep feedforward neural networks and CNNs, which are widely used to predict stock market prices. In addition, Mehtab and Jaydip [45] proposed a stock price movement prediction framework with a set of statistical, machine learning, and deep learning models to predict stock price movement. A study was conducted to predict stock price movements using meta-learning. Further, Zhan et al. [46] proposed meta-adaptive stock movement prediction with two-stage representation learning based on self-supervised learning and meta-learning for stock movement prediction.

B. RESEARCH ON FUNDAMENTAL ANALYSIS

Fundamental analyses use other variables, such as text and company event information, along with historical prices [6], [17]. First, social mood determines the investment behavior of stock investors and corporate managers [38]; thus, text information analyses that determine buying and selling by stock investors must be studied. Investors primarily gauge social mood via text information, such as news or social media, and reflect it in their investment behavior. Therefore, sentiment extracted from text data is one of the critical variables of fundamental analysis for stock price movement prediction [22].

Many studies have assessed the stock price movement by extracting the social atmosphere of social media or news through sentiment analysis. Stankeviciute et al. [22] built a model to predict stock price movement using sentiments extracted from social media and proposed a new topic model, topic sentiment latent Dirichlet allocation, that simultaneously captures social media topics and sentiments. In addition, Pagolu et al. [25] applied sentiment analysis and supervised machine learning principles to a tweet analysis method and analyzed the correlation between corporate stock price movements and tweet sentiments.

Second, news events affect investors' decisions, and the fluctuation in stock prices is influenced by investor trading; thus, events can influence the stock market [26]. Ding et al. [16] proposed a deep learning method for event-driven stock price prediction. Events are extracted from news texts, and long- and short-term impact analyses of events on stock price movement are modeled using the CNN. In addition, Zhou et al. [27] studied an event-driven trading strategy to predict stock price movement based on corporate event detection in news articles. They proposed a bi-level event detection model using global and local information to identify corporate events. Wu et al. [38] used a financial event stream to train classification neural networks containing a combined event extraction method, BERT/ALBERT, and extended hierarchical attention networks to detect latent event-stock links and systematic stock market behavior. Additionally, Xu et al. [47] proposed a relational event-driven stock-trend forwarding framework by constructing a stock graph considering the influence of event information in the stock market and designing a new propagation layer to spread the event information effect from related stocks.

Third, some studies [39], [40] have performed stock price movement prediction reflecting the feature of temporal factors for time-sensitive stock markets. For example, Jang et al. [3] proposed a stock price movement prediction approach through stance detection with a textual and financial signal framework. The framework includes time-sensitive and target-aware investment stage detection, expert-based dynamic stage aggregation, and stock movement prediction to enhance stock movement prediction. In addition, Sawhney et al. [40] proposed a ranked approach, the spatiotemporal hypergraph attention network for stock ranking, to jointly model the temporal evolution of stock interdependencies and prices. The approach also customized a new spatiotemporal attention hypergraph network architecture to rank shares based on profits.

III. METHOD

This paper proposes a hybrid information mixing module to predict stock price movement. The overall structure of the hybrid information mixing module is presented in Fig. 1, illustrating that it is divided into feature embedding, a hybrid information mixing module, and a binary classifier. Feature embedding consists of price and text embedding, and the hybrid information mixing module consists of a feature-mixing MLP and an interaction-mixing MLP. The binary classifier classifies whether stock price movement prediction is up or down.

A. FEATURE EMBEDDING

1) PRICE EMBEDDING

Historical price data serve as an index that can capture the fluctuation pattern of the stock market. We used the historical price data for each company to extract time-series characteristics of the stock market. In price embedding, time-series characteristics are extracted by analyzing and capturing stock market fluctuation signals of the historical price data using the LSTM and GRU [28]. The LSTM and GRU are variants of RNN algorithms that are typically used to analyze time-series data effectively. In RNN algorithms, the hidden state, as the memory of the networks, captures the previous temporal information history, leading to computed outputs that rely on the memory of the networks [29].

In this paper, to apply the RNN algorithm to obtain higher stock price movement prediction accuracy, we conducted a comparative experiment using LSTM and GRU. The LSTM method is suitable for learning long-term dependencies compared to existing RNNs and can process time-series data efficiently. The GRU method was proposed to improve LSTM performance, reduce the number of LSTM parameters, and simplify the design.

In addition, GRU, like LSTM, is an algorithm suitable for time-series data analysis and is commonly used for time-series data analysis with various variables. Therefore, we predicted stock price movement by applying GRU to a historical price data analysis to capture the sequential dependency during the trading day.

