제 138 회 석사학위 논문
지 도 교 수 이 재 성

멀티모달 트랜스포머 기반 주식 방향 예측 연구

A study on stock direction prediction based on multimodal transformer

중앙대학교 대학원
AI학과 AI 응용 전공
이 태 원

2023년 2월

멀티모달 트랜스포머 기반 주식 방향 예측 연구

A study on stock direction prediction based on multimodal transformer

이 논문을 석사학위논문으로 제출함

2023년 2월

중앙대학교 대학원
AI학과 AI 응용 전공
이 태 원

이 태 원의 석사학위 논문으로 인정함

심 사 위 원 장 ________________

심 사 위 원 ________________

심 사 위 원 ________________

중앙대학교 대학원

2023년 2월

Contents

1 Introduction **1**

2 Related Work **5**

2.1 Stock price prediction 6

2.1.1 Statistical stock price prediction 6

2.1.2 Machine learning stock price prediction 6

2.1.3 Deep learning stock price prediction 7

2.1.4 Deep learning data fusion stock price prediction 8

3 Multimodal Transformer for stock price prediction **11**

3.1 materials and method 12

3.1.1 Motivation 12

3.1.2 Proposed Method 14

4 Experiments **25**

4.1 Experimental settings 26

4.2 Comparison results 31

4.3 In-depth Analysis 34

5 Results **45**

Bibliography **48**

국 문 초 록 **57**

ABSTRACT 58

List of Figures

3.1 The overview of the proposed method. The proposed method comprises the feature concatenation, transformer encoder, and classification layer. 20

3.2 The process of feature concatenation and multi-head attention in the proposed method. 23

3.3 Scaled Dot-Product Attention. 23

3.4 The process of point wise feed forward network and classifier in the proposed method. 24

4.1 Blocked time series Cross-Validation. 26

4.2 Result of Bonferroni-Dunn test between the proposed method and comparison models. 31

4.3 Result of Bonferroni-Dunn test between the proposed method and variants of transformer model. 31

4.4 The sectors in which early fusion methods perform better than the late fusion method, and the proportion of how the sectors perform well compared to the late fusion method 35

4.5 The sectors in which late fusion methods perform better than early fusion methods, and the proportion of how the sectors perform well in late fusion method 35

4.6 The sectors in which early fusion and late fusion methods perform respectively better than each other method, and the proportion of how the sectors perform well compared to each other method . . . 36
4.7 The attention map in the ealry fusion method. (1) is stock price modality. (2) is the months and days of the week modality. (3) is the macroeconomic indicator modality. . . . 37
4.8 The attention map in the late fusion method. (1) is stock price modality. (2) is the months and days of the week modality. (3) is the macroeconomic indicator modality. . . . 38

List of Tables

3.1 Stock price modality 21
3.2 Months and day of the week modality 22
3.3 Macroeconomic indicator modality 22
4.1 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 30
4.2 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 31
4.3 Comparison results of the classification accuracy between the proposed method and comparison models 41
4.4 Comparison results of the classification accuracy between the proposed method and the variants of the transformer model . . 42
4.5 The summary of the abbreviated sector 43
4.6 Ablation study about modality in terms of accuracy. 43
4.7 Comparison results of the classification accuracy between early fusion method and late fusion method 44

Chapter 1

Introduction

As the world economy grows, 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 [Lu et al., 2021]. 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 level of noise of the stock price data [Cootner, 1962, Chen et al., 2008]. Thus, using only stock price data may not be suitable for stock price prediction because the stock market is affected by external factors sensitively, such as the world economy, domestic politics, accidental events, and even by months or days of the week [Rogalski, 1984].

Researchers use statistical and machine learning models for modeling the stock price. The ARIMA model, which combines the autoregressive model, moving average model, and differencing, is one of the most representative statistical models to analyze the time series data [Ariyo et al., 2014]. Traditional machine learning models, such as the Support Vector Machine and Hidden Markov Model (HMM), also have been applied in this field [Kim, 2003, Hassan and Nath, 2005, Khaidem et al., 2016]. Deep neural networks (DNNs) recently demonstrated an excellent prediction performance [Agostinelli et al., 2014, Hu et al., 2021] owing to their information fusion capability that helps capture the non-linear relationships among various financial data sources [Gao et al., 2020].

In addition to the stock price data, there is another information source, called the economic indicators, that can represent the economic situations of the market [Kyereboah-Coleman and Agyire-Tettey, 2008]. Thus, by exploiting the economic indicators for predicting the stock price, an improvement in accuracy can be expected. In addition, because the stock price can be correlated to months and days of the week, the accuracy of the stock price prediction can be enhanced [Rogalski, 1984]. In addition, the internal structure of information fusion should be carefully designed to achieve the best prediction power [Thakkar and Chaudhari, 2021]. However, conventional studies suffer from performance degradation because of the limited number of information sources considered and the structure design for the information fusion conducted intrinsically.

The information fusion strategy in DNN can be roughly divided into two folds according to the location where the fusion is started in the network: early fusion, which specializes in capturing inter-modality correlation, and late fusion, which specializes in capturing intra-modality correlation [Wei et al., 2020]. Because these two fusion strategies have different strengths in information processing, it is still in the veil which approaches 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 for achieving accurate stock price prediction. As the fusion method, feature concatenation, one of the methods in multimodal early fusion, is used to fuse the information between each modality effectively.
- Twenty-six information sources from three domains, such as stock price, months and date of the week, and the macroeconomic indicator modalities, are considered in this study. The greatest number of information sources are considered concurrently.
- An in-depth analysis is conducted regarding the fusion strategy. This paper's proposed method indicated that the early fusion strategy gives the best prediction performance. However, there are a group of stocks with the late fusion strategy that offers a better prediction performance.

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

Chapter 2

Related Work

2.1 Stock price prediction

2.1.1 Statistical stock price prediction

In the work of [Bartholomew, 1971], authors proposed the ARIMA model that is combined with the Auto Regressive model, Moving Average model, and Differencing. The ARIMA model is used to predict the time series data, such as power generation, traffic flow, and sensor data [Chen et al., 2009,Van Der Voort et al., 1996, Han et al., 2010]. Another researcher used the ARIMA model to predict financial time series [Ariyo et al., 2014]. The statistical time series model, like ARIMA, assumes that the data have stationary and linearity [Nelson, 1998]. It cannot be easy to apply such assumptions throughout financial time series data analysis because the data has traits of non-linearity and randomness [Cootner, 1962, Chen et al., 2008].

2.1.2 Machine learning stock price prediction

Despite the excellent performance of statistical stock price prediction using, for example, the ARIMA model, there is a series of attempts to exploit the machine learning models for stock price prediction. For instance, in the work of [Henrique et al., 2018], the authors used support vector regression to predict stock minute prices with technical indicators. In addition to the support vector regression, the random forest was also used for building the stock price prediction model [Khaidem et al., 2016]. The continuous Hidden Markov Model (cHMM), one of the most popular modeling techniques for time series data analysis, was used in the work of [Somani et al., 2014]. In this work, the emission probabilities are modeled as Gaussian mixture models.

