

Deep Learning-based Information Fusion Frameworks for Stock Price Movement Prediction

Farnoush Ronaghi

A Thesis

in

The Department

of

Concordia Institute for Information System Engineering (CIISE)

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Quality System Engineering) at

Concordia University

Montréal, Québec, Canada

August 2021

© Farnoush Ronaghi, 2021

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Farnoush Ronaghi**

Entitled: **Deep Learning-based Information Fusion Frameworks for Stock Price
Movement Prediction**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Quality System Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. Amin Hammad

_____ External Examiner
Dr. Hossein Hashemi Doulabi

_____ Examiner
Dr. Amin Hammad

_____ Supervisor
Dr. Farnoosh Naderkhani

_____ Co-supervisor
Dr. Arash Mohammadi

Approved by

Dr. Abdessamad Ben Hamza, Chair
Department of Concordia Institute for Information System Engi-
neering (CIISE)

_____ 2021

Dr. Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Deep Learning-based Information Fusion Frameworks for Stock Price Movement Prediction

Farnoush Ronaghi

The challenges of modeling behaviour of financial markets, such as its high volatility, poor predictive behaviour, and its non-stationary nature, have continuously attracted attention of the researchers to employ advanced engineering methods. Within the context of financial econometrics, stock market movement prediction is a key and challenging problem. The research works reported in this thesis are motivated by the potentials of Artificial Intelligence (AI) and Machine Learning (ML)-based models, especially Deep Neural Network (DNN) architectures, for stock movement prediction. Considering recent progress in design and implementation of advanced DNN-based models, there has been a surge of interest in their application for predicting stock trends. In particular, the focus of the thesis is on utilization of information fusion to combine Twitter data with extended horizon market historical data for the task of price movement prediction. In this regard, the thesis made a number of contributions, first, the Noisy Deep Stock Movement Prediction Fusion (ND-SMPF) framework is proposed to extract news level temporal information; identify relevant words with highest correlation and effects on the stock trends, and; perform information fusion with historical price data. A real dataset is incorporated to evaluate performance of the proposed ND-SMPF framework. In addition, given that recent COVID-19 pandemic has negatively affected financial econometrics and stock markets across the globe, a unique COVID-19 related PRICE MOVEMENT prediction (COVID19 PRIMO) dataset is constructed. The objective is to incorporate effects of social media trends related to COVID-19 on stock market price movements. A novel hybrid and parallel DNN-based framework is then designed that integrates different and diversified learning architectures. Referred to as the COVID-19 adopted Hybrid and Parallel deep fusion framework

for Stock price Movement Prediction (COVID19-HPSMP), innovative fusion strategies are used to combine scattered social media news related to COVID-19 with historical market data and perform accurate price movement prediction during a pandemic crisis.

Acknowledgments

I am extending my heartfelt thanks to my supervisors, Prof. Farnoosh Naderkhani and Prof. Arash Mohammadi, for their support, motivation, patience during our discussions, and spirit of adventure in research that made this project possible.

Besides my supervisors, I would like to thank my thesis committee members: Prof. Amin Hammad and Prof. Hossein Hashemi Doulabi for their valuable time and input, and also for their great personality and friendly advice which incited me to widen my research from various perspectives.

I would like to thank my family, especially my parents, my sister and brother for their endless support and faith in me throughout my years of study in my life. I cannot express how I am grateful for all the sacrifices that my family have made for me.

A warm and special thanks to my husband, Mohammad, for his helpful comments, suggestions, and enlightening discussions, which have been instrumental in the progress of the project. His unfailing support and continuous encouragement throughout my master's study and through the process of researching have provided me with moral and emotional support. This accomplishment would not have been possible without him.

Contents

List of Figures	viii
List of Tables	x
1 Thesis Introduction	1
1.1 Contributions	3
1.2 Thesis Organization	5
2 Literature Review and Background	6
2.1 Literature Review	6
2.1.1 Stock Movement Prediction	6
2.1.2 Portfolio Optimization	8
2.2 Financial Signal Processing	10
2.2.1 Financial Terms and Concepts	10
2.2.2 Technical Indicators	11
2.3 Learning-based Models with Application to Financial Data	13
2.3.1 GloVe	13
2.3.2 Convolutional Neural Networks	14
2.4 Summary	15
3 ND-SMPF: A Noisy Deep Neural Network Fusion Framework for Stock Price Movement Prediction	16
3.1 Problem Definition and Data Description	17

3.2	Proposed ND-SMPF Model	18
3.2.1	Hybrid Attention Network	18
3.2.2	Bidirectional GRU	20
3.2.3	Temporal Attention	21
3.3	Experimental Results	22
3.3.1	Baseline and Proposed Models	23
3.3.2	Performance Evaluation Results	24
3.4	Summary	26
4	COVID19-HPSMP: COVID-19 Adopted Hybrid and Parallel Deep Information Fusion Framework for Stock Price Movement Prediction	27
4.1	Introduction	28
4.2	Problem Definition and COVID19 PRIMO	30
4.2.1	Data Visualization	32
4.3	Proposed COVID19-HPSMP Framework	34
4.3.1	The CNN Local/Global Path	36
4.3.2	CNN-LSTM Attention based Model	39
4.3.3	Fusion Path	41
4.4	Experiments	43
4.4.1	COVID19-HPSMP LSTM-based hybrid attention Model	43
4.4.2	Performance Evaluation/Results	45
4.5	Summary	58
5	Summary and Future Research Directions	59
5.1	Summary of Thesis Contributions	59
5.2	Future Research	61
	Bibliography	63

List of Figures

Figure 3.1	Proposed DNN-based Stock Movement Prediction Model (the ND-SMPF).	19
Figure 3.2	Accuracy of the Models	25
Figure 3.3	Loss of the Models	25
Figure 3.4	Accuracy and Loss contrast of the Models	26
Figure 4.1	Block diagram of the procedure designed to collect and prepare Tweet component of the COVID19 PRIMO.	31
Figure 4.2	Illustrative raw Tweets samples in the COVID19 PRIMO dataset before the pre-processing step.	32
Figure 4.3	(a) The Distribution of The COVID-19 Tweets and Target. (b) The Number of COVID-19 Tweets. (c) Correlation matrix between different variables.	33
Figure 4.4	Grid of scatter plots. The orange dots show COVID-19 related data points, while the blue dots represent lack of COVID-19 related data.	35
Figure 4.5	The proposed COVID19-HPSMP framework.	36
Figure 4.6	The CNN Local/Global path of the COVID19-HPSMP framework.	37
Figure 4.7	The CNN-LSTM path of the COVID19-HPSMP framework.	39
Figure 4.8	(a) Accuracy of the Models. (b) Loss of the Models.	45
Figure 4.9	Accuracy and loss contrast of the different price movement prediction models.	45
Figure 4.10	Confusion matrices for binary predictions. (a) The COVID19-HPSMP Framework. (b) Standalone CNN Local/Global Model, (c) Standalone CNN-LSTM Model.	46

Figure 4.11 (a) Buy/Sale Return based on the 50-50 approach, i.e., whenever the movement prediction value is less than 0.5 the “SELL” signal will be sent to the trader. Alternatively, when the model predicts a boost in the market price (movement prediction value is greater than 0.5) “BUY” signal will be generated. (b) Buy/Sale Return based on 60-40 approach, i.e., when the movement prediction value is less than 0.4 we consider a decrease in the market price sending the “SELL” signal. When the prediction values are greater than 0.6, the “BUY” signal will be sent. For the values between 0.4 and 0.6, the HOLD signal will be offered.	47
Figure 4.12 (a) Per-stock returns for the first trading mechanism over the following six prominent companies: <i>Apple, WaltDesney, IBM, Visa, Caterpillar,</i> and <i>Coca – Cola.</i> (b) Results of MC simulations over all companies in Dow Jones. . .	50
Figure 4.13 MACD Signal	55

List of Tables

Table 3.1 Accuracy comparison between 5 variants of the proposed ND-SMPF framework and 4 Baseline models.	22
Table 4.1 Accuracy, Sensitivity, and Specificity comparisons between the proposed COVID19-HPSMP framework, those of three stand-alone models, and accuracies of algorithms from References [Hu, 2019, Xu, 2018].	43
Table 4.2 Experiment results for trading based on the COVID19-HPSMP framework following the 50-50 approach.	48
Table 4.3 Experiment results for trading based on the COVID19-HPSMP framework following the 60-40 approach.	49
Table 4.4 Return comparisons between the proposed COVID19-HPSMP framework and stand-alone models.	54
Table 4.5 Return comparisons between the proposed COVID19-HPSMP framework (50-50), stand-alone models and two commonly used trading indicators.	54
Table 4.6 Return comparisons between the proposed COVID19-HPSMP framework (60-40), stand-alone models and two commonly used trading indicators.	55

Chapter 1

Thesis Introduction

Stock market movement prediction is a key and challenging problem in financial econometrics as such has attracted extensive recent research focus [Frankel, 1995, Ronaghi, 2020, Arash, 2017, Edwards, 2007, Bollen, 2011, Jiang, 2020, Hu, 2019, Koshiyama, 2020, Schumaker, 2009]. Capitalizing on recent advancements in communication, internet, and social networking technologies, there has been a growing interest across academia and financial sector to identify/collect different valuable data sources to accurately predict market movements. The main objective is to construct new information fusion models to analyze and interpret ever-changing trends in the globally interconnected market space. It is widely acknowledged that investors need high-quality data to make informed and accurate decisions. Particularly, in times of market crisis, specifically during the recent COVID-19 pandemic, investors need advanced Big-Data Analytic and Information Technologies [Choudrie, 2021] to acquire timely and accurate datasets. Using high-quality data, investors can perform fast analysis of the market volatility and react quickly to the ever changing financial conditions. Any positive or negative news related to the stock market can have a ripple effect on the investors' decision-making process. During the pandemic area, stock movement prediction became significantly challenging as stock markets faced high fluctuations. Consequently, it is of paramount importance to develop innovative and advanced processing and learning solutions to accurately predict stock movements for achieving maximum potential profit. This has resulted in a recent surge of interest in Machine Learning (ML)-based prediction techniques [Hu, 2019, Anik, 2020] and fusion

of multi-modal information sources. In particular, there has been a recent surge of interest in information fusion [Narkhede, 2021] and sentiment analysis [Choudrie, 2021] based on stock market data [Patel, 2015, Gite, 2021].

It should be noted that stock movement prediction is a time-series phenomenon inherently affected by temporal dependencies. For example, when a company faces a major economic crisis on a specific trading day, typically, its stock will face recession showing a decrease in the price slope over the next trading days. Another example is when a company's probable wrongdoings leak to public. Neglecting such events can potentially affect performance of financial forecasts [Luss, 2015]. To address these issues, there is an urgent need to design and implement advanced fusion strategies [Xu, 2018] to provide robust stock movement predictions. In the context of stock price movement prediction, historical stock prices are typically fused with information obtained from media news. For the latter, in addition to the conventional news platforms, recently, extensive interest is shown towards utilization of Internet-based news resources, such as social media, for development of ML predictive models. Recent advancements and developments in the field of Deep Neural Networks (DNNs) have motivated different research works to incorporate such advanced modeling techniques for prediction and forecasting tasks in stock markets [Tetlock, 2007].

In brief, the focus of the thesis is on the following two main research areas:

- (i) Applying advanced DNN models to fuse market historical data with public news and social media information resources. The main goal is to increase the overall accuracy of stock movement prediction models. In this regard, the thesis proposes a Noisy Deep Stock Movement Prediction Fusion framework (COVID19-HPSMP) that uses information fusion to combine Twitter data with extended horizon market historical data, and;
- (ii) Consider the effect of coronavirus on stock movement prediction. The novel coronavirus (COVID-19) has suddenly and abruptly changed the world as we knew, causing market volatilities rocketing upward around the world [Mazue, 2020, Baek, 2020]. In particular, the pandemic has negatively triggered several sectors including but not limited to stock markets, global supply chains, labor markets, and consumption behaviors. Disruptions of such sectors, especially the stock markets [Al-Awadhi, 2020, Ansari, 2020], can adversely affect the global

economy. It is expected that the emerging markets, the ones with restricted resources to cope with negative impacts of the COVID-19, more substantially feel the COVID-19 pressure due to having slower economic growth insufficient capital inflows. In these sad and unfortunate pandemic times, Financial Signal Processing (FSP) and ML-based [Zhang, 2020] stock market movement prediction solutions can potentially help in controlling negative effects of the pandemic crisis on financial econometrics.

1.1 Contributions

The main objective of this thesis is development of Stock Movement Prediction (SMP) models to answer the following question; “How market react to Twitter news and how this information can be used for SMP?”. Generally speaking, SMP is beneficial for daily investors, financial analysts and users of financial news. To investigate the influence of tweets on market movement, two different deep learning models are implemented. The main contributions [Ronaghi, 2021, Ronaghi, 2020] of my thesis research work are briefly outlined below:

- (1) **ND-SMPF** [Ronaghi, 2020]: This chapter proposes a Noisy Deep Stock Movement Prediction Fusion framework (ND-SMPF) that uses information fusion to combine Twitter data with extended horizon market historical data. To completely utilize market information, a Noisy Bi-directional Gated Recurrent Unit (NBGRU) is utilized coupled with a Hybrid Attention Network (HAN) to extract news level temporal information. We consider a lag of time for the stock movement in order to construct the proposed ND-SMPF fusion framework with the most reliable information. A two level attention layer is used to identify relevant words of highest importance to price movements, which are then fused with historical price data to perform the prediction task. A real dataset is incorporated to evaluate performance of the proposed ND-SMPF framework, which illustrates its superior performance in comparison to its recently developed counterparts.
- (2) **COVID-19 Adopted Hybrid and Parallel Deep Information Fusion Framework for Stock Price Movement Prediction**): This chapter makes the following contributions.

- First, a unique and real COVID-19 related PRIce MOvement prediction (COVID19 PRIMO) dataset [PRIMO, 2021]¹” is constructed to incorporate effects of internet-based and social media trends related to COVID-19 on stock market price movements. The main component of the constructed COVID19 PRIMO dataset is based on Twitter messages. It is well known that news and media move stock prices [Fama, 1998], [Huang, 2020], [Yun, 2019]. Nowadays, information reaches out to the public via different news platforms ranging from newspaper, radio, and television to social media and Internet-based venues. In this area, social media, especially Twitter, is a popular and widely used platform to share personalized opinion on different topics. Twitter is also used extensively by politicians who potentially have high impact on stock price movements. Based on a survey on Statista [Clement, 2020], from the first quarter of 2017 to 2020, Twitter had 186 million active users worldwide.
- Based on the constructed COVID19 PRIMO dataset, we proposed a data-driven (DNN-based) COVID-19 adopted Hybrid and Parallel deep fusion framework for Stock price Movement Prediction (COVID19-HPSMP) that uses information fusion to combine COVID-19 related Twitter data with extended horizon market historical data. More specifically, in contrary to the existing data-driven movement prediction models, where a single DL model is used [Ronaghi, 2020], the proposed COVID19-HPSMP is a hybrid framework with two parallel paths, i.e., one based on Convolutional Neural Network (CNN) with Local/Global Attention modules, and one integrated CNN and Bi-directional Long Short term Memory (BLSTM). The former path is incorporated within the COVID19-HPSMP framework to extract temporal features, while the latter path is used to extract spatial features. The two parallel paths are followed by a multilayer fusion layer acting as a fusion centre that combines localized features extracted in each of the two parallel paths. The COVID19 PRIMO dataset is used to evaluate the performance of the proposed COVID19-HPSMP framework, which illustrates its superior performance compared to its stand-alone (non-hybrid) counterparts.

¹The COVID19 PRIMO dataset is accessible through the following page:
https://github.com/MSBeni/COVID19_PRIMO#COVID19-PRIMO.

1.2 Thesis Organization

The rest of the thesis is organized as follows:

- Chapter 2 provides an overview of the literature on stock movement prediction. In addition, this chapter provides the background required to follow developments presented in the remainder of the thesis.
- Chapter 3 presents the proposed Noisy Deep Neural Network Fusion Framework ND-SMPF for Stock Price Movement Prediction.
- Chapter 4 presents the proposed Hybrid and Parallel Deep Information Fusion Framework COVID19-HPSMP for Stock Price Movement Prediction during COVID-19 pandemic.
- Chapter 5 concludes the thesis and explains some directions for future research works.

Chapter 2

Literature Review and Background

As stated previously, there has been a recent surge of interest in applying Deep Learning approaches for learning and processing financial time series data [Heaton, 2016]. The effectiveness of DNN-based models for different financial prediction tasks is compared with that of traditional techniques in Reference [Khare, 2017]. In this chapter, first, the recent literature on stock movement prediction and portfolio optimization are presented. Then, we present a brief overview of signal processing and deep learning models.

2.1 Literature Review

2.1.1 Stock Movement Prediction

When it comes to stock movement prediction, two measurement approaches can be utilized, i.e., technical or fundamental analysis. From the fundamental viewpoint, stock price movement prediction is performed via measuring derivative and security by examining earnings and other financial factors. On the other hand, technical viewpoint highly relies on charts, plots, and analysis based on historical data in order to predict future outcomes [Schumaker, 2009]. Along this line, some technical methods rely on numerical data like open, close and adjusted prices on trading market, while others focus on the data in financial statements of each company to forecast the market trend, which, however, is time consuming and, typically, fails to reach accurate results.

Technical approaches deal with time series data in order to predict market patterns. One of the

practical models that is widely used for linear and time series data is Auto Regressive (AR) models [Li, 2016]. However, the worldwide stock price is behaving non-linearly and is non-stationary rendering application of AR models impractical. Therefore, recent studies attempt to implement non-linear models [Nayak, 2015] to better catch the underlying complex patterns in market trends. With the emergence of Deep Learning, more efforts have been made to develop DNN algorithms to predict financial trends [Adebiyi, 2014, Kim, 2010, Patel, 2015]. Among different widely used DNN architectures such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), the latter approaches are best fitted for financial time series modeling/prediction problems. Specifically, Long-Short Term Memory (LSTM) is considered as one of the most common approaches used in the prediction of the stock market [Akita, 2016, Gao, 2016].

Dependency of technical approaches on the historical data, which is the main focus of the existing works, makes them incapable of handling high fluctuations and stochastic markets. On the other hand, fundamental models seek to find information such as news, business information and/or financial trends outside of the market. In this regard, there is an increasing trend to use social media resulting in availability of information sources of high volume, variety, and variability. Consequently, news channels as well as different contents extracted from social media (such as Twitter) are fruitful resources that can be leveraged to predict financial trends. Ding *et al.* [Ding, 2015] investigated a deep learning model for extracting news to be used in stock price prediction. They further expanded their model by incorporating a news embedding for learning and triggering events to be used for event embedding. Wang *et al.* [Wang, 2014] predict the volatility of stock prices by implementing a text regression model. Another investigation introduced a novel tree representation to train predictive models by using Support Vector Machine [Xie, 2013]. Li *et al.* [Li, 2014] proposed a model to predict stock price by using sentiment analysis. Although extensive efforts have been put into exploiting the news to predict stock movements, a limited focus is devoted to reliability and quality of such information sources.

Development of DNN-based stock predictive solutions is still in its infancy due to its high dependence on the reliability and quality of the information available through Internet-based news channels and social media resources [Hu, 2019]. Furthermore, such data sources can not be directly used for prediction tasks [Luss, 2015] due to the highly correlated nature of stock market price

movements. To tackle the aforementioned issues, there is an unmet and timely quest to develop and design: (i) Hybrid processing/learning models based on different and diversified learning architectures to capture underlying correlations and variabilities of the data sources, and; (ii) Smart fusion strategies to combine scattered social media news with historical market data. The thesis aims to take the first step towards addressing this gap.