The five market variables of opening, high, low, and closing prices and trading volume for a stock on trading day i are concatenated and set as $p_i \in \mathbb{R}^5$. The time-series characteristics of p_i are analyzed using GRU. The time-series feature $g_i \in \mathbb{R}^{d_G}$ for the stock data is extracted using the market variables of trading day i as the output value of the hidden layer of GRU. In addition, g_i is the last hidden state of GRU for trading day i, and d_G is the hidden dimension in GRU.

Moreover, p_i^0, p_i^H, p_i^L, p_i^C and p_i^V represent the opening, high, low, closing prices and trading volume, respectively. Finally, h denotes the hidden state:

$$p_i = [p_i^0, p_i^H, p_i^L, p_i^C, p_i^V], \tag{1}$$

$$g_i = GRU_p(p_i, h_{i-1}), \;\; t - T \leq i \leq t - 1. \tag{2}$$

2) TEXT EMBEDDING

The semantic information contained in the text plays a vital role in understanding the context. In addition, a vast amount of stock-trend information is extracted from a massive amount of text data and quantitative information included in the analysis [4].

Tweets used as text data convey factual information and represent user sentiment for stocks [1], [6], [30]. Daily tweets and historical prices influence stock price movement differently [6], [30].

All tweets $[t_1, t_2, \ldots, t_N]$ uploaded on transaction day i for each firm capture the trend of the stock market. We applied BERT [9] to extract contextual information [31]. The BERT output is extracted for contextual word embedding $[e_1, e_2, \ldots, e_N]$ for all text data on trading day i, $e_v \in \mathbb{R}^{N\times d}$.

By averaging the sum of the contextual word embedding extracted from all text data on trading day i via BERT by the number of text data N on trading day i, the semantic feature $s_v \in \mathbb{R}^{d_B}$ of the text data corresponding to a specific trading day i is generated. In addition, s_v is the last hidden state of BERT, and N is the number of text data on each corporation on trading day i, and d_B is the hidden dimension of BERT:

$$e_v = BERT_s(t_v), \;\; v \in [1, N], \tag{3}$$

$$s_i = \frac{1}{N}\left(\sum_{v=1}^{N} e_v\right). \tag{4}$$

B. HYBRID INFORMATION MIXING MODULE

1) MIXED FEATURE

Stocks are affected by multimodal time-series market signals [7]. Stock market information may be strengthened by combining price and text data characteristics to effectively reflect market signals from each data type in predicting stock price movement. In addition, in computer vision tasks, an MLP-based model assumes an input of fixed dimensions, which is necessary because the parameters must be shared across all examples [11].

