Modeling Financial Uncertainty with Multivariate Temporal Entropy-based Curriculums

Ramit Sawhney*[1] **Arnav Wadhwa*[1]** **Ayush Mangal[2]** **Vivek Mittal[3]** **Shivam Agarwal[4]** **Rajiv Ratn Shah[1]**

[1]IIIT Delhi
[2]Indian Institute of Technology Roorkee
[3]Indian Institute of Technology Mandi
[4]Manipal Institute of Technology

Abstract

In the financial realm, profit generation greatly relies on the complicated task of stock prediction. Lately, neural methods have shown success in exploiting stock affecting signals from textual data across news and tweets to forecast stock performance. However, the dynamic, stochastic, and variably influential nature of text and prices makes it difficult to train neural stock trading models, limiting predictive performance and profits. To transcend this limitation, we propose a novel multimodal curriculum learning approach: FinCLASS, which evaluates stock affecting signals via entropy-based heuristics and measures their linguistic and price-based complexities in a time-aware, hierarchical fashion. We show that training financial models can benefit by exposing neural networks to easier examples of stock affecting signals early during the training phase, before introducing samples having more complex linguistic and price-based temporal variations. Through experiments on benchmark English tweets and Chinese financial news spanning two major indexes and four global markets, we show how FinCLASS outperforms state-of-the-art across financial tasks of stock movement prediction, volatility regression, and profit generation. Through ablative and qualitative experiments, we set the case for FinCLASS as a generalizable framework for developing natural language-centric neural models for financial tasks.

1 INTRODUCTION

The ever-growing $60 trillion stock markets around the world present incredibly lucrative opportunities to make monetary profits. However, profit generation heavily relies on financial tasks such as forecasting stock movements and analyzing associated risk, which are complex due to the highly stochastic and dynamic nature of stock prices [Adam et al., 2016]. Prices are greatly influenced by factors such as investor sentiments, opinions about companies expressed across social media platforms (Twitter, Stocktwits), financial news, and much more. The abundance of stock affecting information present on the web helps investors analyze market trends and inspires the adoption of AI-based methods to study the interplay between online textual data and stock prices [Xu and Cohen, 2018, Du and Tanaka-Ishii, 2020].

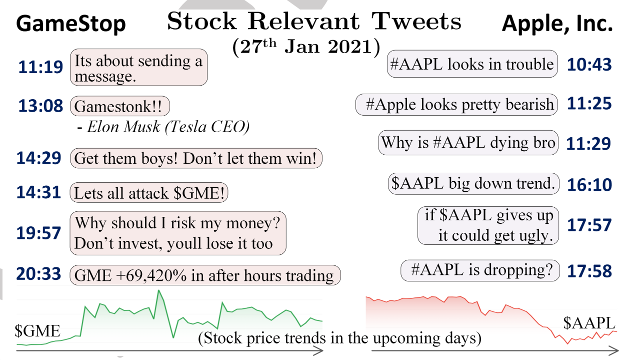

Figure 1: Tweets for GameStop reflect greater linguistic complexity, temporal variations, and context-dependence. In contrast, tweets for Apple are relatively easier to comprehend. Consequently, predicting stock performance from tweets for Apple is easier when compared to GameStop.

However, analyzing and modeling the complex unstructured textual data involves numerous challenges. Stock affecting signals across textual sources like financial news and tweets exhibit *sequential context dependencies* and manifest a *variably influential nature* on the market [Hu et al., 2018]. Further, online text possesses inherent *dynamic timing irregularities*, which makes it challenging to model the influence of their temporal and linguistic variations on market trends.

*Indicates equal contribution to this work.

Accepted for the 37th Conference on Uncertainty in Artificial Intelligence (UAI 2021).

Consider Figure 1, which presents two instances of a time-series of tweets exhibiting varying linguistic complexities to examine future stock trends. Studying the influence of such complicated textual signals (GameStop) on the market is arduous not only for human traders but also for neural stock prediction methods, thus limiting their predictive performance and practical applicability to real-world trading.