2.1.2 Portfolio Optimization

The volatile nature of the stock market is hard to capture using only the financial indicators, i.e., the stock data, as the market is dependent on a variety of social, political, and economical factors. There are three forms of efficient markets (i.e., weak, semi-strong, and strong), defined based on the nature of information utilized to determine asset prices. In the weak form, asset prices referring to all past prices and trading volume information fully reflect all market data. In the semi-strong form, asset prices reflect all publicly known and available information. In the strong form, asset prices fully reflect all information, including public (e.g., news) and private information. We consider semi-strong form of efficiency for the markets, in which prices reflect the news data and financial indicators. There is a vast literature on portfolio optimization, therefore, we aim to highlight key concepts popular in this area as outlined below:

- One popular practical approach is the reallocation strategy [Wild, 2008] adopted by many pension funds (e.g., LifeStrategy Equity Fund, Vanguard). This approach constructs a portfolio by only investing in stocks and bonds. A typical risk moderate portfolio would, for example, comprise of 60% equities and 40% bonds, and the portfolio needs to be only rebalanced semi-annually or annually to maintain this allocation ratio. The method delivers good performance over the long term, however, the fixed allocation ratio means that investors with a preference for more weight on stocks need to tolerate potentially large drawdowns during dull markets.
- Mean-variance analysis [Markowitz, 1952] is used by many institutional portfolios to solve a constraint optimization problem and derive portfolio weights. Despite its popularity, the theory's assumptions are under criticism as they are often not obeyed in real financial markets.

In particular, returns are assumed to follow a Gaussian distribution. Therefore, investors only consider the expected return and variance of the portfolio return to make decisions. However, it is widely accepted (for instance [Zhang, 2019]) that returns tend to have fat tails, and extreme losses are more likely to occur in practice, leading to severe drawdowns that are not bearable.

- The Maximum Diversification (MD) portfolio is another promising method introduced in [Choueifaty, 2008] that aims to maximize the diversification of a portfolio, thereby, aiming to have minimally correlated assets so the portfolio can achieve higher returns (and lower risk) than other classical methods.
- Stochastic Portfolio Theory (SPT) was recently proposed in [Robert, 2002, Ioannis, 2009]. Unlike other methods, SPT aims to achieve relative arbitrage, meaning to select portfolios that can outperform a market index with high probability. Such investment strategies have been studied in [Daniel, 2010, Johannes, 2013, Leonard, 2015]. However, the number of relative arbitrage strategies remains small, as theory does not suggest how to construct such strategies.

To reflect the effect of this information, models proposed in the thesis use attention mechanisms for the task of news-driven stock price movement prediction. The method of weighting news articles in portfolio optimization is similar to the above works. The stock-specific information was encoded in the neural network, however, making it focused on the price prediction task. In contrast, we represent such stock-specific information by the stock embedding, i.e., a vector, which is easy to interpret geometrically and extract for other applications. For one such application, we evaluated our stock embedding in terms of portfolio optimization. We used the mean-variance minimization portfolio model proposed in Markowitz [Markowitz, 1952], which directly led to the capital asset pricing model Sharpe [Sharpe, 1964].

2.2 Financial Signal Processing

In this Section, we briefly review the basic financial terminologies that are used across the thesis to better communicate ideas and gain insight into the underlying financial problems targeted in this thesis.

2.2.1 Financial Terms and Concepts

Below, we briefly describe different terms commonly used in the financial domain:

- **Asset** is an item that has economic value that individuals, corporation or country owns or control with expectation of future gain. Examples of assets are cash (in hand or in a bank), stocks, loans and advances, accrued incomes, to name but a few. We mainly focus on publicly traded stocks as an asset, but general principles apply to all kinds of assets. We assume that the assets under consideration are liquid, hence they can be converted into cash quickly, with little or no loss in value. Moreover, the selected assets have available historical data for analysis.
- **Portfolio** is a collection of multiple financial assets like stocks, bonds, commodities, cash, and cash equivalents. A portfolio characterized by M assets is defined as follow

$$\begin{aligned} \mathbf{w}_t &= [w_{1,t}, w_{2,t}, w_{3,t}, \dots, w_{M,t}]^T \in \mathbb{R}^M \\ &\text{with } \sum_{i=1}^M w_{i,t} = 1, \end{aligned} \tag{2.1}$$

where superscript T denotes transcript operator. For fixed constituents and portfolio vector \mathbf{w}_t , a portfolio can be treated as a single master asset. Therefore, analysis of single simple assets can be applied to portfolios upon determination of the constituents and the corresponding portfolio vector. Portfolio is a powerful representation of diversification, which tries to reduce risk and maximize returns by investing among various financial instruments, industries, and/or other categories.

- **Return** is the price changes of the asset over time, which reflects the profit or loss of the

investment. The calculation of an asset return is as follow

$$r_{i,t} = \frac{p_t}{p_{t-1}} - 1 \in \mathbb{R}, \quad (2.2)$$

where $r_{i,t}$ is the price change of the i^{th} , for $(1 \leq i \leq M)$, i.e., the change of price p_{t-1} at time $(t - 1)$ to the current price of p_t .

- **Portfolio Return** is calculated as element-wise multiplication of the asset return vector \mathbf{r}_t and the asset's weight vector \mathbf{w}_t defined as

$$R_t^P = \sum_{i=1}^M w_{i,t-1} \cdot r_{i,t}, \quad (2.3)$$

where $r_{i,t}$ is the return of asset i and R_t^P is the total portfolio return. We represent the allocation ratio of Asset i as $w_{i,t} \in [0, 1]$ and $\sum_i^M w_{i,t} = 1$.

2.2.2 Technical Indicators

In this Sub-section, we define technical indicators, which are used to better predict stock movement. Below, we briefly describe main indicator categories:

- **Moving Average Convergence Divergence (MACD)**: The MACD is constructed based on moving average of security's price. It is calculated by subtracting longer Exponential Moving Average (EMA) from the shorter EMA. The EMA calculation is defined as follow

$$EMA_t = EMA_{t-1} + \frac{2}{n} \times (p_t - EMA_{t-1}), \quad (2.4)$$

where EMA_t is the EMA at time t and n is the number of periods used for EMA calculation. The initial EMA is the n -day simple moving average of the series. The most commonly used short and long-period $EMAs$ are 12 and 26-days. When the MACD crosses the zero line from below, a buy signal is triggered, while a sell signal is triggered, when the MACD crosses the zero line from the above [Terence, 2008].

- **Simple Moving Average (SMA)**: The SMA is one of the technical indicators that calculates

the average of a range of prices, by the number of periods in that range, i.e.,

$$SMA = \frac{P_1 + P_2 + \dots + P_n}{n} \quad (2.5)$$

where P is adjusted close price and n is the number of prices in the specific period.

- **Relative Strength Index (RSI):** The RSI is a momentum indicator that measures the magnitude of recent price changes to evaluate overbought or oversold of the stock. The RSI ranges from 0 to 100, where the RSI above 70 means the stock is oversold, while the RSI below 30 is an indication of overbought. When the RSI is above 50, it indicates a “bullish signal”, when the index is below 50, the security is considered “bearish”. The RSI is defined as follows

$$RSI_t(n) = \frac{\sum_{i=0}^{n-1} (p_{t-i} - p_{t-i-1}) 1\{p_{t-1} > p_{t-i-1}\}}{\sum_{i=0}^{n-1} |p_{t-i} - p_{t-i-1}|} \times 100, \quad (2.6)$$

where p_t is the value of adjusted close price at time t , n is the number of RSI period, and $1\{\cdot\}$ is an indicator function, which is equal to 1 when the statement $p_{t-1} > p_{t-i-1}$ is true and is equal to zero otherwise.

- **Stochastic Oscillator (SO):** The SO is one of the popular momentum indicator comparing closing price of the security to a range of its prices over a period of time. The SO is bounded from 0 to 100 and is defined based on the following formula

$$CL_t = p_t - \min(p_t, p_{t-1}, \dots, p_{t-s}) \quad (2.7)$$

$$HL_t = \max(p_t, p_{t-1}, \dots, p_{t-s}) - \min(p_t, p_{t-1}, \dots, p_{t-s}) \quad (2.8)$$

$$RSV = \frac{CL_t}{HL_t} \times 100 \quad (2.9)$$

$$K_t = \frac{2}{3}K_{t-1} + \frac{1}{3}RSV_t \quad (2.10)$$

$$\text{and } D_t = \frac{2}{3}D_{t-1} + \frac{1}{3}K_t, \quad (2.11)$$

where CL_t is the difference between closing price and the lowest closing price during N period, HL_t is measured as the difference between the highest and the lowest closing price, RSV is obtained as CL_t over HL_t . The K_t and D_t value changed due to the highest and

lowest value of closing prices during a period of time. K_t and D_t ranges from 0 to 100. If $K > 80$ and $K > D$, the signal is an overbought signal. The oversold signal occurs when $K < 20$ and $K < D$ [Lixiang, 2004].

- **Rate of Change (ROC)** The ROC is the percentage change in the current closing price and the closing price of n periods ago. The ROC measure is defined as follow

$$ROC = \left(\frac{\text{Close Price}_p - \text{Close Price}_{p-n}}{\text{Close Price}_{p-n}} \right) \times 100 \quad (2.12)$$

2.3 Learning-based Models with Application to Financial Data

In this Sub-section, we briefly describe different learning-based solutions, typically, used within the context of financial econometrics.

2.3.1 GloVe

GloVe is one of the unsupervised learning algorithms that build vectors to represent words in a corpus [Pennington, 2014]. The GloVe algorithm is an unsupervised method for learning word representations in the Euclidean space from statistics of word co-occurrences in a text corpus, with the aim to geometrically capture the words meaning and relations. GloVe starts by constructing a global word-word co-occurrence matrix within a context window of a specific size, e.g., ± 10 . Term $X_{i,j}$ is the number of times word j occurs in the context of word i ; $X_i = \sum_k X_{ik}$; $P_{ij} = X_{ij}/X_i$ refers to the probability that word j appears in the context of word i . An embedding of word i written w_i and an embedding of a context word k is written by \tilde{w}_k . The initial formula of GloVe $w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - \log(X_i)$. Since X_{ik} is symmetric in (i, k) but P_{ik} is not. We constructed vectors with a soft constraint that for each word pair of word i and word j [Pennington, 2014]

$$\vec{w}_i^T \vec{w}_j + b_i + b_j = \log X_{ij} \quad (2.13)$$

where b_i and b_j are scalar bias term associated with word i and j respectively. Finally, we enforce this equality by optimizing a weighted least-square loss

$$J = \sum_{i,j=1}^V F(X_{ij})(\vec{w}_i^T \vec{w}_j + b_i + b_j - \log X_{ij})^2 \quad (2.14)$$

where w is the size of the vocabulary and f down-weights the signal coming from frequent words which is typically chosen by $f(x) = \min\{1, (x/x_m)^\alpha\}$.

2.3.2 Convolutional Neural Networks

Convolutional Neural Network (CNN) is a class of Artificial Neural Networks (ANN) that introduced by Le Cun *et al.* [Boser, 1990] and has become dominant in various computer vision tasks. CNN is designed to automatically and adaptively learn spatial hierarchies of features through back-propagation by using multiple building blocks, such as convolution layers, pooling layers, and fully connected layers. The input signal x_j is computed as follows

$$x_j = \rho(W_j x_{j-1}) \quad (2.15)$$

where W_j denotes a stack of convolutional filters, and term ρ is rectifier (i.e., $\max(x, 0)$). Therefore, the layers are filter maps and each layer can be written as a sum of convolutions of the previous layer [Koushik, 2016], i.e.,

$$x_j(u, k_j) = \rho\left(\sum_k (x_{j-1}(\cdot, k) * W_{j,j_k}(\cdot, k)(u))\right), \quad (2.16)$$

where $*$ is the discrete convolution operator, which is defined as

$$(f * g)(x) = \sum_{u=-\infty}^{\infty} f(u)g(x - u). \quad (2.17)$$

The optimization problem defined by a CNN is highly non-convex. Therefore, typically, the weights W_j are learned by stochastic gradient descent, using the backpropagation algorithm to compute the underlying gradients. In recent years, CNNs have shown success in many Natural Language

Processing (NLP) tasks such as text classification and sentiment analysis [kim, 2015, Aliaksei, 2015, Xiang, 2015]. These works represent that CNNs can extract more general contextual features from texts and the features are successfully used to build recommender systems with less prediction error.

2.4 Summary

In this chapter, an overview of the existing solutions proposed previously in the literature for both stock movement predication and portfolio optimization are presented. Furthermore, the basic financial terminologies that are used across the thesis are briefly outlined to better communicate ideas related to the financial problems targeted in this thesis. Finally, different learning-based solutions used within the context of financial econometrics are outlined.

Chapter 3

ND-SMPF: A Noisy Deep Neural Network Fusion Framework for Stock Price Movement Prediction

As stated previously, there is an urgent need to design and implement advanced fusion strategies [Xu, 2018] to provide robust stock movement predictions. In this regard, the chapter proposes a Noisy Deep Stock Movement Prediction Fusion framework (ND-SMPF) that uses information fusion to combine Twitter data with extended horizon market historical data. To completely utilize market information, a Noisy Bi-directional Gated Recurrent Unit (NBGRU) is utilized coupled with a Hybrid Attention Network (HAN) to extract news level temporal information. We consider a lag of time for the stock movement in order to construct the proposed ND-SMPF fusion framework with the most reliable information. Unlike HAN, which only exploited the news articles, the proposed ND-SMPF model uses both news articles and market prices for the task of SMP. Moreover, to reduce overfitting and improve robustness of the ND-SMPF, zero-centered Gaussian Noise layers are used within the fusion framework. A two level attention layer is used to identify relevant words of highest importance to price movements, which are then fused with historical price data to perform the prediction task. A real dataset is incorporated to evaluate performance of the proposed

ND-SMPF framework, which illustrates its superior performance in comparison to its recently developed counterparts.

The remainder of the chapter is organized as follows: Section 3.1 formulates the stock movement prediction problem. Section 3.2 presents the proposed ND-SMPF framework. Evaluation results are presented and discussed in Section 3.3. Finally, Section 3.4 concludes the chapter.

3.1 Problem Definition and Data Description

Closely observing market movements, can reveal presence of a great amount of trading targets with minor movement ratios. Various methods are offered to monitor the stock price movement, e.g., Mekayel *et al.* [Mekayel, 2019] considers next day close price to set the target. A typical practice is to categorize the changes by setting 1 for movements which face increase known as “Upper Changes,” and 0 for lower thresholds on the price changes [Hu, 2019]. In this chapter, we consider the stock movement prediction problem as a binary classification task based on the adjusted closing price. The adjusted closing price is used in calculating stock earnings and dividends [Xie, 2013]. Moreover, it is used in predicting financial fluctuations or stock price movements [Rekabsaz, 2017].

Our dataset is divided in the two parts consisting of historical prices and Twitters. For design and construction of the proposed ND-SMPF model, we make use of the historical data gathered from stock market. Inherently, those stocks, which are most likely to be traded, would grab a larger share of attention in social media and are more discussed in Twitter. As a result, along with the data of historical prices, we also collect and work on related tweets in the dates from 01/01/2014 to 01/01/2016 for 88 different stocks [Xu, 2018]. We split the dataset into training set from September 2014 to August 2015 and the dataset from 01/08/2015 to 01/10/2015 is separated for the validation phase. Finally, data from 01/10/2015 to 01/01/2016 are kept to be used for testing the model [Xu, 2018].

It is worth mentioning that, the human brain (the most complex part of body) builds a network, which can pay attention differently to various parts of the news according to their importance and influence. Consequently, Twitter news quality are different from each other requiring a diverse user

attention feedback. Even if a news by itself is not important, human brain can completely investigate its different parts and also can differentiate between its important parts and the ordinary negligible ones [Hu, 2019]. We use NLTK package to process Tweets with particular Twitter mode, including tokenization, hashtags and @ identifier. NLTK is an open source algorithm that provides ready to use packages in order to help natural language processing (NLP) to solve linguistic problems [Loper, 2002]. As stated previously, predicting stock price movement depends on its previous trading days. We noticed that there are some other days in a period of time, which imitate the prediction movement really close to a specific date [Caruana, 1998] motivating us to put into consideration those other trading days as well. For example, for a sample movement on 07/08/2012 and a lag of 5 days, 03/08/2012 and 06/08/2012 are qualified as trading days in that period of time and we make predictions for them by using market information [Xu, 2018]. More precisely, we use calendar in our research to split the trading dates from others. We then create our dataframe by including Tweet corpora and historical prices by matching them to these trading days.

3.2 Proposed ND-SMPF Model

3.2.1 Hybrid Attention Network

The proposed ND-SMPF fusion framework interprets and analyzes news within a temporal attention context by concentrating more on important time periods, i.e., the focus is given to distinguish more significant news from the available news corpus. In particular, the proposed model is designed based on “Sequential Context Dependency” and “Diverse Influence” principles. In this regard, the ND-SMPF is built based on a Hybrid Attention Network (HAN) formalization [Hu, 2019] to utilize both news-level and temporal-level attentions. To achieve these goals, a multi-modal fusion framework is constructed by coupling extracted information from social media and stock market prices.

Fig. 3.1 illustrates the overall structure of the proposed ND-SMPF architecture. Each temporal input is defined as

$$x^{(t)} \triangleq \{C^{(t)}, P^{(t)}\} \quad (3.1)$$

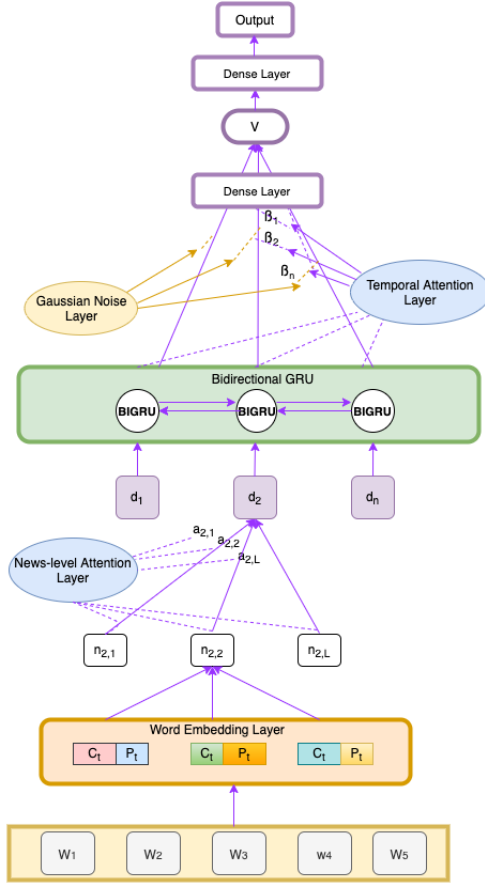


Figure 3.1: Proposed DNN-based Stock Movement Prediction Model (the ND-SMPF).

where $\mathcal{C}^{(t)}$ is the set of corpus embedding, and $\mathcal{P}^{(t)}$ is the set consisting of historical prices [Hu, 2019]. Feeding raw prices cause unusual effects on our model, therefore, as the first step we normalized the open, high, low, adjusted close, and close prices with its last adjusted closing price as

$$P^{(t)} = \frac{P^{(t)}}{P^{(t-1)}} - 1, \quad (3.2)$$

where $P^{(t)}$ is the adjusted close price of the current day and $P^{(t-1)}$ is the adjusted close price of previous day.

As the second step, the available messages are provided as input to a “Message Embedding Layer”, where for each word within the news corpus, we use word embedding layer to calculate the embedded vector, which is then averaged across all the word vectors to construct a news vector.

A pre-trained unsupervised Glove model [Pennington, 2014] is used as the word embedding layer within the ND-SMPF fusion framework. We assign a news level attention to each news vector in a date and calculated the weighted of these corpus. Afterward, these corpus and prices encoded by NBGRU. Finally, another temporal attention is assigned to extract the complete sequential context information. Next, we present the “News Level Attention” layer in more details.