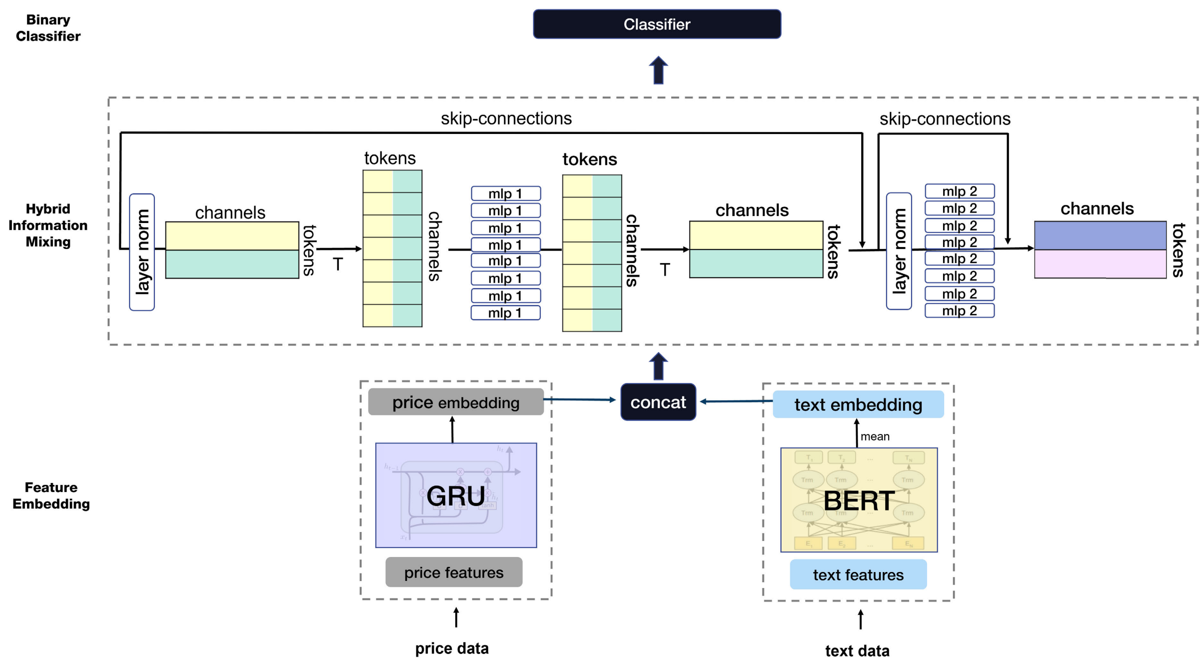

FIGURE 1. Overall procedures for hybrid information mixing module. It is divided into a feature embedding, hybrid information mixing module, and binary classifier. The feature embedding consists of price and text embedding to extract time-series and semantic features, using the GRU and BERT methods, respectively. The hybrid information mixing module consists of a feature-mixing MLP and interaction-mixing MLP to learn the mixed feature x_i created by combining the time-series feature of the price data and the semantic feature of the text data. The input of the feature-mixing MLP transposes x_i, which is the shape of the channels×tokens, taken as the input matrix. The interaction-mixing MLP is used to transpose back to the tokens×channels shape.The binary classifier classifies whether stock price movement prediction is up or down.

Unlike images, texts generally take variable dimensions as input in NLP tasks. Therefore, potential problems may arise if an MLP-based model is applied to NLP tasks [14]. To solve these problems, we combined the time-series feature g_i of the price data extracted using GRU and the semantic feature s_i extracted from the text data using BERT in the feature embedding to create a mixed feature $x_i \in \mathbb{R}^{F \times d_F}$. The mixed feature is a fixed-dimensional pairwise feature that contains multimodal information. The hybrid information mixing module takes a fixed-dimensional mixed feature as input, where F is the number of mixed features, and d_F is the hidden dimension of the fusion features:

$$x_i = CONCAT(g_i, s_i). \tag{5}$$

In the mixed feature, the stock market information is strengthened by creating a mixed feature with multimodal information by combining the two characteristics that affect stock price fluctuations. The market signals for each data type are reflected by capturing the multimodal interaction of information possessed by two different characteristics.

2) HYBRID INFORMATION MIXING MODULE

In this paper, to learn the mixed feature created by combining the time-series feature g_i of the price data and the semantic feature s_i of the text data, we devised a hybrid information mixing module comprising two MLP blocks: the feature-mixing MLP and interaction-mixing MLP. The MLP block consists of a fully connected layer, a GELU nonlinearity, and a fully connected layer in that order. Using the hybrid information mixing module, the information for mixed feature x_i is independently double-learned row-wise and column-wise to generate mixed features.

- **Feature-Mixing MLP:** Features extracted from the feature embedding contain discriminative information captured from the price and text data. Mixed feature x_i, which mixes the two features, is the input of the feature-mixing MLP. A row of the mixed feature matrix contains the channel information (i.e., the time-series information and meaningful information) of each feature, and a column of the mixed feature matrix is an embedding vector. The feature-mixing MLP acts on all columns of x_i. The input of the feature-mixing MLP transposes x_i, and $x_i^T \in \mathbb{R}^{d_F \times F}$, which is the shape of the channels×tokens, taken as the input matrix.

  Each token is input through the same feature-mixing MLP with a shared weight across all columns in the input matrix. We performed dense matrix multiplication that shares weights in the same channel for different tokens. Through all columns of x_i, the module was built to mix and learn channel information effectively by sharing the time-series and meaningful information contained in the time-series and semantic features.

This process enables the feature-mixing MLP to communicate globally between different spatial locations (i.e., tokens) through matrix transposition.

Mixed feature x_i is passed through the feature-mixing MLP to produce the output $x_{Feature}^T \in \mathbb{R}^{d_M \times F}$. Then, $x_{Feature} \in \mathbb{R}^{F \times d_M}$ transposed from $x_{Feature}^T$ takes the interaction MLP as input, where *GELU* denotes the GELU nonlinearity, and d_M denotes the hidden dimension of $x_{Feature}$. In addition, W_1 and $W_2 \in \mathbb{R}^{F \times d_F}$ are weights of the first and second fully connected layers of the interaction-mixing MLP, respectively. Finally. d_F is a hidden dimension in the feature-mixing MLP, and skip connections and layer normalization are used for learning stabilization [8]:

$$x_{Feature} = x_i + W_2(GELU(W_1(LayerNorm(x_i)^T)))^T. \tag{6}$$

Channel information is mixed through the feature-mixing MLP, allowing communication between different tokens.