To overcome this limitation of neural models, we draw inspiration from human learning paradigms [Krueger and Dayan, 2009] and machine learning research, which indicates that presenting training data in a deliberated scheme, starting from easy to difficult samples, can greatly help models learn trends in datasets [Bengio et al., 2009]. We introduce **FinCLASS** – **Fin**ancial **C**urriculum **L**earning for natural language-based **A**lgorithmic **S**tock trading **S**trategies, a novel framework for defining multi-modal financial curriculums (**§3**) for model training to enhance the profitability of neural stock trading methods (**§4**). FinCLASS assigns a difficulty score to each data-point in the training set via a blend of model-based and multi-modal (numerical and text) entropy-based heuristics, and re-orders training data based on sample difficulty levels. We first train the model over *easy* samples and as the training progresses, we gradually introduce more complex data samples. Such a training paradigm provides better model training and significant performance gains across financial tasks. Through experiments (**§5**) on English and Chinese text corresponding to the NASDAQ, Shanghai, Shenzhen, and Hong Kong markets, we show that FinCLASS helps outperform state-of-the-art stock prediction methods across three financial tasks: stock movement prediction, volatility regression, and profit generation (**§6.1**,**§6.2**,**§6.3**). Through exploratory (**§6.4**,**§6.5**) and qualitative (**§6.6**) analyses, we show the efficacy of FinCLASS as a curriculum learning paradigm for algorithmic trading.

The contributions of this study can be summarized as:

- We define a novel multi-modal entropy-based algorithm that measures the linguistic and price-based temporal variations across stock affecting data (text, prices) in a time-aware, hierarchical fashion. The proposed algorithm is flexible and generalizable across numerous tasks involving multi-modal multivariate data.

- We introduce FinCLASS, an end-to-end curriculum learning framework to enhance the training process and practical applicability of language-based neural methods for a variety of real-world financial tasks.

- We define a text-based stock trading model THA-Net, and show how FinCLASS helps THA-Net outperform state-of-the-art baselines on three financial tasks using stock affecting Chinese financial news and English tweets pertaining to stocks in four global markets.

2 RELATED WORK

2.1 FINANCE: STOCK PREDICTION

Conventional Methods Stock prediction spans various methods, commonly formulated as either regression or classification tasks [Jiang, 2020]. Conventional stock prediction methods rely on numeric features like historical prices [Kohara et al., 1997, Lin et al., 2009], technical indicators [Shynkevich et al., 2017], and macroeconomic indicators [Hoseinzade et al., 2019]. These include discrete [Bollerslev, 1986], continuous [Andersen, 2007], and neural approaches [Feng et al., 2019]. Despite their success, a fundamental shortcoming of these methods is that they are limited to numerical features and do not study crucial stock influencing factors across online textual media.

Contemporary Methods Newer models based on the Efficient Market Hypothesis [Malkiel, 1989], leverage natural language features extracted from investor sentiments [Schumaker and Chen, 2009], public earnings calls [Qin and Yang, 2019], online news [Hu et al., 2018, Du and Tanaka-Ishii, 2020] and social media [Xu and Cohen, 2018] for stock prediction. Recent work also demonstrates the benefits of modeling stock affecting signals from sources belonging to multiple modalities [Sawhney et al., 2020, Zhou et al., 2020]. These works show how natural language sources can complement price-based methods in capturing the effect of events like market surprises, mergers and acquisitions over stock returns. However, a major limitation of these methods is that they do not account for the irregularities in release times of stock affecting text [Sawhney et al., 2021b,a]. For trading, the timing plays a critical role, as stock prices rapidly factor all available public market information. Firms may even exploit perception of information [Forbes, 2009] by timing the release of negative news between positive ones to minimize losses [Segal and Segal, 2016].

2.2 CURRICULUM LEARNING

The idea of training neural models based on an easy-to-difficult fashion, was initially advanced as curriculum learning (CL) by Bengio et al. [2009]. Since then, CL found numerous applications in computer vision [Pentina et al., 2015], finance [Koenecke and Gajewar, 2019], NLP [Xu et al., 2020], reinforcement learning [Narvekar et al., 2020], and more. However, a prime limitation of existing NLP and finance-based CL methods is that they do not define the sample-difficulty by considering the complex linguistic and price-based temporal variations across a sequence of stock influencing modalities in a time-aware, hierarchical fashion. Building on the limitations of existing stock prediction and CL methods, we propose a new multi-modal time-aware hierarchical entropy-based CL framework that enhances the training and profitability of stock prediction models.

