

Deep Attentive Learning for Stock Movement Prediction From Social Media Text and Company Correlations

Ramit Sawhney*

Netaji Subhas Institute of Technology
ramits.co@nsit.net.in

Arnav Wadhwa

MIDAS, IIT Delhi
arnavw96@gmail.com

Shivam Agarwal*

Manipal Institute of Technology
shivamag99@gmail.com

Rajiv Ratn Shah

IIIT Delhi
rajivrtn@iiitd.ac.in

Abstract

In the financial domain, risk modeling and profit generation heavily rely on the sophisticated and intricate stock movement prediction task. Stock forecasting is complex, given the stochastic dynamics and non-stationary behavior of the market. Stock movements are influenced by varied factors beyond the conventionally studied historical prices, such as social media and correlations among stocks. The rising ubiquity of online content and knowledge mandates an exploration of models that factor in such multimodal signals for accurate stock forecasting. We introduce an architecture that achieves a potent blend of chaotic temporal signals from financial data, social media, and inter-stock relationships via a graph neural network in a hierarchical temporal fashion. Through experiments on real-world S&P 500 index data and English tweets, we show the practical applicability of our model as a tool for investment decision making and trading.

1 Introduction

Stock prices have an intrinsically volatile and non-stationary nature, making their rise and fall hard to forecast (Adam et al., 2016). Investment in stock markets involves a high risk regarding profit-making. Prices are driven by diverse factors that include but are not limited to company performance (Anthony and Ramesh, 1992), historical trends (Kohara et al., 1997), investor sentiment (Neal and Wheatley, 1998). Uninformed trading decisions can leave traders and investors prone to financial risk and experience monetary losses. On the contrary, careful investment choices can maximize profits (de Souza et al., 2018). Conventional research focused on time series and technical analysis of a stock, i.e., using patterns from historical price signals to forecast stock movements (B et al.,

2013). However, price signals alone fail to capture market surprises and impacts of sudden unexpected events. Social media texts like tweets can have huge impacts on the stock market. For instance, US President Donald Trump shared tweets expressing negative sentiments against Lockheed Martin, which led to a loss of around \$5.8 Billion to the company’s market capitalization.¹

The Efficient Market Hypothesis (EMH) (Malkiel, 1989) states that financial markets are *informationally efficient*, such that stock prices reflect all known information. Existing works (Sec. 2) mainly focus on subsets of stock relevant data. Although useful, they do not jointly optimize learning over modalities like social media text and inter stock relations limiting their potential to capture a broader scope of stock movement affecting data, as we show in Sec. 6. Multimodal stock prediction involves multiple challenges (Hu et al., 2018). Both price signals and tweets exhibit *sequential context dependencies*, where singular samples may not be informative enough but can be considered a sequence for a unified context. Tweets often have *diverse influence* on stock prices, based on their intrinsic content, such as breaking news as opposed to noise like vague comments. Fusing multiple modalities of vast stock related data generated with varying characteristics (*frequency, noise, source*) is complex and mandates the careful design of joint optimization over modality-specific components.

Building on the EMH and prior work (Sec. 2), we propose MAN-SF: Multipronged Attention Network for Stock Forecasting that jointly learns from historical prices, social media, and inter stock relations. MAN-SF through hierarchical attention captures *relevant* signals across diverse data to train a Graph Attention Network (GAT) for stock prediction (Sec. 3). MAN-SF (Sec. 4) jointly learns from

¹<https://medium.com/scoop-markets/7-tweets-which-wiped-40-billion-off-the-stock-market>

* Equal contribution.

price and tweets over graph-based models for stock prediction. Through varied experiments (Sec. 5), we show the predictive power of MAN-SF along with profitability analysis (Sec. 6) and qualitatively analyze MAN-SF in high risk scenarios (Sec. 7).

2 Related Work

Predicting stock movements spans multiple domains (Jiang, 2020); 1) theoretical: quantitative models like *Modern Portfolio Theory* (Elton et al., 2009), *Black-Scholes model* (Black and Scholes, 1973), etc. and, 2) practical: investment strategies (Blitz and Van Vliet, 2007), portfolio management (Hocquard et al., 2013), and beyond the world of finance (Erb et al., 1994; Rich and Tracy, 2004). Financial models conventionally focused on technical analysis (TA) relying only on numerical features like past prices (Ding and Qin, 2019; Nguyen et al., 2019) and macroeconomic indicators like GDP (Hoseinzade et al., 2019). Such TA methods include discrete: GARCH (Bollerslev, 1986), continuous (Andersen, 2007), and neural approaches (Nguyen and Yoon, 2019; Nikou et al., 2019).

Newer models based on the EMH that are categorized under fundamental analysis (FA) (Dichev and Tang, 2006), account for stock affecting factors beyond numerical ones such as investor sentiment through news, etc. Work in natural language processing (NLP) from sources such as news (Hu et al., 2018), social media data (Xu and Cohen, 2018), earnings calls (Qin and Yang, 2019; Sawhney et al., 2020b) shows the merit of FA in capturing market sentiment, surprises, mergers, acquisitions that traditional TA based methods fail to account. A limitation of existing NLP methods for stock prediction is that they assume stock movements to be independent of each other, contrary to true market function (Diebold and Yilmaz, 2014). This assumption hinders NLP centric FA’s ability to learn latent patterns for the study of interrelated stocks.

Another line of FA revolves around employing graph-based methods to improve TA (e.g., price-based models) by augmenting them with inter stock relations (Feng et al., 2019b; Sawhney et al., 2020a). Matsunaga et al. (2019) combine historical prices with stock graphs through Graph Convolution Networks (GCNs), outperforming price-only models. Similarly, Kim et al. (2019) further improve graph neural network methods by weighing stock relations through attention mechanisms, as not all stock movements are equally correlated.

Despite the popularity of NLP and graph-based stock prediction, multimodal methods that capture inter stock relations and market sentiment through linguistic cues are seldom explored. Jue Liu (2019) combines feature extraction from news *sentiment scores*, financial information (*price-earnings ratio*, etc.) along with knowledge graph embeddings through TransR. However, such existing approaches (Deng et al., 2019) are unable to represent textual signals from social media and prices temporally, as they only utilize sentiment scores and do not account for stock correlations. To cover this gap in prior research, MAN-SF captures a broader set of features as opposed to both conventional TA and FA that singularly focus on either text or graph modalities, but not both together.