- **Interaction-Mixing MLP:** In contrast to the feature-mixing MLP, the interaction-mixing MLP acts on all rows of x_i. The interaction-mixing MLP is used to transpose back to the tokens×channels shape and share weights across all tokens. The interaction-mixing MLP allows communication between channels by learning the correlation between the embedding vectors of each feature. In other words, the interaction-mixing MLP acts on each token independently and uses individual rows of x_i as input, allowing communication on different channels. The interaction-mixing MLP effectively uses dense matrix multiplication applied independently to each spatial location. This method enhances interaction by sharing information between channels within each spatial location. The input $x_{Feature}$ to the interaction-mixing MLP generates $x_{interaction} \in \mathbb{R}^{F \times d_I}$. Moreover, d_I denotes the hidden dimension of the interaction-mixing MLP, and W_3 and $W_4 \in \mathbb{R}^{F \times d_E}$ are weights of the third and fourth fully connected layers of the interaction-mixing MLP, respectively. In addition, d_E is a hidden dimension in the interaction-mixing MLP:

$$x_{interaction} = x_{Feature} + W_4(GELU(W_3(LayerNorm(x_{Feature})))). \tag{7}$$

By learning the correlation of two embedding vectors through the interaction-mixing MLP, the correlation between channels of $x_{Feature}$ is mixed into $x_{interaction}$.

3) BINARY CLASSIFIER

We derived the output $x_{interaction}$ through the hybrid information mixing module. The final output layer is passed on to a layer normalization, a global average pooling layer, and a fully connected layer:

$$x_{norm} = LayerNorm(x_{interaction}), \tag{8}$$

$$x_{GAP} = GAP(x_{norm}), \tag{9}$$

$$x_{output} = FC(x_{GAP}). \tag{10}$$

LayerNorm is the layer normalization that takes $x_{interaction}$ as input, and x_{norm} represents the output of layer normalization. In addition, *GAP* denotes global average pooling, and x_{GAP} is its output. Further, *FC* denotes the fully connected layer, and x_{output} represents the final classification result.

IV. CLASSIFICATION OF STOCK MOVEMENT PREDICTION

Stock price fluctuation prediction classification is conducted in the binary classifier. For model training, the label is set by comparing the closing price p_t^C of the corresponding t date with the closing price p_{t-1}^C of the previous date *t-1*:

$$y = \begin{cases} 1, & p^C \geq p_{t-1}^C \\ 0, & p^C < p_{t-1}^C \end{cases} \tag{11}$$

where $t \in \{1, 2, 3, \ldots, T\}$, and T denotes the number of trading days during the period $[t - T, T - 1]$. In addition, *1* indicates that the stock price is 'up' compared to the previous day's closing price, and *0* indicates that the stock price is 'down' compared to the previous day's closing price:

$$\hat{y} = \sigma(x_{output}), \tag{12}$$

$$BCELoss(\hat{y}, y) = -[y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})]. \tag{13}$$

where σ denotes the sigmoid function, y is the actual correct label value of the sample data, and $\hat{y}$ denotes the predicted value estimated by the neural network that passed the sigmoid function. The loss function is calculated using the binary cross-entropy loss.

V. EXPERIMENTS

A. DATASET

In this paper, the learning and performance evaluation of the hybrid information mixing module is performed using the StockNet dataset [1], which comprises a historical price dataset and Twitter dataset. The historical price dataset contains highly traded stock data in Standard and Poor's (S&P) 500 index of the New York Stock Exchange (NYSE) and NASDAQ markets. A Twitter dataset is extracted using regex queries made with NASDAQ ticker symbols, such as Google's $GOOG.

We adjusted the number of trading days and companies by removing samples that lacked price data or tweets. This paper uses 503 trading days from January 1, 2014, to January 1, 2016, for 85 companies in the StockNet dataset. Price data include historical price data for five market variables: opening, high, low, closing prices and trading volume. Text data comprise tweets for each company for the text analysis. In this paper, the StockNet dataset for the experiment is split into training and validation datasets at a ratio of 8:2. Although we used only a single validation dataset, performance can be improved using cross-validation. In this experiment, the window size is set to 30, and the stock price

TABLE 1. Comparison of information mixing method results.

Model	Acc. (%)	Precision(%)	Recall(%)	Specificity(%)	F1 score(%)	MCC
Long short-term memory [35]	53.53	53.29	93.49	11.17	67.89	0.08
Gated recurrent units [28]	53.02	52.80	86.78	16.82	65.66	0.05
Transformer encoder [36]	51.61	51.56	98.96	1.38	67.79	0.01
Hybrid Information Mixing Module (Ours)	69.20	63.49	95.19	41.37	76.17	0.43

movement for the closing price of the next trading day is predicted based on the data for the previous 30 days.