2.1.3 Deep learning stock price prediction

The DNNs can be effective alternatives to conventional econometric and statistical models with weaknesses in modeling financial time series data with non-linear traits [Yu and Yan, 2020]. For example, a convolutional neural network (CNN) used for image data analysis can be applied for stock price prediction [Mehtab and Sen, 2020]. In addition, Long Short-Term Memory (LSTM), Bi-directional LSTM, and attention mechanism are the models that 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 the work of [Sunny et al., 2020]. Not stock prices; limit order book data can be used to predict stock prices. The model combined with CNN and LSTM is utilized to extract the feature of limit order book and stock price data. The CNN module of the model extracts the features in the limit order book, and the LSTM module of the model extracts the time series features in stock price data [Tsantekidis et al., 2020]. The transformer model was devised to improve the attention encoder-decoder LSTM model [Vaswani et al., 2017], which its variant is widely used in the field of sequence modeling [Kalyan et al., 2021, Dama and Sinoquet, 2021]. A stock price prediction model combined with CNN and attention Bi-directional LSTM was also considered [Lu et al., 2021]. Recently, a transformer model for stock market index prediction was suggested in [Wang et al., 2022]. The attention mechanism in the transformer can learn the correlation of many stocks dynamically with the market index and helps each other to predict each stock [Yoo et al., 2021]. The authors in the work of [Feng et al., 2018] insisted that the general machine training ignores the stochastic property of stock that changes over time. As a result, they proposed adversarial training to learn the stochastic property in stock price using attention LSTM.

2.1.4 Deep learning data fusion stock price prediction

Several recent financial studies combined various data with stock price data [Thakkar and Chaudhari, 2021]. The information fusion can be enhanced by conducting parallel operations that contain inter-intra crossover and adaptive mutation in the genetic algorithm (GA) [Thakkar and Chaudhari, 2022]. The proposed information fusion-based GA approach can optimize the parameters of an extended short-term memory stock price prediction model and selects a set of features. The authors in the work of [Lahmiri, 2018] addressed the problem of technical analysis information fusion in improving stock market index-level prediction. Multiple predictive systems based on different categories of technical analysis metrics are devised. The system has multiple prediction models with different technical analysis metrics data categories. Each prediction model is based on an ensemble of neural networks with particle swarm optimization. The authors in the work of [Lee and Yoo, 2020] proposed the multi-modal deep learning architecture where the architecture learns the cross-correlation of the United States and Korean stock prices.

With recent progress in natural language processing research, text data such as news and Twitter data is used for more accurate stock price prediction [Jing et al., 2021]. An event-driven approach was used to predict stock price prediction with stock price data in [Zhang et al., 2019]. The event-driven method was elaborated by extracting the stock-related event from the news titles. Moreover, the method enhancing the joint impacts was also devised by calculating the two stocks' similarity 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 to predict stock market prediction [Zhang et al., 2018]. The stock quantitative feature matrix and stock correlation matrix were made using the stock's quantitative and social media data. The stock movement tensor was built by events and sentiment extraction from news articles and social media. The stock quantitative feature matrix and the stock correlation matrix are factorized, and the stock movement tensor was decomposed to predict stock price prediction. The bag-of-words and named entity approach using a large corpus of freely available financial reports are used to predict the volatility of the stock returns and stock market prediction with support vector regression [Kogan et al., 2009, Schumaker and Chen, 2009]. These data fusion studies commonly reported that combining various data sources could improve prediction accuracy.

Chapter 3

Multimodal Transformer for stock price prediction

3.1 materials and method

3.1.1 Motivation

First, data that could be combined had to be considered. A lot of papers using natural language processing in stock price prediction put forth their efforts to find a good vector representation that represents the text data well [Zhang et al., 2019, Zhang et al., 2018, Kogan et al., 2009, Schumaker and Chen, 2009]. However, problems can arise in the natural language processing approach in stock price prediction. In the application section, using news data can cause legal issues such as copyright. Therefore, macroeconomic indicators and months and the day of the week were considered alternatives to text data. The macroeconomic indicator is a representative factor that affects the stock market [Adam and Tweneboah, 2008, Jareño Cebrián and Negrut, 2016]. Some fundamental macroeconomic indicators such as exchange rate, interest rate, industrial production, and inflation have associations with the stock price [Adam and Tweneboah, 2008]. In addition to the macroeconomic indicators, the months and the day of the week effect can also be considered to get information affecting the stock market [Rogalski, 1984]. However, stock prices may only partially reflect the macroeconomic indicator's information and the effects of the months and the day of the week. Therefore, it needs to let the model know the information about macroeconomic indicators and the months and the day of the week to instantly reflect the macroeconomic indicators and the months and the day of the week information. So, data fusion can be a valid method for accurate stock price prediction since the stock markets are affected by many factors [Gursida, 2017]. An accurate stock direction prediction is a challenging problem because of the randomness in the stock price data [Cootner, 1962].

Multimodal deep learning that fuses various information sources internally can be considered to solve the randomness problem in data. Late fusion and early fusion are methods to combine plural data in multimodal deep learning. The early fusion method extracts inter-modality correlation more efficiently than the intra-modality correlation [Lee and Yoo, 2020]. In this paper, the model that learns the correlation well between significant data affecting the stock price is designed to solve randomness problems in the stock price data itself. The transformer encoder is a model advantageous for extracting the information in sequential data [Vaswani et al., 2017]. 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 in the query, key, and value. The multimodal early fusion transformer method is used in this paper to utilize the characteristic of the transformer encoder that learns the correlation between model input features.

In this study, the problem of the stock direction prediction is considered that predicts whether the stock prices will 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 [Ravikumar and Saraf, 2020, Khuat and Le, 2017]. Specifically, the proposed method predicts the next day's stock up and down direction using the previous ten days' data as a binary classification task. Thus, the stock direction prediction in this study can be formulated as the binary classification problem where the prediction performance can be evaluated using conventional accuracy metrics. As a result, an early fusion neural network is designed with feature concatenation, transformer encoder, and classifier layer. Figure 3.1 shows the overview of the proposed method.

3.1.2 Proposed Method

The entire experiment is conducted by setting three modalities: stock price modality, months and day of the week modality, and macroeconomic indicator modality. Table 3.1 shows the details of the stock price modality. The KDD17 public dataset is used to add to the reliability of this experiment [Zhang et al., 2017]. The features of the KDD17 dataset are open, close, high, low, volume, and adj close. Moreover, the ten technical indicators are added, and two additional features in the stock price modality are extracted through feature engineering to extract significant features in stock price data [Kara et al., 2011]. Table 3.2 shows the details of the months and day of the week modality. In this modality, the months and day of the week features are extracted using one-hot encoding because months 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 $[0, 1, \cdots, 0]$. Lastly, in the macroeconomic indicator modality described in Table 3.3, six features are selected, such as NASDAQ100, US 2-year, 10-year, 30-year Bond Yield, US Dollar Index, and WTI Oil Price. The close column in macroeconomic indicators is only used in the data even if the macroeconomic indicators had plural columns such as open, high, low, close, and volume. Before conducting the training and test process, the min-max normalization is applied to each ten days sequence respectively because financial data varies the scale over a long-time period.