3 FINCLASS

Algorithm 1: Multi-modal Multivariate Hierarchical Time-Aware Entropy-based Stock Complexity S_d

Input: Text $X_{\tau-T,\tau-T+1,\ldots,\tau-1}$, and Price Vectors $P_{\tau-T,\tau-T+1,\ldots,\tau-1} = [P_{open}, P_{high}, P_{low}, P_{close}]$ during a Lookback Window T, Financial BERT

Output: Stock Complexity S_d

Step 1: Extract Financial Sentiment Time-Series

for *each day* i *in* $\tau - T$ *to* $\tau - 1$ **do**
  for *each tweet* $X_{it} \in X_i$ **do**
    Get sentiment vector p_{it} $[p_{bearish}, p_{bullish}, p_{neutral}] \leftarrow FinBERT(X_{it})$
  end
end

Step 2: Compute Intraday Multivariate Permutation Entropy (MPE) [Capturing intraday variations]

for *each day* $i \in \tau - T$ *to* $\tau - 1$ **do**
  $mpe_i \leftarrow MPE(p_{i0}, p_{i1}, p_{i2}, \ldots, p_{iK})$
end

Compute the Permutation Entropy (PE) of the obtained MPE values for all days

$\mathbf{E_{intraday}} \leftarrow PE(mpe_{\tau-T}, mpe_{\tau-T+1}, \ldots, mpe_{\tau-1})$

Step 3: Compute Interday Dynamic Time Warping (DTW) distance [Capturing interday variations]

for *each consecutive day-pair* $(i, j) \in T$ **do**
  $dtw_{i,j} \leftarrow DTW([p_{i0}, \ldots, p_{iK}], [p_{j0}, \ldots, p_{jK}])$
end

Compute the Permutation Entropy (PE) of the obtained DTW distances

$\mathbf{E_{interday}} \leftarrow PE(dtw_{\tau-T,\tau-T+1}, \ldots, dtw_{\tau-2,\tau-1})$

Step 4: Compute the Multivariate Permutation Entropy of stock price series across days [Capturing interday price variations]

$\mathbf{E_{price}} \leftarrow MPE(P_{\tau-T}, P_{\tau-T+1}, \ldots, P_{\tau-1})$

Step 5: Fuse the text and price entropies to obtain the stock complexity S_d

$\mathbf{S_d} \leftarrow \mathbf{E_{price}} + \mathbf{E_{intraday}} + \mathbf{E_{interday}}$

return S_d

Consider a dataset where each data sample comprises stock-relevant financial news items or tweets over a lookback of T days in range $[\tau - T, \tau - 1]$ to study the performance of stock on day τ. We define FinCLASS as an approach to re-arrange the dataset according to a learning curriculum based on data-sample complexities to train neural models for financial tasks. To develop a neural model Φ for stock prediction, let D be the set of training examples in a dataset. We compute a difficulty score χ corresponding to each training sample $d_s \in D$ and re-arrange the training dataset based on difficulty ranging from easy to hard. To determine the scores χ, we compute the stock complexity S_d, and the model complexity M_d for each data sample as follows.