B. TRAINING SETUP

In the GRU used to extract the time-series characteristics of historical price data, the hidden dimension was set to 768, and the number of layers was set to four. The hidden layer dimensions of BERT were set to 768 to extract semantic characteristics of text data. In the hybrid information mixing module, the hidden dimensions of the MLP were set to 768, and the number of layers was set to eight. The number of layers for each GRU and MLP was designed as four and eight, respectively, to achieve high performance under more optimized experimental conditions. For effective information mixing, the hidden dimensions of each stage were designed to be the same as 768. The AdamW was used as the optimization method. The learning rate was set to 20^{-5}, and ϵ was set to 10^{-8} for AdamW. The batch size was set to 32, and we used one NVIDIA GeForce RTX 3090 in all the experiments.

C. EVALUATION METRICS

We used evaluation metrics, such as the accuracy, precision, recall (also called sensitivity) [48], specificity, F1 score, and MCC, based on the confusion matrix $\begin{pmatrix} TP & FN \\ FP & TN \end{pmatrix}$ [32] as an evaluation index for stock price movement prediction:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$Specificity = \frac{TN}{FP + TN} \tag{17}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{18}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{19}$$

The elements of the confusion matrix for computing accuracy are the true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The MCC evaluates the prediction in a more balanced manner than accuracy [43]. The F1 score is a kind of harmonic mean of the precision and recall of the model.

D. RESULTS OF STOCK MOVEMENT PREDICTION

A comparative experiment was conducted with the LSTM, GRU, and transformer for stock price movement prediction task to evaluate the performance of the hybrid information mixing module. As listed in Table 1, the hybrid information mixing module is the best stock price movement prediction model in accuracy, MCC, F1 score, precision, and Specificity. In contrast, transformer encoder is the best baseline in terms of Recall. the hybrid information mixing module accuracy is 69.20%. The MCC for the hybrid information mixing module is 0.43. The hybrid information mixing module achieves the best performance of 76.17% in the F1 score, 63.49% in precision, and 95.19% in recall. This result proves the capability to effectively reflect the market signals for each data type in stock price movement prediction by capturing the multimodal interaction of the information possessed by price and text data features through the proposed hybrid information mixing module.

1) EFFECTS OF HYBRID INFORMATION MIXING MODULE

The LSTM and GRU networks have often been used for effective time-series analyses of the stock market in stock price prediction tasks where time-series data analysis is the key [1], [4], [24], [33], [34]. The hybrid information mixing module is replaced with LSTM and GRU networks to evaluate the performance and compare the effectiveness of the LSTM and GRU networks as classical methods in time-series analysis and the hybrid information mixing module itself. In addition, like the MLP-based computer vision model, we compare the performance of the transformer encoder to verify that attention mechanisms can be replaced with simple MLP architectures in NLP.

Table 1 reveals that the hybrid information mixing module perfoms better than LSTM, GRU, and transformer, which are comparative models. This result proves that the two MLP blocks of the hybrid information mixing module effectively mix the multimodal information of the mixed feature in the stock price movement prediction task. The hybrid information mixing module improves the accuracy, MCC, and F1 score by about 15.67%, 0.35, and 8.28%, and 16.18%, 0.38, and 10.51%, respectively, compared to the performance of LSTM and GRU. The hybrid information mixing module is more effective in predicting stock price movement than the other models. In addition, the hybrid information mixing module has a simpler structure than the transformer-based method using self-attention, but the accuracy, MCC, and

TABLE 2. Comparison of price embedding.

Model	Acc. (%)	Precision(%)	Recall(%)	Specificity(%)	F1 score(%)	MCC
Long short-term memory [35]	68.53	63.21	94.42	40.49	75.73	0.41
Gated recurrent units [28]	69.20	63.49	95.19	41.37	76.17	0.43

F1 score improved by about 17.59%, 0.42, and 8.38%, respectively.

This experiment confirms that the multimodal interaction of each row and each column of information contained in the mixed feature matrix, which mixes time-series and semantic features using the hybrid information mixing module, can be effectively analyzed and captured. This result proves that the proposed MLP-based hybrid information mixing module has higher stock price movement prediction efficiency.

E. ABLATION STUDY

At the time of price embedding, to analyze the market signals on the volatility of the historical prices, the rules of market volatility were identified by analyzing stock data through the RNN-type models, LSTM and GRU.

The hidden state, which serves as the memory of the variant of the RNN, captures the temporal information of the previous history that leads to computed outputs relying on the network memory [29]. In general, the performance of GRU is similar to that of LSTM. Moreover, compared to LSTM, GRU is better suited for model training when the dataset is not very large due to its fewer parameters and simpler structure [37].

For these reasons, LSTM and GRU are used to analyze the price data. The performance difference between these two algorithms depends on the task. We conducted a performance comparison experiment between LSTM and GRU for price embedding using an algorithm with better performance in the stock price movement prediction. The experimental results are in Table 2.