News Level Attention

Capitalizing on the fact that not all the news have the same importance, we construct an attention layer to sum the news weighted by an attention value. In order to reward the information content of news, at first we feed news to the news level attention layer, i.e.,

$$\mathbf{u}_i^{(t)} = \text{sigmoid}(\mathbf{W}_n n_i^{(t)} + \mathbf{b}_n) \quad (3.3)$$

where superscript t denotes the date of the news, while subscripts i refers to the i^{th} news within the available set. Afterwards, the normalized attention weight $a_i^{(t)}$ is calculated through a softmax function as follows

$$a_i^{(t)} = \frac{\exp(\mathbf{u}_i^{(t)})}{\sum \exp(\mathbf{u}_i^{(t)})}. \quad (3.4)$$

Finally, we sum all the news vectors as follows

$$\mathbf{d}^{(t)} = \sum_i a_i^{(t)} \mathbf{n}_i^{(t)}. \quad (3.5)$$

As the result, a temporal sequence vector is computed as the output of the news level attention layer.

3.2.2 Bidirectional GRU

For encoding temporal information embedded within the news corpus and financial time-series, Gated Recurrent Neural Network (GRU) [Chung, 2014] is integrated within the ND-SMPF fusion framework. The GRU is a type of RNN that uses gating mechanism to check the state of sequences without requiring separate memory cells. To encode temporal layer, we adopt Noisy Bidirectional

GRU (NBDGRU) to feed $x^{(t)}$. At Date t , the GRU computes the news state denoted by $h^{(t)}$ by linearly appending the previous state backward $h^{(t-1)}$ and the current updated state \tilde{h}_t given by

$$\begin{aligned}
h_t &= (1 - \alpha^{(t)}) \times h^{(t-1)} + \alpha^{(t)} \times \tilde{h}^{(t)} \\
\tilde{h}_t &= \tanh \left(W_h d^{(t)} + r_t \times (U_h h^{(t-1)}) + b_h \right) \\
r_t &= \sigma(W_r d^{(t)} + U_r h^{(t-1)} + b_r) \\
z_t &= \sigma(W_\alpha d^{(t)} + U_\alpha h^{(t-1)} + b_\alpha)
\end{aligned} \tag{3.6}$$

where $h^{(t)}$ is the news state at Date t ; $\tilde{h}^{(t)}$ is a current updated state, which is a combination of corpus input vector and the previous state; $r^{(t)}$ is a reset gate that controls how much information should be passed for the new state, and; the $\alpha^{(t)}$ is the updated cell, which decides on how much information should be kept from the past and how much new information should be added. For more convenience, we concatenate the latent vector from both directions to form a Bidirectional GRU, i.e.,

$$\begin{aligned}
\vec{h}_i &= GRU(d_i) \quad i \in [1, L] \\
\overleftarrow{h}_i &= GRU(d_i) \quad i \in [1, L] \\
h_i &= [\vec{h}_i, \overleftarrow{h}_i]
\end{aligned} \tag{3.7}$$

where h_i is information from both surrounding context and itself, therefore, encoding the temporal sequence of corpus vectors [Hu, 2019].

3.2.3 Temporal Attention

Since news are published at different dates contribute to stock trend, we design a temporal level mechanism, which incorporates both inherent temporal patterns and the news content to distinguish the temporal differences. The temporal attention layer is designed as follows

$$\mathbf{o}_i = \text{sigmoid}(\mathbf{W}_h \mathbf{h}_i + \mathbf{b}_n) \tag{3.8}$$

Table 3.1: Accuracy comparison between 5 variants of the proposed ND-SMPF framework and 4 Baseline models.

Model Variations	Accuracy	Baselines	Accuracy
<i>ND-SMPF(M-1), using Twitter Data Only</i>	50.26	STOCKNET	58.23
<i>ND-SMPF(M-2), with only News-Level Attention Layer</i>	57.58	RAND-FOREST	53.08
<i>ND-SMPF(M-3), Fusion of 2 Years Twitter data with 2 Years Price Data</i>	58.63	HAN	47.8
<i>ND-SMPF(M-4), with Bidirectional LSTM used instead of BiGRU</i>	54.42	RAND	50.89
<i>ND-SMPF(M-5), Fusion of 2 Years Twitter data with 5 Years Price Data</i>	65.96		

where o_i is the encoded corpus vector. In order to form the attention vector β_i , Parameter θ is used through a softmax layer as follows

$$\beta_i = \frac{\exp(\theta_i o_i)}{\sum_j \exp(\theta_j o_j)} \quad (3.9)$$

Finally, we calculate the weighted sum of v as follows

$$v = \sum_i \beta_i h_i, \quad (3.10)$$

which is used for classification of the sequential news context. The output of the Temporal Attention Layer is then provided as input to fully connected fusion layers where we have incorporated Gaussian Noise layers to improve on robustness of model. This completes presentation of the proposed ND-SMPF fusion framework.

3.3 Experimental Results

In this section, we focus on experimental evaluations p of the proposed stock movement prediction fusion framework. The problem at hand is a classification one, where the expected output is defined as follows: (i) If the adjusted stock price of a specific day is more than the previous day, the output would be 1 denoting the rise, and; (ii) If the expected value for the adjusted price of a day is less than that of the previous day, the output will be 0 showing a fall in the stock value. For evaluation purposes, a 5 days window lag is utilized with batch size of 32 within 15 epochs. We tokenized all the twitter news, remove stop words with the exploitation of the NLTK tool, and also

extracted words occurring less than 5 times in a news in order to build the vocabulary. As long as all the tweets fed into the model simultaneously, we limit the word embedding size to 50. It is worth mentioning that, the latter is considered instead of letting larger sizes in order to control memory cost for the architecture.

We used 100 hidden neurons and use an unsupervised Glove in the message embedding layer and train our model with Adam optimizer [Kingma, 2014] with learning rate of 0.001. During training of the ND-SMPF model, the training inputs are changed due to the changes in parameters of the previous layer and consequently this may slow down the training process [Ioffe, 2015]. To address this issue, we normalized our layers with Batch Normalization for accelerating the training process. Moreover, to reduce overfitting and improve robustness, we utilized zero-centered Gaussian Noise layers within the fusion framework. Finally, Tensorflow [Abadi, 2016] is used to construct the computational graph of our model and find best values for the underlying hyper-parameters.

3.3.1 Baseline and Proposed Models

To contrast our model's performance, we implement the Han [Hu, 2019] model along with our implementations of three different baseline approaches. The results obtained from different variants of the ND-SMPF framework are compared with those computed based on these baseline models. More specifically, the following baseline models are considered for evaluation purposes:

- **RAND** [Xu, 2018]: A simple model that randomly make a guess for price ups and downs. The Rand model selected two year prices of 88 stocks from 01/01/2014 to 01/01/2016.
- **RANDOM FOREST** [Pagolu, 2016]: The random forest classifier proposed in Reference [Pagolu, 2016] is implemented, which uses word2vec text representations. The total number of 250,000 tweets of Microsoft Company are selected from 31/08/2015 to 25/08/2016.
- **HAN** [Hu, 2019]: A recently proposed DNN-based model with Hybrid Attention mechanism, which in contrary to the proposed fusion framework only uses news information. The 1,271,442 data are collected from economic news from 2014 to 2017.
- **STOCKNET** [Xu, 2018]: Another recently DNN model with hybrid attention and temporal

auxiliary mechanism to capture predictive dependencies. In Stocknet, the stocks and tweets are selected from 01/01/2014 to 01/01/2016 for 88 companies.

We have implemented and evaluated the following 5 variants of the proposed ND-SMPF, referred to as $M1$. – $M5$.:

- ***ND-SMPF(M-1), using Twitter Data Only***: As we mentioned previously methods belonging to the fundamental analysis category, only use news information. In this variant of the ND-SMPF framework, we excluded price data and only used Twitter data.
- ***ND-SMPF(M-2), with only News-Level Attention Layer***: In this variant of the ND-SMPF framework, we implement the ND-SMPF model without incorporation of the temporal attention layer.
- ***ND-SMPF(M-3), Fusion of 2 Years Twitter data with 2 Years Price Data***: As we discussed in model definition we run the model directly to see its performance.
- ***ND-SMPF(M-4), with Bidirectional LSTM used instead of BiGRU***.
- ***ND-SMPF(M-5), Fusion of 2 Years Twitter data with 5 Years Price Data***: We had access to 5 years stock price data and in this variant of the ND-SMPF framework, we fed all the price data from 01-09-2012 to 01-09-2017 and then fused them with two years twitter message information.

3.3.2 Performance Evaluation Results

The accuracy of all five models are shown in Fig. 3.2. The accuracy is a fraction of correct predictions to the total number of predictions. The loss function associated with the evaluated models is illustrated Fig. 3.3. Loss function demonstrates the distinction between the output of the model and the target value in order to show the probability of misclassification. We demonstrate the performance of the baseline models and a different variation of the proposed ND-SMPF fusion framework is shown in Table 3.1. The results of Rand, Stocknet and random forest models are obtained from Xu and Cohen [Xu, 2018] and the result of HAN model is obtained from Hu *et al.* [Hu, 2019].

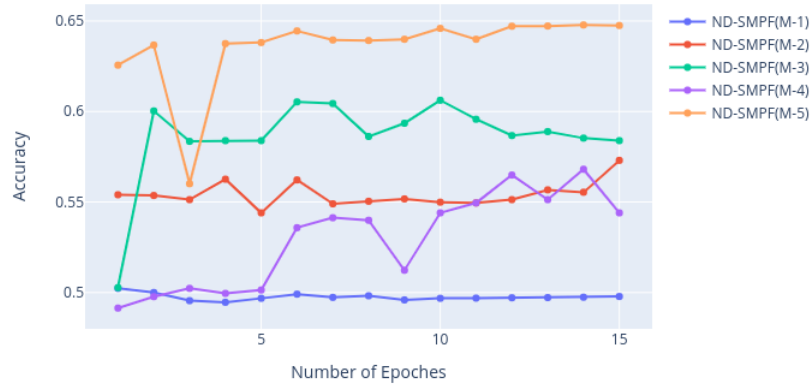


Figure 3.2: Accuracy of the Models

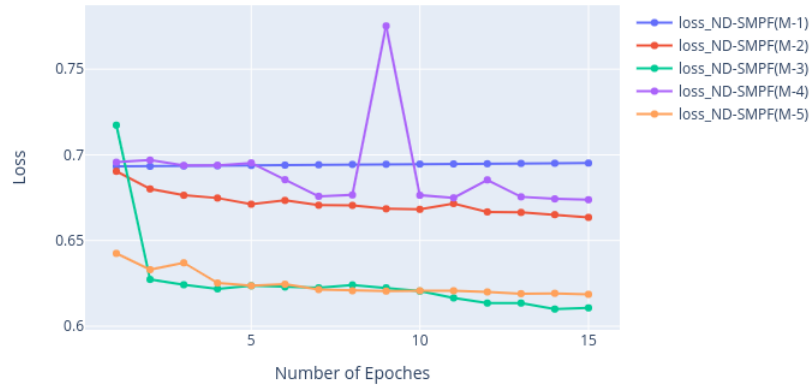


Figure 3.3: Loss of the Models

The accuracy of a news sentiment analysis approach such as the HAN’s model [Hu, 2019], which only uses news information obtained through Twitter (i.e., not incorporating information fusion into account) is 47.8%. The proposed ND-SMPF framework fuses historical prices and twitter news improving the accuracy to 58.63%, which also outperforms the model proposed in Reference [Xu, 2018]. To evaluate effects of the “Temporal Attention State” layer, we run the ND-SMPF model by excluding this layer and only based on news level attention layer achieving accuracy of 57.58%. It can be observed that accuracy of using two level attention layer is slightly higher than its one layer counterpart. To evaluate effects of using Bidirectional GRU, we changed the sequential step of our model and instead of Bidirectional GRU, used to Bidirectional LSTM.

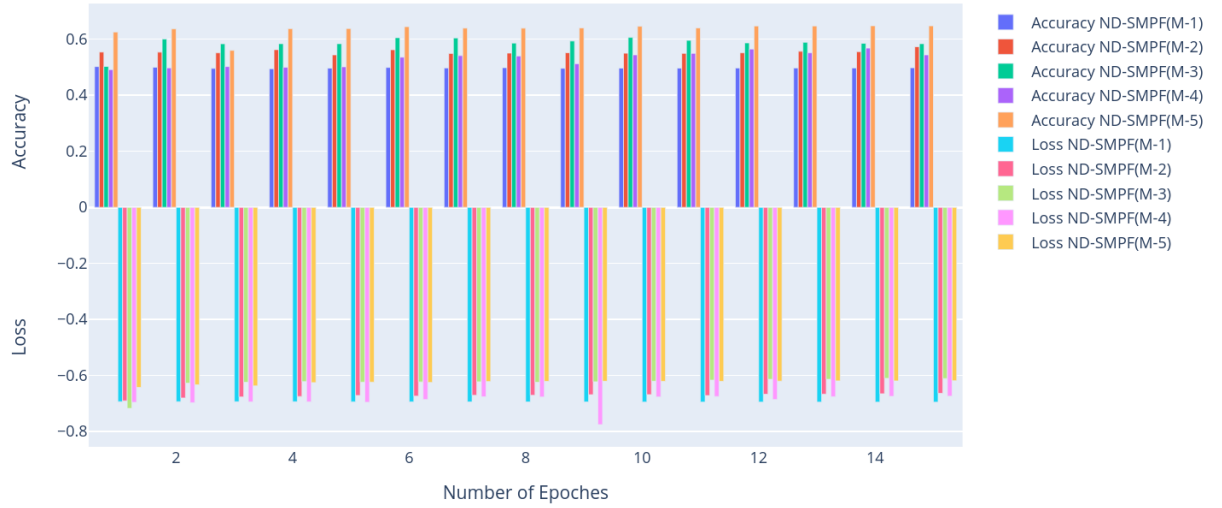


Figure 3.4: Accuracy and Loss contrast of the Models

A relatively high accuracy drop is observed reducing the accuracy to 54.42%. Finally, we fused all 5 years price data with the 2 years twitter news and considering the years without the news information to be zero. A better result is achieved with the accuracy of 65.16% indicating effects of fusing more price data with information extracted from news.

3.4 Summary

Developments of this chapter are motivated by the limited focus devoted in the literature to reliability fusing different available information resources for the task of stock trend movement predictions. In this regard, we proposed the ND-SMPF framework, which is a predictive model that uses information fusion to combine twitter data with extended horizon market historical prices to boost the accuracy of the stock movement prediction task. To completely utilize market information, within the proposed framework, the NBGRU model is coupled with the HAN architecture to extract news level's temporal information. A lag of time is considered for the stock movement to construct the proposed ND-SMPF fusion framework with the most reliable information. Finally, a two level attention layer is used to identify relevant words of highest importance to the price movements, which are then fused with historical price data to perform the prediction task. Evaluation results based on a real dataset illustrate potentials of the proposed ND-SMPF framework.

Chapter 4

COVID19-HPSMP: COVID-19 Adopted Hybrid and Parallel Deep Information Fusion Framework for Stock Price Movement Prediction

The novel of coronavirus (COVID-19) has suddenly and abruptly changed the world as we knew at the start of the 3rd decade of the 21st century. Particularly, COVID-19 pandemic has negatively affected financial econometrics and stock markets across the globe. Artificial Intelligence (AI) and Machine Learning (ML)-based prediction models, especially Deep Neural Network (DNN) architectures, have the potential to act as a key enabling factor to reduce the adverse effects of the COVID-19 pandemic and future possible ones on financial markets. In this regard, first, a unique COVID-19 related PRIce MOvement prediction (COVID19 PRIMO) dataset is introduced in this chapter, which incorporates effects of social media trends related to COVID-19 on stock market price movements. Afterwards, a novel hybrid and parallel DNN-based framework is proposed that integrates different and diversified learning architectures. Referred to as the COVID-19 adopted Hybrid and Parallel deep fusion framework for Stock price Movement Prediction (COVID19-HPSMP), innovative fusion strategies are used to combine scattered social media news related to COVID-19

with historical mark data. The proposed COVID19-HPSMP consists of two parallel paths (hence hybrid), one based on Convolutional Neural Network (CNN) with Local/Global Attention modules, and one integrated CNN and Bi-directional Long Short term Memory (BLSTM) path. The two parallel paths are followed by a multilayer fusion layer acting as a fusion centre that combines localized features. Performance evaluations are performed based on the introduced COVID19 PRIMO dataset illustrating superior performance of the proposed framework.

4.1 Introduction

The novel of coronavirus (COVID-19) has suddenly and abruptly changed the world as we knew at the end of the 2nd decade of the 21st century. The global COVID-19 pandemic caused market volatility [Mazue, 2020, Baek, 2020] rocketing upward around the world. In particular, the pandemic has negatively triggered several sectors including but not limited to stock markets, global supply chains, labor markets, and consumption behaviors. Disruptions of such sectors, especially the stock markets [Bustos, 2020, Al-Awadhi, 2020, Ansari, 2020], can adversely affect the global economy. The United States volatility levels in the mid-March of 2020 are similar to those last seen during October 1987; after 1929 to 1939, and; during in 2008. In September 2008, the Dow Jones Industrial Average fell 777.68 points in intraday trading. During the recent pandemic, in the latter part of March 2020, volatility began to retreat and, by late April, fell sharply but remained well above pre-pandemic levels. It is expected that the emerging markets, the ones with restricted resources to cope with negative impacts of the COVID-19, more substantially feel the COVID-19 pressure due to having slower economic growth and not having sufficient capital inflows. In these sad and unfortunate pandemic times, Artificial Intelligence (AI) and Machine Learning (ML)-based [Radojicic, 2020, Rezaei, 2020, Hoseinzade, 2019, Chong, 2019, Seong, 2021, Zhang, 2021] stock market movement prediction solutions can potentially prevent the pandemic crisis that negatively affecting the market across the world causing unexpected havocs.

Contributions: The chapter makes the following contributions. First, a unique and real COVID-19 related PRIce MOvement prediction (COVID19 PRIMO) dataset [PRIMO, 2021]¹ is constructed

¹The COVID19 PRIMO dataset is accessible through the following page:

to incorporate effects of internet-based and social media trends related to COVID-19 on stock market price movements. The main component of the constructed COVID19 PRIMO dataset is based on Twitter messages. It is well known that news and media move stock prices [Fama, 1998, Huang, 2020, Yun, 2019]. Nowadays, information reaches out to the public via different news platforms ranging from newspaper, radio and television to social media and Internet-based venues. In this area, social media, especially Twitter, is a popular and widely used platform to share personalized opinion on different topics. Twitter is also used extensively by politicians who potentially have high impact on stock price movements. Based on a survey on Statista [Clement, 2020], from the first quarter of 2017 to 2020, Twitter had 186 million active users worldwide.

Based on the constructed COVID19 PRIMO dataset, the chapter proposes a data-driven (DNN-based) COVID-19 adopted Hybrid and Parallel deep fusion framework for Stock price Movement Prediction (COVID19-HPSMP) that uses information fusion to combine COVID-19 related Twitter data with extended horizon market historical data. More specifically, in contrary to the existing data-driven movement prediction models, where a single DL model is used [Ronaghi, 2020], the proposed COVID19-HPSMP is a hybrid framework with two parallel paths, i.e., one based on Convolutional Neural Network (CNN) with Local/Global Attention modules, and one integrated CNN and Bi-directional Long Short term Memory (BLSTM). The former path is incorporated within the COVID19-HPSMP framework to extract temporal features, while the latter path is used to extract spatial features. The two parallel paths are followed by a multilayer fusion layer acting as a fusion centre that combines localized features extracted in each of the two parallel paths. The COVID19 PRIMO dataset is used to evaluate the performance of the proposed COVID19-HPSMP framework, which illustrates its superior performance compared to its stand-alone (non-hybrid) counterparts.

The remainder of the chapter is organized as follows: Section 4.2 introduces the COVID19 PRIMO dataset and formulates the stock movement prediction task. The COVID19-HPSMP hybrid framework is presented in Section 4.3. The implementation study and results are presented in Section 4.4. Finally, Section 4.5 concludes the chapter.

https://github.com/MSBeni/COVID19_PRIMO#COVID19-PRIMO.