Stock-complexity S_d For each data sample $d_s \in D$, S_d reflects the linguistic complexity of the stock affecting textual signals and the variations in price signals present in d_s. To capture the stock complexity of a sample, we propose a novel multi-modal multivariate hierarchical time-aware entropy-based algorithm, as presented in Algorithm 1. Entropy indicates the volatility and associated risk based on the complexity of financial time-series [Pincus and Kalman, 2004]. For a given lookback T, we exploit the price and text information available corresponding to a stock s. First, we compute the bearish, neutral, and bullish intent of each news item or tweet in the lookback using class probabilities obtained via fine-tuned Financial BERT for English tweets [Araci, 2019] and Chinese financial text [Rao et al., 2021]. We then form a time series corresponding to all three intents separately for each day in T. To measure the three intents' variations and trends over a day, we compute the Multivariate Permutation Entropy (MPE) [Morabito et al., 2012] of the three time-series of intents, which reflects the linguistic complexity involved in studying financial news or tweets over the day. For multiple channel signals, each time series is usually considered separately while computing the entropy involved. Such a procedure may be acceptable for uncorrelated signals, but the multivariate financial sentiment time series we compute are highly correlated in nature. Thus we use MPE, which effectively captures the cross channel complexities as well. In MPE, the original multivariate time series is transformed into a time dependent matrix from which the relevant statistics and entropies are extracted. For days where the financial texts exhibit greater linguistic and intent variations, MPE would be higher compared to days where texts indicate relatively consistent sentiments towards a stock. To analyze how the temporal evolution of stock-affecting signals varies across days, we adopt Dynamic Time Warping (DTW) distance [Müller, 2007] which measures the similarities between time-series of the three intents across consecutive days in a time-aware fashion. For data-samples where financial sentiments across news and tweets show large inter-day variations, DTW distance would be higher, indicating greater difficulty in analyzing stock behavior across days. Next, for each day in the lookback, we form a price vector $p_i = [p_i^o, p_i^h, p_i^l, p_i^c]$ comprising the stock's opening, highest, lowest, and adjusted closing prices for trading day i. Financial research indicates that irregular price trends lead to increased difficulties and complexities in predicting future stock performance [Adam et al., 2016]. To this end, we calculate the MPE of all price vectors p_i to measure the irregularities in the time-series of historic prices across days. We then fuse the MPE of sentiment trends for each day, the DTW distance of sentiment trends across days,

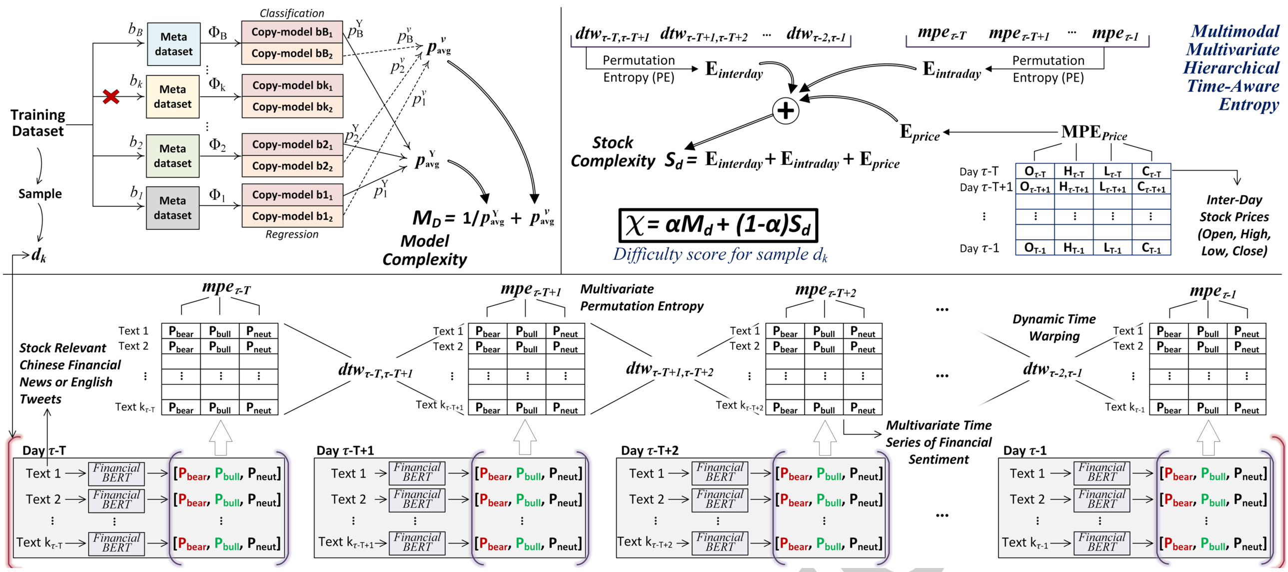

Figure 2: A high-level overview of FinCLASS sample difficulty score computation. Input: Data sample comprising stock-relevant financial news or tweets, and prices over a lookback period of T days; Output: sample difficulty score χ.

and the MPE of price features to obtain stock-complexity S_d, as shown in Figure 2. Given that entropy is an additive quantity, we add them to find S_d. We experiment with other fusion types such as dot product and weighted sum, but could find no significant improvement over addition.