The experiment results reveal a slight difference in accuracy, precision, recall, specificity, F1 score, and MCC: GRU scores 69.20%, 63.49%, 95.19%, 41.37%, 76.17%, and 0.43, respectively, and LSTM scores 68.53%, 63.21%, 94.42%, 40.49%, 75.73%, and 0.41, respectively. In this paper, price embedding is performed using GRU.

VI. CONCLUSION AND FUTURE RESEARCH

In this paper, we focus on the stock price movement prediction. After extracting time-series and semantic features, we proposed creating a mixed feature by mixing two characteristics in a hybrid information mixing module and mixing the multimodal information in the mixed feature. The feature-mixing and interaction-mixing MLPs of the hybrid information mixing module operate independently in a row-wise and column-wise manner. This learning process strengthens the interaction between the two data characteristics in the row and column information in the mixed feature to predict stock price movement. The proposed hybrid information mixing module predicts stock price movements better than the other models. The experiment results confirm that the accuracy, MCC, and F1 score of the hybrid information mixing module are 69.20%, 0.43, and 76.17%, which is improved compared to the previous model, exhibiting high performance.

In future research, we intend to use additional data that affect stock market volatility and improve the hybrid information mixing module to analyze the influence of additional variables affecting stock market volatility. We can use company relationship data as an additional data source and analyze the company relationship data to extract the degree of relationship and pattern of influence between each company. By combining price, text data, and company relationship data, we can generate multimodal information with three types of information in the stock market. We can also mix the multimodal information through the improved hybrid information mixing module. We could conduct a study to predict fluctuations in the stock market by extracting three types of interactions of multimodal information and capturing the correlation of dynamic markets.

REFERENCES

[1] 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, 2018, pp. 1970–1979.

[2] T. Lin, T. Guo, and K. Aberer, "Hybrid neural networks for learning the trend in time series," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Melbourne, VIC, Australia, Aug. 2017, pp. 2273–2279.

[3] E. Jang, H. R. Choi, and H. C. Lee, "Stock prediction using combination of BERT sentiment analysis and macro economy index," *J. Korea Soc. Comput. Inf.*, vol. 25, no. 5, pp. 47–56, 2020.

[4] X. Li, Y. Li, H. Yang, L. Yang, and X.-Y. Liu, "DP-LSTM: Differential privacy-inspired LSTM for stock prediction using financial news," 2019, *arXiv:1912.10806*.

[5] P. Sonkiya, V. Bajpai, and A. Bansal, "Stock price prediction using BERT and GAN," 2021, *arXiv:2107.09055*.

[6] R. Sawhney, S. Agarwal, A. Wadhwa, and R. R. Shah, "Deep attentive learning for stock movement prediction from social media text and company correlations," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 8415–8426. [Online]. Available: https://aclanthology.org/2020.emnlp-main.676

[7] Y. Zhao, H. Du, Y. Liu, S. Wei, X. Chen, F. Zhuang, Q. Li, J. Liu, and G. Kou, "Stock movement prediction based on bi-typed hybrid-relational market knowledge graph via dual attention networks," 2022, *arXiv:2201.04965*.

[8] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, D. Keysers, J. Uszkoreit, M. Lucic, and A. Dosovitskiy, "MLP-mixer: An all-MLP architecture for vision," in *Proc. Adv. Neural Inf. Process. Sys.*, vol. 34, 2021, pp. 24261–24272. [Online]. Available: https://proceedings.neurips.cc/paper/2021/hash/cba0a4ee5ccd02fda0fe3f9a3e7b89fe-Abstract.html

[9] H. Liu, Z. Dai, D. So, and Q. V. Le, "Pay attention to MLPs," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 9204–9215. [Online]. Available: https://openreview.net/pdf?id=KBnXrODoBW

