

Received 6 December 2022, accepted 21 January 2023, date of publication 30 January 2023, date of current version 2 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3240422

RESEARCH ARTICLE

Effective Exploitation of Macroeconomic Indicators for Stock Direction Classification Using the Multimodal Fusion Transformer

TAE-WON LEE¹, PAWEŁ TEISSEYRE^{2,3}, AND JAESUNG LEE^{1,4}

¹Department of Artificial Intelligence, Chung-Ang University, Seoul 06974, South Korea

²Institute of Computer Science, Polish Academy of Sciences, 00-901 Warsaw, Poland

³Faculty of Mathematics and Information Sciences, Warsaw University of Technology, 00-661 Warsaw, Poland

⁴AI/ML Innovation Research Center, Chung-Ang University, Seoul 06974, South Korea

Corresponding author: Jaesung Lee (curseor@cau.ac.kr)

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (2021-0-01341, Artificial Intelligence Graduate School Program(Chung-Ang University)).

ABSTRACT An enormous ripple effect can occur in financial data mining if it accurately predicts stock prices. However, predicting stock prices using only stock price data is difficult because of the random nature of stock price data. This paper attempts to fuse data to solve the stock price prediction problem. The following data affecting the stock price are added to the proposed method as an additional modality: macroeconomic indicators and the months and day of the week. The multimodal early fusion method is used, which learns the intermodality correlation of features. The proposed model in this paper outperformed the comparison models and achieved statistically significant results. Specifically, 27 out of 50 stocks achieved higher classification accuracy than the comparative model. In addition, the in-depth analysis indicates that the early fusion strategy achieved better classification accuracy in 30 of 50 datasets than the late fusion strategy for stock price prediction.

INDEX TERMS Stock price prediction, multimodal learning, information fusion, macroeconomic data, technical indicator.

I. INTRODUCTION

As the world economy expands, equity markets have grown, and the number of market participants has increased. Stock price prediction has become one of the most popular subjects of financial data mining [1]. Although an accurate stock price prediction can help investors to make the right decision, it is difficult to achieve because of randomness, non-linearity, and the high noise level of the stock price data [2], [3]. Thus, using only the stock price data may be unsuitable for stock price prediction because the stock market is sensitively affected by external factors, such as the world economy, domestic politics, accidental events, and even by the months or day of the week [4].

Researchers use statistical and machine learning models to model the stock price. The autoregressive integrated moving

average (ARIMA) model, which combines the autoregressive model, moving average model, and differencing, is one of the most representative statistical models to analyze time-series data [5]. Traditional machine learning models, such as the support vector machine and hidden Markov model, have also been applied in this field [6], [7], [8]. Deep neural networks (DNNs) have recently demonstrated excellent prediction performance [9], [10] owing to their information fusion capability that helps capture nonlinear relationships between various financial data sources [11].

In addition to the stock price data, another information source, economic indicators, can represent economic situations in the market [12]. Thus, by exploiting the economic indicators for predicting stock prices, an improvement in accuracy can be expected. In addition, because the stock price can be correlated to the months and day of the week, the accuracy of the stock price prediction can be enhanced [4]. In addition, the internal structure of information fusion

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

should be carefully designed to achieve the best prediction power [13].

However, conventional studies suffer from performance degradation because of the limited number of considered information sources and the intrinsically conducted structure design of the information fusion. The information fusion strategy in DNNs can be roughly divided into two types according to the location where fusion starts in the network: early fusion, specializing in capturing the intermodality correlation, and late fusion, specializing in capturing the intramodality correlation [14]. Because these two fusion strategies have different strengths in information processing, it is still unclear which fusion strategy will lead to the best prediction power for stock price prediction. This paper proposes a novel multimodal fusion transformer for stock price prediction. Contributions from this study can be summarized as follows:

- A novel multimodal early fusion transformer is proposed to achieve accurate stock price prediction. In the proposed network, the early fusion strategy is used to fuse the information between each modality effectively.
- Twenty-five information sources are considered in this study from three domains: stock price, months and day of the week, and macroeconomic indicator modalities.
- An in-depth analysis is conducted regarding the fusion strategy. The proposed method indicated that the early fusion strategy provides the best prediction performance. However, a group of stocks offers better prediction performance using the late fusion strategy.

The experimental results on 50 stock price datasets indicated that the proposed multimodal early fusion transformer significantly outperforms conventional methods.

II. RELATED WORK

In the field of time-series prediction, the ARIMA model was frequently employed for developing the model of power generation, traffic flow, and sensor data [15], [16], [17]. Another researcher used the ARIMA model to predict financial time series [5]. The statistical time-series model, like ARIMA, assumes that the data have linearity [18]. It is challenging to apply such assumptions throughout a financial time-series data analysis because the data have traits of nonlinearity and randomness [2], [3].

Despite the excellent performance of a statistical stock price prediction using, for example, the ARIMA model, a series of attempts have been made to exploit machine learning models for stock price prediction. For instance, in [19], the authors used support vector regression to predict stock minute prices using technical indicators. In addition to the support vector regression, the random forest was also used to build a stock price prediction model [8]. The continuous hidden Markov model, one of the most popular modeling techniques for time-series data analysis, was used in [20]. In this work, the emission probabilities are modeled as Gaussian mixture models.

The DNNs can be effective alternatives to conventional econometric and statistical models with weaknesses in modeling financial time-series data with nonlinear traits [21]. For example, a convolutional neural network (CNN) used for image data analysis can be applied to stock price prediction [22]. In addition, long short-term memory (LSTM), bidirectional LSTM, and attention mechanisms model sequential data, such as natural language and time-series data. In detail, the bidirectional attention LSTM model for stock price prediction was proposed in [23]. The model combined with the CNN and LSTM is employed to extract the features of the limit order book and stock price data. The CNN module extracts the features in the limit order book, and the LSTM module extracts the time-series features in stock price data [24].

The transformer model was devised to improve the attention encoder-decoder LSTM model [25], whose variant is widely used in the field of sequence modeling [26], [27]. A stock price prediction model combined with the CNN and attention bidirectional LSTM was also considered [1]. Recently, a transformer model for stock market index prediction was suggested [28]. The attention mechanism in the transformer can learn the correlation of many stocks dynamically with the market index and helps them predict each stock [29]. Some authors [30] insisted that conventional machine learning algorithms ignore the stochastic property of a stock that changes over time. Thus, they proposed adversarial training to learn the stochastic property of the stock price using the attention LSTM.

Recent financial studies have combined various data with stock price data [13]. The information fusion can be enhanced by conducting parallel operations that contain inter/intra crossovers and adaptive mutations in the genetic algorithm [31]. The proposed information fusion-based genetic algorithm approach can optimize the parameters of an extended short-term memory stock price prediction model and selects a set of features. Other authors [32] addressed the problem of technical analysis information fusion in improving stock market index-level prediction. Multiple predictive systems based on various technical analysis metric categories have been devised. The system has multiple prediction models with various technical analysis metric data categories. Each prediction model is based on an ensemble of neural networks using particle swarm optimization. Other authors [33] proposed multimodal neural networks where the architecture learns the cross-correlation of the United States and Korean stock prices.

With recent progress in natural language processing techniques, text data, such as news and Twitter data, were used for feeding the overall atmosphere of the market to the prediction model that each stock data is unable to deliver [34]. An event-driven approach was used with stock price data [35]. The event-driven method was elaborated by extracting the stock-related event from the news titles. Moreover, the method enhancing the joint effects was also devised by calculating the similarity of the two stocks using

their p -change values and Pearson correlation coefficient. Historical stock quantitative data, social media data, and web news data could be used by expressing the information as a matrix and tensor for stock market prediction [36]. The stock quantitative feature matrix and stock correlation matrix were created using the quantitative and social media data for the stock. The stock movement tensor was built using events and sentiment extraction from news articles and social media. The stock quantitative feature matrix and stock correlation matrix were factorized, and the stock movement tensor was decomposed for stock price prediction. The bag-of-words and named entity approach using a large corpus of freely available financial reports were used to predict the volatility of the stock returns and stock market prediction using support vector regression [37], [38], [38]. However, these approaches suffer from several issues; the complex nature of natural language processing techniques, the subjectivity in the news sentences or social media, and legal problems such as copyright issues. As a result, the empirical results cannot be reproduced, making the application of those methods to practice extremely difficult. In this paper, we consider a simpler alternative, the macroeconomic indicators, to feed the tendency of the overall market.

III. MATERIALS AND METHOD

A. MOTIVATION

First, we discuss the data we used in this study. The macroeconomic indicator is a representative factor affecting the stock market [40]. Some fundamental macroeconomic indicators, such as the exchange rate, interest rate, industrial production, and inflation, have associations with the stock price [39]. In addition to the macroeconomic indicators, the effect of the months and day of the week can also be considered to obtain information affecting the stock market [4]. However, stock prices may only partially reflect the macroeconomic indicator information and the effects of the months and day of the week. Therefore, the model must have information about the macroeconomic indicators and the months and day of the week to reflect this information instantly. Thus, data fusion can be a valid method for accurate stock price prediction because the stock markets are affected by many factors [41] and randomness [2].