4.2 Problem Definition and COVID19 PRIMO

In this section, first, the COVID19 PRIMO dataset is introduced, which is constructed based on the Dow Jones stock market index and its associated Twitter messages for the period of 01/01/2016 to 30/07/2020. The focus is on the problem of stock price movement prediction as close observation of market movements can reveal presence of a significant amount of trading targets with minor movement ratios. More specifically, the chapter focuses on investigating effects of COVID-19 pandemic on stock price movement prediction. In this chapter, stock movement prediction is modeled as a two-class classification problem based on the adjusted closing price of the underlying stocks. The adjusted closing price is commonly utilized to compute the associated stock dividends and earnings [Xie, 2013]. Furthermore, the adjusted closing price is beneficial to learn and predict fluctuations in the stock market [Li, 2014, Rekabsaz, 2017].

We have prepared a new dataset for the aforementioned prediction problem, which can facilitate analysis and evaluation of potential impacts of a pandemic on stock market and can provide priceless insights to combat future possible pandemic. The constructed COVID19 PRIMO dataset consists of two components, i.e., historical prices and Twitter messages. The first component, historical data, is obtained from Dow Jones stock market. The Dow Jones index is a popular and widely watched index in the US market because of its national reputation for quality, reliability, and the ability to operate profitably in high fluctuations. It is a price weighting index that tracks 30 large capital and publicly-owned companies that trade on New York Stock Exchange and the NASDAQ. With the ticker of DJI. Dow Jones is a stock market index that measures the performance of 30 large companies like Apple, Boeing, and Microsoft. Historical stock market prices are obtained from the Yahoo finance. For this task, we used the Yahoo finance library in Python² to collect the data from the Yahoo API. For some of the stocks, we also used Alpha Vantage APIs³. The data is prepared based on three different temporal resolutions, i.e., daily; weekly, and; monthly. The daily prices are used in our model described later for the task of stock movement prediction.

Capitalizing on the facts identified in Section 4.1, for the news component of the COVID-19 price movement prediction dataset, we focused on Twitter. Fig. 4.1 shows the block diagram of the

²<http://pypi.org/project/yahoo-finance/>

³<http://pypi.org/project/alpha-vantage/>

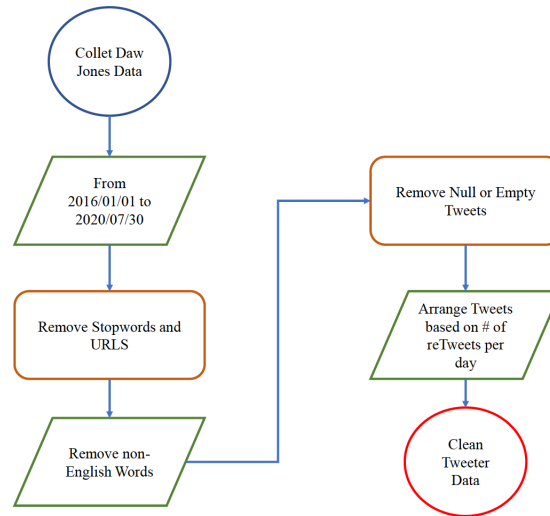


Figure 4.1: Block diagram of the procedure designed to collect and prepare Tweet component of the COVID19 PRIMO.

approach followed to collect and analyze Twitter messages. Web scraping from the Twitter search engine is utilized to build the Twitter dataset. The official API of the Twitter has some limitations that restricts the extent of text that can be extracted. Additionally, the official API of the Twitter cuts the tweets at times, which in turn results in items with missing data. We have developed a localized API to address the aforementioned issues. The localized API uses Twitter search engine and directly collects the required dataset from Twitter. We set up our data collection platform based on scraping the twitter website. The twitter web scraping returns the Tweet text content with a range of useful attributes, for example, *Tweet – ID*, *Tweet Created at*, *Retweet*, *Text*, *Favorite Count*, *Hashtag Text*, *User ID*, *Followers Count*, *Friends Count*, *Statuses Count*, *User Created at*, and *Location*. To collect informative public Tweets, we added a constraint to our implementation to collect tweets retweeted more than once. Many other unnecessary attributes regarding a tweet were also removed from the data gathering session to focus on the essential information such as date, tweet text, and number of retweets. Fig. 4.2 illustrates an illustrative example of raw tweets collected by web scraping.

A critical challenge is scraping the raw content of Twitter data. Such a process takes extensive time and needs manual and cumbersome pre-processing procedures. The preprocessing step consists of the following stages: First, the NLTK package is used to clean the text. Secondly, URL,

Date	Tweets
2019-12-29	All the markets are going mad pump right now. Do you think it's normal not to pump Crypto? \$ dji \$ xau \$ btc
2020-01-23	\$ SPY \$ DJI - MArket perking up. Did the coronavirus get cured?
2020-04-18	Markets trust @realDonaldTrump White House response to coronavirus plague #coronavirus #COVID #COVID19 #POTUS
2020-05-18	Wall Street rises amid hopes of a Covid-19 Vaccine. The Dow Jones, US30 Index is now up 3.64% for the day and possibly aiming higher. Analysis by @analystphil \$ DJI
2020-07-24	Will be covering over the weekend #video the outlook for the #trend in #StockMarket for #futures #stocks #forex #Commodities #gold #CrudeOil #btc \$ spx \$ ndx \$ dji \$ rut #es_f 3203.25 C #nq_f 10457.75 C \$ gc_f \$ cl_f #algorithm #AI #TRADERS #enjoy #weekend be safe #Covid_19

Figure 4.2: Illustrative raw Tweets samples in the COVID19 PRIMO dataset before the pre-processing step.

stopwords and unimportant words are removed and finally, tweets are prioritized based on the number of its retweets to differentiate between the widely seen tweets and the less important ones. We retrieved Dow Jones and COVID – 19 tweets by querying symbols \$*DJI*, #*DOW*, *Covid19*, #*Covid19*, #*Covid*–19 and *CoronaVirus*. Additionally, the corresponding data associated with historical prices are collected. The constructed dataset includes related tweets from 01/01/2016 to 30/07/2020 for the Dow Jones stock index. Not every day is considered as a trading day, i.e., weekends and holidays are not among the trading dates and ought to be out of the analysis scope. To better organize and use the input, we subtract the number of days in a year from the number of weekends, the number of half trading days, and the number of market holidays. More specifically, our dataframe is created by combining historical prices and Tweet corpora and matching them to the trading days. Consequently, we considered 1, 152 trading days from January 2016 to July 2020 to build our dataset. The COVID19 PRIMO dataset is then divided into a training set from January 2016 to January 2020, and a validation set from January 2020 to February 2020. Data from 01/03/2020 to 30/07/2020 is kept to be used for test purposes.

4.2.1 Data Visualization

In this sub-section, we visualize the existing relations between different parameters of the COVID19 PRIMO dataset, particularly that of COVID-19 tweets with other parameters. As stated

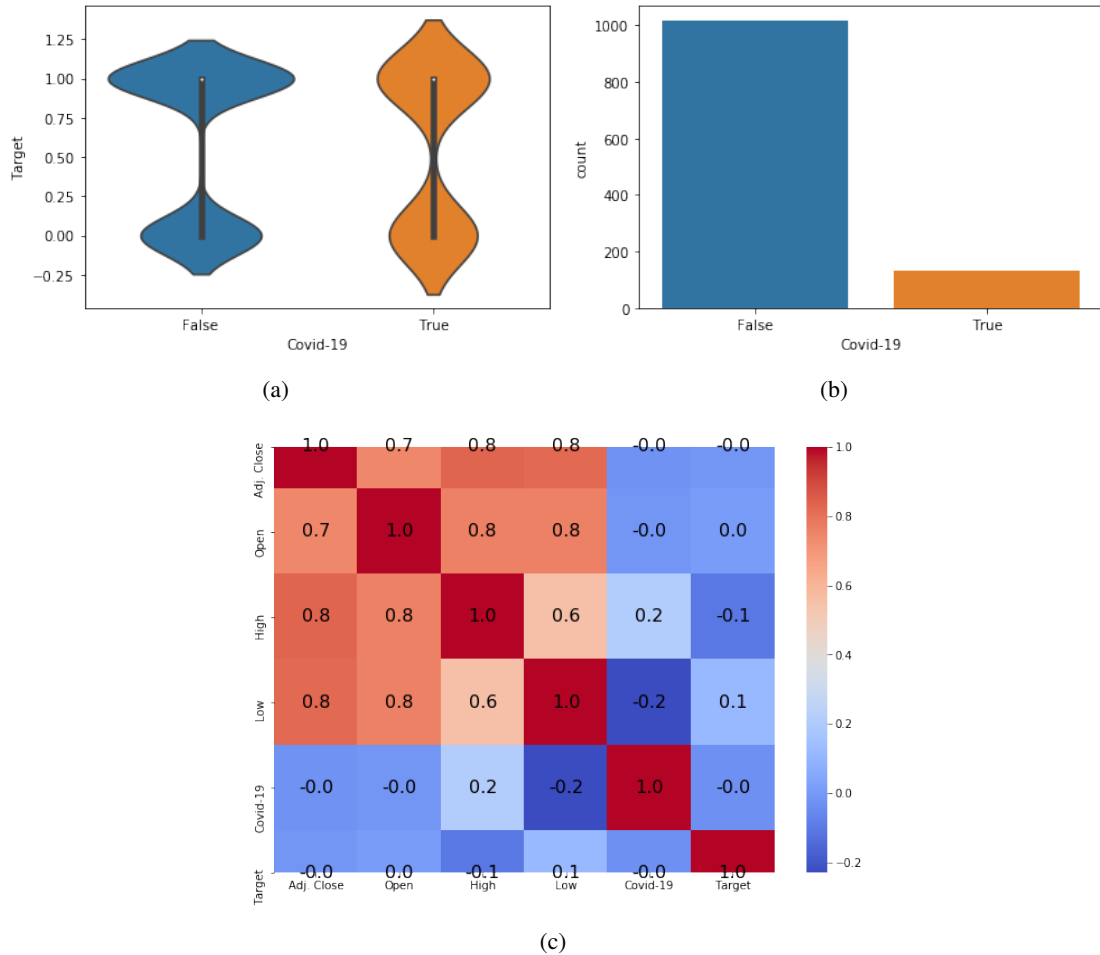


Figure 4.3: (a) The Distribution of The COVID-19 Tweets and Target. (b) The Number of COVID-19 Tweets. (c) Correlation matrix between different variables.

above, the introduced dataset is constructed based on the Dow Jones stock market index and its associated Twitter messages for the period of 01/01/2016 to 30/07/2020. The COVID-19 pandemic crisis covers a fraction of the data represented in the COVID19 PRIMO dataset but plays an essential role in predicting the pandemic's effects on the stock market movements. The COVID-19 related tweets appeared in 2020 (from February to July), specifically starting to show up from the end of February. The dates containing the COVID-19 are stamped with True, making it possible to consider its distribution in the whole frame. Fig. 4.3 visualizes different aspects of the COVID19 PRIMO dataset and illustrates relation of COVID-19 with other parameters of the dataset. Violin plot is shown in Fig. 4.3(a), where the distribution of COVID-19 tweets and the market movements are illustrated. As shown in Fig. 4.3(b), the COVID-19 tweets are less than 200 days, which is expected given the

recent emergence of the pandemic. Fig. 4.3(c) shows the correlation between several different variables affecting the COVID19 PRIMO dataset. Market prices, including Adjusted close price (Adj. Close); Open price (Open); High price (High); Low price (Low); calculated target (Target), and; Presence of COVID-19 in tweets are depicted in this figure. For example, the correlation between normalized adjusted close price and normalized high price is 0.83. Finally, Fig. 4.4 is a grid of scatter plot used to visualize bivariate relationships between combinations of variables. Fig. 4.4 is included to have a big picture of the distribution of the data and better understand existing relations between different parameters in the dataset. Fig. 4.4 shows the relationship for a different combination of variables in a DataFrame as a matrix of plots. The orange dots, show the COVID-19 related data in the dataset, while the blue dots, represent the lack of COVID-19 related data. Fig. 4.4 can potentially depict the bivariate relationships between different market price data and COVID-19 together with the relation between the Target recognized in this period of time with the pandemic data.

4.3 Proposed COVID19-HPSMP Framework

In this section, we describe the constituent components used in the development of the proposed COVID19-HPSMP framework. As stated previously, the main architecture of the COVID19-HPSMP is developed based on DNNs. The prominent advantage of DNNs is their ability to extract meaningful patterns from raw data through multiple non-linear transformations and approximation of complex non-linear functions [Al-Dulaimi, 2019]. More specifically, the proposed COVID19-HPSMP is a data-driven (deep learning model) designed based on hybrid or multiple-model strategies. The COVID19-HPSMP framework extracts and interprets the available news corpus via temporal attention modeling based two key principles, i.e., “Diverse Influence” and “Sequential Context Dependency”. To achieve these objectives, the COVID19-HPSMP is designed as a hybrid multi-modal fusion framework that integrates information obtained from stock market historical data and social media (Twitter data). The proposed hybrid framework consists of three paths, two parallel paths, i.e., the CNN Local/Glocal path, and; the CNN-LSTM path, together with a fusion path. Fig. 4.5

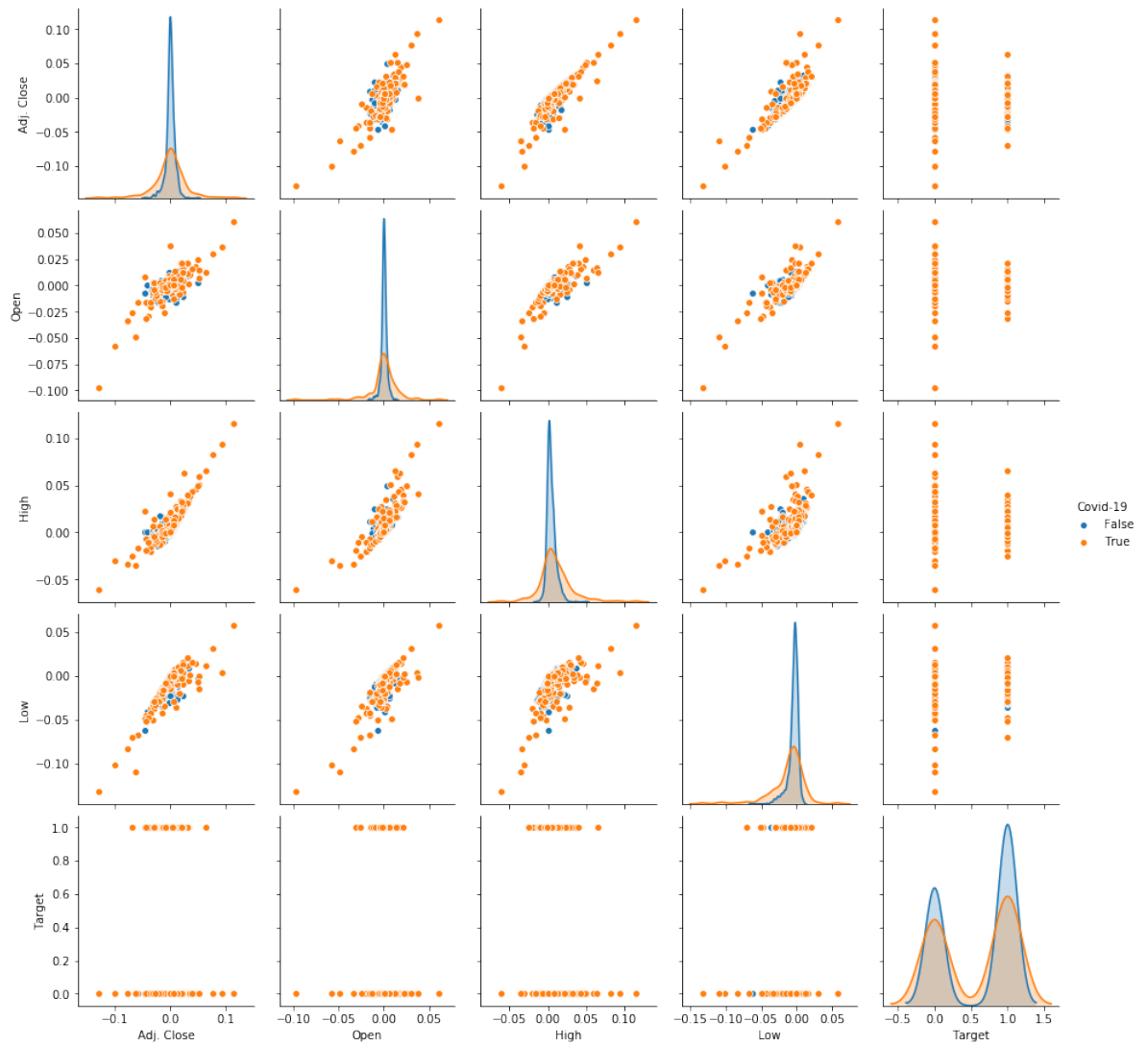


Figure 4.4: Grid of scatter plots. The orange dots show COVID-19 related data points, while the blue dots represent lack of COVID-19 related data.

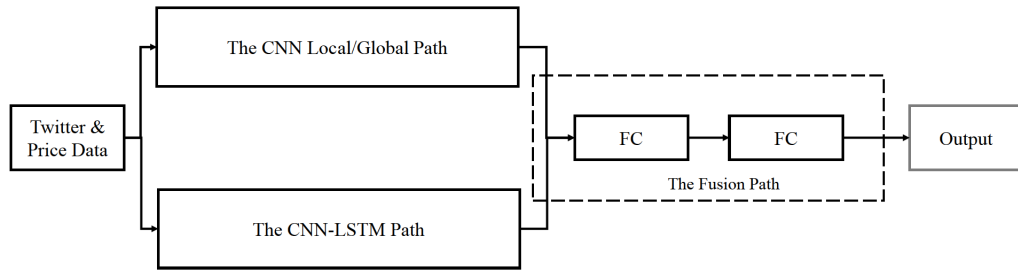


Figure 4.5: The proposed COVID19-HPSMP framework.

illustrates the overall structure of the proposed framework. The fusion path composed of fully connected layers that combine extracted features from each of the two parallel paths. Each of the two parallel paths within the COVID19-HPSMP framework are constructed based on the following two main components:

- (i) *Word Embedding Module*: This module is used to calculate embedded vectors for Twitter data. For this purpose, Glove [Pennington, 2014], as a pre-trained unsupervised model, is used within the word embedding module of each of the two parallel paths within the proposed COVID19-HPSMP framework.
- (ii) *Attention Module*: The main objective of this module is to extract specific words with highest attention weight. The COVID19-HPSMP is a hybrid model where each of the two parallel paths (i.e., the CNN Local/Glocal path, and; the CNN-LSTM path) is a unique attention module extracting different related features. The rationale behind such a hybrid and parallel structure is the significance of the attention network and the intuition that extracting different attention-related features would improve the overall performance of the model.

In what follows, we present each of the three constituent paths of the proposed COVID19-HPSMP framework.

4.3.1 The CNN Local/Global Path

The first parallel path of the proposed COVID19-HPSMP framework is a CNN-based local/global attention model designed to capture and extract spatial features from the input data. More specifically, the CNN-based path consists of Local and Global attention layers, which are described in details below.

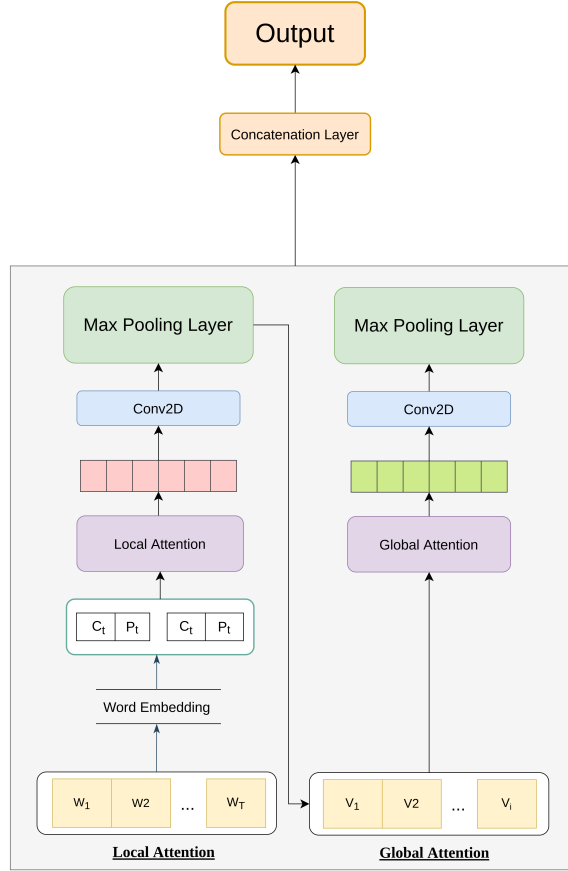


Figure 4.6: The CNN Local/Global path of the COVID19-HPSMP framework.