[10] M. Li, X. Zhao, C. Lyu, M. Zhao, R. Wu, and R. Guo, "MLP4Rec: A pure MLP architecture for sequential recommendations," in *Proc. 31st Int. Joint Conf. Artif. Intell.*, Vienna, Austria, Jul. 2022, pp. 2138–2144.
[11] H. Touvron, P. Bojanowski, M. Caron, M. Cord, A. El-Nouby, E. Grave, G. Izacard, A. Joulin, G. Synnaeve, J. Verbeek, and H. Jégou, "ResMLP: Feedforward networks for image classification with data-efficient training," 2021, *arXiv:2105.03404*.
[12] F. Mai, A. Pannatier, F. Fehr, H. Chen, F. Marelli, F. Fleuret, and J. Henderson, "HyperMixer: An MLP-based green AI alternative to transformers," 2022, *arXiv:2203.03691*.
[13] R. Kim, C. H. So, M. Jeong, S. Lee, J. Kim, and J. Kang, "HATS: A hierarchical graph attention network for stock movement prediction," 2019, *arXiv:1908.07999*.
[14] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Deep learning for event-driven stock prediction," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, Buenos Aires, Argentina, 2015, pp. 2327–2333.
[15] Z. Hu, W. Liu, J. Bian, X. Liu, and T.-Y. Liu, "Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction," in *Proc. 11th ACM Int. Conf. Web Search Data Mining*, New York, NY, USA, Feb. 2018, pp. 261–269.
[16] Q. Liu, X. Cheng, S. Su, and S. Zhu, "Hierarchical complementary attention network for predicting stock price movements with news," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, Italy, Oct. 2018, pp. 1603–1606.
[17] X. Du and K. Tanaka-Ishii, "Stock embeddings acquired from news articles and price history, and an application to portfolio optimization," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, Seattle, WA, USA, 2020, pp. 3353–3363.
[18] L. Zhang, C. Aggarwal, and G.-J. Qi, "Stock price prediction via discovering multi-frequency trading patterns," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Halifax, NS, Canada, Aug. 2017, pp. 2141–2149.
[19] T. H. Nguyen and K. Shirai, "Topic modeling based sentiment analysis on social media for stock market prediction," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, Beijing, China, vol. 1, 2015, pp. 1354–1364.
[20] X. Li, H. Xie, L. Chen, J. Wang, and X. Deng, "News impact on stock price return via sentiment analysis," *Knowl.-Based Syst.*, vol. 69, pp. 14–23, Oct. 2014.
[21] C. Rui and Q. Li, "Modeling the momentum spillover effect for stock prediction via attribute-driven graph attention networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35 no. 1, 2021, pp. 55–62. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/16077
[22] K. Stankeviciute, A. M. Alaa, and M. van der Schaar, "Conformal time-series forecasting," in *Proc. Adv. Neural Inf. Process. Sys.*, vol. 34, 2021, pp. 6216–6228. [Online]. Available: https://proceedings.neurips.cc/paper/2021/hash/312f1ba2a72318edaaa995a67835fad5-Abstract.html
[23] C. Li, D. Song, and D. Tao, "Multi-task recurrent neural networks and higher-order Markov random fields for stock price movement prediction: Multi-task RNN and higer-order MRFs for stock price classification," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Anchorage, AK, USA, Jul. 2019, pp. 1141–1151.
[24] D. M. Q. Nelson, A. C. M. Pereira, and R. A. de Oliveira, "Stock market's price movement prediction with LSTM neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Anchorage, AK, USA, May 2017, pp. 1419–1426.
[25] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, "Sentiment analysis of Twitter data for predicting stock market movements," in *Proc. Int. Conf. Signal Process., Commun., Power Embedded Syst. (SCOPES)*, Paralakhemundi, Odisha, India, Oct. 2016, pp. 1345–1350.
[26] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Using structured events to predict stock price movement: An empirical investigation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Doha, Qatar, 2014, pp. 1415–1425.
[27] Z. Zhou, L. Ma, and H. Liu, "Trade the event: Corporate events detection for news-based event-driven trading," 2021, *arXiv:2105.12825*.
[28] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Doha, Qatar, 2014, pp. 1724–1734.
[29] P. Murugan, "Learning the sequential temporal information with recurrent neural networks," 2018, *arXiv:1807.02857*.
[30] G. P. C. Fung, J. X. Yu, and W. Lam, "News sensitive stock trend prediction," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Berlin, Germany: Springer, 2002, pp. 481–493.
[31] Z. Lu, D. Pan, and J.-Y. Nie, "VGCN-BERT: Augmenting BERT with graph embedding for text classification," in *Proc. 42nd Eur. Conf. Inf. Retrieval.*, Lisbon, Portugal: Springer, 2020, pp. 369–382.
[32] F. J. Provost, T. Fawcett, and R. Kohavi, "The case against accuracy estimation for comparing induction algorithms," in *Proc. ICML*, vol. 98, 1998, pp. 445–453.
[33] K. Bandara, C. Bergmeir, and S. Smyl, "Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach," *Exp. Syst. Appl.*, vol. 140, Feb. 2020, Art. no. 112896.
[34] K. A. Althelaya, E.-S.-M. El-Alfy, and S. Mohammed, "Stock market forecast using multivariate analysis with bidirectional and stacked (LSTM, GRU)," in *Proc. 21st Saudi Comput. Soc. Nat. Comput. Conf. (NCC)*, Riyadh, Saudi Arabia, Apr. 2018, pp. 1–7.
[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[36] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Sys.*, Long Beach, CA, USA, 2017, pp. 5998–6008.
[37] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust anomaly detection for multivariate time series through stochastic recurrent neural network," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Anchorage, AK, USA, Jul. 2019, pp. 2828–2837.
[38] X. Wu, "Event-driven learning of systematic behaviours in stock markets," in *Proc. Findings Assoc. Comput. Linguistics. (EMNLP)*, 2020, pp. 2434–2444. [Online]. Available: https://aclanthology.org/2020.findings-emnlp.220
[39] H. Wang, T. Wang, and Y. Li, "Incorporating expert-based investment opinion signals in stock prediction: A deep learning framework," in *Proc. AAAI Conf. Artif. Intell.*, New York, NY, USA, vol. 34, no. 1, 2020, pp. 971–978.
[40] R. Sawhney, S. Agarwal, A. Wadhwa, T. Derr, and R. R. Shah, "Stock selection via spatiotemporal hypergraph attention network: A learning to rank approach," in *Proc. AAAI Conf. Artif. Intell.*, New York, NY, USA, vol. 35, no. 1, 2021, pp. 497–504.
[41] H. Wu, W. Zhang, W. Shen, and J. Wang, "Hybrid deep sequential modeling for social text-driven stock prediction," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, Italy, Oct. 2018, pp. 1627–1630.
[42] J. Yoo, Y. Soun, Y.-C. Park, and U. Kang, "Accurate multivariate stock movement prediction via data-axis transformer with multi-level contexts," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Singapore, Aug. 2021, pp. 2037–2045.
[43] G. Wang, L. Cao, H. Zhao, Q. Liu, and E. Chen, "Coupling macro-sector-micro financial indicators for learning stock representations with less uncertainty," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 5, 2021, pp. 4418–4426. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/16568
[44] S. Mukherjee, B. Sadhukhan, N. Sarkar, D. Roy, and S. De, "Stock market prediction using deep learning algorithms," in *Proc. CAAI Trans. Intell. Technol.*, 2021, pp. 1–13.
[45] S. Mehtab and J. Sen, "A time series analysis-based stock price prediction using machine learning and deep learning models," 2020, *arXiv:2004.11697*.
[46] D. Zhan, Y. Dai, Y. Dong, J. He, Z. Wang, and J. Anderson, "Meta-adaptive stock movement prediction with two-stage representation learning," in *Proc. NeurIPS Workshop Distrib. Shifts, Connecting Methods Appl.*, New Orleans, LA, USA, 2022, pp. 1–16.
[47] W. Xu, W. Liu, C. Xu, J. Bian, J. Yin, and T.-Y. Liu, "REST: Relational event-driven stock trend forecasting," in *Proc. Web Conf.*, Ljubljana, Slovenia, Apr. 2021, pp. 1–10.
[48] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," 2020, *arXiv:2010.16061*.