Figure 3.2 shows the detailed structure of the former half of the proposed method, including feature concatenation and multi-head attention model. Each modality is concatenated by feature dimension to fuse data using the early fusion method. Let F_1, F_2 and F_3 be 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, the months and day of the week modality, and the macroeconomic indicator modality where l be the sequence length, respectively. These multiple information sources can be concatenated alongside the feature dimension to create input modality matrix $M \in \mathbb{R}^{l\times(F_1+F_2+F_3)}$, represented as

$$M = [x_1; x_2; x_3]. \tag{3.1}$$

In this paper, F_1, F_2, and F_3 are 17, 17, and 6, respectively, because the number of each feature is 17, 17, and 6. Which day affects the next day's stock price prediction can be different. The positional encoding module is used to let the model know the position information of the data. The sine and cosine functions are used to express different frequencies [Vaswani et al., 2017]. pos is the position, and i is the dimension. d_{model} is the feature dimension of modality data concatenated by feature dimension. The M concatenated three modalities are added to the positional encoding.

$$PE_{(pos,2i)} = sin(pos/10000^{2i/d_{model}}) \tag{3.2}$$

and

$$PE_{(pos,2i+1)} = cos(pos/10000^{2i/d_{model}}) \tag{3.3}$$

Scaled dot-product attention comprises query, key of dimension d_k, and value of dimension d_v, its dot-product [Niu et al., 2021]. The query, key, and value are calculated by multiplying each other linear layer of dimension d_v to the output of the positional encoding. Then, the query and key are multiplied using the dot-product operation. After the dot-product, the multiplied output of the dot-product is divided by $\sqrt{d_k}$ and multiplied by the value. Scaled dot-product self-attention represents which parts of the data are and how relevant they are to each other. Figure 3.3 shows the overview of the scaled dot-product attention, represented as

$$Attention(Q; K; V) = softmax(\frac{QK^T}{\sqrt{d_k}})V. \tag{3.4}$$

Using only a single attention mechanism can be challenging to learn various features in three modalities. Multi-head attention can jointly attend to different representation subspaces that express a variety of features in data. It expresses the diverse subspaces representation in modality. Let $Head_i = Attention(QW_i^q; KW_i^k; VW_i^v)$ be the ith head of multi-head attention. Let the $W_i^q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^k \in \mathbb{R}^{d_{model} \times d_k}$, and $W_i^v \in \mathbb{R}^{d_{model} \times d_k}$ be the linear layer. Let $W^0 \in \mathbb{R}^{d_{model} \times d_k}$ be a weighted matrix multiplied after all the heads are concatenated. $h = 8$ is the number of heads. d_{model} is set to $d_k = d_{model}/h =$

40.

$$MultiHead(Q; K; V) = [Head_1; ...; Head_h]W^0 \tag{3.5}$$

Next, after the input data is passed through the feature concatenation and multi-head attention, feature compression feed-forward networks extract the information between features extracted by multi-head attention. Figure 3.4 show the latter half of the proposed method. The point-wise feed-forward networks have a ReLU activation function that extracts the non-linear feature. W_1 and b_1 are weight and bias in point-wise feed-forward networks. W_2 and b_2 are a weigh and bias in point-wise feed-forward networks. Each point-wise Feed-Forward Network is applied to each position separately and identically to the output of the multi-head attention.

$$FFN(x) = max(0; xW_1 + b_1)W_2 + b_2 \tag{3.6}$$

After getting through the feature compression feed-forward networks, the feature dimension can be one. In addition, time sequence compression feed-forward networks extract the information between the time and the time sequence dimension can be one. Let x be the output of the point-wise feed-forward networks. $W_f \in \mathbb{R}^{d_k \times 1}$ and $b_f \in \mathbb{R}^{d_k}$ are the weight and bias of feature compress feed-forward networks. $W_s \in \mathbb{R}^{l \times 1}$ and $b_s \in \mathbb{R}^1$ are the weight and bias of sequence compress feed-forward networks.

$$FFN_f(x) = max(0; xW_f + b_f) \tag{3.7}$$

and

$$FFN_s(x) = max(0; xW_s + b_s) \tag{3.8}$$

The output of the sequence compress feed-forward networks gets through the sigmoid function. The sigmoid function is the non-linear function that makes the input value in the range of $[0, 1]$. The output value of the sigmoid function is used for the input of Binary Cross Entropy Loss with the target value. Thus, trainable parameters are updated based on the Binary Cross Entropy Loss.

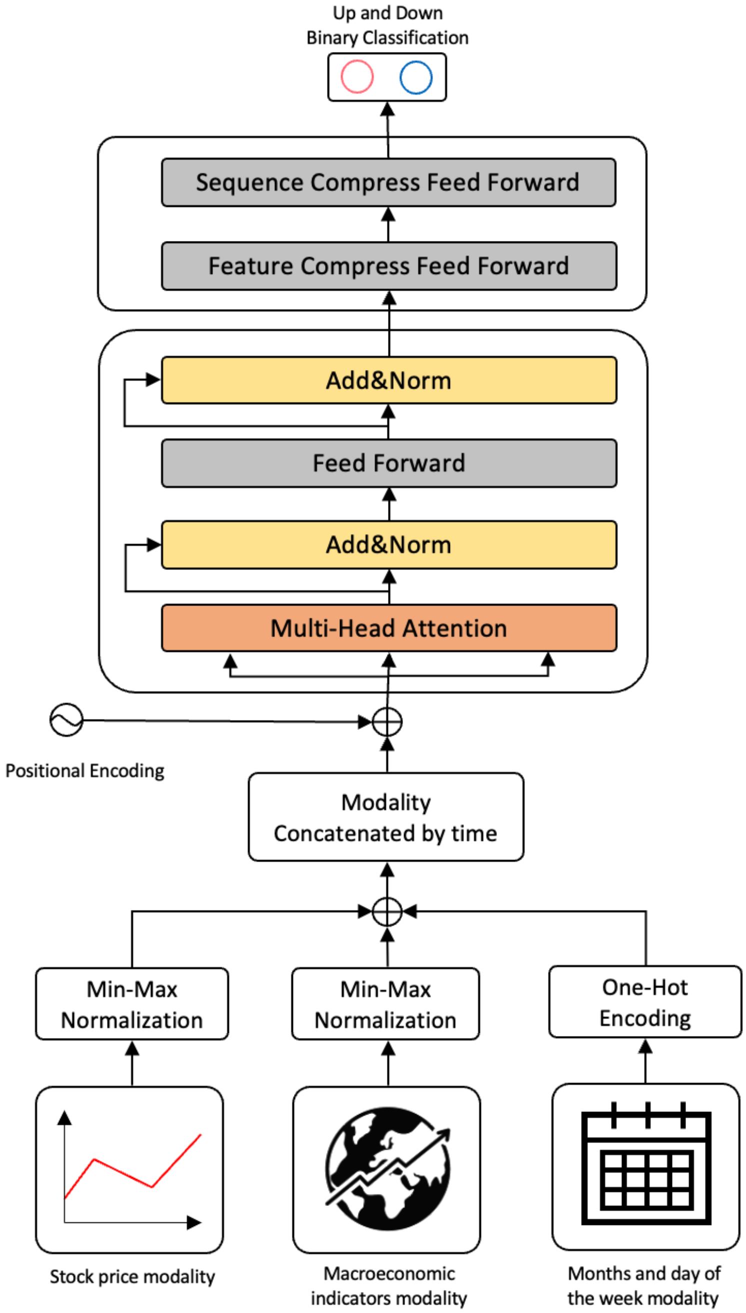

Figure 3.1: The overview of the proposed method. The proposed method comprises the feature concatenation, transformer encoder, and classification layer.

Table 3.1: 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.
Close	The price at which the financial security closes in the market when trading ends.
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	$\frac{C_t + C_{t-1} + \cdots + C_{t-9}}{10}$
Weighted 10-day Moving Average	$\frac{(n) \times C_t + (n-1) \times C_{t-1} + \cdots + C_{t-9}}{(n + (n-1) + \cdots + 1)}$
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)	$100 - \frac{100}{1 + \frac{(\sum_{i=0}^{n-1} Up_{t-i}/n)}{(\sum_{i=0}^{n-1} Dw_{t-i}/n)}}$
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,
Dw_t: the downward price change at time t

Table 3.2: 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 3.3: 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.