Model-complexity M_d For a sample $d_s \in D$, M_d reflects the difficulty faced by a neural model in accurately mapping d_s to the ground truth given a financial task. We compute the model-complexity for each sample $d_s \in D$ based on the confidence score (movement classification) and regression (volatility) scores obtained from the model Φ. To compute M_d, we first scatter the training set D into B uniform meta-datasets, each having $1/B$ of the total samples in D. We then train $2 \times B$ identical copies of the model Φ, two over each meta-dataset, for price movement classification and volatility regression, respectively. For each example d_k belonging to meta-dataset b, we compute the classification (softmax probabilities for price movement) and regression (predicted volatilities) scores for corresponding tasks from all copies of the model except the ones which were trained on meta-dataset b. We then identify the class Y predicted by a majority of the classification copy-models for sample d_k, and compute the average p^Y_{avg} of the confidence scores generated for class Y by each classification copy-model. Next, we compute the average p^v_{avg} of magnitude of difference between the true v_τ and the predicted $\hat{v_\tau}$ volatilities for sample d_k as obtained by each regression copy-model. We then compute the model-complexity as $M_d = p^v_{avg} + 1/p^Y_{avg}$. Lastly, for each training sample $d_s \in D$, we define and compute the ***overall difficulty score*** χ as:

$$\chi = \alpha M_d + (1 - \alpha) S_d \tag{1}$$

where α is a learnable parameter for the two complexities.

Curriculum definition We re-arrange the training dataset based on difficulty scores χ and expose the model to more difficult samples with each consecutive epoch. Formally, we first sort the training samples d_s by their overall difficulty scores χ and divide the scores into B equal sized ranges. Next, we arrange each sample $d_s \in D$ into the B groups, based on their difficulty scores. After this arrangement, we train the stock prediction model Φ, for ten epochs each on samples from every group. For the first epoch group, we train the model on samples belonging to the group having the lowest difficulty scores. With each consecutive set of epochs, we train the model on the data-groups having increasingly difficulty samples. After $10 \times B$ epochs, when the model has seen all the ordered samples, we expose it to the original data distribution in entire training set D, and train it until convergence. *Note*: Different curriculums can be arranged based on difficulties computed across combinations of price, text, and model complexities.

4 NEURAL STOCK TRADING

We derive inspiration from Hu et al. [2018], Sawhney et al. [2021a], and define a model that takes as input stock relevant texts in lookback $T \in [\tau - T, \tau - 1]$, and predicts the price movement or volatility for day τ. While any trading model can be used, we propose **THA-Net**: **T**ime-aware **H**ierarchical **A**ttention **Net**work for these financial tasks.

Intra-Day Encoder First, THA-Net encodes the texts (news or tweets) t for stock s released in a day via an embedding layer: BERT [Devlin et al., 2019] as $m = \text{BERT}(t) \in \mathbb{R}^d$, d=768, by averaging the final outputs from BERT.