3 Problem Formulation

MAN-SF’s main objective is to learn *temporally relevant information* jointly from tweets and historical price signals and make use of corporate relations among stocks to predict movements. Following Xu and Cohen (2018), we formalize movement based on the difference between the adjusted closing prices of the stock $s \in S$ on trading days d and $d - 1$. We formulate stock movement prediction as a binary classification problem.

Problem Statement: Given stock $s \in S$, and historical price data and tweets for stock s over a lookback window of T days over the day range $[t - T, t - 1]$, we define the price movement of stock s from day $t - 1$ to t as:

$$Y_t = \begin{cases} 0, & p_d^c < p_{d-1}^c \\ 1, & p_d^c \geq p_{d-1}^c \end{cases} \quad (1)$$

where p_d^c represents the widely used (Yang et al., 2020; Qin and Yang, 2019) adjusted closing price² of a given stock on day t . Here, 0 represents a price downfall, and 1 represents a rise in the price.

4 MAN-SF: Components and Learning

In this section, we first give an overview of MAN-SF, followed by a detailed explanation of each component. As shown in Figure 1, MAN-SF first encodes market data for each stock over a fixed period. Formally, we encode stock features $x_t \in \mathbb{R}^w$ for each trading day t as, $x_t = \mathcal{B}(c_t, q_t)$; where, $c_t \in \mathbb{R}^u$ represents a social media feature that we

²Source: https://www.investopedia.com/terms/a/adjusted_closing_price.asp

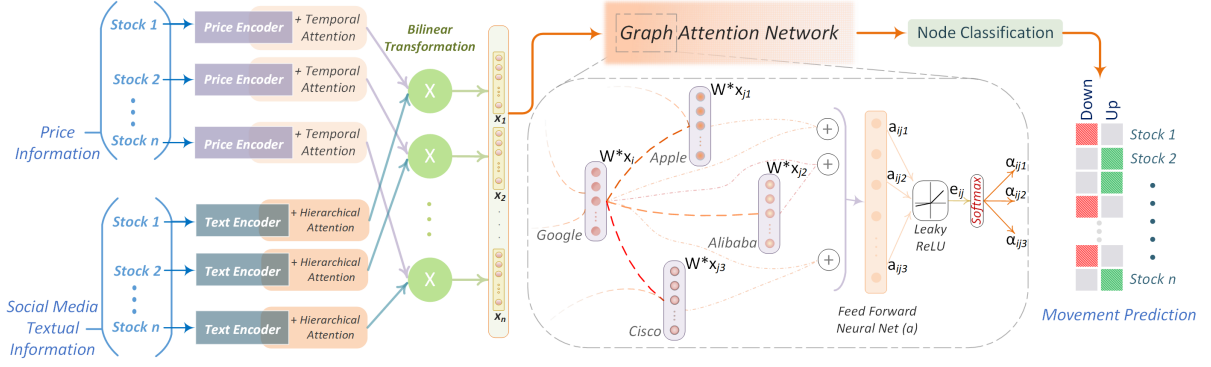


Figure 1: An overview of MAN-SF: Encoding Mechanisms, GAT Mechanism, Joint Optimization.

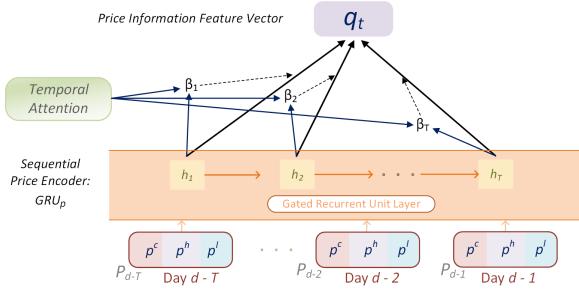


Figure 2: An overview of the Price Encoder.

obtain by encoding tweets over the lag window for each stock $s \in \mathcal{S} = \{s_1, s_2, \dots, s_S\}$. Similarly, $q_t \in \mathbb{R}^v$ are the features obtained from historical prices for a stock in the lag window. We detail these encoders first, and then explain the fusion $\mathcal{B}(\cdot)$ over c_t and q_t to obtain $x_t \in \mathbb{R}^w$. We then describe the graph to represent the inter stock relations. Lastly, we explain the GAT to which the fused feature vector x_t is passed to propagate features based on inter-stock relations along with the joint optimization of MAN-SF.

4.1 Price Encoder

Technical Analysis shows that historical price information is a strong indicator of future trends (Jeanblanc et al., 2009). Therefore, price data from each day is a crucial input to MAN-SF. The Price Encoder shown in Figure 2 encodes historical stock price movements to produce price feature, q_t . It takes in a per-day price feature from the lookback of T days and encodes the temporal trend in prices. To capture such sequential dependencies across trading days, we use a Gated Recurrent Unit (GRU) (Cho et al., 2014; Giles et al., 2001). The output of the GRU on day i is denoted by:

$$h_i = GRU_p(p_i, h_{i-1}) \quad t - T \leq i \leq t \quad (2)$$

where, $p_i \in \mathbb{R}^{d_p}$ is the price vector on day i for each stock s in the lookback. The raw price vector, $p_i = [p_i^c, p_i^h, p_i^l]$ comprises of a stock's adjusted closing price, highest price and lowest price for a trading day i . Since it is the price change that determines the stock movement rather than the absolute price value, we normalize it with its last adjusted closing price, $p_i = p_i / p_{i-1}^c$.

It has been shown that the stock trend of each day has a different impact on stock trend prediction (Feng et al., 2019a). Towards this end, we employ temporal attention $\zeta(\cdot)$ (Li et al., 2018) that learns to weigh critical days and forms an aggregated feature representation across all hidden states of the GRU (Qin et al., 2017). The temporal attention mechanism yields $q_t = \zeta(\bar{h}_p)$; where, $\bar{h}_p \in \mathbb{R}^{d_p \times T}$ is the concatenated hidden states of GRU_p for each stock s . This temporal attention mechanism $\zeta(\cdot)$ rewards days with more impactful information and aggregates it from all days in the lag window to produce price features $q_t \in \mathbb{R}^v$.

Temporal Attention We use a temporal attention mechanism that is a form of additive attention (Bahdanau et al., 2014). The mechanism $\zeta(\cdot)$ aggregates all the hidden representations of the GRU across different time-steps into an overall representation with learned adaptive weights (Feng et al., 2019a). We formulate this mechanism $\zeta(\cdot)$ as:

$$\beta_i = \frac{\exp(h_i^T W \bar{h}_z)}{\sum_{i=1}^T \exp(h_i^T W \bar{h}_z)} \quad (3)$$

$$\zeta(\bar{h}_z) = \sum_i \beta_i h_i \quad (4)$$

where, $\bar{h}_z \in \mathbb{R}^{T \times d_m}$ denotes the concatenated hidden states of the GRU. β_i represents the learned attention weights for trading day i , and W is a learnable parameter matrix.