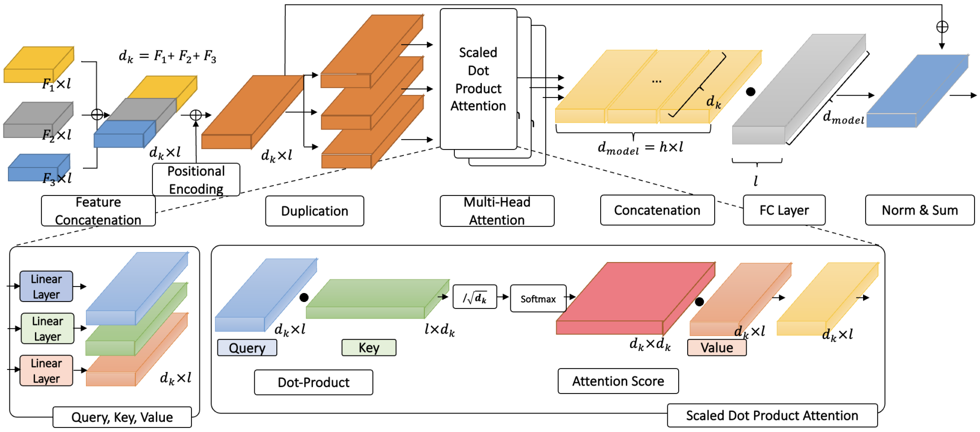

Figure 3.2: The process of feature concatenation and multi-head attention in the proposed method.

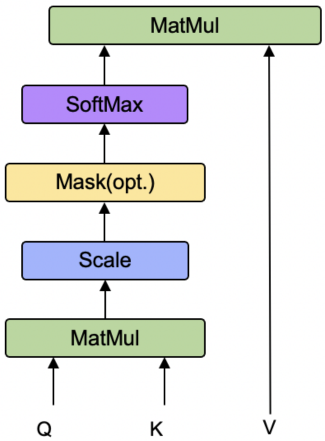

Figure 3.3: Scaled Dot-Product Attention.

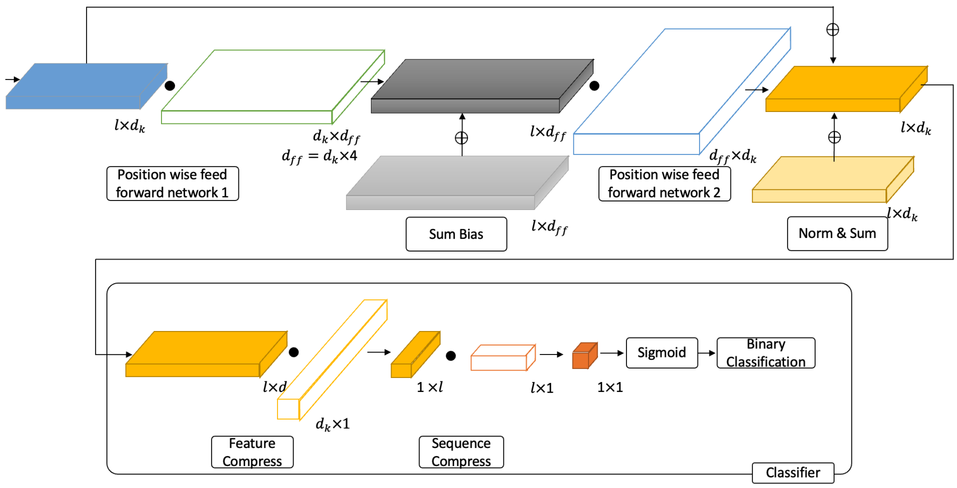

Figure 3.4: The process of point wise feed forward network and classifier in the proposed method.

Chapter 4

Experiments

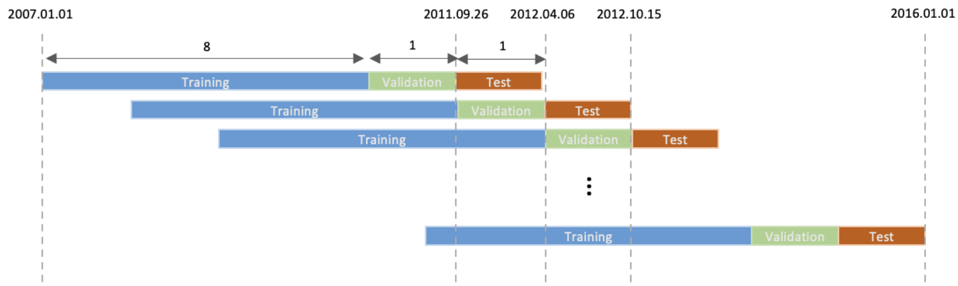

Figure 4.1: Blocked time series Cross-Validation.

4.1 Experimental settings

KDD17 datasets [Zhang et al., 2017] and macroeconomic indicators are used for stock direction prediction. The macroeconomic indicators were collected on Investing.com. The length of historical price in KDD17 datasets and macroeconomic indicators is the 2518 days ranging from Jan-01-2007 to Jan-01-2016. The KDD17 dataset has 50 stocks in U.S. markets. This paper's macroeconomic indicators include NASDAQ100, U.S. 2-year, 10-year, and 30-year Bond Yield, U.S. Dollar Index, and WTI Oil Price.

A blocked time series cross-validation is used as a data split strategy to consider the trait of the time series data [Bao et al., 2017]. Data leakage can occur if the random split strategy is used. Because the data random split strategy in time series can make the model predict the next day by looking at the future value. In addition, a 10-fold split is used for the entire data length and sets the training, validation, and test data ratio as 8:1:1. Figure 4.1 shows the overview of the blocked time series cross-validation.

The proposed model is compared to three comparison models by accuracy. For three comparison models, the parameter is set to the value recommended by each research. A brief review of three comparison models is as follows:

- CNN-Attention Bi-LSTM [Lu et al., 2021]: CNN-Attention Bi-LSTM model combines CNN and Attention Bi-LSTM to predict stock price prediction. CNN is used to extract local perception and improve the efficiency of model training by reducing the number of parameters. Attention Bi-LSTM is used to learn sequential features of the stock data. They used the Shanghai Composite Index (000,001) stock as a dataset.
- Adversarial LSTM [Feng et al., 2018]: 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 [Yoo et al., 2021]: DTML model is combined with the Attention LSTM and Transformer model. They 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 plural stocks in the concatenated Attention LSTM feature.

In this paper, the hyperparameters are set as follows. The batch size is set as 32. The learning rate is set to 0.001. The training is progressed as 100 epochs, and a model with the lowest validation loss between epochs is used for testing. Adam optimizer is used for the training [Kingma and Ba, 2014]. The number of transformer model layers is one. The transformer model dimension is 320, and the number of transformer heads is set to 8. The loss function is set to Binary Cross Entropy Loss. Binary Cross Entropy Loss is defined as

$$BCE(x) = -\frac{1}{N}\sum_{i=1}^{N} y_i log(h(x_i)) + (1 - y_i)log(1 - h(x_i)) \tag{4.1}$$

where N, y_i, and $h(x_i)$ are the number of data, 0/1 binary target value is the prediction of the proposed model, and the final output of the model gets through the sigmoid function, respectively.

Accuracy is employed as an evaluation measure.

$$Accuracy(\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{4.2}$$

where TP, TN, FP, and FN refer to True Positives, True Negates, False Positives, and False Negates.