Multimodal neural networks that fuses various information sources internally can be considered to solve the randomness problem in data. The early fusion method extracts the intermodality correlation more efficiently than the intramodality correlation [33]. In this paper, the model that learns the correlation between significant data affecting the stock price well is designed to solve randomness problems in the stock price data. The transformer encoder is advantageous for extracting information from sequential data [25]. The query, key, and value have the same value in the transformer encoder. The scaled dot-product self-attention in the transformer encoder learns the correlation between the query, key, and value. The multimodal early fusion transformer method is used in this

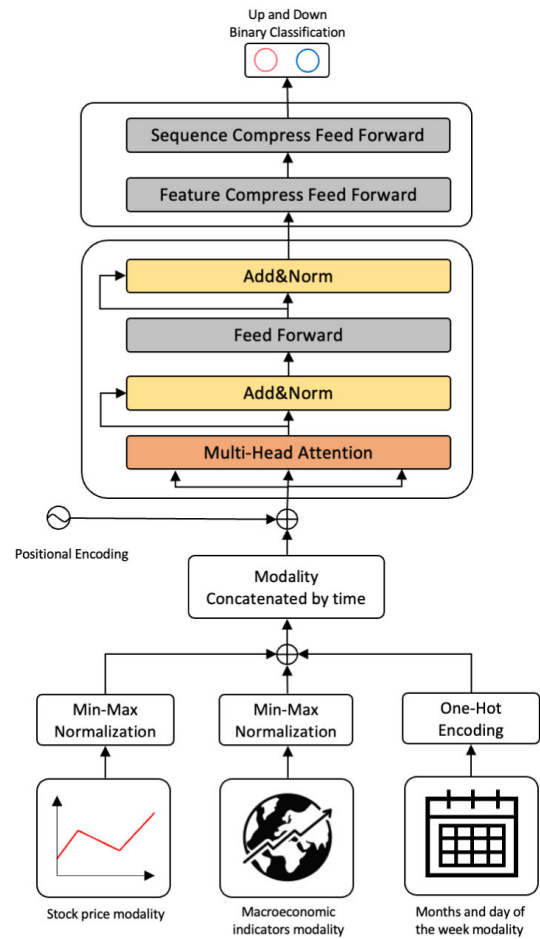


FIGURE 1. Overview of the proposed method comprising feature concatenation, a transformer encoder, and a classification layer.

paper to employ the characteristics of the transformer encoder that learn the correlations between model input features.

In this study, the problem of the stock direction classification is considered that predicts whether stock prices rise or fall in the future because it guides investors to buy a stock before the price rises or sell it before its value declines [42], [43]. Specifically, the proposed method predicts the next day's stock up and down direction using the previous ten days of data as a binary classification task. Thus, the stock direction classification in this study can be formulated as the binary classification problem based on about two weeks of stock market history, where the performance can be evaluated using a conventional accuracy metric. As a result, an early fusion neural network is designed with feature concatenation, a transformer encoder, and a classifier layer. Fig. 1 presents the overview of the proposed method.

B. PROPOSED METHOD

The entire experiment was conducted by setting three modalities: stock price modality, months and day-of-the-week modality, and macroeconomic indicator modality. Table 1

TABLE 1. Stock price modality.

Column Name	Description
Open	The price at which the financial security opens in the market when trading begins.
High	Intraday highest trading price.
Low	Intraday lowest trading price.
Volume	The number of shares transacted every day.
Adjusted Close	A stock's closing price to reflect that stock's value after accounting for any corporate actions.
Updown	$C_t - C_{t-1}$
Change	$\frac{C_t - C_{t-1}}{C_{t-1}} \times 100$
10-day Moving Everage	$\frac{C_t + C_{t-1} + \dots + C_{t-9}}{10}$
Weighted 10-day Moving Average	$\frac{(n) \times C_t + (n-1) \times C_{t-1} + \dots + C_{t-9}}{(n + (n-1) + \dots + 1)}$
Momentum	$C_t - C_{t-n}$
Stochastic K%	$\frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} \times 100$
Stochastic D%	$\frac{\sum_{i=0}^{n-1} K_{t-i} \%}{n}$
RSI (Relative Strength Index)	$100 - \frac{100}{1 + \frac{(\sum_{i=0}^{n-1} Up_{t-i}/n)}{(\sum_{i=0}^{n-1} Dwn_{t-i}/n)}}$
MACD (Moving Everage Convergence Divergence)	$MACD(n)_{t-1} + 2/n + 1 \times (DIFF_t - MACD(n)_{t-1})$
Larry William's R%	$\frac{H_n - C_t}{H_n - L_n} \times 100$
A_D (Accumulation /Distribution) Oscillator	$\frac{H_t - C_{t-1}}{H_t - L_t}$
CCI (Commodity Channel Index)	$\frac{M_t - SM_{t-1}}{0.015 D_t}$

C_t : the adjacent close price, L_t : the low price,

H_t : the high price at time t ,

$DIFF$: $EMA(12)_t - EMA(26)_t$,

EMA : exponential moving average,

$EMA(k)_t$: $EMA(k)_{t-1} + \alpha \times (C_t - EMA(k)_{t-1})$,

α smoothing factor: $2/1+k$,

k : time period of k day exponential moving average LL_t

HH_t : lowest low and highest high in the last t days,

M_t : $H_t + L_t + C_t/3$,

SM_t : $(\sum_{i=1}^n M_{t-i+1})/n$,

D_t : $(\sum_{i=1}^n |M_{t-i+1} - SM_t|)/n$,

UP_t : the upward price change,

Dwn_t : the downward price change at time t

lists the details of the stock price modality. The KDD17 public dataset was used to add to the reliability of this experiment [44]. The features of the KDD17 dataset are open, close, high, low, volume, and adjusted close. Moreover, ten

TABLE 2. Months and day of the week modality.

Column Name	Description
Months	Encoded months data using One-Hot Encoding.
Day of the week	Encoded day of the week data using One-Hot Encoding.

TABLE 3. Macroeconomic indicator modality.

Column Name	Description
NASDAQ100	A stock market index made up of 102 equity securities issued by 101 of the largest non-financial companies listed on the Nasdaq stock exchange.
US 2year Bond Yield	The yield received for investing in a US government issued treasury security that has a maturity of 2 years.
US 10year Bond Yield	The yield received for investing in a US government issued treasury security that has a maturity of 10 years.
US 30year Bond Yield	The yield received for investing in a US government-issued treasury security that has a maturity of 30 years.
US Dollar Index	An index of the value of the United States dollar relative to a basket of foreign currencies, often referred to as a basket of U.S. trade partners' currencies.
WTI Oil Price	A specific grade of crude oil and one of the main three benchmarks in oil pricing, along with Brent and Dubai Crude.

technical indicators were added, and two additional features in the stock price modality were extracted through feature engineering to extract significant features in stock price data [45]. Table 2 details the months and day-of-the-week modality. In this modality, the months and day-of-the-week features were extracted using one-hot encoding because the month is a categorical variable with 12 classes, and the day of the week is also a categorical variable with seven classes. The extracted one-hot encoding vector can be expressed as $[0, 1, \dots, 0]$. Last, in the macroeconomic indicator modality described in Table 3, six features were selected: NASDAQ 100; the US 2-, 10-, and 30-year bond yields; the US Dollar Index; and the WTI oil price. Only the close column in macroeconomic indicators was used in the data, even if the macroeconomic indicators had several columns, such as open, high, low, close, and volume. Before conducting the training and testing processes, the min-max normalization was applied to each ten-day sequence because financial data vary the scale over a long period.