For each stock s on a day i, a variable number (K) of tweets (t) are posted at irregular times (k). Studying a *sequence* of tweets over the day provides a more unified context to understand a stock, as compared to a single tweet alone [Barber and Odean, 2007]. LSTMs are a natural way to capture such *sequential context dependencies* over time. However, LSTMs assume inputs to be equally spaced in time whereas the time interval between release of consecutive news or tweets can vary widely, from a few seconds to many hours, which can have a drastic impact on their influence on the market [O'Hara, 2015]. To this end, we use a **time-aware LSTM** (TLSTM) [Baytas et al., 2017] to model stock-relevant texts over a day. We feed the time between texts to the TLSTM to capture the temporal irregularities in their release times. We encode the financial news or tweets for a stock s on a day i using the TLSTM as:

$$h_t = \text{TLSTM}(m_t, \Delta k, h_{t-1}); t \in [1, K] \tag{2}$$

where h_t represents the hidden state for text t. Owing to the variably influential nature of news and tweets, we use an attention mechanism [Luong et al., 2015] to emphasize texts that have a higher impact on the stock, as shown in Equation 3. This attention (intra-day attention) learns to aggregate the hidden states of the TLSTM into an *intra-day vector* x_i:

$$x_i = \sum_t \gamma_t h_t, \gamma_t = \frac{\exp\left(W_1(W_2 h_t + W_3 h_K)\right)}{\sum_{t=1}^{K} \exp\left(W_1(W_2 h_t + W_3 h_K)\right)} \tag{3}$$

where γ_t denotes the attention weights, and W_1, W_2 and W_3 are learnable network parameters.

Inter-Day Encoder We combine the representations learned from texts in each day across multiple days in a lookback in a *hierarchical* fashion using the sequence of intra-day vectors x. We feed the vectors x to an LSTM:

$$h_i = \text{LSTM}(x_i, h_{i-1}); \quad \tau - T \leq i \leq \tau - 1 \tag{4}$$

where, h_i is the hidden state representation for day i. Further, tweets and news published across different days have shown to have a varying impact on stock prices, due to financial phenomena such as calendar anomalies, the week-day effect, etc. Thus to selectively weigh critical days, we employ an inter-day attention mechanism to aggregate all days into an *overall representation* z_τ. The inter-day and the intra-day attention together form a **hierarchical attention**, allowing the model to emphasize crucial textual signals within and across days in the lookback in an attentive fashion. We now define the following financial tasks to evaluate the benefits of using FinCLASS to train THA-Net.

Stock Movement Classification We define the price movement of stock $s \in S$ from day $\tau - 1$ to τ as:

$$Y_\tau = \begin{cases} 0, & p^c_\tau < p^c_{\tau-1} \\ 1, & p^c_\tau \geq p^c_{\tau-1} \end{cases} \tag{5}$$

where p^c_τ is the closing price of the stock on day τ, and 0 and 1 denote price downfall and rise. We feed z_τ to a feed-forward layer followed by softmax that outputs the predicted price movement $\hat{Y_\tau}$ for the stock s on day τ.

Stock Volatility Regression Next, we define the single day log volatility using the daily log of absolute returns as:

$$v_\tau = ln\left(|\frac{p^c_\tau - p^c_{\tau-1}}{p^c_{\tau-1}}|\right) \tag{6}$$

We input the representation z_τ to a feed-forward layer followed by a linear activation that outputs the predicted volatility $\hat{v_\tau}$ for stock s on day τ.

Stock Trading Strategy To assess the profitability of THA-Net, we propose a rule-based trading strategy defined using the predicted movement and volatility for stock s on day τ. We execute trades only if THA-Net classifies the price movement with a confidence higher than 70% and the predicted volatility lies within the 1^{st} standard deviation from the mean of true volatilities across the training dataset. Note that true volatility is computed by parsing the ground truth closing price values of the data in the training set through Equation 6.

5 EXPERIMENTS AND SETUP

5.1 PUBLIC DATASETS

US S&P 500 [Xu and Cohen, 2018][1] Comprises 109,915 *English tweets* from social media platform Twitter spanning January 2014 to December 2015, related to 88 high-trade-volume-stocks from the *NASDAQ* stock exchange forming the *S&P 500* index. Xu and Cohen [2018] extract stock specific tweets using regex queries made of stock tickers (e.g., $AAPL for Apple, where $ is a *cashtag* on Twitter).

China & Hong Kong [Huang et al., 2018][2] Comprises 90,361 financial *news headlines* in *Chinese* spanning January to December 2015, aggregated by *Wind* related to 85 top *China A-shares* stocks in Shanghai, Shenzhen and Hong Kong Exchanges. Huang et al. [2018] extract corporate news from major Chinese financial websites. We extract historic stock prices from Yahoo Finance for both the datasets.