Local Attention Layer (LAL)

Intuitively speaking, a word embedding model produces representations for each word in the Twitter corpus. Let us denote the l^{th} Tweet among the available set of N_T Tweets with $\mathcal{T}^{(l)}$, for $(1 \leq l \leq N_T)$. Furthermore, consider that Tweet $\mathcal{T}^{(l)}$ contains $N_W^{(l)}$ number of words. The embedding can be thought of as a linear operator (function) that takes as an input a one-hot vector $e_i^{(l)} \in \mathbb{R}^{N_W}$ corresponding to the i^{th} word of Tweet $\mathcal{T}^{(l)}$, for $(1 \leq i \leq N_W^{(l)})$. Note that, here N_W denotes the number of words in the overall vocabulary. The embedding then maps the one-hot vector into a dense feature vector $\mathbf{x}_i^{(l)} = [x_{i,1}^{(l)}, \dots, x_{i,N_F}^{(l)}]^T \in \mathbb{R}^{N_F}$, which consists of N_F scalar features for the i^{th} word of the l^{th} Tweet within the Twitter corpus. The feature vector $\mathbf{x}_i^{(l)}$ is obtained based on a trainable weighting matrix (to be learned) of the embedding layer as follows

$$\mathbf{x}_i^{(l)} = \mathbf{W}^{(Emb)} e_i^{(l)}. \quad (4.1)$$

The embedding layer’s output together with price values are provided as a concatenated input to the Local Attention Layer (LAL). The LAL focuses on the words, which are more informative within a localized window. More specifically, let the l^{th} Tweet be represented by $N_W^{(l)}$ word embedding as $(\mathbf{x}_1^{(l)}, \dots, \mathbf{x}_m^{(l)}, \dots, \mathbf{x}_{N_W^{(l)}}^{(l)})$, where $\mathbf{x}_m^{(l)}$ is the middle (center) word within the embedding sequence of the l^{th} Tweet. Local attention process is achieved via a sliding window of length W rolling over the word embedding sequence of $\mathcal{T}^{(l)}$. Attention score $s_i^{(l)}$ for the i^{th} word of $\mathcal{T}^{(l)}$ is computed based on an attention weighting score $\mathbf{W}_i^{(l,LA)} \in \mathbb{R}^{W \times N_F}$, for $(1 \leq i \leq N_W^{(l)})$, and its associated bias vector $\mathbf{b}_i^{(l,LA)}$ as follows

$$s_i^{(l,LA)} = \sigma\left(\mathbf{X}_i^{(l,LA)} \circ \mathbf{W}_i^{(l,LA)} + \mathbf{b}_i^{(l,LA)}\right), \quad (4.2)$$

where \circ denotes the Hadamard product (element-wise multiplication), $\sigma \cdot$ is the sigmoid activation function, and

$$\mathbf{X}_i^{(l,LA)} \triangleq [\mathbf{x}_{m+\frac{-W+1}{2}}^{(l)}, \dots, \mathbf{x}_i^{(l)}, \dots, \mathbf{x}_{m+\frac{W+1}{2}}^{(l)}]^T, \quad (4.3)$$

where superscript T denotes transpose operator. The attention score $s_i^{(l)}$ is used as a weight for the words to form localized word embedding as follows $\hat{\mathbf{x}}_i^{(l,LA)} = s_i^{(l,LA)} \mathbf{x}_i^{(l)}$. A higher attention score can be interpreted as higher importance associated with that specific word than the others. The weighted sequences then go through a Convolutional layer with a kernel size of 15, which is designed to avoid overfitting. A Max-Pooling layer is then implemented after the convolution one to create position invariance over larger local regions and down-sample the input. Addition of the Max-Pooling layer also leads to a faster convergence rate by selecting superior invariant features, which in turn improves generalization performance.

Global Attention Layer (GAL)

The output of the LAL is provided as input to a Global Attention layer (GAL). This scoring process of the GAL is similar in nature to that of the LAL (Eqs. (4.2)-(4.3)). However, the attention

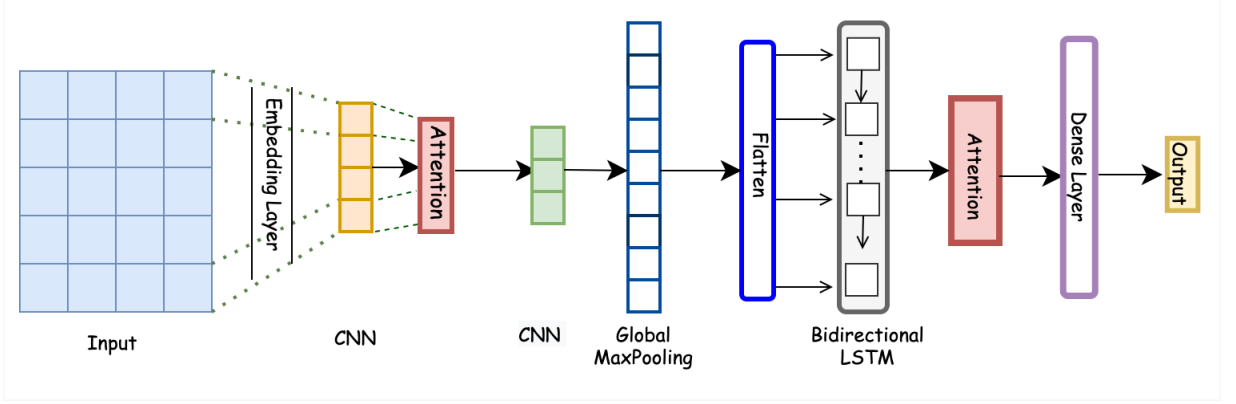


Figure 4.7: The CNN-LSTM path of the COVID19-HPSMP framework.

score, now denoted by $s_i^{(l,GA)}$, is computed through the entire input, i.e.,

$$s_i^{(l,GA)} = \sigma\left(\mathbf{X}^{(l,GA)} \circ \mathbf{W}_i^{(l,GA)} + \mathbf{b}_i^{(l,GA)}\right), \quad (4.4)$$

where $\mathbf{W}_i^{(l,GA)} \in \mathbb{R}^{N_W^{(l)} \times N_F}$, and

$$\mathbf{X}^{(G,Att)} \triangleq [\mathbf{x}_1, \dots, \mathbf{x}_{N_W^{(l)}}]^T. \quad (4.5)$$

By applying global attention, the effect of uninformative words will be diminished, and the global semantic meaning will be captured more precisely through the CNN path. This completes description of the CNN Local/Global path of the proposed COVID19-HPSMP *framework*. Next, we present the CNN-BLSTM path.

4.3.2 CNN-LSTM Attention based Model

The second parallel path of the proposed COVID19-HPSMP is a hybrid CNN and BLSTM attention model, referred to as the CNN-BLSTM path. Similar to the CNN Local/Global path, in the first step, the Twitter messages are provided as input to a “Word Embedding layer”. As stated previously, a pre-trained unsupervised Glove model [Pennington, 2014] is used as the word embedding layer within the COVID19-HPSMP framework. Afterwards, the corpus and prices are encoded by a CNN layer to extract general contextual features. An attention layer is assigned across all the vectors to calculate the weighted corpus. At the next step, a second CNN layer is

implemented to capture and learn more fine-tuned features. The first CNN layer has 50 number of filters with a window size of 25. The second CNN layer has 100 filters with a window size of 25. The first Attention layer is used to capture essential and unique features to provide insight into the vector of the data including tweets and prices. The second Attention layer acts on each vector and calculates the weighted mean of these encoded corpus vectors to represent the overall sequential context information. A global Max-Pooling layer is then applied to capture the essential features and reduce the framework's complexity. Global Max-Pooling is similar to the regular version but with pool size equals to the size of the input. At the next stage, an attention-based BLSTM layer is designed to remember what has been previously learned to better understand the input. The attention-based BLSTM layer is described next.

Attention-based Bidirectional LSTM:

To encode temporal information based on the available set of news corpus and financial time-series data, BLSTM is incorporated within the COVID19-HPSMP hybrid framework. Learning based on financial time-series data is a sequence learning task for which BLSTMs are considered as the state-of-the-art DNN architectures. The LSTM architecture is initially developed by Hochreiter and Schmidhuber [[Hochreiter, 1997](#)] to address the vanishing and exploding gradient problem of conventional Recurrent Neural Networks (RNNs). Since then, LSTM models have gained significant popularity owing to their extensions, advancements and successful applications in different domains. Generally speaking, LSTM is a memory-based architecture that uses different gating functions and a memory state to manage process if information through time [[Di, 2018](#)]. LSTM works

based on the following update model at each time step (denoted by t)

$$\mathbf{i}_t = \sigma(\mathbf{W}_i x^{(t)} + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i) \quad (4.6)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f x^{(t)} + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f) \quad (4.7)$$

$$\mathbf{g}_t = \tanh(\mathbf{W}_c x^{(t)} + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c) \quad (4.8)$$

$$\mathbf{c}_t = \mathbf{f}_t \circ \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t \quad (4.9)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o x^{(t)} + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o) \quad (4.10)$$

$$\mathbf{h}_t = \mathbf{o}_t \circ \tanh(\mathbf{c}_t) \quad (4.11)$$

where \mathbf{W}_i , \mathbf{W}_o , \mathbf{W}_f , \mathbf{W}_c , and \mathbf{U}_i , \mathbf{U}_o , \mathbf{U}_f , \mathbf{U}_c are weight matrices; Terms \mathbf{b}_i , \mathbf{b}_o , \mathbf{b}_f , \mathbf{b}_c are bias vectors, and; $\tanh(\cdot)$ represents element-wise hyperbolic tangent activation function. Furthermore, $\mathbf{h}^{(t)}$ and \mathbf{h}_{t-1} represent the current and previous hidden states, respectively. In the context of the proposed COVID19-HPSMP and to encode the temporal layer, we adopt Bidirectional version of the LSTM (BDLSTM) to feed the i^{th} word embedding. BLSTM can access both the preceding and succeeding contexts. It separates the hidden layer into two parts, forward state sequence and backward state sequence based on an iterative process.

4.3.3 Fusion Path

The final component of the proposed COVID19-HPSMP framework is the Fusion Path with three fully connected layers for fusing features extracted from each of the two underlying parallel paths and performing the final price movement prediction task. The first fusion layer has 100 number of neurons and uses “tan” activation function, while the second fusion layer has 50 number of neurons with the same activation function. The final layer of the Fusion Path, has 1 neuron and uses Rectified Linear Unit (ReLU) as its activation function to produce the price movement predictions. The input to the Fusion Path is constructed by concatenating the output of the CNN Local/Global path, which is a flattened 1-Dimensional feature vector, with that of the CNN-BLSTM path.

As a final note, we would like to briefly focus on trading from financial perspective to better

link the COVID19 PRIMO (presented in Section 4.2) and the COVID19-HPSMP framework (discussed in the section). The COVID19 PRIMO provides means to trade based on 5-day in-advance predictions (provided by the COVID19-HPSMP). Generally speaking, a common reason for purchasing, conventionally used by traders and stockholders, is to only buy a stock when its prediction is increasing as this provides the chance to sell at a higher price in future and gain profit. This future to sell a stock, varies for different investors, e.g., some consider a short period trade, some are more eager to mid-term trades, and others focus on long-term trades. A sentimental engine that can infer the market's sense and predict its movements is a significant asset. In other words, the "BUY" and "SELL" decisions can be made more successfully when an accurate movement prediction model is available.

For use of COVID19-HPSMP as a trading engine over the COVID19 PRIMO, we consider the following two trading approaches:

- *The 50-50 Trading Approach:* Whenever the movement prediction value is less than 0.5 (i.e., the model senses a plunge in the market's future trends) the "SELL" signal will be sent to the trader. Alternatively, when the movement prediction value is greater than 0.5 (i.e., the model predicts a boost in the market price) it will generate a "BUY" signal. We refer to this algorithm as the 50-50 approach. The traders have a chance to buy stocks at a lower price before the price goes up and sell them at a higher price. Based on these signals, provided by the COVID19-HPSMP framework, the traders have insightful predictions for the next five days. In other words, based on the COVID19 PRIMO, 5-day in advance predictions can be computed allowing decisions to buy a stock for their portfolio (when predictions show that market will face a "surge") or vice versa.
- *The 60-40 Trading Approach:* As an alternative approach, we consider a BUY and SELL with a HOLD scenario, by defining new thresholds over the market movement predictions. Based on this new approach, if the movement prediction value is less than 0.4 we consider a decrease in the market price. For the values greater than 0.6, the increase in the market price is predicted. For the values between 0.4 and 0.6, the HOLD signal will be offered. We refer to this model as the 60-40 approach.

Table 4.1: Accuracy, Sensitivity, and Specificity comparisons between the proposed COVID19-HPSMP framework, those of the stand-alone models, and accuracies of algorithms from References [Hu, 2019, Xu, 2018].

Model	Sensitivity	Specificity	Accuracy
The COVID19-HPSMP Framework	60	71.11	66.48
Standalone CNN Local/Global Model	56	72.5	64.65
Standalone CNN-LSTM Model	54.35	66.67	62.06
Han [Hu, 2019]	-	-	47.8
Stocknet [Xu, 2018]	-	-	58.23

4.4 Experiments

Experimental results and comparisons are presented in this section to evaluate the proposed hybrid COVID19-HPSMP framework for the task of stock movement prediction. As stated previously, the problem at hand is a classification one with the following expected outputs: (i) On one hand, within a 5 days prediction horizon, if the adjusted stock price of a specific day is more than that of the previous day, the output of that specific day would be 1. Then, the sum of the output values is computed over the 5 days horizon and if the sum is greater than a pre-defined threshold of 3, we consider the final output for that 5 day horizon to be 1, denoting a rise, and; (ii) On the other hand, when the adjusted stock price associated with a specific day is less than its previous day, value 0 is assigned as the output of that specific day. When the number of such 0 output values within the 5 days window is more than 3, we consider the final output to be 0, representing the fall prediction/state.

4.4.1 COVID19-HPSMP LSTM-based hybrid attention Model

To perform the evaluations, the available Twitter news corpora is tokenized and words occurring less than 5 times are removed to construct the vocabulary. It is worth noting that removing words with limited usage will reduce the associated memory cost of the DNN models. As stated above, we consider a five day horizon and used a batch size of 64 within 15 epoch. In addition, Glove, which is an unsupervised word embedding algorithm, is used within the embedding modules of the two parallel paths of the COVID19-HPSMP. For comparison purposes, three different models are implemented as follows:

- (ii) *The proposed COVID19-HPSMP Framework*: The proposed hybrid COVID19-HPSMP framework developed in Section 4.3 is the first implemented stock movement prediction model. The COVID19-HPSMP consists of 2 parallel paths and a fusion path integrating extracted features of each of the two parallel paths.
- (ii) *Stand-Alone CNN Local/Global Model*: The second implemented movement prediction model is the CNN Local/Global path implemented independently (stand alone as a single model). To implement the stand alone version of the CNN Global/Local model, initialization is performed following the guideline provided in Reference [Seo, 2017]. A pre-trained Glove [Pennington, 2014] is used for weighting corpus within the word embedding layer. In the LAL, we use window of size 5 with a sigmoid function (σ). Total of 80 filters are implemented within the LAL. In the GAL, we used 50 filters of length 2 and 3. Finally, a fully connected layer with 0.5 dropout is designed to form the output.
- (iii) *Stand-Alone CNN-BLSTM Model*: The third implemented movement prediction model is the CNN-BLSTM path implemented independently as a single model. Similar to the stand-alone CNN Global/Local model, a pre-trained Glove [Pennington, 2014] is used for weighting corpus within the embedding layer. A convolutional layer with a 64 number of filters and a window size of 25 is followed by an attention layer. To extract essential features and reduce the framework's complexity, a max-pooling layer is designed. The output of max-pooling layer is the input of next layer which is attention-based Bidirectional LSTM with 250 hidden layers. Finally, two fully connected layers are considered with 300 and 1 number of hidden neurons, respectively, to form the price movement prediction results.

These three implemented models are trained with Adam optimizer [Kingma, 2014] with a learning rate of 0.001. To reduce the training times of the implemented models, Batch Normalization [Ioffe, 2015] is utilized to normalize the underlying layers. Furthermore, to avoid overfitting issues and improve the overall robustness of the implementations, 0.5 dropout is used within the fully connected layers. Finally, the computational graphs of the implemented models are constructed via Tensorflow [Abadi, 2016] to fine tune different hyper-parameters.

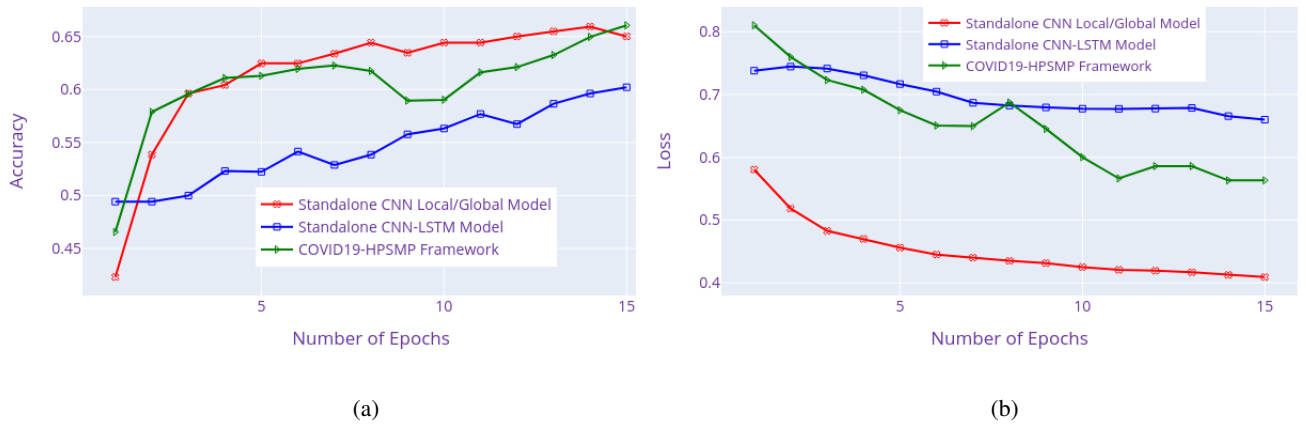


Figure 4.8: (a) Accuracy of the Models. (b) Loss of the Models.

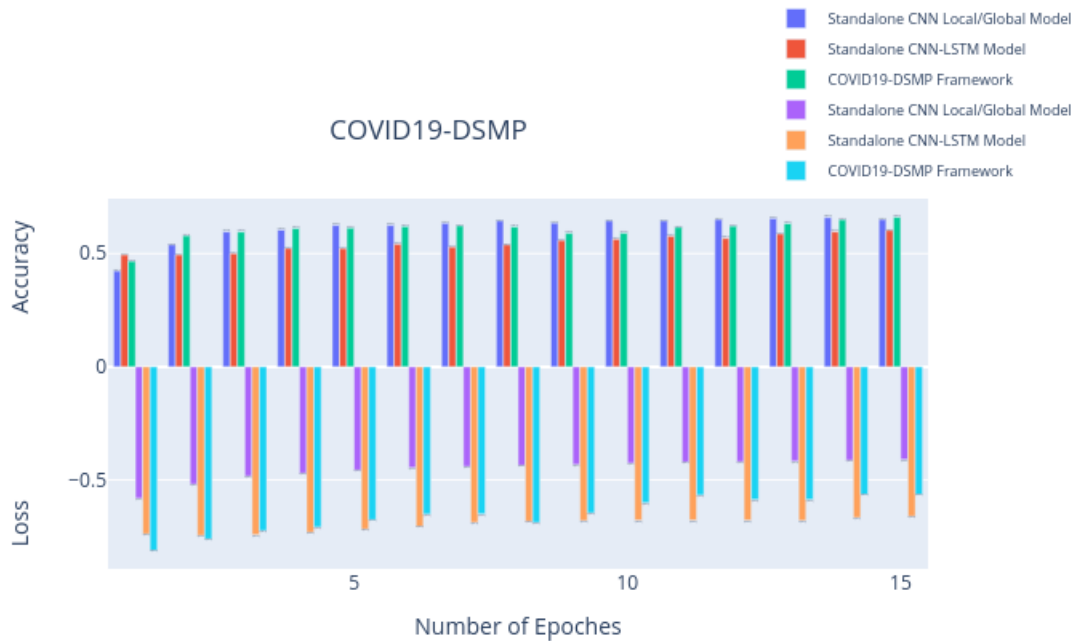


Figure 4.9: Accuracy and loss contrast of the different price movement prediction models.