Statistical tests are performed to validate the superiority among models for the stock direction prediction. The Friedman test proceeded, which is the statistical test for non-parametric multiple comparisons of the methods with multiple datasets [Demšar, 2006]. Given k methods and N datasets, let r_i^j denote the rank of j-th methods on the i-th dataset (mean ranks are used in case of ties). The Friedman test compares the average ranks of the method and $R_j = \frac{1}{N}\sum_{i=1}^{N} r_i^j$ is the average ranks on the j-th method. Furthermore, the F_F, Friedman statistic, are distributed according to the F-distribution with k-1 and $(k-1)(N-1)$ degrees of freedom:

$$F_F = \frac{(N-1)x_F^2}{N(k-1) - x_F^2} \tag{4.3}$$

Table 4.1: 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

where

$$x_F^2 = \frac{12N}{k(k+1)}\left[\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right] \tag{4.4}$$

In the Frieman test, the null hypothesis of the multiple comparisons is rejected when F_F statistic is larger than the critical value under significance level α. In this case, the post-hoc test can proceed, which allows comparing which methods have the difference. The Bonferroni-Dunn test is selected as a post-hoc test Because it is generally used after rejecting the Friedman test's null hypothesis [Trawiński et al., 2012]. The difference between the average ranks of the methods is compared with the Critical Difference(CD).

$$CD = q_a\sqrt{\frac{k(k+1)}{6N}} \tag{4.5}$$

Table 4.2: 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

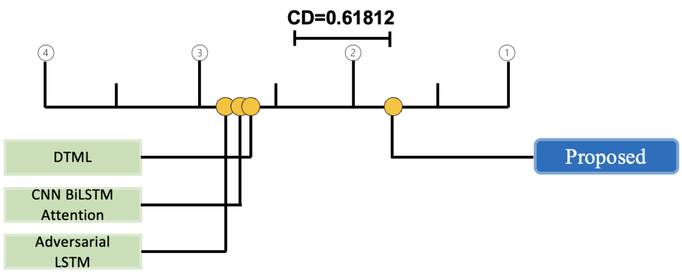

Figure 4.2: Result of Bonferroni-Dunn test between the proposed method and comparison models.

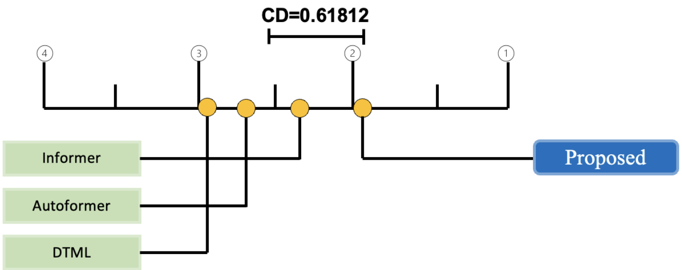

Figure 4.3: Result of Bonferroni-Dunn test between the proposed method and variants of transformer model.

4.2 Comparison results

Table 4.3 shows experimental results for the entire data in terms of accuracy.

The table consists of a ticker, sector, proposed method's experiment results, and three comparison model's experiment results. The ticker represents the name of a certain stock. The sector in stock is a unit consisting of similar industries. The experiment results are expressed as an average accuracy and standard deviation. Table 4.5 shows the abbreviation of each sector. The proposed method, 27 out of 50 stock data, showed the best accuracy compared to comparison models. Compared to the comparison model, the stock that the proposed method yielded the best performance was Verizon Communications (VZ), and the difference in accuracy was 8.85%p, which can be regarded as a significant difference in the stock direction prediction domain. Friedman test with a 95% confidence interval was used to verify whether differences exist between groups. Chi-squared statistic of the Friedman test 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 is proceeded with a 95% confidence interval as a post hoc test of the Friedman test to test which groups differ. The Critical Difference(CD) value was 0.6181 under a significance level of 0.05 in the Bonferroni-Dunn test. The differences in average rank between the proposed method and comparison models were greater than the CD. Figure 4.2 shows the result 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 the comparison methods. The quantitative analysis confirmed that the proposed method has a numerical difference compared to the comparison model. As a result of the statistical tests, it was found that the proposed model had a statistically significant difference compared to the comparison models.

Additional comparison experiments are conducted on transformer variant models to determine whether the result of the proposed model is the effect of transformer or modality fusion. Table 4.4 shows the experimental result between the proposed method and transformer variant models in terms of accuracy. The proposed method gained the best average rank of 1.96 compared to transformer variant models. To see if there is a statistical difference, the Friedman test and the Bonferroni-Dunn test are proceeded under a significance level of 0.05. Chi-squared statistic of the Friedman test was 18.595, and the p-value was 0.00033. There were statistically significant differences between the proposed model and comparison models. The critical difference value was 0.6181 under a significance level of 0.05 in the Bonferroni-Dunn test. When comparing the difference in average rank between the proposed model 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. As a result of the statistical tests, it was found that the proposed model had a statistically significant difference compared to the comparison models except for the Informer model. Table 4.2 shows the result of the Friedman test and Figure 4.3 shows the result of the Bonferroni-Dunn test.

4.3 In-depth Analysis

The classification accuracy is analyzed with each stock sector to demonstrate the fusing modalities' effect on the classification accuracy. The number of sectors is ten in total. Table 4.3 and 4.7 show which stock belongs to which sector. Higher classification accuracy is expected in the proposed model than in comparison models on the financial and energy sector because the macroeconomic indicator modality that includes U.S. bond yields, the U.S. dollar index, and WTI oil prices is used as input. The stocks in the financial sector are affected by the base rate. U.S. bond yields change every time but are fundamentally affected by the Base Rate. Moreover, the stocks in the energy sector are affected by oil or natural gas prices. Therefore, the proposed model is expected to perform better than the comparison models. In the financial sector, the four stocks achieved the best performance among the six stocks. In the energy sector, the six stocks achieved the best performance among the eight stocks. Table 4.3 shows the sector-related comparison results with comparison

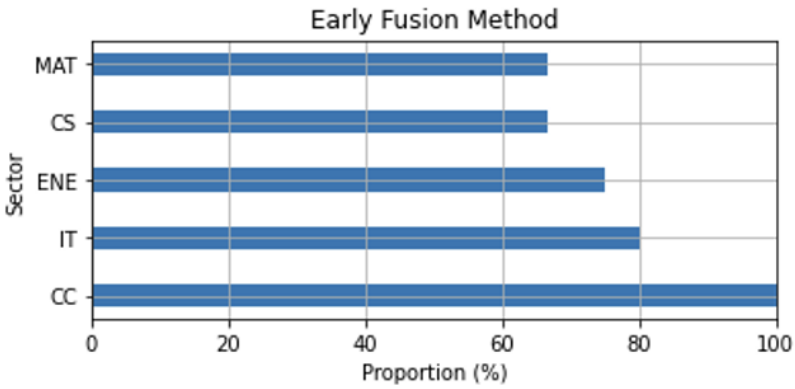

Figure 4.4: The sectors in which early fusion methods perform better than the late fusion method, and the proportion of how the sectors perform well compared to the late fusion method

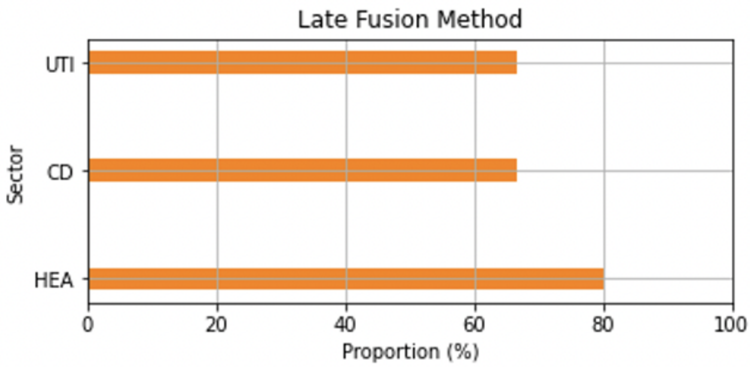

Figure 4.5: The sectors in which late fusion methods perform better than early fusion methods, and the proportion of how the sectors perform well in late fusion method

models.