Fig. 2 depicts the detailed structure of the former half of the proposed method, including the feature concatenation and multihead attention model. Each modality was concatenated by a feature dimension to fuse data using the early fusion method. We let F_1, F_2 and F_3 be the feature sizes of the stock price modality, the months and day-of-the-week modality, and the macroeconomic indicator modality, respectively. Then, $x_1 \in \mathbb{R}^{l \times F_1}$, $x_2 \in \mathbb{R}^{l \times F_2}$, and $x_3 \in \mathbb{R}^{l \times F_3}$ are the stock price modality, months and day-of-the-week modality, and macroeconomic indicator modality, where l is the sequence

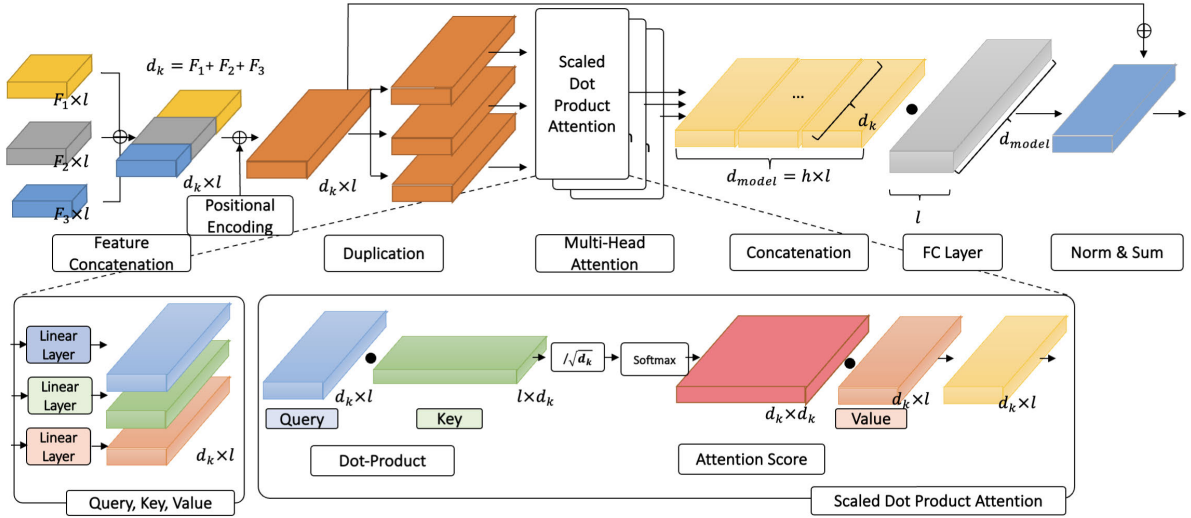


FIGURE 2. Process of feature concatenation and multihead attention in the proposed method.

length, respectively. These multiple information sources can be concatenated alongside the feature dimension to create an input modality matrix $M \in \mathbb{R}^{l \times (F_1 + F_2 + F_3)}$, represented as follows:

$$M = [x_1; x_2; x_3]. \quad (1)$$

In this paper, the months and day-of-the-week modality is represented using a one-hot encoding scheme, resulting in twelve months features and five day-of-the-week features. Thus, F_1 , F_2 , and F_3 were set to 17, 17, and 6, respectively because the numbers of features in each modality were 17, 17, and 6. The day that affects the next day's stock price prediction can be different. The positional encoding module was used to let the model know the position information of the data. The sine and cosine functions express different frequencies [25]. In addition, pos denotes the position, and i represents the dimension. Further, d_{model} denotes the feature dimension of the modality data concatenated along with the feature dimension, and M concatenates three modalities that are added based on the positional encoding:

$$PE_{(pos, 2i)} = \sin(pos/10000^{2i/d_{model}}) \quad (2)$$

and

$$PE_{(pos, 2i+1)} = \cos(pos/10000^{2i/d_{model}}) \quad (3)$$

Scaled dot-product attention comprises the query, key of dimension d_k , and value of dimension d_v , its dot-product [46]. The query, key, and value were calculated by multiplying each other's linear layer of dimension d_v by the output of the positional encoding. Then, the query and key were multiplied using the dot-product operation. After the dot-product, the multiplied output of the dot-product was divided by $\sqrt{d_k}$ and multiplied by the value. Scaled dot-product self-attention represents the data parts and their relevance to each other. Fig. 3

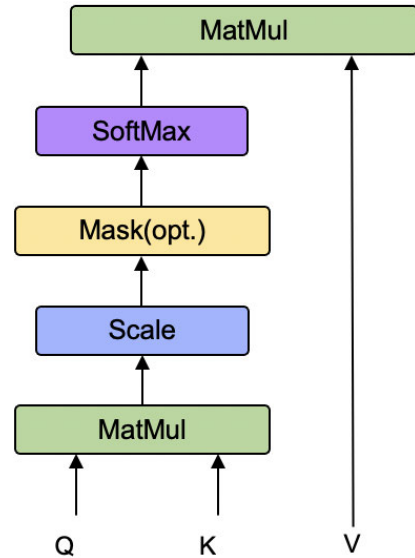


FIGURE 3. Scaled dot-product attention.

illustrates the overview of the scaled dot-product attention, represented as

$$Attention(Q; K; V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (4)$$

Using only a single attention mechanism can be challenging to learn various features in three modalities. Multihead attention can jointly address different representation subspaces that express a variety of features in data. It expresses the diverse subspace representation in the modality. We let $Head_i = Attention(QW_i^q; KW_i^k; VW_i^v)$ be the i th head of the multihead attention and $W_i^q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^k \in \mathbb{R}^{d_{model} \times d_k}$, and $W_i^v \in \mathbb{R}^{d_{model} \times d_k}$ comprise the linear layer. We let $W^0 \in \mathbb{R}^{d_{model} \times d_k}$ be a weighted matrix multiplied after all heads are

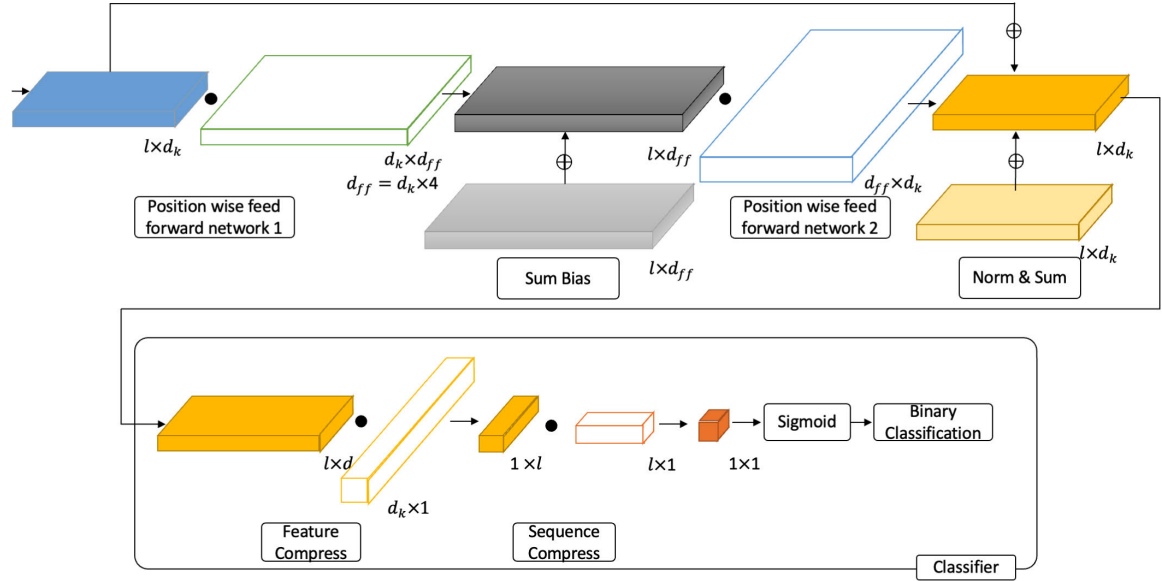


FIGURE 4. Process of the point-wise feedforward network and classifier in the proposed method.

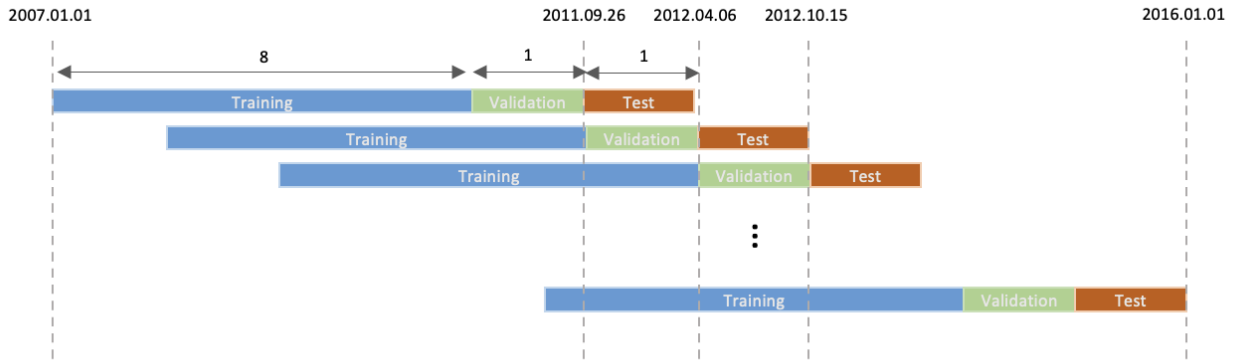


FIGURE 5. Blocked time-series cross-validation.

concatenated. In addition, $h = 8$ was the number of heads, and d_{model} was set to $d_k = d_{model}/h = 40$. Then the multihead in the proposed architecture can be defined as

$$MultiHead(Q; K; V) = [Head_1; \dots; Head_h]W^0. \quad (5)$$

Next, after the input data were passed through the feature concatenation and multihead attention, feature compression feedforward networks extract the information between features extracted using multihead attention. Fig. 4 depicts the latter half of the proposed method. The point-wise feedforward networks have a rectified linear unit (ReLU) activation function that extracts nonlinear features. In addition, W_1 and b_1 and W_2 and b_2 are the weight and bias, respectively, in point-wise feedforward networks. Each point-wise feedforward network was applied to each position separately and identically to the output of the multihead attention:

$$FFN(x) = \max(0; xW_1 + b_1)W_2 + b_2. \quad (6)$$

After the feature compression feedforward networks, the feature dimension is 1. In addition, time-sequence compression feedforward networks extract the information through the time-sequence dimension. We let x be the output of the point-wise feedforward networks. Moreover, $W_f \in \mathbb{R}^{d_k}$ and $b_f \in \mathbb{R}^{d_k}$ are the weight and bias values, respectively, of the feature-compressed feedforward networks. Further, $W_s \in \mathbb{R}^l$ and $b_s \in \mathbb{R}^1$ are the weight and bias of the sequence-compressed feedforward networks:

$$FFN_f(x) = \max(0; xW_f + b_f) \quad (7)$$

and

$$FFN_s(x) = \max(0; xW_s + b_s). \quad (8)$$

The output of the sequence-compressed feedforward networks enters the sigmoid function. The sigmoid function is the nonlinear function that makes the input value range $[0, 1]$.