4.4.2 Performance Evaluation/Results

In this sub-section, we represent different experimental results to evaluate the performance of the proposed COVID19-HPSMP framework for the stock movement prediction. The accuracy of the proposed models areas follows: 64.65% for the stand-alone CNN-based local/global; 62.06% for the stand-alone CNN-LSTM, and 66.48% for the hybrid attention model, i.e., the COVID19-HPSMP

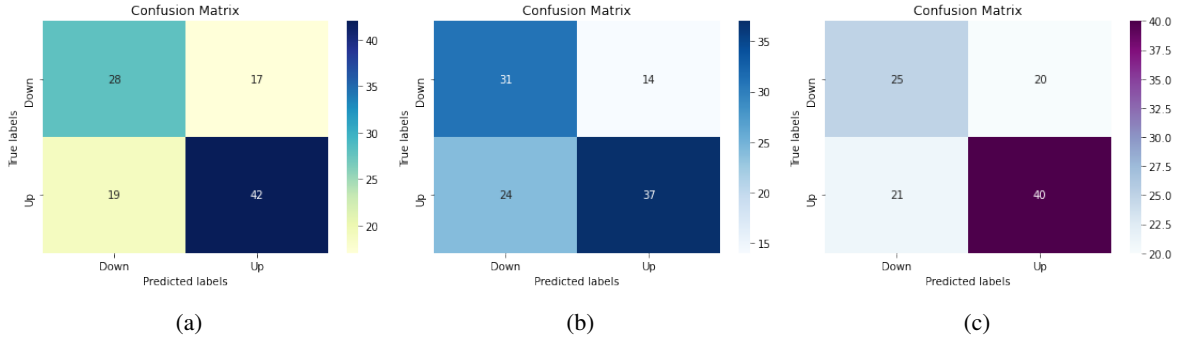


Figure 4.10: Confusion matrices for binary predictions. (a) The COVID19-HPSMP Framework. (b) Standalone CNN Local/Global Model, (c) Standalone CNN-LSTM Mode.

framework. The accuracy of all three implemented models are shown in Fig. 4.8(a). The accuracy is a fraction of correct predictions to the total number of predictions. The loss function associated with the evaluated models is illustrated Fig. 4.8(b). Loss function demonstrates the distinction between the output of the model and the target value in order to show the probability of misclassification. In addition to the accuracy measure, we have computed sensitivity and specificity metrics to evaluate different error types (i.e., False Negative (FN), and False Positive (FP)). Sensitivity represents the number of days that stock price has an upward trend correctly predicted by the model. Specificity, on the other hand, is illustrating model's performance in identifying negative conditions. More specifically, if the stock movement test indicates that the price movement has an upward trend while the trend is downward in the actual scenario, the test result is considered to be a FP. Likewise, if the test result indicates a descending trend, but the actual market price is increasing, the result is FN. Sensitivity and specificity are then defined as follows

$$\begin{aligned}
 \textit{Sensitivity} &= TP/(TP + FN) & (4.12) \\
 &= (\# \text{ of True GoUp Assessments})/(\# \text{ of all GoUp Assessments})
 \end{aligned}$$

$$\begin{aligned}
 \textit{Specificity} &= TN/(TN + FP) & (4.13) \\
 &= (\# \text{ of True GoDown Assessments})/(\# \text{ of all GoDown Assessments}),
 \end{aligned}$$

where True Positive (TP) happens when a stock price goes up and the model also indicates the up-trend. On the other hand, True Negative (TN) refers to the scenario where the market indicates that

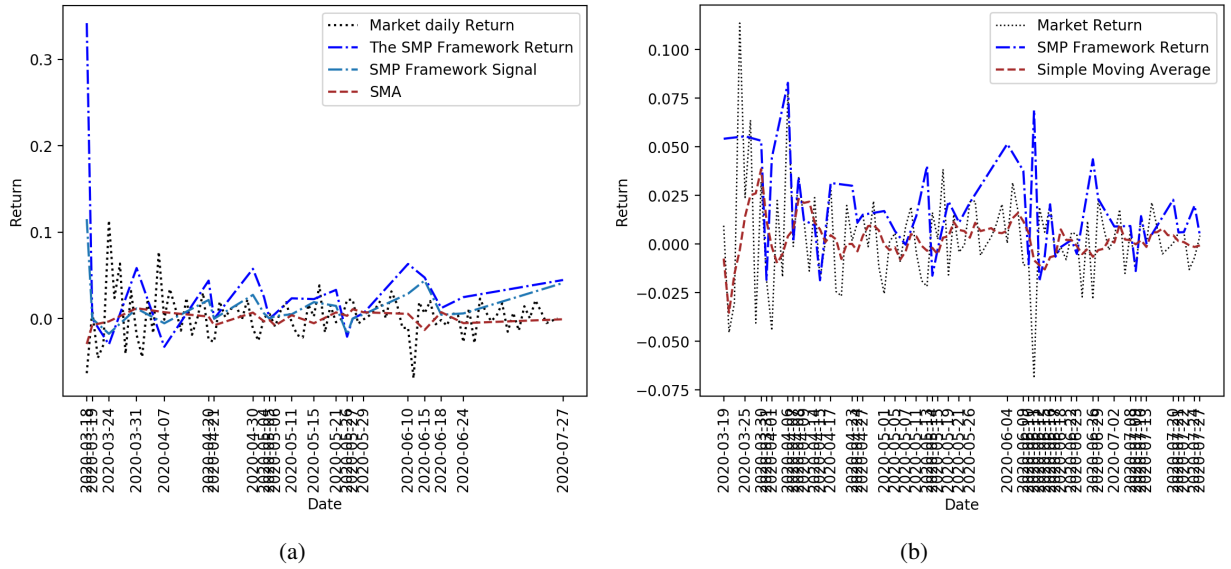


Figure 4.11: (a) Buy/Sale Return based on the 50-50 approach, i.e., whenever the movement prediction value is less than 0.5 the “SELL” signal will be sent to the trader. Alternatively, when the model predicts a boost in the market price (movement prediction value is greater than 0.5) “BUY” signal will be generated. (b) Buy/Sale Return based on 60-40 approach, i.e., when the movement prediction value is less than 0.4 we consider a decrease in the market price sending the “SELL” signal. When the prediction values are greater than 0.6, the “BUY” signal will be sent. For the values between 0.4 and 0.6, the HOLD signal will be offered.

stock goes down and the model shows the downward movement as well. The result of our proposed models are shown in the revised Table 4.1. In particular, we should point out that the higher the numerical value of sensitivity, the less likely the stock movement model returns FP results. The results show that the proposed hybrid model provides considerable improvement in terms of sensitivity, compared to its stand-alone counterparts. Fig. 4.10 illustrates confusion matrices [Bhatt, 2021] associated to the proposed COVID19-HPSMP framework and its two stand-alone components (implemented individually).

As it can be observed, the hybrid model (the COVID19-HPSMP) outperforms its counterparts. It is worth mentioning that the achieved accuracy of 66.48% is significant, although in absolute terms it seems to be low. First, please note that average accuracies achieved in the literature for the task of price movement prediction is around 50%. Second, these lower accuracies are obtained based on a much wider window of information compared to the limited duration of the introduced COVID19 PRIMO dataset. The limited duration of the dataset is due to recent emergence of the COVID-19 pandemic.

Table 4.2: Experiment results for trading based on the COVID19-HPSMP framework following the 50-50 approach.

The COVID19-HPSMP Framework					
Companies	Profit	Return	Weight	E(R)	Std(R)
Coca-Cola	8,009	44%	8%	3.5%	11.5%
Caterpillar	3,790	25%	18%	4.5%	8.7%
Visa	10,606	41%	28%	11.5	15.6
Int. Business Machines Corp.	7,291	39%	18%	7%	13.6%
Walt Disney	3,712	19%	17%	3%	6.2%
Apple	4,804	36%	11%	4%	10.6%
Standalone CNN Local/Global Model					
Coca-Cola	1,954	20%	8%	1.6%	5.2%
Caterpillar	566	26%	18%	4.75%	9.2%
Visa	2,269	24%	28%	6.7%	9%
Int. Business Machines Corp.	1,192	22%	18%	3.9%	7.6%
Walt Disney	483	6%	17%	1.02%	2.2%
Apple	1,531	16%	11%	1.8%	4.7%
Standalone CNN-LSTM Model					
Coca-Cola	1,651	17%	8%	1.4%	4.4%
Caterpillar	478	22%	18%	4.1%	7.3%
Visa	1,917	20%	28%	5.6%	8.6%
Int. Business Machines Corp.	1,007	19%	18%	3.4%	6%
Walt Disney	408	5%	17%	9%	1.8%
Apple	1,293	14%	11%	1.5%	4.1%

Table 4.3: Experiment results for trading based on the COVID19-HPSMP framework following the 60-40 approach.

The COVID19-HPSMP Framework					
Companies	Profit	Return	Weight	E(R)	Std(R)
Coca-Cola	10,448	46%	8%	3.7%	4.9%
Caterpillar	6,896	41%	18%	7.3%	10.2%
Visa	12,590	62%	28%	17.3%	23.6%
Int. Business Machines Corp.	8,742	43%	18%	8%	14.8%
Walt Disney	3,549	19%	17%	3.2%	3.5%
Apple	7,779	59%	11%	6.5%	10.4%
Standalone CNN Local/Global Model					
Coca-Cola	4,978	37%	8%	2.9%	9.6%
Caterpillar	720	28%	18%	5%	9.8%
Visa	3,711	26%	28%	7.3%	10%
Int. Business Machines Corp.	2,304	32%	18%	5.7%	7.1%
Walt Disney	1,414	9%	17%	1.5%	1%
Apple	3,236	0.24	11%	2.6%	7.1%
Standalone CNN-LSTM Model					
Coca-Cola	3,931	21%	8%	1.7%	5.4%
Caterpillar	764	25%	18%	4.5%	11.7%
Visa	2,027	26%	28%	7.2%	9.9%
Int. Business Machines Corp.	2,812	23%	18%	4.1%	8.1%
Walt Disney	880	8%	17%	1.4%	2.7%
Apple	2,472	17%	11%	1.9%	5%

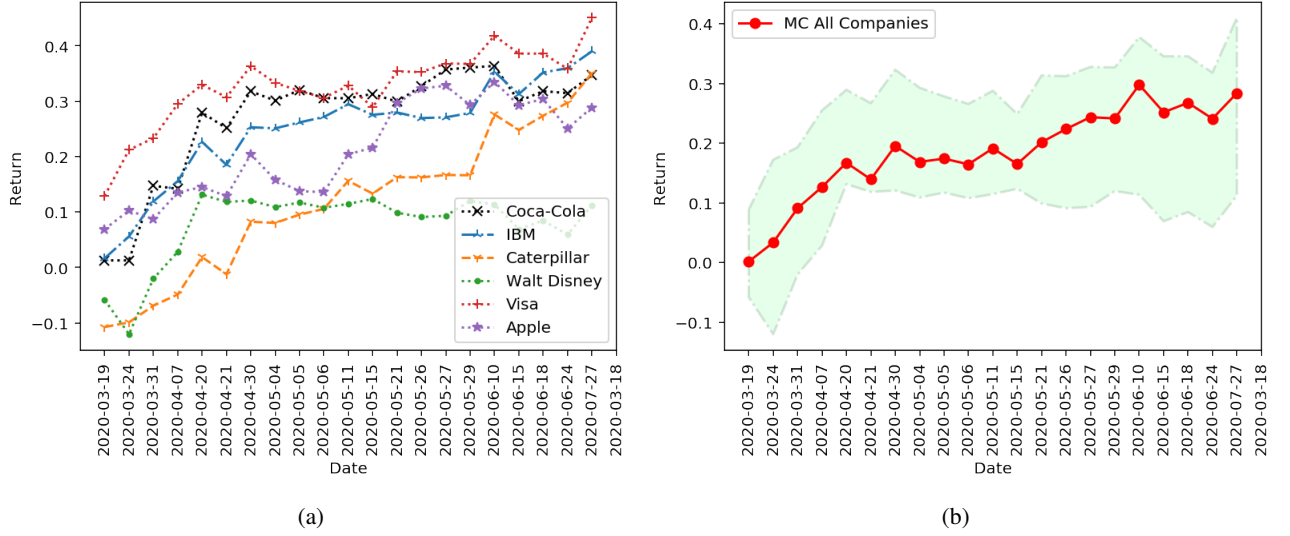


Figure 4.12: (a) Per-stock returns for the first trading mechanism over the following six prominent companies: *Apple*, *Walt Disney*, *IBM*, *Visa*, *Caterpillar*, and *Coca – Cola*. (b) Results of MC simulations over all companies in Dow Jones.

Stock Trading Experiments: In this part, to evaluate effects of the classification prediction mechanism on the investors’ return, we conduct a back-testing experiment by simulating stock trading for March 2020 to the end of July 2020, which is during the COVID-19 financial crisis period. Investors’ return [Zhang, 2020] is computed as follows

$$R_t = w_{t-1} \cdot r_t \quad (4.14)$$

$$\text{with } r_t = \frac{p_t}{p_{t-1}} - 1, \text{ and } w_t \in [0, 1] \quad (4.15)$$

where R_t denotes the final return of the day, and r_t is the return of the day based on the adjusted close price. In the proposed COVID19-HPSMP framework, term w_t represents the baseline prediction of the market’s movement. For evaluation purposes, we consider the following three trading mechanisms:

- **First Trading Mechanism:** We consider a scenario with 100 stock shares of a specific company on 01/03/2020. Total of 88 stocks were already chosen in the literature [Xu, 2018] and we added more than 12 stocks considering the Dow Jones market to be analyzed as well. All these stocks are selected based on different market sectors, i.e., Energy, Materials, Industrials, Utilities, Healthcare, Financial, Consumer Discretionary, Consumer Staples, Information

Technology, Communication Services and Real Estate. In this regard, first, we perform a selective test over these companies from different sectors, and then conduct a Monte Carlo (MC) simulation to choose six profitable ones. In both cases and to better simulate a real-world trading scenario, we consider a transaction cost of 0.3% for each trading. We have calculated risk-adjusted return based on the Sharpe Ratio (SR) approach. In summary, the final profit and SR of our COVID19-HPSMP engine is 6,369 and 1.07%, respectively, for the 50-50 approach. The final profit and SR for the second trading mechanism (60-40) is 8,813 and 2.49%, respectively. Generally speaking, SRs greater than one are preferred with higher the SR value, the better the risk to return scenario for the investors. Furthermore, we consider that the stock shares will be available to buy and sell at the moment they are requested. For the first scenario, the following six prominent companies are selected: *Apple*, *WaltDesney*, *IBM*, *Visa*, *Caterpillar*, and *Coca – Cola*. Fig. 4.11(a) shows the return obtained based on the COVID19-HPSMP framework and the 50-50 trading approach, compared to the market return. Fig. 4.11(b) illustrates the results of this 60-40 trading algorithm. As can be seen, the returns obtained based on the 60-40 approach are higher than the return values gained from the market, standalone CNN Local/Global Model, and standalone CNN-LSTM Model. The average market return for the 60-40 approach is 0.497 which shows improvements over the 50-50 algorithm.

In Fig. 4.11, the “BUY” and “SELL” signals are also shown with green and red triangles, respectively. As can be seen, the COVID19-HPSMP framework closely follows the historical data showing the real market return. Furthermore, the COVID19-HPSMP engine, in most scenarios, correctly predicts the trend of the market in the future and produces timely “BUY” and “SELL” signals. We should point out that the profit is calculated as the difference between the Adjusted Close price of Sell and Buy signals after deducting the transaction cost and risk-adjusted return. Further details on the per-stock returns are provided in Table 4.1 for the 50-50 trading approach, and Table 4.2 for the 60-40 trading approach. In brief, the final profit and Sharpe Ratio (SP) of the proposed COVID19-HPSMP engine based on the 50-50 approach is 6,369 and 1.07% respectively. The 60-40 trading approach provides total profit of 8,813 with SR of 2.49%, which means that, overall, the proposed model performs considerably well.

The COVID19-HPSMP model makes the highest profit for Visa Company, which is 10,606\$ and the lowest profit for Walt Disney Company is 3,712\$. Based on the results, we construct a portfolio to measure the overall profitability of the six selected stocks. It is assumed that 100 shares of each stock are bought, and the trade is made based on the proposed COVID19-HPSMP buy/sell signals. The final profit of our COVID19-HPSMP engine is 6,369\$, which means that, overall, our model performs considerably well. In a second scenario, MC simulation of 100 runs is performed considering the profit of all the companies in the Dow Jones. In each MC run, six stocks are randomly selected and the procedure outlined above for the first scenario is applied. As can be seen in Fig. 4.12(b), based on the MC results, the proposed method offers a positive return over this COVID-19 crisis period and can provide the users with a profit despite being in a pandemic area.

With regards to Figs. 4.11 and 4.12, we would like to mention that the market return is computed based on daily trading while the return calculated by the COVID19-HPSMP Framework is based on Buy/Sell signals that the proposed model has produced. In other words, the results shown in the aforementioned figures include the market daily return ($r_t = \frac{p_t}{p_{t-1}} - 1$), which corresponds to using a daily buy/sell signal. While at times daily trading results in higher return, the proposed model achieved average return of 0.453, which is significantly higher than the average daily market return of 0.004. It is interesting to observe that the COVID19-HPSMP Framework performs better as the model sees more market data. For instance, due to the COVID-19 crisis, abnormal market fluctuations started at the middle of March 2020 and continued until April 2020. Some false predictions in those periods are due to these fluctuations since our model has not seen these type of market reactions during its training phase. After learning from these initial abnormal market fluctuations, the model gradually performs better in predicting the market trend. For example, it can be observed that there is a sharp drop in daily returns in mid of June 2020 where the proposed model's adaptive nature (actively learning) has resulted in achieving positive reruns during that period.

- **Second Trading Mechanism:** In the first trading mechanism, we considered an equal weight for all the stocks in our portfolio. Here, we construct a portfolio based on market price

weighting [Plyakha, 2017] as follows

$$w_i^p = \frac{p_i}{\sum_{i=1}^N p_i}, \quad (4.16)$$

where w_i^p represents the weight of each stock based on its market price, and p_i denotes the market price (we assume that market prices are adjusted close price of the stock on the first day of March). Weights are applied based on each stock's market price on the date that the transaction is going to be placed. Further details on market price weighting is provided in Table 4.1. To evaluate the performance of our model Expected return ($E(R)$) and standard deviation of the return ($Std(R)$) are computed. The results show that Visa has a profit of 10,606\$ with standard deviation of 0.156, which means that visa stock is riskier than others. While the Walt Disney has lower profit and lower risk than the other stocks. Overall, the sum of the profit of the portfolio is equal to 6,764\$, expected return $E(R)$ is 0.337 with standard deviation $Std(R)$ of 0.079.