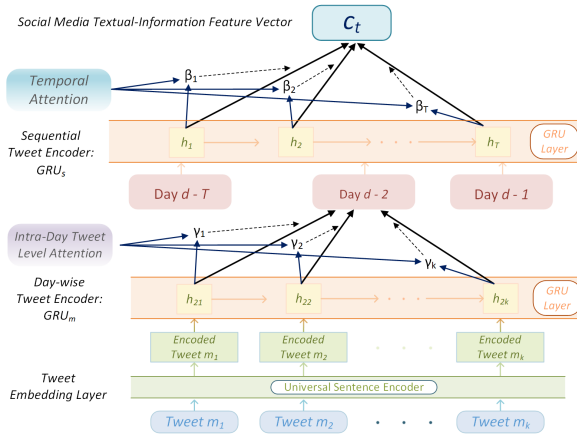


Figure 3: Social Media Information Encoder.

4.2 Social Media Information Encoder (SMI)

Xu and Cohen (2018) suggest that tweets not only convey factual data, but also portray user sentiment towards stocks that influence financial prediction (Bollen et al., 2011). A variety of market factors beyond historical prices drive stock trends (Abu-Mostafa and Atiya, 1996). With the rising ubiquity of the Internet, social media platforms, such as Twitter, influence investors to follow market trends (Tetlock, 2007; Hu et al., 2018). Tweets not only convey factual information but also portray user sentiment towards stocks (Xu and Cohen, 2018; Fung et al., 2002). To this end, MAN-SF uses the SMI encoder to extract a feature vector c_t using tweets. The encoder shown in Figure 3 extracts social media features, c_t , by first encoding tweets for a day and then over multiple days using a hierarchical attention mechanism (Yang et al., 2016).

Tweet Embedding For any given tweet t_w , we generate an embedding vector $m \in \mathbb{R}^d$. We explored word and sentence level embedding methods to learn tweet representations: Global Vectors for Word Representation (GloVe) (Pennington et al., 2014), Fasttext (Joulin et al., 2017), and Universal Sentence Encoders (USE) (Cer et al., 2018). Empirically, sentence-level embeddings generated using a deep averaging network encoder variant of the USE³ gave us the most promising results. Thus, we encode each tweet t_w using USE.

Learning Representations for one day On any day i , a variable number tweets $[t_{w1}, t_{w2}, \dots, t_{wK}]$ for each stock s are posted, and these capture and influence the stock trends (Fung et al.,

³Implementation used: <https://tfhub.dev/google/universal-sentence-encoder/2>

2002). For each tweet, we obtain a representation using the Tweet Embedding layer (USE) as $[m_1, m_2, \dots, m_K]$ where $m_j \in \mathbb{R}^d$ and K is the number of tweets per stock on day i . To model the sequence of tweets within a day, we use a GRU. For stock s on each day i :

$$h_j = GRU_m(m_j, h_{j-1}); j \in [1, K] \quad (5)$$

The influence of online tweets on the market can vary greatly (Hu et al., 2018). To identify tweets that are likely to have a more substantial influence on the market, we use an intraday tweet level attention. For each stock s on each day i the mechanism can be summarized as:

$$\gamma_j = \frac{\exp(h_j^T W \bar{h}_m)}{\sum_{j=1}^K \exp(h_j^T W \bar{h}_m)} \quad (6)$$

$$r_i = \sum_j \gamma_j h_j \quad (7)$$

where, $\bar{h}_m \in \mathbb{R}^{K \times d_m}$ denotes a concatenation of all hidden states from GRU_m and d_m is the dimension of each hidden state. γ_j represents the attention weights and r_i represents the features obtained from several published tweets on day i for each stock s . W is a learned linear transformation.

Learning Representations across days Analyzing a temporal sequence of tweets and combining them can provide a more reliable assessment of market trends (Zhao et al., 2017). We learn a social media representation from the sequence of day level tweet representations r_i . This feature vector encodes all the information in a lookback window. We then feed temporal day level tweet vectors to a GRU for sequential modeling given by:

$$h_i = GRU_s(r_i, h_{i-1}) \quad t - T \leq i \leq t \quad (8)$$

where, h_i summarizes the tweets on day i for stock s as well as tweets from preceding days while focusing on day i . Like historical prices, tweets from each day have a different impact on stock movements. Hence, the previously described temporal attention mechanism used for historical prices is also used for social media. This mechanism learns a procedure to aggregate impactful information to form SMI features c_t over a lookback of T days for each stock s . The temporal attention mechanism yields $c_t = \zeta(\bar{h}_s)$; $\bar{h}_s \in \mathbb{R}^{T \times d_s}$ represents the concatenated hidden states of GRU_s and d_s is the size of output space of the GRU. This temporal

attention $\zeta(\cdot)$, along with the intraday tweet-level attention, forms a hierarchical attention mechanism. This mechanism captures the fact that tweets are differently informative and have varied impacts during different market phases. The obtained SMI and price features for each stock are then blended to obtain a joint representation.

4.3 Blending Multimodal Information

Signals from different modalities often carry complementary information about different events in the market (Robert P. Schumaker, 2019). Direct concatenation treats information from Price and SMI encoders equally (Li et al., 2016). Furthermore, the interdependencies between price and tweets are not appropriately captured, damping the framework’s capacity to learn their correlations to market trends (Li et al., 2014). We use a bilinear transformation that learns the pairwise feature interactions from historical price features and tweets. Formally, $q_t \in \mathbb{R}^v$ and $c_t \in \mathbb{R}^u$ are obtained from the Price Encoder and SMI Encoder, respectively. The output $x_t \in \mathbb{R}^w$ is given by:

$$x_t = \mathcal{B}(c_t, q_t) = \text{ReLU}(q_t^T W c_t + b) \quad (9)$$

where, $W \in \mathbb{R}^{w \times v \times u}$ is the weight matrix, and $b \in \mathbb{R}^w$ is the bias. Methods like direct mean and attention-based aggregation (Bahdanau et al., 2014) do not account for pair-wise interactions as shown in the results (Sec. 6). Other methods like factorized bilinear pooling (Yu et al., 2017), reduce computational complexity; however, we empirically find that the generalized bilinear layer outperforms these techniques. This layer learns an optimum blend of features from prices and tweets in a translationally invariant manner.