TABLE 4. Comparison of the classification accuracy (%) between the proposed method and comparison models.

Ticker	Sector	Proposed Method	CNN-Attention BiLSTM [1]	Adversarial LSTM [30]	DTML [29]
AAPL	IT	54.16±4.14✓	48.43±6.98	47.65±1.49	51.09±6.67
AMZN	CD	51.66±3.39✓	50.46±6.13	50.78±6.71	47.81±5.38
BA	IND	52.60±5.53	52.97±5.16✓	49.53±7.10	49.53±6.56
BAC	FIN	54.16±5.29✓	49.06±8.54	50.78±7.10	51.09±4.84
BHP	MAT	48.33±4.47	48.28±5.01	49.84±5.01✓	45.62±6.87
BRK-B	FIN	47.91±5.70	48.28±5.43	50.00±5.63✓	47.96±10.15
CHL	FIN	48.12±2.58	50.00±6.70	50.62±7.66	52.96±6.18✓
CMCSA	CS	52.39±3.05	51.56±6.21	48.43±4.89	53.12±8.69✓
CVX	ENE	50.83±3.32✓	49.06±6.10	49.68±5.31	47.50±5.89
D	UTI	51.66±2.91	53.75±7.27✓	50.31±5.12	49.84±6.79
DCM	IND	51.71±9.56	50.31±5.17	48.12±5.21	52.50±4.93✓
DIS	CS	53.12±4.16✓	52.19±4.90	51.09±4.31	48.12±5.07
DOW	-	52.08±5.59✓	46.56±5.80	48.43±5.67	51.09±3.28
DUK	UTI	52.39±3.89	55.00±3.94✓	47.96±6.55	49.53±5.23
EXC	ENE	50.83±4.92✓	50.78±5.98	46.87±6.51	47.50±6.78
GE	ENE	52.50±4.93✓	47.96±4.63	52.44±6.74	48.90±4.94
GOOGL	CS	51.66±3.52	52.34±6.54✓	50.78±7.40	48.28±7.07
HD	CC	52.18±5.15	51.87±8.35	53.12±5.80	53.43±9.10✓
INTC	IT	51.45±2.60✓	51.09±4.36	48.43±8.64	48.59±4.16
JNJ	HEA	51.14±4.71	49.21±4.85	48.12±5.21	52.03±7.81✓
JPM	FIN	52.50±5.75✓	51.25±6.08	50.00±3.04	52.18±5.72
KO	CD	51.04±3.86	52.18±3.50✓	47.65±6.33	50.78±7.97
MA	FIN	53.33±4.41✓	50.09±5.50	49.37±7.75	51.09±6.44
MMM	IND	52.08±3.89✓	47.03±6.68	51.71±6.42	49.06±7.75
MO	CD	54.37±3.86	55.46±4.32✓	51.87±3.87	48.90±4.37
MRK	HEA	51.66±4.44✓	49.06±3.84	49.68±5.21	49.37±4.37
MSFT	IT	51.25±3.51✓	50.46±6.84	49.30±5.46	50.78±5.55
NGG	UTI	54.27±4.25✓	51.71±5.69	51.56±5.76	52.65±10.00
NTT	CS	50.62±4.68	48.43±4.84	51.40±6.02✓	50.65±7.26
NVS	HEA	51.77±5.18✓	50.15±4.86	51.25±5.66	51.25±4.98
ORCL	IT	51.25±2.22	52.03±3.95✓	50.62±8.00	50.46±4.25
PEP	CD	51.56±5.12✓	47.90±5.08	50.78±6.81	48.12±4.67
PFE	HEA	50.83±2.86✓	50.46±5.23	48.28±6.22	47.65±4.80
PG	CD	51.25±4.76	48.75±2.50	52.50±6.56✓	52.18±6.33
PTR	ENE	51.87±5.58✓	51.40±7.03	49.37±2.81	51.56±8.64
RDS-B	ENE	52.70±4.29✓	51.40±4.86	49.84±6.60	48.75±4.57
RIO	MAT	50.31±4.32	48.75±3.40	51.56±6.21	52.03±4.79✓
SO	ENE	53.12±3.01✓	50.62±5.28	51.71±5.25	48.75±4.79
SPY	-	55.20±4.24✓	52.96±9.09	50.31±3.93	47.18±5.21
SYT	IT	48.54±5.19	49.53±5.23	50.93±6.85	51.40±5.56✓
T	CS	51.66±5.14✓	50.62±7.88	50.93±5.59	51.56±8.75
TM	CC	50.83±6.74✓	47.96±5.59	50.78±4.52	50.78±4.02
TOT	ENE	51.97±4.64	51.40±5.69	52.03±6.36✓	48.59±7.44
UNH	HEA	52.50±4.70✓	50.93±5.55	51.71±4.49	44.37±5.81
UPS	IND	51.97±3.20	47.65±4.64	50.00±4.73	52.18±6.01✓
VALE	MAT	50.20±4.51	51.71±3.45	49.06±9.03	52.03±6.17✓
VZ	CS	55.10±5.63✓	52.34±3.29	51.71±5.65	46.25±4.90
WFC	FIN	53.85±3.81✓	49.21±7.16	47.65±7.75	49.68±5.53
WMT	CD	49.37±4.39	51.25±7.11	51.40±4.49✓	46.87±5.63
XOM	ENE	46.87±3.05	51.71±6.45✓	48.90±5.72	49.53±4.99
Avg. Rank		1.72✓	2.72	2.78	2.68

The output value of the sigmoid function was used for the input of the binary cross-entropy loss with the target value. Thus, trainable parameters were updated based on the binary cross-entropy loss.

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL SETTINGS

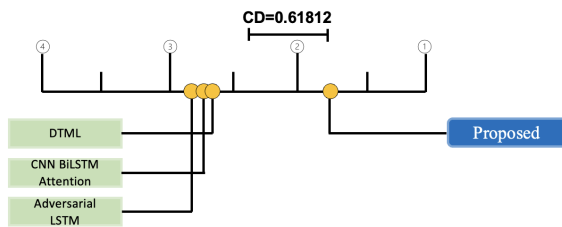
The KDD17 datasets [44] and macroeconomic indicators were used for stock direction classification. The

TABLE 5. Summary of the abbreviated sector names.

Abbreviated sector	Description
IT	Information Technology
CD	Consumer Defensive
FIN	Financial
MAT	Materials
CS	Communication Services
ENE	Energy
UTI	Utility
IND	Industrial
HEA	Healthcare
CC	Consumer Cyclical

TABLE 6. Summary of the friedman statistic F_F ($k = 4$, $N = 50$) and critical value in terms of accuracy measure between the proposed method and comparison models.

Evaluation Measure	F_F	Critical Value ($\alpha = 0.05$)
Accuracy	24.084	2.557

**FIGURE 6.** Result of the Bonferroni-Dunn Test between the Proposed and Comparison Models.

macroeconomic indicators were collected on Investing.com. The period of the historical price in KDD17 datasets and macroeconomic indicators was 2,518 days from January 1, 2007, to January 1, 2016. The KDD17 dataset has 50 stocks in US markets. This paper's macroeconomic indicators include the NASDAQ 100; the US 2-, 10-, and 30-year bond yield; the US Dollar Index; and the WTI oil price.

A blocked time-series cross-validation was used as a data split strategy to consider the trait of the time-series data [47]. Data leakage can occur if the random split strategy is used because the random split strategy in a time series can make the model predict the next day by examining a future value. In addition, a ten-fold split was used for the entire data length, setting the training, validation, and testing data ratio to 8:1:1. Fig. 5 presents the overview of the blocked time-series cross-validation.