- **Third Trading Mechanism:** We consider the scenario where 100 shares in the Dow Jones market are bought on 01/03/2020 and held until the end of July. Consequently, the cumulative daily return R shows that we get 0.086 return based on the COVID19-HPSMP framework. From a finance perspective, the risk also needs to be presented as a model can lead to more returns but higher risk. The risk may be defined as the variability of returns that are expected from a given investment. The greater the variability, the riskier the stock. Investment in the Treasury bond is risk-free because of the governmental guarantees. In contrast, investment in equity shares is risky because of the uncertainty on return. Some shares are riskier than others, and even in years when the overall stock market goes up, many individual shares decline in their price. Therefore, investing in one share is extremely risky. Investment risk, then, is related to the probability of actually earning a low or negative return. Therefore, the probability of low earning or negative return is higher in riskier investments. Consequently, it is necessary to measure the degree of risk associated with the investment. For this purpose, we consider the scenario where 100 shares in the Dow Jones market are bought on 01/03/2020 and held until the end of July. As a measure of risk, we calculated Sharpe Ratio (SR). The

Table 4.4: Return comparisons between the proposed COVID19-HPSMP framework and stand-alone models.

Model	Return		
	60-40 Strategy	Hold Strategy 1	Hold Strategy 2
The COVID19-HPSMP Framework	49.7%	27.1%	8.6%
Standalone CNN Local/Global Model	41.8%	21%	2.8%
Standalone CNN-LSTM Model	27.3%	25.5%	2.6%

Table 4.5: Return comparisons between the proposed COVID19-HPSMP framework (50-50), stand-alone models and two commonly used trading indicators.

Model Return	Cumulative Return	Sharpe Ratio
The COVID19-HPSMP Framework	8.6%	1.85%
Standalone CNN Local/Global Model	2.8%	0.48%
Standalone CNN-LSTM Model	2.6%	0.41%
Simple Moving Average (SMA)	0.9%	0.1%
Moving average convergence divergence (MACD)	3.8%	0.55%

cumulative return and sharpe ratio of our baseline models based on two the trading approaches are illustrated in Tables 4.5 and 4.6. The proposed COVID19-HPSMP based on the 60-40 trading approach obtains the highest cumulative return and SR compared to its counterparts. It is worth mentioning that a higher SR shows superiority of a portfolio compared to its counterparts.

As a final experiment, we evaluate the performance based on the following three trading strategies: (i) *60-40 Strategy*: First, we consider the performance of our models where 100 shares in the Dow Jones market are bought on 01/03/2020 based on buy, sell and hold scenarios (the so called 60-40 approach); (ii) *Hold Strategy 1*: Second, we consider the “BUY” and “HOLD” strategy where 100 shares in the Dow Jones market are bought and held for a long period regardless of the fluctuations in the market. For this purpose, we bought the stock on the day that the COVID19-HPSMP framework predicted a “BUY” signal and “HOLD” it until the end of the period. Based on the signal generated by our model, we buy the stock on 2020-03-19, and sell it on 2020-07-29 when the models predicted a “SELL” signal, and; (iii) *Hold Strategy 2*: Finally, we consider the scenario where 100 shares in the Dow Jones market are bought on 01/03/2020 and held until the end of July. The results are shown in Table 4.4 illustrating that the proposed COVID19-HPSMP framework achieves higher return across different strategies compared to its counterparts.

Table 4.6: Return comparisons between the proposed COVID19-HPSMP framework (60-40), stand-alone models and two commonly used trading indicators.

Model Return	Cumulative Return	Sharpe Ratio
The COVID19-HPSMP Framework	29.8%	2.19%
Standalone CNN Local/Global Model	17.5%	0.91%
Standalone CNN-LSTM Model	14.9%	0.98%
Simple Moving Average (SMA)	0.9%	0.1%
Moving average convergence divergence (MACD)	3.8%	0.55%

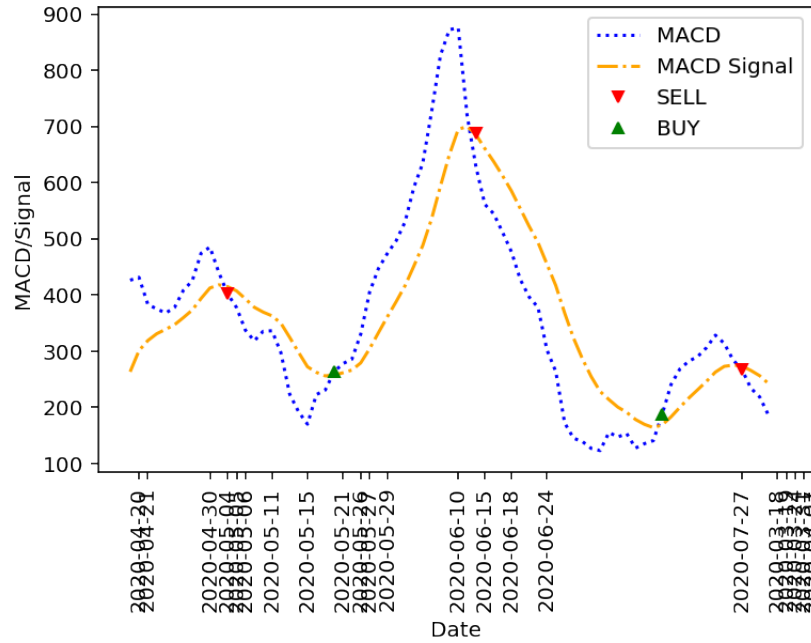


Figure 4.13: MACD Signal

Comparisons with Trading Indicators: In this part, we compare the COVID19-HPSMP engine with two common trading indicators, i.e., the Simple Moving Average (SMA) [Ansari, 2017] and the Moving Average Convergence/Divergence (MACD) [Vaidya, 2020], which are among the most widely used technical indicators by market technicians to identify short, medium, and long-term price movements. First we focus on the SAM approach, which is also known as moving mean or rolling mean in statistics. The SMA is one of the oldest and most straightforward strategies for analysing signals and trade based on the results. In this approach, the moving average of the price or return is calculated to estimate the trends in the market price. The SMA is calculated by creating a series of averages of the market data’s time-ordered subsets. While different window times can be used, a 5-day moving average is calculated for comparison with the proposed COVID19-HPSMP

framework. Fig. 4.11(b) compares returns obtained from the COVID19-HPSMP framework with those obtained from the SMA approach and that of the market. Fig. 4.11(b) demonstrates that our COVID19-HPSMP framework outperforms the SMA in prediction of the market's trend. A second indicator utilized for evaluation purposes is the MACD method, which is another important trading indicator widely used by traditional technical analysis and algorithmic trading. By leveraging the close price's historical value, MACD can be calculated as a collection of three time-series: (i) The MACD value series, which is the difference between the shorter period (fast) and the longer period (slow) of the exponential moving average (EMA); (ii) The MACD Signal, which is the EMA of the MACD value series, and; (iii) A divergence series, which is the difference between Items (i) and (ii). Three parameters a, b, c are used to calculate the MACD, where a is related to the fast EMA; b relates to slow EMA, and; c is related to EMA. To compare the proposed framework with the *MACD* indicator, we used $a = 12$, $b = 26$, and $c = 9$, which are the commonly used values for these parameters. The BUY/SELL signals generated based on the MACD indicator is shown in Fig. 4.13. The MACD provides five different Buy/Sell signal as shown in Fig. 4.13, which is much less than the Buy/Sell signals provided by the COVID19-HPSMP framework. The cumulative return of different models are illustrated in Tables 4.5 and 4.6. Based on the results shown in Tables 4.5 and 4.6, the COVID19-HPSMP framework outperforms the other two indicators in terms of the cumulative return.

Significance Tests: As a final experiment, we evaluate the hypothesis that predictions made by the COVID19-HPSMP framework are more profitable than those obtained via a standalone CNN Local/Global model (potentially illustrating benefits of a hybrid design), and the Moving Average Convergence/Divergence (MACD) method, a common trading indicator (potentially illustrating benefits in comparison to a commonly utilized indicator). In this regard, we have conducted statistical significance analysis of predictive models given their rate of return through a null hypothesis test. In the first test, the null hypothesis is to use the stand-alone CNN Local/Global engine for investment, and the alternative hypothesis is to use the COVID19-HPSMP framework as a recommendation engine for investment. The price movement prediction in stock market is assumed to be aleatory (null is not rejected) until it is proven the results of a recommendation engine (in this case, the COVID19-HPSMP framework) has no relationship with the results of a random trader

(null is rejected). To define a confidence threshold, a cut point of $\alpha=2.145$ corresponding to 95% of confidence is established. The resulting t -value is equal to 3.0526, which is associated with p -value of 0.004303 showing a significant result as $p < 0.05$. This suggests that the null hypothesis is not accepted, therefore, predictions made by COVID19-HPSMP are more profitable. Similarly, in a second experiment, we check if predictions of the COVID19-HPSMP framework is more profitable than those of the MACD indicator. The computed t -value is 2.8433 (i.e., p -value of 0.00651), which means that the COVID19-HPSMP framework reaches better results compared to the MACD, which is a common trading indicator.

As a final note, we would like to mention that the variance between the actual return from a given portfolio and the expected return over a given period of time is referred to as an “Abnormal Return”. A positive abnormal return refers to the scenario that the actual return is higher than its expected return. Abnormal return can be achieved in stock market, however, the volatile nature of the market is hard to earn abnormal return using only the financial indicators, i.e., the stock data, as the market is dependent on a variety of social, political, and economical factors. There are three forms of efficient markets (i.e., weak, semi-strong, and strong), defined based on the nature of information utilized to determine asset prices. In the weak form, asset prices referring to all past prices and trading volume information fully reflect all the market data. In the semi-strong form, asset prices reflect all publicly known and available information. In the strong form, asset prices fully reflect all information, including public (e.g., news) and private information. We consider semi-strong form of efficiency for the markets, in which prices reflect the news data and financial indicators. The proposed COVID19-HPSMP framework incorporates effects of social media trends related to COVID-19 on stock movement prediction. The proposed framework uses information fusion to combine COVID-19 related Twitter data with historical data to trade 5-day in-advance in order to predict the movement of the market. Our model can predict the stock market’s fluctuations with more than 66% accuracy, which will hopefully be a metric to earn abnormal return.

4.5 Summary

Motivated by abrupt, sudden, and negative effects of COVID-19 pandemic on stock markets, first, the chapter introduced a unique COVID-19 related PRICE MOVEMENT prediction (COVID19 PRIMO) dataset. The constructed dataset incorporates effects of social media trends related to COVID-19 on stock market price movements. Based on the constructed COVID19 PRIMO dataset, the chapter then proposed a novel data-driven (DNN-based) COVID-19 adopted Hybrid and Parallel deep fusion framework for Stock price Movement Prediction (COVID19-HPSMP). The proposed framework uses information fusion to combine COVID-19 related Twitter data with extended horizon market historical data. More specifically, in contrary to the existing data-driven stock price movement prediction models, where a single DNN model is used, the COVID19-HPSMP framework is a hybrid model consisting of two parallel paths (i.e., the CNN Local/Glocal path, and; the CNN-LSTM path) and a fusion path that combines localized features. Each of the two parallel paths is a unique attention module extracting different attention related features. The rationale behind such a hybrid and parallel structure is the significance of the attention network and the intuition that extracting different attention-related features would improve the overall performance of the model. The proposed COVID19-HPSMP architecture can predict the stock price movements during the pandemic crisis to forecast sudden sharp movements (fall or rise) in the stock market. Based on the results of the COVID19-HPSMP architecture, we can predict the stock market's fluctuations with more than 66% accuracy, which will hopefully be a metric to be more prepared for the unexpected havocs.

Chapter 5

Summary and Future Research

Directions

This chapter concludes the thesis with a list of important contributions made in this dissertation and some proposed directions for future work.

5.1 Summary of Thesis Contributions

The research works performed in this thesis are motivated by potentials of Machine Learning (ML) and Deep Neural Network (DNN) approaches in stock movement prediction. Considering recent progress in design and implementation of advanced processing and learning models, especially DNN-based techniques, researchers focused on their application for predicting stock trends. The focus of the thesis is on utilization of information fusion to combine Twitter data with extended horizon market historical data for the task of price movement prediction. In this regard, the thesis made a number of contributions as briefly outlined below:

- (1) **The ND-SMPF Framework:** We proposed, the Noisy Deep Stock Movement Prediction Fusion framework (ND-SMPF) for stock price movement prediction. The proposed ND-SMPF predictive model uses information fusion to combine twitter data with extended horizon market historical prices to boost the accuracy of the stock movement prediction task. More specifically, Noisy Bi-directional Gated Recurrent Unit (NBGRU) is utilized coupled with

a Hybrid Attention Network (HAN) to extract news level temporal information. A two level attention layer is used to identify relevant words with highest correlation and effects on the stock trends, which are then fused with historical price data to perform the prediction task. More specifically, the proposed ND-SMPF consists of the following three components:

- **Hybrid Attention Network**, which interprets and analyzes news within a temporal attention context by concentrating more on important time periods, i.e., the focus is given to distinguish more significant news from the available news corpus. In particular, the proposed model is designed based on “Sequential Context Dependency” and “Diverse Influence” principles. In this regard, the ND-SMPF is built based on a Hybrid Attention Network (HAN) formalization [Hu, 2019] to utilize both news-level and temporal-level attentions. To achieve these goals, a multi-modal fusion framework is constructed by coupling extracted information from social media and stock market prices.
- **News Level Attention**: We construct an attention layer to sum the news weighted by an attention value. In order to reward the information content of news, at first we feed news to the news level attention layer and a temporal sequence vector is computed as the output of the news level attention layer.
- **Bidirectional GRU**: For encoding temporal information embedded within the news corpus and financial time-series, Gated Recurrent Neural Network (GRU) [Chung, 2014] is integrated within the ND-SMPF fusion framework. The GRU is a type of RNN that uses gating mechanism to check the state of sequences without requiring separate memory cells. To encode temporal layer, we adopt Noisy Bidirectional GRU (NBDGRU).

(2) **The COVID19-HPSMP Framework**: The novel of coronavirus (COVID-19) has suddenly and abruptly changed the world as we knew at the end of the 2nd decade of the 21st century. The global COVID-19 pandemic caused market volatility [Mazue, 2020, Baek, 2020] rocketing upward around the world. The proposed COVID19-HPSMP adopted Hybrid and Parallel deep fusion framework for Stock price Movement Prediction (COVID19-HPSMP) incorporates innovative fusion strategies to combine scattered social media news related to COVID-19 with historical market data. More specifically, the proposed COVID19-HPSMP

framework is composed of the following two data-driven components:

- ***The CNN Local/Global Path:*** The first parallel path of the proposed COVID19-HPSMP framework is a CNN-based local/global attention model designed to capture and extract spatial features from the input data.
- ***CNN-LSTM Attention based Model:*** The second parallel path of the proposed COVID19-HPSMP is a hybrid CNN and BLSTM attention model, referred to as the CNN-BLSTM path. Similar to the CNN Local/Global path, in the first step, the Twitter messages are provided as input to a “Word Embedding layer”. As stated previously, a pre-trained unsupervised Glove model [Pennington, 2014] is used as the word embedding layer within the COVID19-HPSMP framework. Afterwards, the corpus and prices are encoded by a CNN layer to extract general contextual features. An attention layer is assigned across all the vectors to calculate the weighted corpus. At the next step, a second CNN layer is implemented to capture and learn more fine-tuned features.

5.2 Future Research

- (1) For the purpose of tweet sentiment analysis, the Bidirectional Encoder Representations Transformers (BERT) [Devlin, 2019] can be adapted to simultaneously consider the context of a word from both left and right sides. BERT can extract more context features from a sequence compared to existing modeling solutions.
- (2) Portfolio optimization is an essential component of a trading system, which aims to select the best asset distribution within a portfolio in order to maximise returns at a given risk level. The main benefit of constructing such a portfolio comes from the promotion of diversification that smoothes out the equity curve, leading to a higher return per risk than trading an individual asset. This observation has been proven that the risk (volatility) of a long-only portfolio is always lower than that of an individual asset, for a given expected return, as long as assets are not perfectly correlated. Despite the undeniable power of such diversification, it is not

straightforward to select the “right” asset allocations in a portfolio, as the dynamics of financial markets change significantly over time. Assets that exhibit, for example, strong negative correlations in the past could be positively correlated in the future. This adds extra risk to the portfolio and degrades subsequent performance. On the other hand, news articles influence the dynamics of financial markets. For example, after the release of breaking news, the share prices of related stocks are often observed to move. This suggests the possibility of using NLP to aid traders by analyzing this influence between news articles and prices. To deal with these issues, one potential direction for future research is to target advances in sequential decision making, mainly through the concept of Reinforcement Learning.

- (3) Last but not least, another fruitful direction for future research is adapting a deep learning framework to forecast expected returns based on stock movement prediction and directly optimize portfolio weights by updating model parameters.

Bibliography

- [Abadi, 2016] Abadi, M., *et al.* (2016). Tensorflow: Large-scale Machine Learning on Heterogeneous Distributed Systems. *arXiv:1603.04467*.
- [Al-Awadhi, 2020] Al-Awadhi, A.M., Alsaifi, K., Al-Awadhi, A., & Alhammadi, S. (2020). Death and Contagious Infectious Diseases: Impact of the COVID-19 Virus on Stock Market Returns. *Journal of Behavioral and Experimental Finance*, 27.
- [Al-Dulaimi, 2019] Al-Dulaimi, A., Zabihi, S., Asif, A., & Mohammadi, A. (2019). A Multimodal and Hybrid Deep Neural Network Model for Remaining Useful Life Estimation. *Computers in Industry*, 108, 186-196.
- [Anik, 2020] Anik, M.M., Arefin, M.S., & Dewan, M.A.A. (2020). An Intelligent Technique for Stock Market Prediction. *Proceedings of International Joint Conference on Computational Intelligence*, Singapore.
- [Ansari, 2020] Ansari, S.A., & Del Val, E.B. (2020). SutteARIMA: Short-term Forecasting Method, a Case: Covid-19 and Stock Market in Spain. *Science of the Total Environment*, 729.
- [Ansari, 2017] Ansari, A. (2017). Sutte Indicator: A Technical Indicator in Stock Market. *International Journal of Economics and Financial Issues*, 7.
- [Baek, 2020] Baek, S., Mohanty, S.K., & Glambosky, M. (2020). COVID-19 and Stock Market Volatility: An Industry Level Analysis. *Finance Research Letters*, 37.
- [Beckmann, 2017] Beckmann, M. (2017). Stock Price Change Prediction Using News Text Mining. *Federal University of Rio de Janeiro*.

- [Bhatt, 2021] Bhatt, A. R., Ganatra, A., & Kotecha, K. (2021) Cervical Cancer Detection in Papsmear whole slide Images using convNet with Transfer Learning and Progressive Resizing. *PeerJ Computer Science* 7:e348.
- [Bollen, 2011] Bollen, J., & Mao, H. (2011). Twitter Mood as a Stock Market Predictor. in *Computer*, 44, 91-94.
- [Bustos, 2020] Bustos, O., & Pomares-Quimbaya, A. (2020). Stock Market Movement Forecast: A Systematic Review. *Expert Systems with Applications*, 156.
- [Caruana, 1998] Caruana, R. (1998). Multitask Learning. *Machine Learning*, 28, 41-75.
- [Chong, 2019] Chong, E., Han, C., Park, F.C. (2019). Deep Learning Networks for Stock Market Analysis and Prediction: Methodology, Data Representations, and Case Studies. *Expert Systems with Applications*, 83.
- [PRIMO, 2021] COVID19 PRIMO Dataset: https://github.com/MSBeni/COVID19_PRIMO#COVID19-PRIMO.
- [Choudrie, 2021] Choudrie, J., Banerjee, S., Kotecha, K., Walambe, R., Karende, H., & Ameta, J. (2021). Machine Learning Techniques and Older Adults Processing of online Information and Misinformation: A COVID-19 Study. *Computers in Human Behavior*, 119.
- [Clement, 2020] Clement, J. (2020). Number of Monetizable Daily Active Twitter Users (mDAU) Worldwide from 1st Quarter 2017 to 1st Quarter 2020. <https://www.statista.com/statistics/970920/monetizable-daily-active-twitter-users-worldwide/>.
- [Di, 2018] Di, W., Bhardwaj, A., & Wei, J. (2018). Deep Learning Essentials: your Hands-on Guide to the Fundamentals of Deep Learning and Neural Network Modeling. *Packt Publishing Inc*.
- [Edwards, 2007] Edwards, R. D., Bassetti, W., & Magee, J. (2007). Technical Analysis of Stock Trends(4th ed.). in *CRC Press*, 134-140.
- [Fama, 1998] Fama, E.F. (1998). Market Efficiency, Long-term Returns, and Behavioral Finance. *Journal of Financial Economics*, 49, 283-306.
- [Frankel, 1995] Frankel, J. A. (1995). Financial Markets And Monetary Policy. *MIT Press*, 87-95.