JOOWEON CHOI (Student Member, IEEE) received the B.A. degree in library and information science from Duksung Women's University, Seoul, Republic of Korea, in 2014, and the M.A. degree in media management from Fudan University, Shanghai, China, in 2021. She is currently pursuing the M.S. degree with the Graduate School of Artificial Intelligence, Chung-Ang University, Seoul. Her research interests include stock movement prediction, time series data analysis, and natural language processing.

SHIYONG YOO received the B.S. and M.S. degrees in economics (agricultural economics) from Seoul National University, Seoul, Republic of Korea, in 1991 and 1993, respectively, and the Ph.D. degree in applied economics and finance from Cornell University, Ithaca, NY, USA, in 2003. He is currently a Professor of finance with the CAU Business School, Chung-Ang University, Seoul. His current research interests include asset pricing and machine learning in financial markets.

XIAO ZHOU received the B.S. degree in financial statistics from Hunan University, Changsha, Hunan, in 1996, the M.S. degree in science of business administration from the Chinese Academy of Social Sciences, Beijing, in 2001, and the Ph.D. degree in science of business management from Fudan University, Shanghai. She is currently a Professor with the School of Journalism, Fudan University. She is a Visiting Scholar with the Massachusetts Institute of Technology (MIT) and The City University of New York (CUNY). Her research interests include the development of new media technology and industry, new media culture, and management.

YOUNGBIN KIM (Member, IEEE) received the B.S. and M.S. degrees in computer science and the Ph.D. degree in visual information processing from Korea University, in 2010, 2012, and 2017, respectively. From August 2017 to February 2018, he was a Principal Research Engineer with Linewalks. He is currently an Assistant Professor with the Graduate School of Advanced Imaging Science, Multimedia and Film, Chung-Ang University. His current research interests include data mining and deep learning.