The proposed model was compared to three comparison models via accuracy. The parameter was set for the three comparison models to the value recommended by each research. A brief review of the three comparison models follows.

- CNN-Attention Bi-LSTM [1]: The CNN-attention Bi-LSTM model combines the CNN and attention Bi-LSTM to predict stock price prediction. The CNN extracts the local perception and improves the efficiency of model training by reducing the number of parameters. The attention Bi-LSTM learns the sequential features

of the stock data. The authors used the stock dataset of Shanghai Composite Index.

- Adversarial LSTM [30]: The adversarial training method is a module that captures the stochastic property in stock data. The adversarial training method has been used for computer vision tasks at the data level. Still, this paper concatenates feature-level perturbations during training to express inherent stochastic properties in stock data to overcome the stochastic property in stock price data for stock price prediction.
- DTML [29]: The DTML model is combined with the attention LSTM and transformer model. The authors used the attention LSTM model to extract the sequential feature in the stock data and concatenated the model output. Furthermore, the transformer model is used to learn the correlation between the multiple stocks in the concatenated attention LSTM feature.

In this paper, the hyperparameters were set as follows. The batch size was set to 32. The learning rate was set to 0.001. The training progression used 100 epochs, and the model with the lowest validation loss between epochs was used for testing. The Adam optimizer was used for training [48]. The number of transformer model layers was 1. The transformer model dimensions were 320, and the number of transformer heads was set to eight. The loss function was set to binary cross-entropy loss, which is defined as follows:

$$BCE(x) = -\frac{1}{n} \sum_{i=1}^n y_i \log(h(x_i)) + (1 - y_i) \log(1 - h(x_i)) \quad (9)$$

where n , y_i , and $h(x_i)$ are the number of data, 0/1 binary target value for the prediction of the proposed model, and the final output of the model through the sigmoid function, respectively. In addition, Accuracy is employed as an evaluation measure:

$$Accuracy(\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (10)$$

where TP, TN, FP, and FN refer to true positive, true negative, false positive, and false negative. For each dataset, ten accuracy values can be obtained each representing the accuracy of the corresponding fold among ten splits. These values are averaged to represent the performance of the corresponding method. Based on the average accuracy of ten splits, the superiority among compared methods for each dataset can be determined and represented as the rank value. Lastly, the average rank can be obtained by averaging the rank value of each method over all the datasets.

Statistical tests were performed to validate the superiority of the proposed method for stock direction classification. The Friedman test is a statistical test for multiple nonparametric comparisons of the methods with multiple datasets [49]. Given k methods and N datasets, r_i^j denotes the rank of the j th method on the i th dataset (mean ranks are used in case of ties). The Friedman test compares the average ranks of the methods, and $R_j = \frac{1}{N} \sum_{i=1}^N r_i^j$ represents the average

TABLE 7. Comparison of the classification accuracy (%) between the proposed method and transformer model variants.

Ticker in Dataset	Sector	Proposed Method	Informer [53]	Autoformer [54]	DTML [29]
AAPL	IT	54.16±4.14✓	48.54±4.97	49.47±5.37	51.09±6.67
AMZN	CD	51.66±3.39✓	50.31±3.09	51.35±4.68	47.81±5.38
BA	IND	52.60±5.53✓	52.39±4.27	50.62±6.32	49.53±6.56
BAC	FIN	54.16±5.29✓	50.00±4.63	48.95±4.77	51.09±4.84
BHP	MAT	48.33±4.47	47.08±4.51	49.37±3.67✓	45.62±6.87
BRK-B	FIN	47.91±5.70	51.66±4.56✓	48.22±3.57	47.96±10.15
CHL	FIN	48.12±2.58	50.00±4.63	48.64±5.51	52.96±6.18✓
CMCSA	CS	52.39±3.05	53.54±1.81✓	50.31±4.32	53.12±8.69
CVX	ENE	50.83±3.32✓	50.72±5.94	50.10±4.59	47.50±5.89
D	UTI	51.66±2.91✓	51.35±5.01	48.33±4.59	49.84±6.79
DCM	IND	51.71±9.56	47.60±3.92	52.81±5.94✓	52.50±4.93
DIS	CS	53.12±4.16	53.64±5.87✓	49.06±5.48	48.12±5.07
DOW	-	52.08±5.59	47.50±4.84	54.06±4.78✓	51.09±3.28
DUK	UTI	52.39±3.89	52.70±3.33	53.54±3.81✓	49.53±5.23
EXC	ENE	50.83±4.92	51.14±5.00✓	50.20±6.58	47.50±6.78
GE	ENE	52.50±4.93	51.66±4.22	52.70±4.32✓	48.90±4.94
GOOGL	CS	51.66±3.52✓	49.89±4.12	50.00±2.94	48.28±7.07
HD	CC	52.18±5.15	53.42±4.29	50.52±3.39	53.43±9.10✓
INTC	IT	51.45±2.60	52.81±2.75✓	52.50±5.29	48.59±4.16
JNJ	HEA	51.14±4.71	50.83±4.28	52.18±5.21✓	52.03±7.81
JPM	FIN	52.50±5.75✓	51.04±6.16	51.35±4.61	52.18±5.72
KO	CD	51.04±3.86✓	50.93±4.25	49.68±5.31	50.78±7.97
MA	FIN	53.33±4.41	53.54±2.76✓	50.00±3.22	51.09±6.44
MMM	IND	52.08±3.89✓	50.83±2.82	51.66±5.39	49.06±7.75
MO	CD	54.37±3.86✓	53.33±4.48	50.00±3.22	48.90±4.37
MRK	HEA	51.66±4.44	54.16±5.49✓	51.14±3.96	49.37±4.37
MSFT	IT	51.25±3.51✓	48.02±5.48	48.54±3.23	50.78±5.55
NGG	UTI	54.27±4.25✓	50.52±3.39	50.83±4.94	52.65±10.00
NTT	CS	50.62±4.68	51.14±4.64✓	49.37±8.30	50.65±7.26
NVS	HEA	51.77±5.18	53.75±5.23✓	49.89±4.01	51.25±4.98
ORCL	IT	51.25±2.22✓	48.64±5.59	49.47±5.00	50.46±4.25
PEP	CD	51.56±5.12	52.29±4.46✓	51.25±5.42	48.12±4.67
PFE	HEA	50.83±2.86	51.97±3.82✓	51.66±3.49	47.65±4.80
PG	CD	51.25±4.76	53.12±3.72✓	49.06±3.53	52.18±6.33
PTR	ENE	51.87±5.58	50.31±5.23	52.50±6.00✓	51.56±8.64
RDS-B	ENE	52.70±4.29	53.33±4.00✓	52.08±4.00	48.75±4.57
RIO	MAT	50.31±4.32	47.39±3.40	46.97±5.02	52.03±4.79✓
SO	ENE	53.12±3.01✓	49.79±5.81	48.33±5.33	48.75±4.79
SPY	-	55.20±4.24✓	54.16±3.12	50.00±4.90	47.18±5.21
SYT	IT	48.54±5.19	50.52±6.83	47.91±5.43	51.40±5.56✓
T	CS	51.66±5.14✓	49.06±3.65	50.00±3.05	51.56±8.75
TM	CC	50.83±6.74	48.85±2.96	54.79±6.81✓	50.78±4.02
TOT	ENE	51.97±4.64	53.33±8.05✓	52.39±5.70	48.59±7.44
UNH	HEA	52.50±4.70✓	52.05±3.49	51.39±6.28	44.37±5.81
UPS	IND	51.97±3.20	52.29±4.79✓	50.20±1.38	52.18±6.01
VALE	MAT	50.20±4.51	49.37±6.99	49.58±6.37	52.03±6.17✓
VZ	CS	55.10±5.63✓	51.97±3.93	49.06±3.46	46.25±4.90
WFC	FIN	53.85±3.81✓	51.66±3.42	51.14±3.14	49.68±5.53
WMT	CD	49.37±4.39	49.89±4.30	50.00±3.69✓	46.87±5.63
XOM	ENE	46.87±3.05	53.22±4.44✓	48.22±5.93	49.53±4.99
Avg. Rank		1.96✓	2.3	2.74	2.98

rank on the j th method. The null-hypothesis states that all the methods are equivalent and so their ranks R_j should be equal. Furthermore, the Friedman test statistic is

$$F_F = \frac{(N-1)x_F^2}{N(k-1) - x_F^2} \quad (11)$$

where

$$x_F^2 = \frac{12N}{k(k+1)} \left[\sum_{j=1}^k R_j^2 - \frac{k(k+1)^2}{4} \right]. \quad (12)$$

Under the null hypothesis, the statistic F_F is distributed according to the F -distribution with $k-1$ and $(k-1)(N-1)$

TABLE 8. Summary of the Friedman statistic F_F ($k = 4$, $N = 50$) and Critical Value in terms of Accuracy measure between the proposed method and variants of the transformer model.