4.4 Graph Attention Network (GAT)

Stocks are often interlinked with one another, and thus, we model stocks and their relations as a graph.

Graph Creation Following Feng et al. (2019b), we make use of Wiki company-based relations. Using Wikidata⁴, we extract first and second-order relations between the company stocks in the S&P 500 index. A first-order relation is defined as $X \xrightarrow{R1} Y$ where X and Y denote entities in Wikidata that correspond to the two stocks. A second-order relation is defined by $X \xrightarrow{R2} Z \xleftarrow{R3} Y$ where Z denotes another entity connecting the two entities X

⁴https://www.wikidata.org/wiki/Wikidata:List_of_properties/all

and Y . R1, R2, and R3, defined in Wikidata, are different types of entity-relations. For instance, Wells Fargo and Bank of America are related to Berkshire Hathaway via a first-order company relation "owned by." Another example is Microsoft and Berkshire Hathaway that are related through Bill Gates (second-order relation: "owned by" - "is a board member of") since Bill Gates possesses ownership over Microsoft and is a Board member of Berkshire Hathaway. We define the stock relation network as a graph $G(S, E)$ where S denotes the set of nodes, and E is the set of edges. Each node $s \in S$ represents a stock, and two stocks $s_1, s_2 \in S$ are joined by an edge $e \in E$ if s_1, s_2 are linked by a first or second-order relation.

Graph Attention Graph-based representation learning through graph neural networks can be considered as information exchange between related nodes (Gilmer et al., 2017). As each stock has a different degree of influence on another stock, it is essential that the graph encoding suitably weighs more relevant relations between stocks. To this end, we use graph attention networks (GATs), which are graph neural networks with node-level attention (Veličković et al., 2017).

We first describe a single GAT layer that is used throughout the GAT component. The input to the GAT is a set of stock (node) features, $h = [x_1, x_2, \dots, x_{|S|}]$, where x_i is the encoded multi-modal market information (Sec. 4.3). The GAT layer produces an updated set of node features $h' = [z_1, z_2, \dots, z_{|S|}]$; $z_i \in \mathbb{R}^{w'}$ based on the GAT mechanism (shown in Figure 1). We first apply a shared linear transform parameterized by $W \in \mathbb{R}^{w' \times w}$ to all the nodes. Then, we apply a shared self-attention mechanism to each node i in its immediate neighborhood \mathcal{N}_i . For each node $j \in \mathcal{N}_i$, we compute normalized attention coefficients α_{ij} representing the importance of relations among stocks i and j . Formally, α_{ij} is given as:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(a_w^T [W x_i \oplus W x_j]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(a_w^T [W x_i \oplus W x_k]))} \quad (10)$$

where, \cdot^T and \oplus represent transpose and concatenation respectively. $a_w \in \mathbb{R}^{2w'}$ is a learnable weight matrix of a single layer feed forward neural network. The learned attention coefficients α_{ij} are used to weigh and aggregate feature vectors from neighboring with a non-linearity σ . The updated

node feature vector z_i is given as:

$$z_i = \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} W x_j \right) \quad (11)$$

We use multi-head attention to stabilise training (Vaswani et al., 2017). Formally, U independent executors apply the above attention mechanism. Their output features are concatenated to yield:

$$z_i = \bigoplus_{k=1}^U \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^k W^k x_j \right) \quad (12)$$

where, α_{ij}^k and W^k denote normalised attention coefficients and linear transformation parameter matrix computed by the k^{th} attention mechanism.

We use a two-layer GAT, the first layer is followed by Exponential Linear Unit (Clevert et al., 2015), and the second layer outputs a vector y_i for each stock i , which is then used to classify the stock’s future price movements. MAN-SF is trained using the Adam optimiser by optimizing the cross-entropy loss, given as:

$$L_{cse} = - \sum_{i=1}^{|\mathcal{S}|} Y_i \ln(y_i) + (1 - Y_i) \ln(1 - y_i) \quad (13)$$

where, Y_i is the true price movement of stock i .

5 Experiments

5.1 Dataset and Training Setup

We adopt the StockNet dataset (Xu and Cohen, 2018) for the training and evaluation of MAN-SF. The dataset contains data of high-trade-volume stocks in the S&P 500 index in the NYSE and NASDAQ markets. Stock specific tweets are extracted using regex queries made out of NASDAQ ticker symbols, for instance, \$AMZN for Amazon. The price data has been obtained from Yahoo Finance⁵. We shift a 5-day lag window along the trading days to generate samples. We label the samples according to the movement percentage of the closing price such that those $\geq 0.55\%$ and $\leq -0.5\%$ are labeled positive and negative samples, respectively. This leaves us with 26, 614 samples divided as 49.78% and 50.22% in the two classes. We temporally split the dataset in a ratio of Train:Validation:Test in 70:10:20, leaving us with date ranges from 01/01/2014 to 31/07/2015 for

⁵<https://finance.yahoo.com/industries>

training, 01/08/2015 to 30/09/2015 for validation, and 01/10/2015 to 01/01/2016 for testing. Following Xu and Cohen (2018), we align trading days by dropping samples that lack either prices or tweets, and further align the data across trading windows for related stocks to ensure data is available for all trading days in the window for all stocks. The hidden size of all GRUs is 64, and the USE embedding dimension is 512. We use $U = 8$ attention heads for both GAT layers. We use the Adam optimizer with a learning rate set to $5e-4$ and train MAN-SF for 10,000 epochs. It takes 3hrs to train and test MAN-SF on Tesla K80 GPU. We use early stopping based on Matthew’s Correlation Coefficient (MCC) taken over the validation set.

5.2 Evaluation

Following prior research for stock prediction (Ding et al., 2014; Xu and Cohen, 2018), we use accuracy, F1 score, MCC (implementations from *sklearn*⁶) for classification performance. We use MCC because, unlike the F1 score, MCC avoids bias due to data skew as it does not depend on the choice of the positive class and accounts for the True Negatives.