The early fusion and late fusion models are compared in terms of accu-

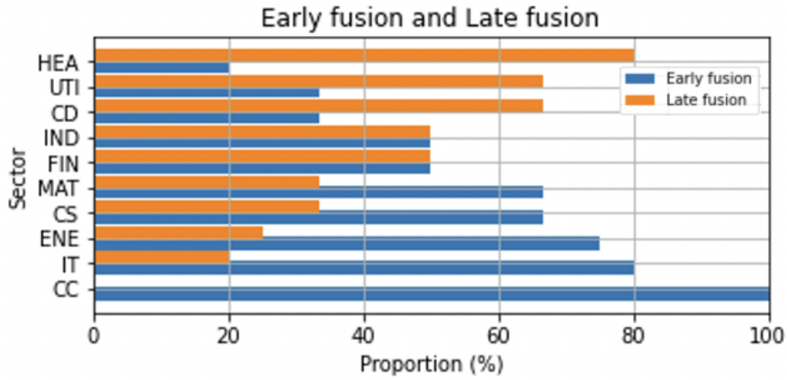

Figure 4.6: The sectors in which early fusion and late fusion methods perform respectively better than each other method, and the proportion of how the sectors perform well compared to each other method

racy to demonstrate the early fusion algorithms' superiority compared to the late fusion model. The proposed model in this paper 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 three out of six stocks in the financial sector, it acquired the best performance in six out of eight stocks compared to the late fusion method in the energy sector in classification accuracy. Figure 4.4 shows the sectors in which early fusion methods perform better than late fusion methods. The proportion of the number of the stocks and how the stocks in each sector perform well. Figure 4.5 shows the sectors in which late fusion methods perform better than early fusion methods. The proportion of the number of the stocks and how the stocks in each sector

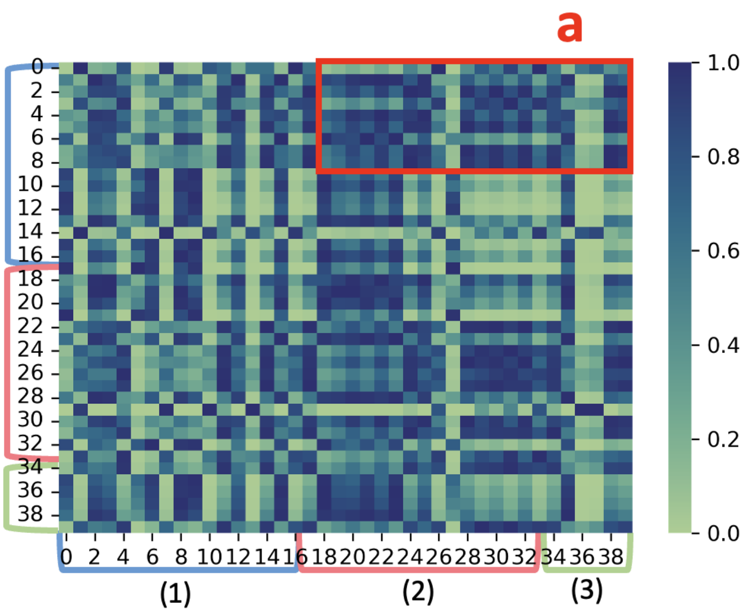

Figure 4.7: The attention map in the ealry fusion method. (1) is stock price modality. (2) is the months and days of the week modality. (3) is the macroeconomic indicator modality.

perform well. Figure 4.6 shows the overall result of the sectors in which early fusion and late fusion methods respectively perform better than each other method, and the proportion of the number of the stocks how the stocks in each sector perform well compared to each other method. The early fusion method achieved higher classification accuracy than late fusion in five sectors: materials, communication services, energy, information technology, and consumer cyclical. The late fusion method achieved higher classification accuracy than the early fusion method in three sectors: utility, consumer defensive, and

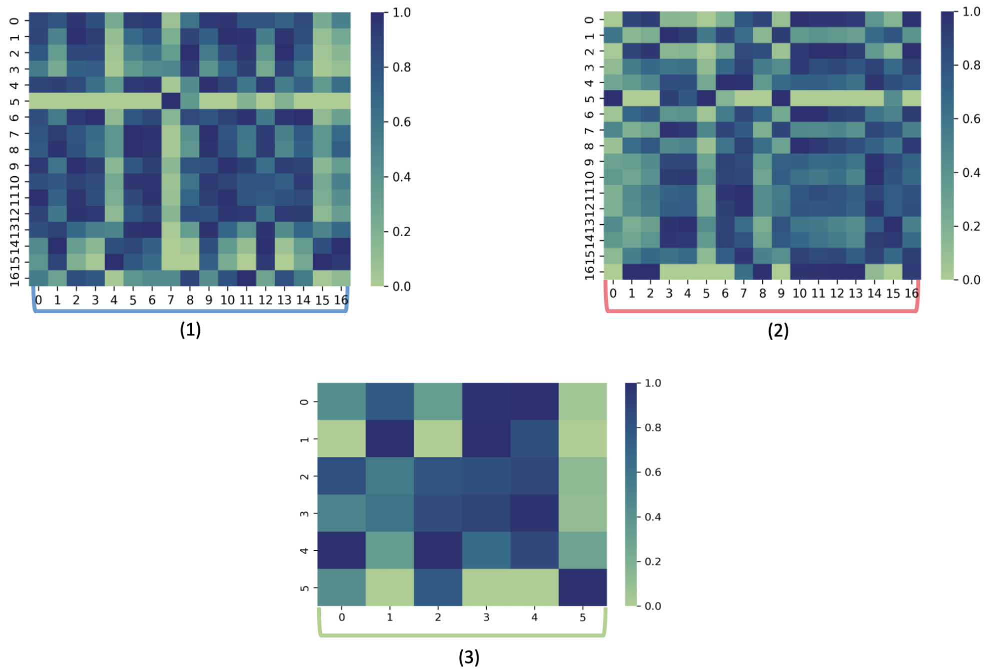

Figure 4.8: The attention map in the late fusion method. (1) is stock price modality. (2) is the months and days of the week modality. (3) is the macroeconomic indicator modality.

healthcare. Sectors with the same ratio of stocks with high classification accuracy in early fusion and late fusion were the industrial and financial sectors.

The attention map is also analyzed in scaled dot-product attention at the inference level. The corporation of the attention map appeared in Figure 4.7, and Figure 4.8 is the Apple Inc., which dataset achieved high classification accuracy compared to comparison models. If the color in the attention map is white, the features corresponding to white are highly correlated. If the color in the attention map is black, the features corresponding to black are weakly correlated. Figure 4.7 shows the attention map in early fusion and Figure 4.8 shows the attention map in late fusion. Part "a" in Figure 4.7 represents that the 11 feature in stock price modality is highly correlated with the months and days of the week modality and macroeconomic indicator modality. The attention map in early fusion represents that each modality (1), (2), and (3) is correlated since 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 late fusion method's features are separated and passed through the scaled dot-product attention. Consequently, the early fusion method learns inter-modality correlation better than the late fusion method because all the modalities get through the scaled dot-product attention together, which learns the relationship between data in the early fusion method.