Evaluation Measure	F_F	Critical Value ($\alpha = 0.05$)
Accuracy	18.595	2.557

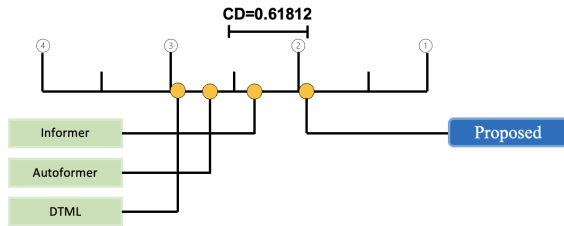


FIGURE 7. Result of the Bonferroni-Dunn Test between the Proposed method and Variants.

degrees of freedom [50]. The null hypothesis of the multiple comparisons is rejected when the F_F statistic is larger than the critical value under significance level α , or equivalently the corresponding p -value is smaller than α . In this case, the post-hoc test can proceed to make pairwise comparisons between the methods. The Bonferroni-Dunn test was selected as a post-hoc test because it is usually recommended after rejecting the Friedman test's null hypothesis [52]. The performance of the two methods is significantly different if the corresponding average ranks differ by at least the critical difference (CD):

$$CD = q_{\alpha/(k-1)} \sqrt{\frac{k(k+1)}{12N}}, \quad (13)$$

where $q_{\alpha/(k-1)}$ is the upper $\alpha/(k-1)$ -th percentile point of the studentized range distribution with (K, ∞) degrees of freedom [51].

B. COMPARISON RESULTS

Table 4 reveals the experimental results for the entire dataset in terms of accuracy. The table consists of the ticker, sector, and experimental results for the proposed and three comparison models. The ticker represents the name of a specific stock. The sector of the stock is a unit consisting of similar industries. The experimental results are expressed as the average accuracy and its standard deviation. Table 5 presents the abbreviation of each sector.

The proposed method, 27 out of 50 stocks, had the best accuracy compared to the comparison models. With the proposed method, the stock that yielded the best performance was Verizon Communications (VZ), and the difference in accuracy was 8.85%p compared to the comparison model, which can be regarded as a significant difference in the stock direction classification. The Friedman test with a 5% significance level was used to verify whether differences exist between groups. As shown in Table 6, the Friedman test statistic was 24.084, and the p -value was 2.3981e-05. There were statistically significant differences between groups under a significance level of 0.05.

As a result, the Bonferroni-Dunn test was applied with a 5% significance level to verify which groups differ. The

TABLE 9. Comparison of the classification accuracy between the early fusion method (proposed) and late fusion method.

Ticker	Sector	Early Fusion Method (Proposed)	Late Fusion Method
AAPL	IT	54.16% ± 4.14% ✓	51.87% ± 4.92%
AMZN	CD	51.66% ± 3.39% ✓	50.72% ± 3.33%
BA	IND	52.60% ± 5.53%	53.12% ± 2.27% ✓
BAC	FIN	54.16% ± 5.29%	54.27% ± 4.76% ✓
BHP	MAT	48.33% ± 4.47%	50.52% ± 4.19% ✓
BRK-B	FIN	47.91% ± 5.70%	51.25% ± 5.18% ✓
CHL	FIN	48.12% ± 2.58%	50.83% ± 5.16% ✓
CMCSA	CS	52.39% ± 3.05% ✓	51.45% ± 5.04%
CVX	ENE	50.83% ± 3.32% ✓	50.56% ± 5.52%
D	UTI	51.66% ± 2.91%	52.50% ± 3.64% ✓
DCM	IND	52.50% ± 4.93% ✓	51.66% ± 2.84%
DIS	CS	53.12% ± 4.16%	53.54% ± 3.20% ✓
DOW	-	52.08% ± 5.59% ✓	47.50% ± 5.08%
DUK	UTI	52.39% ± 3.89%	53.75% ± 5.14% ✓
EXC	ENE	50.83% ± 4.92% ✓	50.20% ± 4.16%
GE	ENE	52.50% ± 4.93% ✓	51.77% ± 3.75%
GOOGL	CS	51.66% ± 3.52% ✓	50.52% ± 6.80%
HD	CC	52.18% ± 5.15% ✓	51.56% ± 6.20%
INTC	IT	51.45% ± 2.60% ✓	49.68% ± 3.22%
JNJ	HEA	51.14% ± 4.71%	51.87% ± 4.60% ✓
JPM	FIN	52.50% ± 5.75% ✓	50.41% ± 5.65%
KO	CD	51.04% ± 3.86%	51.25% ± 5.88% ✓
MA	FIN	53.33% ± 4.41% ✓	53.24% ± 4.38%
MMM	IND	52.08% ± 3.89%	52.50% ± 5.49% ✓
MO	CD	54.37% ± 3.86%	55.72% ± 3.84% ✓
MRK	HEA	51.66% ± 4.44% ✓	51.56% ± 5.08%
MSFT	IT	51.25% ± 3.51% ✓	50.10% ± 6.34%
NGG	UTI	54.27% ± 4.25% ✓	52.18% ± 2.48%
NTT	CS	50.62% ± 4.68%	51.56% ± 4.95% ✓
NVS	HEA	51.77% ± 5.18%	52.29% ± 5.81% ✓
ORCL	IT	51.25% ± 2.22% ✓	50.41% ± 3.98%
PEP	CD	51.56% ± 5.12%	52.08% ± 5.18% ✓
PFE	HEA	50.83% ± 2.86%	51.45% ± 4.91% ✓
PG	CD	51.25% ± 4.76% ✓	49.58% ± 4.37%
PTR	ENE	51.87% ± 5.58%	53.64% ± 6.81% ✓
RDS-B	ENE	52.70% ± 4.29%	54.89% ± 5.57% ✓
RIO	MAT	50.31% ± 4.32% ✓	50.00% ± 3.32%
SO	ENE	53.12% ± 3.01% ✓	51.45% ± 3.58%
SPY	ETF	55.20% ± 4.24% ✓	54.58% ± 3.58%
SYT	IT	48.54% ± 5.19%	49.47% ± 5.69% ✓
T	CS	51.66% ± 5.14% ✓	47.91% ± 4.81%
TM	CC	50.83% ± 6.74% ✓	49.79% ± 4.94%
TOT	ENE	51.97% ± 4.64% ✓	51.77% ± 3.98%
UNH	HEA	52.50% ± 4.70% ✓	52.18% ± 4.62%
UPS	IND	51.97% ± 3.20% ✓	49.79% ± 3.63%
VALE	MAT	50.20% ± 4.51% ✓	47.70% ± 6.19%
VZ	CS	55.10% ± 5.63% ✓	50.62% ± 1.76%
WFC	FIN	53.85% ± 3.81% ✓	50.62% ± 5.21%
WMT	CD	49.37% ± 4.39%	51.04% ± 4.21% ✓
XOM	ENE	46.87% ± 3.05% ✓	46.27% ± 5.27%
Avg. Rank		1.42 ✓	1.52

CD value was 0.6181 under a significance level of 0.05 in the Bonferroni-Dunn test. The differences in the average rank between the proposed and comparison models were greater than the CD value. Fig. 6 presents the results of the Bonferroni-Dunn test. In the figure, the three comparison models are farther to the left than the $CD = 0.61812$ compared to the proposed model. The proposed method achieved the best average rank and outperformed all comparison methods. The quantitative analysis confirmed that the proposed method has a numerical difference compared to the comparison model. The statistical tests indicate that the proposed model had a statistically significant difference compared to the competitive models.

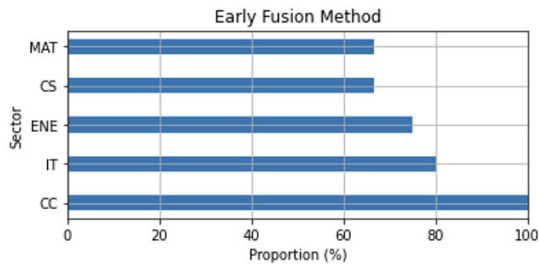


FIGURE 8. Sectors in which the early fusion method performs better than the late fusion method and the proportion of how well the sectors perform compared to the late fusion method.

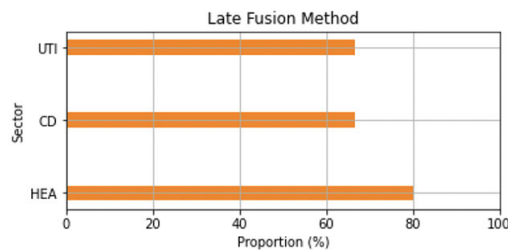


FIGURE 9. Sectors in which the late fusion method performs better than the early fusion method and the proportion of how well the sectors perform compared to the early fusion method.