For a given confusion matrix $\begin{pmatrix} tp & fn \\ fp & tn \end{pmatrix}$:

$$MCC = \frac{tp \times tn - fp \times fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}} \quad (14)$$

Like prior work (Kim et al., 2019; Feng et al., 2019b), to evaluate MAN-SF’s applicability to real-world trading, we assess its profitability on the test data of the S&P 500 index using two metrics: Cumulative Profit and Sharpe Ratio (Sharpe, 1994). We follow a trading strategy where, if MAN-SF predicts a rise in a stock’s value the next day, then one share of that stock is bought (long position) at the closing price of the current trading session and sold on the next day’s closing price. Otherwise, if the strategy speculates a fall in price, a short sell⁷ is performed. We compute the cumulative profit (Krauss, 2018) earned as:

$$\text{Profit}_t = \sum_{i \in \mathcal{S}} \frac{p_i^t - p_i^{t-1}}{p_i^{t-1}} (-1)^{\text{Action}_i^{t-1}} \quad (15)$$

where, \mathcal{S} denotes the set of stocks, p_i^t denotes the price of stock i at day t . Action_i^{t-1} is a binary value $[0, 1]$. The Action_i^{t-1} is 0 if the long position is taken at time t for stock i ; otherwise it is 1.

⁶sklearn: <https://scikit-learn.org>

⁷Short sell: [https://en.wikipedia.org/wiki/Short_\(finance\)](https://en.wikipedia.org/wiki/Short_(finance))

	Model	F1 \uparrow	Accuracy \uparrow	MCC \uparrow
TA	RAND	0.502 \pm 8e-4	0.509 \pm 8e-4	-0.002 \pm 1e-3
	ARIMA (Brown, 2004)	0.513 \pm 1e-3	0.514 \pm 1e-3	-0.021 \pm 2e-3
	Selvin et al. (2017)	0.529 \pm 5e-2	0.530 \pm 5e-2	-0.004 \pm 7e-2
FA	RandForest (Venkata Sasank Pagolu, 2016)	0.527 \pm 2e-3	0.531 \pm 2e-3	0.013 \pm 4e-3
	TSLDA (Nguyen and Shirai, 2015)	0.539 \pm 6e-3	0.541 \pm 6e-3	0.065 \pm 7e-3
	HAN (Hu et al., 2018)	0.572 \pm 4e-3	0.576 \pm 4e-3	0.052 \pm 5e-3
	StockNet - TechnicalAnalyst (Xu and Cohen, 2018)	0.546 \pm -	0.550 \pm -	0.017 \pm -
	StockNet - FundamentalAnalyst (Xu and Cohen, 2018)	0.572 \pm -	0.582 \pm -	0.072 \pm -
	StockNet - IndependentAnalyst (Xu and Cohen, 2018)	0.573 \pm -	0.575 \pm -	0.037 \pm -
	StockNet - DiscriminativeAnalyst (Xu and Cohen, 2018)	0.559 \pm -	0.562 \pm -	0.056 \pm -
	StockNet - HedgeFundAnalyst (Xu and Cohen, 2018)	0.575 \pm -	0.582 \pm -	0.081 \pm -
	HATS (Kim et al., 2019)	0.560 \pm 2e-3	0.562 \pm 2e-3	0.117 \pm 6e-3
	Chen et al. (2018)	0.530 \pm 7e-3	0.532 \pm 7e-3	0.093 \pm 9e-3
	Adversarial LSTM (Feng et al., 2019a)	0.570 \pm -	0.572 \pm -	0.148 \pm -
	MAN-SF (This work)	0.605 \pm 2e-4	0.608 \pm 2e-4	0.195 \pm 6e-4

Table 1: Results compared with baselines. **Bold** shows the best results. Green is indicative of higher performance. TA and FA represent Technical Analysis and Fundamental Analysis models, respectively.

The Sharpe Ratio is a measure of the return of a portfolio compared to its risk. We calculate the Sharpe ratio by computing the ratio of the expected return R_a of a portfolio to its standard deviation as:

$$\text{Sharpe Ratio}_a = \frac{E[R_a]}{\text{std}[R_a]} \quad (16)$$

5.3 Baselines

We compare MAN-SF with the below baselines spanning both technical and fundamental analysis.

Technical Analysis: These methods uses only historical price information.

- **RAND:** Random guess as price rise or fall.
- **ARIMA:** Autoregressive Integrated Moving Average models historical prices as a non-stationary time series (Brown, 2004).
- **Selvin et al. (2017):** Three deep neural architectures (RNN, CNN and LSTM) using prices. We compare with the best performing LSTM.

Fundamental Analysis: These methods use other modalities such as text information and company relationships along with historical prices.

- **RandForest:** Random Forests classifier trained over word2vec (Mikolov et al., 2013) embeddings for tweets.
- **TSLDA:** Topic Sentiment Latent Dirichlet Allocation model is a generative model that uses sentiments and topic modeling on social media (Nguyen and Shirai, 2015).

- **HAN:** A hierarchical attention mechanism to encode textual information during a day and across multiple days (Hu et al., 2018).
- **StockNet:** A variational Autoencoder (VAE) that uses price and text information. Text is encoded using hierarchical attention during and across days. Price features are modeled sequentially (Xu and Cohen, 2018). We compare with all five variants of StockNet.
- **HATS:** A hierarchical graph attention method that uses a multi-graph to weigh different relationships between stocks. It uses only historical price data (Kim et al., 2019).
- **Chen et al. (2018):** GCNs to model inter stock relations with only historical price data.

6 Results and Analysis

We now discuss the experimental results and some findings with their financial implications.

Performance Comparison Table 1 shows the performance of the compared methods on StockNet’s test data split from 01/10/2015 to 31/12/2015 on the S&P 500 index averaged over ten different runs. Using a learned blend of historical price and tweets using corporate relationships, MAN-SF achieves the best performance, outperforming the strongest baselines, StockNet, and Adversarial LSTM. We also note that Fundamental Analysis (FA) techniques outperform numerical only Technical Analysis (TA) methods, reiterating the effectiveness of factoring in social media signals and

Model Component	F1 \uparrow	MCC \uparrow
LSTM + Historical Price	0.521	0.002
GRU + Social Media Text (BERT)	0.539	0.077
GCN + Historical Price	0.532	0.093
GRU + Social Media Text (USE)	0.546	0.101
GCN + Social Media Text (USE)	0.555	0.102
GAT + Historical Price	0.562	0.117
MAN-SF (Concatenation)	0.588	0.156
MAN-SF (Attention Fusion)	0.594	0.173
MAN-SF (Bilinear Transformation)	0.605	0.195

Table 2: Ablation study over MAN-SF’s components.