The ablation study is also progressed to verify if each modality's effect exists. Table 4.6 shows the average rank of each ablation study about modality. Each model's average rank was ranked by accuracy. The model used only stock price modality has an average rank of 2.96, which is the worst value of ablation study results. A model with all modalities received the best average rank of 2.3.

Table 4.3: Comparison results of the classification accuracy between the proposed method and comparison models

Ticker in dataset	Sector	Proposed Method	CNN-Attention BiLSTM [Lu et al., 2021]	Adversarial LSTM [Feng et al., 2018]	DTML [Yoo et al., 2021]
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

Table 4.4: Comparison results of the classification accuracy between the proposed method and the variants of the transformer model

Ticker in dataset	Sector	Proposed Method	Informer [Zhou et al., 2021]	Autoformer [Wu et al., 2021]	DTML [Yoo et al., 2021]
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

Table 4.5: The summary of the abbreviated sector

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 4.6: Ablation study about modality in terms of accuracy.

Modality	Avg. Rank
Stock price modality	2.96
Stock price modality + Month and days of the week modality	2.4
Stock price modality + Macroeconomic modality	2.32
Stock price modality + Month and days of the week modality + Macroeconomic modality	2.3

Table 4.7: Comparison results of the classification accuracy between early fusion method and late fusion method

Ticker in dataset	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

Chapter 5

Results

The task of stock price prediction can be difficult if the algorithm only depends on the stock data because the stock price can be affected by external impacts such as the world economy and policy. A new strategy was considered to fuse multiple data sources related to stock price prediction to solve the problem. Multiple data affecting stock price were added in this study, such as macroeconomic indicators and months and days of the week data as an additional modality. A multimodal early fusion structure was considered when designing the proposed method, which learns the inter-modality correlation in features.

The proposed model in this paper outperformed the comparison models and showed a statistically significant difference. Specifically, 27 out of 50 stocks achieved higher classification accuracy than the comparative model. A series of statistical tests were conducted, such as the Friedman and Bonferroni-Dunn tests. The test results confirmed that the proposed model showed a statistically significant difference compared to the comparison models. An in-depth analysis of the comparison result was also performed between early fusion and late fusion methods. The analysis indicates that the early fusion method achieved high classification accuracy in 30 of 50 datasets. In addition, in terms of the stock sector, in the early fusion method, there were five sectors where many stocks did well of the entire ten sectors, but in the late fusion method, there were only three sectors where many stocks did well in the whole ten sectors.

Potential directions of future works can be expected in additional data exploratory research can be expected in financial data mining. Moreover, data fusion methodology considering the trait between different data is also crucial in multimodal stock direction prediction.

Bibliography

[Adam and Tweneboah, 2008] Adam, A. M. and Tweneboah, G. (2008). Macroeconomic factors and stock market movement: Evidence from ghana. *Available at SSRN 1289842*.

[Agostinelli et al., 2014] Agostinelli, F., Hoffman, M., Sadowski, P., and Baldi, P. (2014). Learning activation functions to improve deep neural networks. *arXiv preprint arXiv:1412.6830*.

[Ariyo et al., 2014] Ariyo, A. A., Adewumi, A. O., and Ayo, C. K. (2014). Stock price prediction using the arima model. In *2014 UKSim-AMSS 16th international conference on computer modelling and simulation*, pages 106–112. IEEE.

[Bao et al., 2017] Bao, W., Yue, J., and Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PloS one*, 12(7):e0180944.

[Bartholomew, 1971] Bartholomew, D. J. (1971). Time series analysis forecasting and control.

[Chen et al., 2009] Chen, P., Pedersen, T., Bak-Jensen, B., and Chen, Z. (2009). Arima-based time series model of stochastic wind power generation. *IEEE transactions on power systems*, 25(2):667–676.

[Chen et al., 2008] Chen, S.-W. et al. (2008). Non-stationarity and non-linearity in stock prices: Evidence from the oecd countries. *Economics Bulletin*, 3(11):1–11.

[Cootner, 1962] Cootner, P. H. (1962). Stock prices: Ramdom vs. systematic changes. *Industrial Management Review (pre-1986)*, 3(2):24.

[Dama and Sinoquet, 2021] Dama, F. and Sinoquet, C. (2021). Time series analysis and modeling to forecast: a survey. *arXiv preprint arXiv:2104.00164*.

[Demšar, 2006] Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine learning research*, 7:1–30.

[Feng et al., 2018] Feng, F., Chen, H., He, X., Ding, J., Sun, M., and Chua, T.-S. (2018). Enhancing stock movement prediction with adversarial training. *arXiv preprint arXiv:1810.09936*.

[Gao et al., 2020] Gao, J., Li, P., Chen, Z., and Zhang, J. (2020). A survey on deep learning for multimodal data fusion. *Neural Computation*, 32(5):829–864.

[Gursida, 2017] Gursida, H. (2017). The influence of fundamental and macroeconomic analysis on stock price. *Jurnal Terapan Manajemen dan Bisnis*, 3(2):222–234.

[Han et al., 2010] Han, P., Wang, P. X., Zhang, S. Y., et al. (2010). Drought forecasting based on the remote sensing data using arima models. *Mathematical and computer modelling*, 51(11-12):1398–1403.

[Hassan and Nath, 2005] Hassan, M. R. and Nath, B. (2005). Stock market forecasting using hidden markov model: a new approach. In *5th International Conference on Intelligent Systems Design and Applications (ISDA'05)*, pages 192–196. IEEE.

[Henrique et al., 2018] Henrique, B. M., Sobreiro, V. A., and Kimura, H. (2018). Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of finance and data science*, 4(3):183–201.

[Hu et al., 2021] Hu, Z., Zhao, Y., and Khushi, M. (2021). A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, 4(1):9.

[Jareño Cebrián and Negrut, 2016] Jareño Cebrián, F. and Negrut, L. (2016). Us stock market and macroeconomic factors.

[Jing et al., 2021] Jing, N., Wu, Z., and Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications*, 178:115019.

[Kalyan et al., 2021] Kalyan, K. S., Rajasekharan, A., and Sangeetha, S. (2021). Ammus: A survey of transformer-based pretrained models in natural language processing. *arXiv preprint arXiv:2108.05542*.

[Kara et al., 2011] Kara, Y., Boyacioglu, M. A., and Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange. *Expert systems with Applications*, 38(5):5311–5319.

[Khaidem et al., 2016] Khaidem, L., Saha, S., and Dey, S. R. (2016). Predicting the direction of stock market prices using random forest. *arXiv preprint arXiv:1605.00003*.

[Khuat and Le, 2017] Khuat, T. T. and Le, M. H. (2017). An application of artificial neural networks and fuzzy logic on the stock price prediction problem. *JOIV: International Journal on Informatics Visualization*, 1(2):40–49.

[Kim, 2003] Kim, K.-j. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1-2):307–319.

[Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[Kogan et al., 2009] Kogan, S., Levin, D., Routledge, B. R., Sagi, J. S., and Smith, N. A. (2009). Predicting risk from financial reports with regression. In *Proceedings of human language technologies: the 2009 annual conference of the North American Chapter of the Association for Computational Linguistics*, pages 272–280.