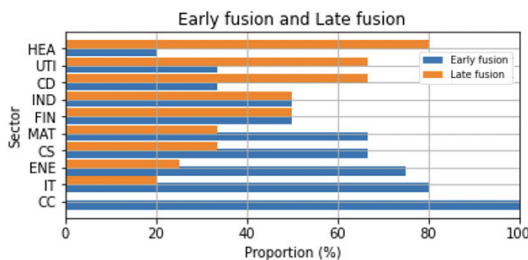


FIGURE 10. Sectors in which early and late fusion methods perform respectively better than each other, and the proportion of how well the sectors perform compared to each other.

Additional comparison experiments were conducted on transformer variant models to determine whether the result of the proposed model is the effect of the transformer or modality fusion. Table 7 presents the experimental results between the proposed and transformer variant models in terms of accuracy. We included DTML [29] in this experiment again because it is based on the transformer architecture. The proposed method gained the best average rank of 1.96 compared to the transformer variant models. The Friedman and Bonferroni-Dunn tests were performed with a significance level of 0.05 to determine whether a statistical difference exists. Table 8 lists the results of the Friedman test, and Fig. 7 displays the results of the Bonferroni-Dunn test. The Friedman test statistic was 18.595, and the p -value was 0.00033. There were statistically significant differences between the proposed model and comparison models. The CD value was 0.6181 with a significance level of 0.05 in the Bonferroni-Dunn test. When comparing the difference in average rank between the proposed and comparison models, among the comparison models, the mean rank difference between the proposed model and the informer model was smaller than that of the CD. The statistical tests found that

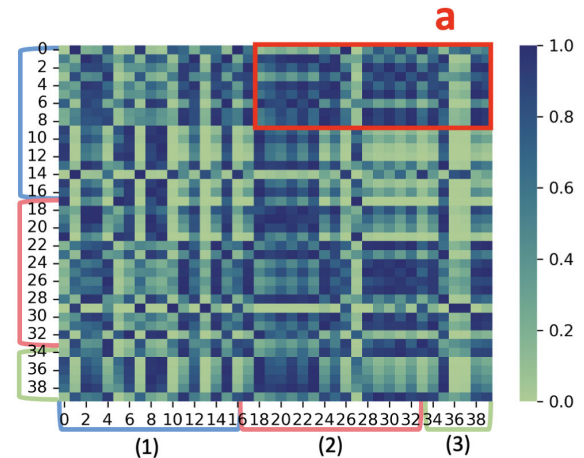


FIGURE 11. Attention map for the early fusion method: (1) stock price modality, (2) months and day-of-the-week modality, and (3) macroeconomic indicator modality.

the proposed model had a statistically significant difference from the comparison models except for the informer model.

C. IN-DEPTH ANALYSIS

In this study, we devised a multimodal transformer based on early fusion strategy. To verify the effectiveness of our choice, the early and late fusion models were compared in terms of accuracy to demonstrate the superiority of early fusion strategy compared to the late fusion model. Table 9 shows the comparison results in terms of accuracy. The proposed early fusion model achieved better classification accuracy in 30 out of 50 stocks compared to the late fusion model. Although the early fusion method achieved the best performance in 3 out of 6 stocks in the financial sector, it acquired the best performance in 6 out of 8 stocks compared to the late fusion method in the energy sector in classification accuracy. Fig. 8 presents the sectors where early fusion methods perform better than late fusion methods, the proportion of the number of stocks, and how well the stocks in each sector perform.

Fig. 9 lists the sectors where late fusion methods perform better than early fusion methods, the proportion of the number of stocks, and how well the stocks in each sector perform. Fig. 10 displays the overall results of the sectors in which early and late fusion methods respectively perform better than each other, the proportion of the number of stocks, and how well the stocks in each sector perform comparatively. The early fusion method achieved higher classification accuracy than late fusion in five sectors: materials, communication services, energy, information technology, and cyclical consumer sectors. The late fusion method achieved higher classification accuracy than the early fusion method in three sectors: utility, consumer defensive, and healthcare. The sectors with the same ratio of stocks with high classification accuracy in early and late fusion were the industrial and financial sectors.

The attention map was also analyzed for scaled dot-product attention at the inference level. The corporation for the attention map in Figs. 11 and 12 is Apple Inc., which achieved high classification accuracy compared to the comparison models.

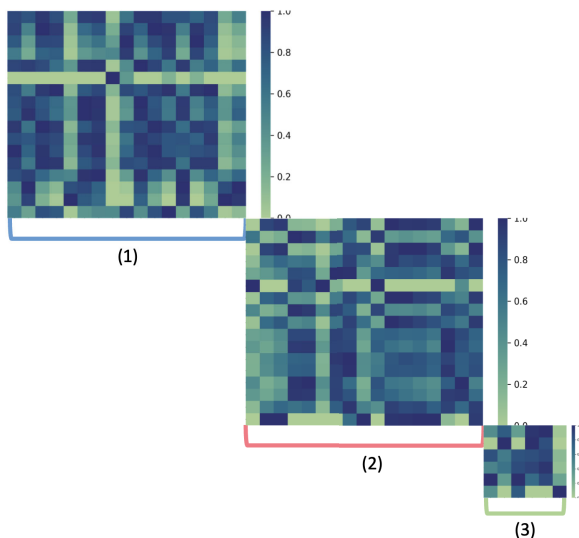


FIGURE 12. Attention map for the late fusion method: (1) stock price modality, (2) months and day-of-the-week modality, and (3) macroeconomic indicator modality.

If the color in the attention map is white, the corresponding features are highly correlated. If the color in the attention map is black, the corresponding features are weakly correlated. Fig. 11 depicts the attention map for early fusion, and Fig. 12 depicts the attention map for late fusion. Part “a” in Fig. 11 represents that the 11 features in stock price modality are highly correlated with the months and day-of-the-week modality and the macroeconomic indicator modality. The attention map for early fusion indicates that each modality (1), (2), and (3) is correlated because the value of each section in the attention map is highly activated. Accordingly, in the early fusion method, each modality learns by referring to the other. However, in late fusion, the attention map is highly activated apart from each modality. The features in the late fusion method are separated and passed through the scaled dot-product attention. Consequently, the early fusion method learns the intermodality correlation better than the late fusion method because all the modalities are processed through the scaled dot-product attention together, which learns the relationship between data in the early fusion method.

V. CONCLUSION

The task of stock price prediction can be challenging if the algorithm depends only on the stock data because the stock price can be affected by external factors, such as the world economy and policy. In this study, in addition to the stock data, two modalities such as macroeconomic indicators and month-and-day-of-the-week provide a global view of the stock market and seasonal/weekly information, respectively. A multimodal fusion architecture was considered when designing the proposed method, which learns the intermodality correlation for features.

The proposed model in this paper outperformed the comparison models and displayed a statistically significant difference. The test results confirmed that the proposed model

exhibited a statistically significant difference compared to the comparison models. An in-depth analysis of the comparison results was also performed between early and late fusion methods. The analysis indicates that the early fusion method achieved high classification accuracy in 30 of 50 datasets. Our analysis showed that each strategy demonstrates different strengths for given sectors.

Potential directions of future work include additional data exploratory research in financial data mining. Moreover, the data fusion methodology considering the traits between various data types is also crucial in multimodal stock direction classification.