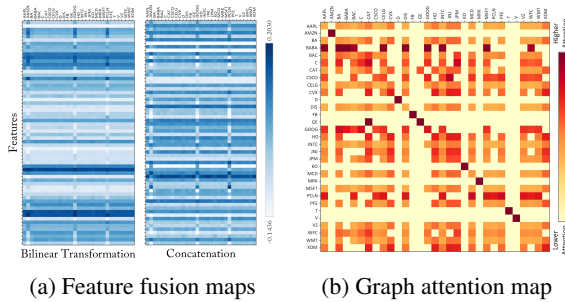


Figure 4: Feature weight heatmaps for MAN-SF

inter stock relations. These results empirically validate the effectiveness of multimodal signals due to a broader capture of stock price influencing information, including tweets and other related stocks.

Ablation Study In Table 2, we observe the ability of price and text models to predict the market trend to an extent using unimodal features. Improvements over individual modalities are noted with the inclusion of a graph-based learning model, i.e., GCN and GAT validating the premise of using inter stock relations for enhanced forecasting. When the text and price signals are fused, and more relevant information is extracted using the attention mechanisms, a performance gain is seen. The ablation study ties up with the EMH, as we add additional modalities, we note an increment in MAN-SF’s ability for stock prediction. Two critical observations from Table 2 are the substantial MCC gains when using GAT over GCN and the contrast between fusing text and prices via concatenation and bilinear transformations. We discuss these next.

Impact of Bilinear Transformations Bilinear blending outperforms concatenation, and attention fusion variants, as seen in Table 2. We postulate that the bilinear transformation can better learn the interplay between the signals compared to other variants. On examining Figure 4a, we observe that the bilinear layer blends highly non-linear relation-

Table 3: Annualized Sharpe Ratio comparison with baselines. **Bold** and *italics* denotes best and second best results, respectively.

Model	Sharpe Ratio \uparrow
Stocknet	0.83
HATS	0.78
MAN-SF	1.05

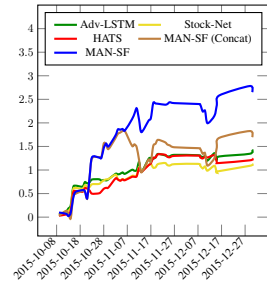


Figure 6: Cumulative profit trend

ships between the two signals leading to a joint representation that captures more specific features noticed by areas of concentrated attention as compared to simple concatenation based fusion.

Analyzing Graph Attention We notice that equally weighing all correlations using GCN-based models leads to smaller performance gains, as shown in Table 2, as compared to GAT (GAT, and MAN-SF variants). To analyze this difference, we first calculate each neighbor’s attention scores in the stock relations graph, as shown in Figure 4b. By analyzing the different stock associations with the highest and lowest attention scores, we observe that some relations between stocks, such as being a part of the same industry or having the same founder, are more critical than other relations like stocks having the same country of origin. For instance, C (CitiCorp) and JPM (JP Morgan) have a relatively high attention score and are a part of the same investment and banking industry, whereas the attention score for JPM and CSCO (Cisco) is relatively low. We also observe that some stocks share hidden correlations captured by the GAT due to the market’s temporal nature. We explain one such example in Section 7.

Profitability We examine MAN-SF’s practical applicability through a profitability analysis on real-world stock data. From Table 3 and Figure 6, we note that MAN-SF achieves higher risk-adjusted returns and an overall profit. MAN-SF outperforms different baselines over the common testing period of three months using the stocks data in the S&P 500 index. These observations show the profitability of MAN-SF over models that do not capture stock correlations (StockNet) and models that do not use the impact of textual data (HATS). We potentially attribute these improvements to MAN-SF’s ability to learn a more concentrated blend of text and price features as opposed to competitive

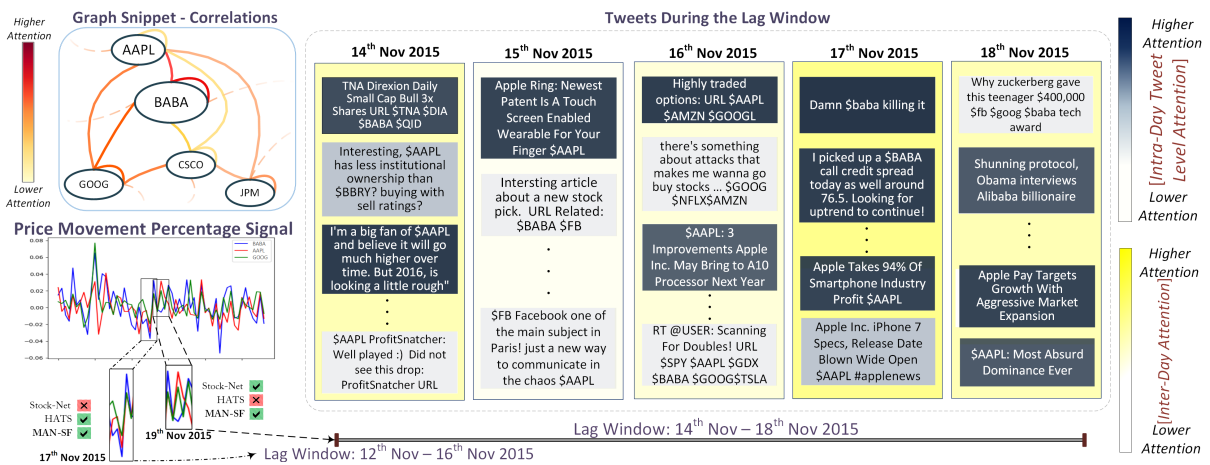


Figure 5: Graph sample showing attention weights for stock correlations (top left); Stock price movement depicting inter-stock relationships (bottom left); Tweets with hierarchical temporal attention weights (right)

models. We extend this analysis in the next section.

7 Qualitative Analysis

We conduct an extended analysis across two high-risk scenarios, as shown in Figure 5, to study the applicability of MAN-SF to investors in the stock market. The study is based on Apple’s (AAPL) trend during 12th Nov - 18th Nov. Figure 5 shows some of the tweets posted and AAPL’s relations with relevant stocks such as Alibaba (BABA), Google (GOOG), and among others during that period.

12th Nov to 16th Nov: Failure of StockNet and models that do not capture inter stock relations: From Figure 5, we see from the price movement that 12th to 16th November 2015 shows a decline in Apple’s stock price. Here, we observe that Stock-Net predicts a further drop in Apple’s price, and similar models that use only price and text are unable to predict the price rise for Apple on 17th November correctly. However, we discover that Apple shares a strong relationship with Alibaba and Google during that time, as indicated by the attention weights. MAN-SF incorporates inter-stock relations through graph attention to learn latent correlations between AAPL, BABA, and GOOG, as shown by the graph snippet in Figure 5. MAN-SF correctly predicts a rise in Apple’s price and makes a profit, unlike StockNet. We attribute this prediction to MAN-SF likely having a broader context by blending multimodal signals.