[Kyereboah-Coleman and Agyire-Tettey, 2008] Kyereboah-Coleman, A. and Agyire-Tettey, K. F. (2008). Impact of macroeconomic indicators on stock market performance: The case of the ghana stock exchange. *The Journal of Risk Finance*.

[Lahmiri, 2018] Lahmiri, S. (2018). A technical analysis information fusion approach for stock price analysis and modeling. *Fluctuation and Noise Letters*, 17(01):1850007.

[Lee and Yoo, 2020] Lee, S. I. and Yoo, S. J. (2020). Multimodal deep learning for finance: integrating and forecasting international stock markets. *The Journal of Supercomputing*, 76(10):8294–8312.

[Lu et al., 2021] Lu, W., Li, J., Wang, J., and Qin, L. (2021). A cnn-bilstm-am method for stock price prediction. *Neural Computing and Applications*, 33(10):4741–4753.

[Mehtab and Sen, 2020] Mehtab, S. and Sen, J. (2020). Stock price prediction using convolutional neural networks on a multivariate timeseries. *arXiv preprint arXiv:2001.09769*.

[Nelson, 1998] Nelson, B. K. (1998). Time series analysis using autoregressive integrated moving average (arima) models. *Academic emergency medicine*, 5(7):739–744.

[Niu et al., 2021] Niu, Z., Zhong, G., and Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing*, 452:48–62.

[Ravikumar and Saraf, 2020] Ravikumar, S. and Saraf, P. (2020). Prediction of stock prices using machine learning (regression, classification) algorithms. In *2020 International Conference for Emerging Technology (INCET)*, pages 1–5. IEEE.

[Rogalski, 1984] Rogalski, R. J. (1984). New findings regarding day-of-the-week returns over trading and non-trading periods: a note. *The Journal of Finance*, 39(5):1603–1614.

[Schumaker and Chen, 2009] Schumaker, R. P. and Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The azfin text system. *ACM Transactions on Information Systems (TOIS)*, 27(2):1–19.

[Somani et al., 2014] Somani, P., Talele, S., and Sawant, S. (2014). Stock market prediction using hidden markov model. In *2014 IEEE 7th joint international information technology and artificial intelligence conference*, pages 89–92. IEEE.

[Sunny et al., 2020] Sunny, M. A. I., Maswood, M. M. S., and Alharbi, A. G. (2020). Deep learning-based stock price prediction using lstm and bi-directional lstm model. In *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, pages 87–92. IEEE.

[Thakkar and Chaudhari, 2021] Thakkar, A. and Chaudhari, K. (2021). Fusion in stock market prediction: a decade survey on the necessity, recent developments, and potential future directions. *Information Fusion*, 65:95–107.

[Thakkar and Chaudhari, 2022] Thakkar, A. and Chaudhari, K. (2022). Information fusion-based genetic algorithm with long short-term memory for stock price and trend prediction. *Applied Soft Computing*, 128:109428.

[Trawiński et al., 2012] Trawiński, B., Smetek, M., Telec, Z., and Lasota, T. (2012). Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms. *International Journal of Applied Mathematics and Computer Science*, 22(4):867–881.

[Tsantekidis et al., 2020] Tsantekidis, A., Passalis, N., Tefas, A., Kanniainen, J., Gabbouj, M., and Iosifidis, A. (2020). Using deep learning for price prediction by exploiting stationary limit order book features. *Applied Soft Computing*, 93:106401.

[Van Der Voort et al., 1996] Van Der Voort, M., Dougherty, M., and Watson, S. (1996). Combining kohonen maps with arima time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies*, 4(5):307–318.

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

[Wang et al., 2022] Wang, C., Chen, Y., Zhang, S., and Zhang, Q. (2022). Stock market index prediction using deep transformer model. *Expert Systems with Applications*, 208:118128.

[Wei et al., 2020] Wei, X., Zhang, T., Li, Y., Zhang, Y., and Wu, F. (2020). Multi-modality cross attention network for image and sentence matching. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10941–10950.

[Wu et al., 2021] Wu, H., Xu, J., Wang, J., and Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, 34:22419–22430.

[Yoo et al., 2021] Yoo, J., Soun, Y., Park, Y.-c., and Kang, U. (2021). Accurate multivariate stock movement prediction via data-axis transformer with multi-level contexts. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 2037–2045.

[Yu and Yan, 2020] Yu, P. and Yan, X. (2020). Stock price prediction based on deep neural networks. *Neural Computing and Applications*, 32(6):1609–1628.

[Zhang et al., 2017] Zhang, L., Aggarwal, C., and Qi, G.-J. (2017). Stock price prediction via discovering multi-frequency trading patterns. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 2141–2149.

[Zhang et al., 2019] Zhang, X., Li, Y., Wang, S., Fang, B., and Yu, P. S. (2019). Enhancing stock market prediction with extended coupled hidden markov model over multi-sourced data. *Knowledge and Information Systems*, 61(2):1071–1090.

[Zhang et al., 2018] Zhang, X., Zhang, Y., Wang, S., Yao, Y., Fang, B., and Philip, S. Y. (2018). Improving stock market prediction via heterogeneous information fusion. *Knowledge-Based Systems*, 143:236–247.

[Zhou et al., 2021] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., and Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 11106–11115.

국 문 초 록

멀티모달 트랜스포머 기반 주식 방향 예측 연구

이 태 원

AI학과 AI 응용 전공

중앙대학교 대학원

금융시장에서 주가를 정확하게 예측할 수 있다면 금융 데이터 마이닝 분야에서 엄청난 파급효과가 발생할 수 있다. 그러나 주가 데이터만을 사용하여 주가를 예측하는 것은 주가 데이터의 무작위적 특성 때문에 어려운 문제이다. 본 논문에서는 주가 예측 문제를 해결하기 위해 다양한 데이터를 융합하려고 한다. 제안된 방법에는 거시경제 지표와 월 및 요일 데이터와 같은 주가에 영향을 미치는 데이터가 추가 모달리티로 추가된다. 특징에서 모달리티 간 상관관계를 효과적으로 학습하기 위해 멀티모달 조기 융합 방법이 사용된다. 본 논문에서 제안된 모델은 비교 모델을 능가하고 통계적으로 유의미한 결과를 달성했다. 구체적으로 50개 종목 중 27개 종목이 비교 모델보다 높은 분류 정확도를 달성했다. 또한 실험 결과에 관한 심층 분석을 진행했고, 분석 결과에 따르면 초기 융합 전략은 50개 데이터 세트 중 30개에서 주가 예측을 위한 후기 융합 전략보다 더 나은 분류 정확도를 달성하였다.

키워드: 주가 예측, 멀티모달 러닝, 정보 융합, 거시경제 데이터, 기술적 지표

ABSTRACT

A study on stock direction prediction based on multimodal transformer

Tae-Won Lee
Major in Artificial Intelligence Application
Department of Artificial Intelligence
The Graduate School, Chung-Ang University

An enormous ripple effect can come in financial data mining if it can predict accurate stock prices. However, predicting stock prices using stock price data only is a difficult task because of the random nature of the stock price data. In this paper, fusing various data is tried to solve the stock price prediction problem. The data affecting stock price is added in the proposed method, such as macroeconomic indicators and months and days of the week data as an additional modality. The multimodal early fusion method is used, which learns the inter-modality correlation in 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, our in-depth analysis indicates that the early fusion strategy achieved a better classification accuracy in 30 of 50 datasets than the late fusion strategy for stock price prediction.

Keywords: Stock Price Prediction, Multimodal Learning, Information Fusion, Macroeconomic data, Technical indicator