- [Gite, 2021] Gite, S., Khatavkar, H., Kotecha, K., Srivastava, S., Maheshwari, P., & Pandey, N. (2021). Explainable Stock Prices Prediction from Financial News Articles using Sentiment Analysis. *PeerJ Computer Science*, 7:e340.
- [Hochreiter, 1997] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.
- [Hoseinzade, 2019] Hoseinzade, E., & Haratizadeh, S. (2019). CNNpred: CNN-based Stock Market Prediction using a Diverse Set of Variables. *Expert Systems with Applications*, 129.
- [Hu, 2019] Hu, Z., Liu, W., Bian, J., & Liu, X. (2019). Listening to Chaotic Whispers: A Deep Learning Framework for News-oriented Stock Trend Prediction. *arXiv:1712.02136*.
- [Huang, 2020] Huang, J.Y., & Li, J.H. (2020). Using Social Media Mining Technology to Improve Stock Price Forecast Accuracy. *Journal of Forecasting*, 39, 104-116.
- [Ioffe, 2015] Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv preprint arXiv:1502.03167*.
- [Jiang, 2020] Jiang, W. (2020). Applications of Deep Learning in Stock Market Prediction: Recent Progress. *arXiv preprint arXiv:2003.01859*.
- [Kingma, 2014] Kingma, D.P., Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*.
- [Koshiyama, 2020] Koshiyama, S., Firoozye, N., & Treleaven, P. (2020). Algorithms in Future Capital Markets. <https://dx.doi.org/10.2139/ssrn.3527511>.
- [Li, 2014] Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News Impact on Stock Price Return via Sentiment Analysis. *Knowledge-Based Systems*, 69, 14-23.
- [Loper, 2002] Loper, E., & Bird, S. (2002). NLTK: The Natural Language Toolkit. *In Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, Philadelphia, PA.
- [Luss, 2015] Luss, R., & D'Aspremont, A. (2015). Predicting Abnormal Returns from News Using Text Classification. *Quantitative Finance*, 15, 999-1012.
- [Mazue, 2020] Mazur, M., Dang, M., & Vega, M. (2020). COVID-19 and the March 2020 Stock Market Crash. Evidence from S&P1500. *Finance Research Letters*, 38.

- [Mohammadi, 2017] Mohammadi, A., Zhang, X., & Plataniotis, K. N. (2017). Interactive Gaussian-sum Filtering for Estimating Systematic Risk in Financial Econometrics. *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Montreal, QC.
- [Narkhede, 2021] Narkhede, P., Walambe, R., Mandaokar, S., Chandel, P., Kotecha, & K., Ghinea, G. (2021). Gas Detection and Identification Using Multimodal Artificial Intelligence based Sensor Fusion. *Applied System Innovation*, 4(3).
- [Patel, 2015] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting Stock and Stock Price Index Movement using Trend Deterministic Data Preparation and Machine Learning Techniques. *Expert Systems with Applications*, 42, 259-268.
- [Radojicic, 2020] Radojicic, D., & Kredatus, S. (2020). The Impact of Stock Market Price Fourier Transform Analysis on the Gated Recurrent Unit Classifier Model. *Expert Systems with Applications*, 159.
- [Rekabsaz, 2017] Rekabsaz, N., *et al.* (2017). Volatility Prediction Using Financial Disclosures Sentiments with Word Embedding-based IR Models. *Association for Computational Linguistics*, Vancouver, BC.
- [Pennington, 2014] Pennington, J., Socher, R., Manning, C.D. (2014). Glove: Glove Vector of Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar.
- [Plyakha, 2017] Plyakha, Y., Uppal, R., & Vilkov, G. (2017). Why Does an Equal-Weighted Portfolio Outperform Value- and Price-Weighted Portfolios?. <http://dx.doi.org/10.2139/ssrn.2724535>.
- [Rezaei, 2020] Rezaei, H., Faaljoui, H., & Mansourfar, G. (2020). Stock Price Prediction using Deep Learning and Frequency Decomposition. *Expert Systems with Applications*, 169.
- [Ronaghi, 2021] Ronaghi, F., Salimibeni, M., Naderkhani, F., & Mohammadi, A. (2020). OVID19-HPSMP: COVID-19 Adopted Hybrid and Parallel Deep Information Fusion Framework for Stock Price Movement Prediction. Minor Revisions Submitted at *Expert Systems With Applications*, July 2021.
- [Ronaghi, 2020] Ronaghi, F., Salimibeni, M., Naderkhani, F., & Mohammadi, A. (2020). NDSMPF: A Noisy Deep Neural Network Fusion Framework for Stock Price Movement Prediction. *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, Sun City, Africa.
- [Schumaker, 2009] Schumaker, R.P., & Chen, H. (2009). Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFin Text System. *ACM Transactions on Information Systems*, 12.

- [Seo, 2017] Seo, S., Huang, J., Yang, H., & Liu, Y. (2017). Interpretable Convolutional Neural Networks with Dual Local and Global Attention for Review Rating Prediction. *Association for Computing Machinery*, New York, NY.
- [Seong, 2021] Seong, N., & Nam, K. (2021). Predicting Stock Movements based on Financial News with Segmentation. *Expert Systems with Applications*, 64.
- [Tetlock, 2007] Tetlock, P.C. (2017). Giving Content to Investor Sentiment: The Role of Media in The Stock Market. *The Journal of Finance*, 62, 1139-1168.
- [Vaidya, 2020] Vaidya, R. (2020). Moving Average Convergence-Divergence (MACD) Trading Rule: An Application in Nepalese Stock Market(NEPSE). *Quantitative Economics and Management Studies (QEMS)*, 1, 6. <https://doi.org/10.35877/454RI.qems197>.
- [Xie, 2013] Xie, B., Passonneau, R.J., Wu, L., & Creamer, G.G. (2013). Semantic Frames to Predict Stock Price Movement. *Association for Computational Linguistics*, Sofia, Bulgaria.
- [Xu, 2018] Xu, Y., & Cohen, S.B. (2018). Stock Movement Prediction from Tweets and Historical Prices. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia.
- [Yun, 2019] Yun, H., Sim, G., & Seok, J. (2019). Stock Prices Prediction using the Title of Newspaper Articles with Korean Natural Language Processing. *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIC)*, Okinawa, Japan.
- [Zhang, 2021] Zhang, Y., Chu, G., & Shen, D. (2021). The Role of Investor Attention in Predicting Stock Prices: The Long Short-Term Memory Networks Perspective. *Finance Research Letters*, 38.
- [Zhang, 2020] Zhang, Z., Zohren, S., & Roberts, S. (2020). Deep Learning for Portfolio Optimization. <https://arxiv.org/abs/2005.13665>.
- [1] J. A Frankel, "Financial Markets And Monetary Policy," *MIT Press*, 1995, pp. 87-95.
- [Arash, 2017] A. Mohammadi, X. Zhang and K. N. Plataniotis, "Interactive Gaussian-sum Filtering for Estimating Systematic Risk in Financial Econometrics," *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Montreal, QC, 2017, pp. 903-907.
- [Jiang, 2020] W. Jiang, "Applications of Deep Learning in Stock Market Prediction: Recent Progress," *arXiv, Statistical Finance*, 2020.

- [Edwards, 2007] R. D. Edwards, W. Bassetti and J. Magee, “Technical Analysis of Stock Trends,” in *CRC Press*, 2007, pp. 134-140.
- [Bollen, 2011] J. Bollen and H. Mao, “Twitter Mood as a Stock Market Predictor,” in *Computer*, vol.44, no. 10 ,2011, pp. 91-94.
- [Hu, 2019] Z. Hu, W. Liu, J. Bian and X. Liu, “Listening to Chaotic Whispers: A Deep Learning Framework for News-oriented Stock Trend Prediction,” In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, ACM, Los Angeles, California, USA, 2019, pp. 261–269.
- [Koshiyama, 2020] S. Koshiyama, N. Firoozye and P. Treleaven, “ Algorithms in Future Capital Markets,” *SSRN*, 2020, <https://ssrn.com/abstract=3527511>.
- [Schumaker, 2009] R. P. Schumaker and H.Chen, “ Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFin Text System,” *ACM Transactions on Information Systems*, 2009, 10.1145/1462198.1462204
- [Tetlock, 2007] P. C. Tetlock, “Giving Content to Investor Sentiment: The Role of Media in The Stock Market,” *Journal of Finance, Forthcoming*, 2007, Available at SSRN: <https://ssrn.com/abstract=685145> or <http://dx.doi.org/10.2139/ssrn.685145>
- [Luss, 2015] R. Luss and A. D’Aspremont, “Predicting Abnormal Returns from News Using Text Classification,” *Quantitative Finance*, vol. 15, no.6, 2015, pp. 999-1012.
- [Xu, 2018] Y. Xu and S.B. Cohen, “Stock Movement Prediction from Tweets and Historical Prices,” *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia, 2018, pp. 1970-1979.
- [Rekabsaz, 2017] N. Rekabsaz, *et al.*, “Volatility Prediction Using Financial Disclosures Sentiments with Word Embedding-based IR Models.” *Association for Computational Linguistics*, Vancouver, Canada, 2017, pp. 1712–1721.
- [Caruana, 1998] R. Caruana, “ Multitask Learning,” *Machine Learning*, 1998 , pp. 41-75.
- [Heaton, 2016] J. B. Heaton, N. G. Polson and J. H. Witte, “ Deep Learning for Finance: Deep Portfolios,” *Applied Stochastic Models in Business and Industry*, 2016.
- [Khare, 2017] K. Khare, O. Darekar, P. Gupta and V. Z. Attar, “ Short Term Stock Price Prediction Using Deep Learning,” *2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT)*, Bangalore, 2017, pp. 482-486.

- [Li, 2016] L. Li, S. Leng, J. Yang, and M. Yu, “ Stock Market Autoregressive Dynamics: A Multinational Comparative Study with Quantile Regression,” *Mathematical Problems in Engineering*, 2016, <https://doi.org/10.1155/2016/1285768>.
- [Nayak, 2015] R. K. Nayak, D. Mishra and A. K. Rath, “ A Naive SVM-KNN based stock market trend reversal analysis for Indian benchmark indices,” *Applied Soft Computing*, vol. 35, 2015, pp. 670-680.
- [Adebiyi, 2014] A. A. Adebiyi, A. O. Adewumi and C. K. Ayo, “Comparison of Arima and Artificial Neural Networks Models for Stock Price Prediction,” *Journal of Applied Mathematics*, 2014.
- [Kim, 2010] K. J. Kim and H. Ahn, “Simultaneous Optimization of Artificial Neural Networks for Financial Forecasting,” *Applied Intelligence*, 2010, pp. 887-898.
- [Patel, 2015] J. Patel, S. Shah, P. Thakkar and K. Kotecha, “ Predicting Stock Market Index Using Fusion of Machine Learning Techniques,” *Expert Systems with Applications*, vol. 42, 2015, pp. 2162-2172.
- [Akita, 2016] R. Akita, A. Yoshihara, T. Matsubara and K. Uehara, “Deep Learning for Stock Prediction Using Numerical and Textual Information,” *IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, Okayama, 2016, pp.1–6.
- [Gao, 2016] Q. Gao, “Stock Market Forecasting Using Recurrent Neural Network,” *Ph.D. Dissertation*, University of Missouri-Columbia, 2016.
- [Ding, 2015] X. Ding, Y. Zhang, T. Liu and J. Duan, “Deep Learning for Event-Driven Stock Prediction,” *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence* , 2015, pp. 2327-2333.
- [Wang, 2014] W. Y. Wang and Z. Hua, ‘A Semiparametric Gaussian Copula Regression Model for Predicting Financial Risks from Earnings Calls,” *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, vol.1, 2014, pp. 1155-1165.
- [Xie, 2013] B. Xie, R. J. Passonneau, L. Wu and G. G. Creamer, “Semantic Frames to Predict Stock Price Movement,” *Association for Computational Linguistics*, Sofia, Bulgaria, 2013, pp. 873-883.
- [Li, 2014] X. Li, H. Xie, L. Chen, J. Wang and X. Deng, “News Impact on Stock Price Return via Sentiment Analysis,” *Knowledge-Based Systems*, vol.69, 2014, pp.14-23.
- [Chung, 2014] J. Chung, C. Gulcehre, K. Cho and Y. Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,” *NIPS’2014 Deep Learning Workshop*, 2014.

- [Huang, 2020] J.Y. Huang and J.H. Li, “Using Social Media Mining Technology to Improve Stock Price Forecast Accuracy,” *Journal of Forecasting*, vol.39, 2020, pp. 104-116.
- [Hochreiter, 1997] S. Hochreiter and J.Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol.9, no.8, 1997, pp. 1735-1780.
- [Grave, 2013] A. Graves, N. Jaitly and A. Mohamed, “Hybrid Speech Recognition with Deep Bidirectional LSTM,” *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, Olomouc, 2013, pp. 273-278.
- [Mekayel, 2019] M. M. Anik, M. S. Arefin and M.A.A Dewan “An Intelligent Technique for Stock Market Prediction,” *Proceedings of International Joint Conference on Computational Intelligence. Algorithms for Intelligent Systems*, Singapore, 2019.
- [Loper, 2002] E. Loper, S. Bird, “NLTK: The Natural Language Toolkit,” *In Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*, Philadelphia, 2002.
- [Yun, 2019] H. Yun, G. Sim and J. Seok, “Stock Prices Prediction using the Title of Newspaper Articles with Korean Natural Language Processing,” *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIC)*, Okinawa, Japan, 2019, pp. 019-021.
- [Kingma, 2014] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [Ioffe, 2015] S. Ioffe nad C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” *arXiv preprint arXiv:1502.03167*, 2015.
- [Abadi, 2016] M. Abadi, *et al.*, “Tensorflow: Large-scale Machine Learning on Heterogeneous Distributed Systems,” *arXiv preprint arXiv:1603.04467*, 2016.
- [Pagolu, 2016] V. S. Pagolu, K. N. Reddy, G. Panda and B. Majhi, “Sentiment Analysis of Twitter Data for Predicting Stock Market Movements,” *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, Paralakhemundi, 2016, pp. 1345-1350.
- [Pennington, 2014] J. Pennington, R. Socher, C. D. Manning, “Glove: Glove Vector of Word Representation,” *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532-1543.

- [conneau, 2017] A. Conneau, H. Schwenk, L. Barrault and Y. Lecun, “Very Deep Convolutional Networks for Text Classification,” <https://arxiv.org/abs/1606.01781>, 2017.
- [kim, 2015] Y. Kim, Y. Jernite, D. Sontag and A.M. Rush, “Character-Aware Neural Language Models,” <https://arxiv.org/abs/1508.06615>, 2015.
- [Aliaksei, 2015] A. Severyn and A. Moschitti, “Twitter Sentiment Analysis with Deep Convolutional Neural Networks,” *In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, 2015, pp. 959–962.
- [Xiang, 2015] X. Zhang, J. Zhao and Y. LeCun, “Character-level Convolutional Networks for Text Classification,” *In Proceedings of the 28th International Conference on Neural Information Processing Systems*, MIT Press, Vol.1, 2017, pp. 649–657.
- [Boser, 1990] Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. DJackel. “Handwritten Digit Recognition with a Back-Propagation Network,” *In Advances in Neural Information Processing Systems*, 1990, pp. 396–404.
- [Koushik, 2016] J. Koushik. “Understanding Convolutional Neural Networks,” <https://arxiv.org/abs/1605.09081>, 2016.
- [Markowitz, 1952] H. Markowitz. “Portfolio Selection,” *The journal of Finance*, 7(1), 1952, pp. 77–91.
- [Wild, 2008] R. Wild. “Index Investing for Dummies,” *John Wiley and Sons*, 2008.
- [Zhang, 2019] Z. Zhang, S. Zohren, and S. Roberts. “Extending Deep Learning Models for Limit Order Books to Quantile Regression,” *Proceedings of Time Series Workshop of the 36 th International Conference on Machine Learning*, Long Beach, CA, PMLR 97, 2019.
- [Choueifaty, 2008] Y. Choueifaty and Y. Coignard. “Toward Maximum Diversification,” *The Journal of Portfolio Management*, 35(1), 2008, pp. 40-51.
- [Robert, 2002] E.R Fernholz. “Stochastic Portfolio Theory,” *In Stochastic Portfolio Theory*, Springer, 2002, pp. 1–24. .
- [Ioannis, 2009] R. Fernholz and L. Karatzas. “Stochastic Portfolio Theory: An Overview,” *Handbook of Numerical Analysis*, 2009, pp. 89-167.
- [Daniel, 2010] D. Fernholz and L. Karatzas. “On Optimal Arbitrage,” *The Annals of Applied Probability*, 2010, pp. 1179–1204.

- [Johannes, 2013] J. Ruf. “Hedging under arbitrage,” *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 2013, pp. 297–317.
- [Leonard, 2015] T. L. Wong. “Optimization of relative arbitrage,” *Annals of Finance*, 2015, pp. 345– 382, .
- [Sharpe, 1964] W. Sharpe. “Capital Asset Prices: A Theory Of Market Equilibrium Under Conditions Of Risk,” *Journal of Finance, American Finance Association*, vol. 19(3), 1964, pp. 425–442.
- [Terence, 2008] T.T Chong and W.K Ng. “Technical Analysis and the London Stock Exchange: Testing the MACD and RSI Rules Using the FT30,” *Applied Economics Letters*, vol. 15, no.14, 2008, pp. 1111–1114.
- [Lixiang, 2004] L. Shen, H.T. Loh. “Applying Rough Sets to Market Timing Decisions,” *Decision Support Systems*, vol. 37, 2004, pp. 583–597.
- [Hutter, 2008] M. Hutter and S. Legg, “Temporal Difference Updating without a Learning Rate,” *Advances in Neural Information Processing Systems*, 2008, pp. 705–712.
- [Sutton, 1996] R. S. Sutton, “Generalization in Reinforcement Learning: Successful Examples using Sparse Coarse Coding,” *Advances in Neural Information Processing Systems*, 1996, pp. 1038–1044.
- [Watkins, 1992] C. J. Watkins and P. Dayan, “Q-learning,” *Machine Learning*, vol. 8, no. 3-4, pp. 279-292, 1992.
- [Salimibeni, 2021] M. Salimibeni, P. Malekzadeh, A. Mohammadi, P. Spachos and K. N. Plataniotis, “Makf-Sr: Multi-Agent Adaptive Kalman Filtering-Based Successor Representations,” *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 8037-8041.
- [Malekzadeh, 2020] Parvin Malekzadeh and Mohammad Salimibeni and Arash Mohammadi and Akbar Assa and Konstantinos N. Plataniotis, “MM-KTD: Multiple Model Kalman Temporal Differences for Reinforcement Learning,” <https://arxiv.org/abs/2006.00195>, 2020.
- [Spano, 2019] S. Spanò, G.C. Cardarilli, L. Di Nunzio, R. Fazzolari, D. Giardino, M. Matta, A. Nannarelli, and M. Re, “An Efficient Hardware Implementation of Reinforcement Learning: The Q-Learning Algorithm,” *IEEE Access*, 7, 2019, pp.186340-186351.
- [Seo, 2019] M. Seo, L.F. Vecchiotti, S. Lee, and D. Har, “Rewards Prediction-Based Credit Assignment for Reinforcement Learning With Sparse Binary Rewards,” *IEEE Access*, 7, 2019, pp.118776-118791.

[Devlin, 2019] J. Devlin, M.W Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, Minnesota, 2019, pp. 4171-4186.