14th Nov to 18th Nov: Failure of HATS and models that do not leverage social media data: Despite Apple’s sharp fall on 18th November, we see tweets with positive sentiment having higher

attention weights during the lookback window, indicating a possible increase in Apple’s price. MAN-SF uses hierarchical attention mechanisms over tweets and inter-stock correlations correctly. Thereby likely predicting a rise in Apple’s stock price, similar to models such as StockNet. As opposed to these, models such as HATS forecast a continual decrease in Apple’s price, potentially due to not factoring in social media data.

8 Conclusion and Future Work

We study stock movement prediction by using natural language, graph-based and numeric features. We propose MAN-SF, a neural model that jointly learns *temporally relevant* signals from *chaotic* multimodal data spanning historical prices, tweets, and inter stock correlations in a hierarchical fashion. Extensive quantitative and qualitative experiments on real market data demonstrate MAN-SF’s applicability for neural stock forecasting. We plan to further use news articles, earnings calls, and other data sources to capture market dynamics better. Another interesting direction of future research is to explore the cold start problem, where MAN-SF could be leveraged to predict stock movements for new stocks. Lastly, we would also like to extend MAN-SF’s architecture to not be limited to model all stocks together (because of its GAT component) to increase scalability to cross-market scenarios.

References

Yaser S. Abu-Mostafa and Amir F. Atiya. 1996. [Introduction to financial forecasting](#). *Applied Intelligence*, 6(3):205–213.

- Klaus Adam, Albert Marcet, and Juan Pablo Nicolí. 2016. [Stock market volatility and learning](#). *The Journal of Finance*, 71(1):33–82.
- Leif B. G. Andersen. 2007. [Efficient simulation of the heston stochastic volatility model](#). *SSRN Electronic Journal*.
- Joseph H Anthony and K Ramesh. 1992. [Association between accounting performance measures and stock prices](#). *Journal of Accounting and Economics*, 15(2-3):203–227.
- Uma Devi B, Sundar D, and Alli P. 2013. [An effective time series analysis for stock trend prediction using ARIMA model for nifty midcap-50](#). *International Journal of Data Mining & Knowledge Management Process*, 3(1):65–78.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. [Neural machine translation by jointly learning to align and translate](#). *arXiv preprint arXiv:1409.0473*.
- Fischer Black and Myron Scholes. 1973. [The pricing of options and corporate liabilities](#). *Journal of Political Economy*, 81(3):637–654.
- David C Blitz and Pim Van Vliet. 2007. [The volatility effect](#). *The Journal of Portfolio Management*, 34(1):102–113.
- J. Bollen, H. Mao, and X. Zeng. 2011. [Twitter mood predicts the stock market](#). *Journal of Computational Science*, 2(1):1–8. Cited By 2072.
- Tim Bollerslev. 1986. [Generalized autoregressive conditional heteroskedasticity](#). *Journal of Econometrics*, 31(3):307–327.
- Robert Goodell Brown. 2004. [Smoothing, forecasting and prediction of discrete time series](#). *Courier Corporation*.
- Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, et al. 2018. [Universal sentence encoder for english](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 169–174.
- Yingmei Chen, Zhongyu Wei, and Xuanjing Huang. 2018. [Incorporating corporation relationship via graph convolutional neural networks for stock price prediction](#). In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM '18*, page 1655–1658, New York, NY, USA. Association for Computing Machinery.
- Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. [On the properties of neural machine translation: Encoder-decoder approaches](#). *arXiv preprint arXiv:1409.1259*.
- Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. 2015. [Fast and accurate deep network learning by exponential linear units \(elus\)](#). *arXiv preprint arXiv:1511.07289*.
- Shumin Deng, Ningyu Zhang, Wen Zhang, Jiaoyan Chen, Jeff Z. Pan, and Huajun Chen. 2019. [Knowledge-driven stock trend prediction and explanation via temporal convolutional network](#). In *Companion Proceedings of The 2019 World Wide Web Conference, WWW '19*, page 678–685, New York, NY, USA. Association for Computing Machinery.
- Iliia D. Dichev and Vicki Wei Tang. 2006. [Earnings volatility and earnings predictability](#). *SSRN Electronic Journal*.
- F.X. Diebold and K. Yilmaz. 2014. [On the network topology of variance decompositions: Measuring the connectedness of financial firms](#). *Journal of Econometrics*, 182(1):119–134. Cited By 416.
- Guangyu Ding and Liangxi Qin. 2019. [Study on the prediction of stock price based on the associated network model of lstm](#). *International Journal of Machine Learning and Cybernetics*.
- Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2014. [Using structured events to predict stock price movement: An empirical investigation](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1415–1425.
- Edwin J Elton, Martin J Gruber, Stephen J Brown, and William N Goetzmann. 2009. *Modern portfolio theory and investment analysis*. John Wiley & Sons.
- Claude B Erb, Campbell R Harvey, and Tadas E Viskanta. 1994. [Forecasting international equity correlations](#). *Financial analysts journal*, 50(6):32–45.
- Fuli Feng, Huimin Chen, Xiangnan He, Ji Ding, Maosong Sun, and Tat-Seng Chua. 2019a. [Enhancing stock movement prediction with adversarial training](#). In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 5843–5849. International Joint Conferences on Artificial Intelligence Organization.
- Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. 2019b. [Temporal relational ranking for stock prediction](#). *ACM Trans. Inf. Syst.*, 37(2).
- Gabriel Pui Cheong Fung, Jeffrey Xu Yu, and Wai Lam. 2002. [News sensitive stock trend prediction](#). In *Advances in Knowledge Discovery and Data Mining*, pages 481–493. Springer Berlin Heidelberg.
- C. Lee Giles, Steve Lawrence, and Ah Chung Tsoi. 2001. [Noisy time series prediction using a recurrent neural network and grammatical inference](#). *Machine Learning*, 44(1/2):161–183.

- Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. 2017. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70, ICML'17*, page 1263–1272. JMLR.org.
- Alexandre Hocquard, Sunny Ng, and Nicolas Papanicolaou. 2013. A constant-volatility framework for managing tail risk. *The Journal of Portfolio Management*, 39(2):28–40.
- Ehsan Hoseinzade, Saman Haratizadeh, and Arash Khoeini. 2019. [U-cnnpred: A universal cnn-based predictor for stock markets](#).
- Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. 2018. 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*, pages 261–269.
- Monique Jeanblanc, Marc Yor, and Marc Chesney. 2009. *Mathematical methods for financial markets*. Springer Science & Business Media.
- Weiwei Jiang. 2020. Applications of deep learning in stock market prediction: recent progress.
- Armand Joulin, Édouard Grave, Piotr Bojanowski, and Tomáš Mikolov. 2017. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 427–431.
- Wei Du Jue Liu, Zhuocheng Lu. 2019. Combining enterprise knowledge graph and news sentiment analysis for stock price volatility prediction. *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
- Raehyun Kim, Chan Ho So, Minbyul Jeong, Sanghoon Lee, Jinkyu Kim, and Jaewoo Kang. 2019. Hats: A hierarchical graph attention network for stock movement prediction. *arXiv preprint arXiv:1908.07999*.
- Kazuhiro Kohara, Tsutomu Ishikawa, Yoshimi Fukuhara, and Yukihiro Nakamura. 1997. Stock price prediction using prior knowledge and neural networks. *Intelligent Systems in Accounting, Finance & Management*, 6(1):11–22.
- Thomas Fischer Christopher Krauss. 2018. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*.
- Hao Li, Yanyan Shen, and Yanmin Zhu. 2018. [Stock price prediction using attention-based multi-input lstm](#). In *ACML*, volume 95 of *Proceedings of Machine Learning Research*, pages 454–469. PMLR.
- Xiaodong Li, Xiaodi Huang, Xiaotie Deng, and Shan-feng Zhu. 2014. Enhancing quantitative intra-day stock return prediction by integrating both market news and stock prices information. *Neurocomputing*, 142:228–238.
- Xiaodong Li, Haoran Xie, Ran Wang, Yi Cai, Jingjing Cao, Feng Wang, Huaqing Min, and Xiaotie Deng. 2016. Empirical analysis: stock market prediction via extreme learning machine. *Neural Computing and Applications*, 27(1):67–78.
- Burton G. Malkiel. 1989. [Efficient market hypothesis](#). In *Finance*, pages 127–134. Palgrave Macmillan UK.
- Daiki Matsunaga, Toyotaro Suzumura, and Toshihiro Takahashi. 2019. Exploring graph neural networks for stock market predictions with rolling window analysis. *ArXiv*, abs/1909.10660.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Robert Neal and Simon M Wheatley. 1998. Do measures of investor sentiment predict returns? *Journal of Financial and Quantitative Analysis*, 33(4):523–547.
- Duc Huu Dat Nguyen, Loc Phuoc Tran, and Vu Nguyen. 2019. [Predicting stock prices using dynamic LSTM models](#). In *Communications in Computer and Information Science*, pages 199–212. Springer International Publishing.
- Thi-Thu Nguyen and Seokhoon Yoon. 2019. [A novel approach to short-term stock price movement prediction using transfer learning](#). *Applied Sciences*, 9(22):4745.
- Thien Hai Nguyen and Kiyooki Shirai. 2015. Topic modeling based sentiment analysis on social media for stock market prediction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1354–1364.
- Mahla Nikou, Gholamreza Mansourfar, and Jamshid Bagherzadeh. 2019. [Stock price prediction using deep learning algorithm and its comparison with machine learning algorithms](#). *Intelligent Systems in Accounting, Finance and Management*, 26.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. [GloVe: Global vectors for word representation](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.
- Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, and Garrison Cottrell. 2017. A

- dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv:1704.02971*.
- Yu Qin and Yi Yang. 2019. What you say and how you say it matters: Predicting financial risk using verbal and vocal cues. In *57th Annual Meeting of the Association for Computational Linguistics (ACL 2019)*, page 390.
- Robert Rich and Joseph Tracy. 2004. Uncertainty and labor contract durations. *Review of Economics and Statistics*, 86(1):270–287.
- Hsinchun Chen Robert P. Schumaker. 2019. Textual analysis of stock market prediction using breaking financial news: The azfin text system. *ACM Transactions on Information Systems*.
- Ramit Sawhney, Shivam Agarwal, Arnav Wadhwa, and Rajiv Ratn Shah. 2020a. Spatiotemporal hypergraph convolution network for stock forecasting. In *2020 IEEE International Conference on Data Mining (ICDM)*.
- Ramit Sawhney, Puneet Mathur, Ayush Mangal, Piyush Khanna, Rajiv Shah, and Roger Zimmermann. 2020b. **Multimodal multi-task financial risk forecasting**. In *Proceedings of the 28th ACM International Conference on Multimedia, MM '20*, New York, NY, USA. Association for Computing Machinery.
- Sreelekshmy Selvin, R Vinayakumar, EA Gopalakrishnan, Vijay Krishna Menon, and KP Soman. 2017. Stock price prediction using lstm, rnn and cnn-sliding window model. In *2017 international conference on advances in computing, communications and informatics (icacci)*, pages 1643–1647. IEEE.
- William F Sharpe. 1994. The sharpe ratio. *Journal of portfolio management*, 21(1):49–58.
- Matheus José Silva de Souza, Danilo Guimarães Franco Ramos, Marina Garcia Pena, Vinicius Amorim Sobreiro, and Herbert Kimura. 2018. Examination of the profitability of technical analysis based on moving average strategies in brics. *Financial Innovation*, 4(1):3.
- Paul C. Tetlock. 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. **Attention is all you need**.
- Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.
- Babita Majhi Venkata Sasank Pagolu, Kamal Nayan Reddy ; Ganapati Panda. 2016. Sentiment analysis of twitter data for predicting stock market movements. *SCOPES*.
- Yumo Xu and Shay B. Cohen. 2018. Stock movement prediction from tweets and historical prices. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1970–1979, Melbourne, Australia. Association for Computational Linguistics.
- Linyi Yang, Tin Lok James Ng, Barry Smyth, and Rihai Dong. 2020. **Html: Hierarchical transformer-based multi-task learning for volatility prediction**. In *Proceedings of The Web Conference 2020, WWW '20*, page 441–451, New York, NY, USA. Association for Computing Machinery.
- Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489.
- Zhou Yu, Jun Yu, Jianping Fan, and Dacheng Tao. 2017. Multi-modal factorized bilinear pooling with co-attention learning for visual question answering. In *Proceedings of the IEEE international conference on computer vision*, pages 1821–1830.
- Z. Zhao, R. Rao, S. Tu, and J. Shi. 2017. Time-weighted lstm model with redefined labeling for stock trend prediction. In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 1210–1217.