REFERENCES

- [1] W. Lu, J. Li, J. Wang, and L. Qin, “A CNN-BiLSTM-AM method for stock price prediction,” *Neural Comput. Appl.*, vol. 33, no. 10, pp. 4741–4753, May 2021.
- [2] P. H. Cootner, “Stock prices: Random vs. systematic changes,” *Ind. Manage. Rev.*, vol. 3, no. 2, p. 24, 1962.
- [3] S.-W. Chen, “Non-stationarity and non-linearity in stock prices: Evidence from the OECD countries,” *Econ. Bull.*, vol. 3, no. 11, pp. 1–11, 2008.
- [4] R. J. Rogalski, “New findings regarding day-of-the-week returns over trading and non-trading periods: A note,” *J. Finance*, vol. 39, no. 5, pp. 1603–1614, Dec. 1984.
- [5] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, “Stock price prediction using the ARIMA model,” in *Proc. UKSim-AMSS 16th Int. Conf. Comput. Modeling Simulation*, Mar. 2014, pp. 106–112.
- [6] K. J. Kim, “Financial time series forecasting using support vector machines,” *Neurocomputing*, vol. 55, nos. 1–2, pp. 307–319, Sep. 2003.
- [7] M. R. Hassan and B. Nath, “Stock market forecasting using hidden Markov model: A new approach,” in *Proc. 5th Int. Conf. Intell. Syst. Design Appl. (ISDA)*, Sep. 2005, pp. 192–196.
- [8] L. Khaideem, S. Saha, and S. R. Dey, “Predicting the direction of stock market prices using random forest,” 2016, *arXiv:1605.00003*.
- [9] F. Agostinelli, M. Hoffman, P. Sadowski, and P. Baldi, “Learning activation functions to improve deep neural networks,” 2014, *arXiv:1412.6830*.
- [10] Z. Hu, Y. Zhao, and M. Khushi, “A survey of forex and stock price prediction using deep learning,” *Appl. Syst. Innov.*, vol. 4, no. 1, p. 9, Feb. 2021.
- [11] J. Gao, P. Li, Z. Chen, and J. Zhang, “A survey on deep learning for multimodal data fusion,” *Neural Comput.*, vol. 32, no. 5, pp. 829–864, May 2020.
- [12] A. Kyereboah-Coleman and K. F. Agyire-Tettey, “Impact of macroeconomic indicators on stock market performance: The case of the Ghana stock exchange,” *J. Risk Finance*, vol. 9, no. 4, pp. 365–378, Aug. 2008.
- [13] A. Thakkar and K. Chaudhari, “Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions,” *Inf. Fusion*, vol. 65, pp. 95–107, Jan. 2021.
- [14] X. Wei, T. Zhang, Y. Li, Y. Zhang, and F. Wu, “Multi-modality cross attention network for image and sentence matching,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10941–10950.
- [15] P. Chen, T. Pedersen, B. Bak-Jensen, and Z. Chen, “ARIMA-based time series model of stochastic wind power generation,” *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 667–676, May 2010.
- [16] M. Van Der Voort, M. Dougherty, and S. Watson, “Combining Kohonen maps with ARIMA time series models to forecast traffic flow,” *Transp. Res. C, Emerg. Technol.*, vol. 4, no. 5, pp. 307–318, 1996.
- [17] P. Han, P. X. Wang, S. Y. Zhang, and D. H. Zhu, “Drought forecasting based on the remote sensing data using ARIMA models,” *Math. Comput. Model.*, vol. 51, nos. 11–12, pp. 1398–1403, 2010.
- [18] B. K. Nelson, “Time series analysis using autoregressive integrated moving average (ARIMA) models,” *Academic Emergency Med.*, vol. 5, no. 7, pp. 739–744, Jul. 1998.
- [19] B. M. Henrique, V. A. Sobreiro, and H. Kimura, “Stock price prediction using support vector regression on daily and up to the minute prices,” *J. Finance Data Sci.*, vol. 4, no. 3, pp. 183–201, 2018.
- [20] P. Somani, S. Talele, and S. Sawant, “Stock market prediction using hidden Markov model,” in *Proc. IEEE 7th Joint Int. Inf. Technol. Artif. Intell. Conf.*, Dec. 2014, pp. 89–92.

- [21] P. Yu and X. Yan, "Stock price prediction based on deep neural networks," *Neural Comput. Appl.*, vol. 32, no. 6, pp. 1609–1628, Mar. 2020.
- [22] S. Mehtab and J. Sen, "Stock price prediction using convolutional neural networks on a multivariate timeseries," 2020, *arXiv:2001.09769*.
- [23] M. A. I. Sunny, M. M. S. Maswood, and A. G. Alharbi, "Deep learning-based stock price prediction using LSTM and bi-directional LSTM model," in *Proc. 2nd Novel Intell. Lead. Emerg. Sci. Conf. (NILES)*, Oct. 2020, pp. 87–92.
- [24] A. Tsantekidis, N. Passalis, A. Tefas, J. Kannianen, M. Gabbouj, and A. Iosifidis, "Using deep learning for price prediction by exploiting stationary limit order book features," *Appl. Soft Comput.*, vol. 93, Aug. 2020, Art. no. 106401.
- [25] 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. Syst.*, vol. 30, 2017, pp. 1–11.
- [26] K. S. Kalyan, A. Rajasekharan, and S. Sangeetha, "AMMUS : A survey of transformer-based pretrained models in natural language processing," 2021, *arXiv:2108.05542*.
- [27] F. Dama and C. Sinoquet, "Time series analysis and modeling to forecast: A survey," 2021, *arXiv:2104.00164*.
- [28] C. Wang, Y. Chen, S. Zhang, and Q. Zhang, "Stock market index prediction using deep transformer model," *Expert Syst. Appl.*, vol. 208, Dec. 2022, Art. no. 118128.
- [29] 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*, Aug. 2021, pp. 2037–2045.
- [30] F. Feng, H. Chen, X. He, J. Ding, M. Sun, and T.-S. Chua, "Enhancing stock movement prediction with adversarial training," 2018, *arXiv:1810.09936*.
- [31] A. Thakkar and K. Chaudhari, "Information fusion-based genetic algorithm with long short-term memory for stock price and trend prediction," *Appl. Soft Comput.*, vol. 128, Oct. 2022, Art. no. 109428.
- [32] S. Lahmiri, "A technical analysis information fusion approach for stock price analysis and modeling," *Fluctuation Noise Lett.*, vol. 17, no. 1, Mar. 2018, Art. no. 1850007.
- [33] S. I. Lee and S. J. Yoo, "Multimodal deep learning for finance: Integrating and forecasting international stock markets," *J. Supercomput.*, vol. 76, no. 10, pp. 8294–8312, Oct. 2020.
- [34] N. Jing, Z. Wu, and H. Wang, "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction," *Expert Syst. Appl.*, vol. 178, Sep. 2021, Art. no. 115019.
- [35] X. Zhang, Y. Li, S. Wang, B. Fang, and P. S. Yu, "Enhancing stock market prediction with extended coupled hidden Markov model over multi-sourced data," *Knowl. Inf. Syst.*, vol. 61, no. 2, pp. 1071–1090, Nov. 2019.
- [36] 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.
- [37] S. Kogan, D. Levin, B. R. Routledge, J. S. Sagi, and N. A. Smith, "Predicting risk from financial reports with regression," in *Proc. Hum. Lang. Technol., Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics (NAACL)*, 2009, pp. 272–280.
- [38] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news: The AZFin text system," *ACM Trans. Inf. Syst.*, vol. 27, no. 2, pp. 1–19, 2009.
- [39] A. M. Adam and G. Tweneboah, "Macroeconomic factors and stock market movement: Evidence from Ghana," *Social Sci. Res. Netw.*, vol. 2008, no. 1289842, pp. 1–26, 2008.
- [40] F. J. Cebrian and L. Negrut, "US stock market and macroeconomic factors," *J. Appl. Bus. Res.*, vol. 32, no. 1, pp. 325–340, 2016.
- [41] H. Gursida, "The influence of fundamental and macroeconomic analysis on stock price," *Jurnal Terapan Manajemen dan Bisnis*, vol. 3, no. 2, pp. 222–234, 2017.
- [42] S. Ravikumar and P. Saraf, "Prediction of stock prices using machine learning (regression, classification) algorithms," in *Proc. Int. Conf. Emerg. Technol. (INCET)*, Jun. 2020, pp. 1–5.
- [43] T. T. Khuat and M. H. Le, "An application of artificial neural networks and fuzzy logic on the stock price prediction problem," *JOIV: Int. J. Informat. Visualizat.*, vol. 1, no. 2, pp. 40–49, 2017.
- [44] 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*, Aug. 2017, pp. 2141–2149.
- [45] Y. Kara, M. A. Boyacioglu, and Ö. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul stock exchange," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5311–5319, May 2011.
- [46] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, Sep. 2021.
- [47] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS ONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0180944.
- [48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [49] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Dec. 2006.
- [50] R. L. Iman and J. M. Davenport, "Approximations of the critical region of the Friedman statistic," *Commun. Statist.-Theory Methods*, vol. 9, no. 6, pp. 571–595, 1980.
- [51] P. B. Nemenyi, "Distribution-free multiple comparisons," Ph.D. thesis, Dept. Math., Princeton Univ., Princeton, NJ, USA, 1963.
- [52] B. Trawiński, M. Smętek, Z. Telec, and T. Lasota, "Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms," *Int. J. Appl. Math. Comput. Sci.*, vol. 22, no. 4, pp. 867–881, Dec. 2012.
- [53] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 12, 2021, pp. 11106–11115.
- [54] J. Xu, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 22419–22430.



TAE-WON LEE received the B.Sc. degree in statistics from Inje University, in 2021. He is currently pursuing the M.Sc. degree with the Department of Artificial Intelligence, Chung-Ang University. His research interest includes multimodal learning methods for time series data.



PAWEŁ TEISSEYRE received the Ph.D. degree from the Institute of Computer Science, Polish Academy of Sciences, in 2013. He works as an Assistant Professor with the Institute of Computer Science, Polish Academy of Sciences and on the Faculty of Mathematics and Information Sciences, Warsaw University of Technology. His research interests include feature selection in high-dimensional supervised problems, multi-label classification, learning from partially labeled data, and applications of machine learning methods in medicine and genetics.



JAESUNG LEE received the B.S., M.S., and Ph.D. degrees in computer science from Chung-Ang University, Seoul, South Korea, in 2007, 2009, and 2013, respectively. He also studies classification, feature selection, especially multilabel learning with information theory. He is currently the Head and an Associate Professor with the Department of Artificial Intelligence, Chung-Ang University, where he is the Chief of the AI/ML Innovation Research Center. His research interests include

machine learning, multilabel learning, model selection, and neural architecture search.

• • •