Preprocessing We align trading days by dropping data samples that do not possess any news or tweets in the 5-day (T) window. We split the US S&P 500 dataset temporally based on date ranges from January 01, 2014 to July 31, 2015 for training, August 01, 2015 to September 30,

[1] US S&P 500 dataset: www.github.com/yumoxu/stocknet-dataset

[2] China & Hong Kong dataset: https://pan.baidu.com/s/1mhCLJJi

2015 for validation, and October 01, 2015 to January 01, 2016 for testing. We split the China & Hong Kong dataset temporally based on date ranges from January 01, 2015 to August 31, 2015 for training, September 01, 2015 to September 30, 2015 for validation, and October 01, 2015 to January 01, 2016 for testing all models and experiments. The (token) length distribution of the news items in China & Hong Kong dataset (minimum=7, average=22, maximum=115) and tweets in the US S&P 500 dataset (minimum=1, average=21, maximum=98) are similar.

5.2 THA-NET AND FINCLASS TRAINING SETUP

We conduct all experiments on an NVIDIA Tesla T4 GPU. We adopt grid search to find optimal hyperparameters based on the validation MCC and MSE (**§5.4**) for all classification and regression models, respectively. We explore the hidden states for both TLSTM and LSTM $d \in [64, 128, 256]$, and find the best performance at $d = 128$ for both encoders. We divide the training dataset in $B = 10$ groups, and present the performance variation across different values of B in **§6.5**. We use a learning rate of $1e{-}4$ and a decay rate of $1e{-}5$ to train the models using the Adam [Kingma and Ba, 2014] optimizer for 500 epochs. We further elucidate on THA-Net and FinCLASS training setup in the supplementary material.

5.3 BASELINE APPROACHES

We contrast the performance of THA-Net and FinCLASS against the following baselines on the public datasets **§5.1**:

- ***WLSTM:*** LSTMs with stacked autoencoders that encode noise-free data obtained through wavelet transform of prices [Bao et al., 2017].
- ***RandForest:*** Random Forest classifiers trained over text embeddings obtained using word2vec [Mikolov et al., 2013].
- ***TSLDA:*** Topic Sentiment Latent Dirichlet Allocation – a generative model that uses sentiments and topics in text [Nguyen and Shirai, 2015].
- ***CH-RNN:*** An RNN-based model with cross-modal attention on price movement trends and texts across days [Wu et al., 2018].
- ***SN-HFA:*** StockNet - HedgeFundAnalyst - a variational autoencoder with attention on text and prices in the lookback period [Xu and Cohen, 2018].
- ***SN-DA:*** StockNet - DiscriminativeAnalyst - a StockNet model that directly optimizes the log likelihood objective [Xu and Cohen, 2018].
- ***Chaotic:*** A hierarchical attention network that uses Gated Recurrent Units with attention across words, texts and days [Hu et al., 2018].
- ***Adv-LSTM:*** An Adversarial LSTM-based model which leverages adversarial training to improve the training process [Feng et al., 2019].
- ***StockEmb:*** Stock embeddings acquired using prices, and dual vector (word-level and context-level vectors) representation of texts [Du and Tanaka-Ishii, 2020].
- ***FAST:*** A BERT-based hierarchical time-aware encoder for financial text using hierarchical attention while modeling stocks together [Sawhney et al., 2021a].

5.4 EVALUATION METRICS

Classification We use Accuracy and Matthew's Correlation Coefficient (MCC) for evaluating the stock movement prediction performance. MCC avoids potential bias due to data skew as it does not depend on the choice of the positive class and also accounts for the true negatives. For a given confusion matrix $\begin{pmatrix} tp & fn \\ fp & tn \end{pmatrix}$:

$$\text{MCC} = \frac{tp \times tn - fp \times fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}} \quad (7)$$

Regression To evaluate the volatility regression performance, we adopt the Mean Squared Error (MSE) to compute the error between the actual and the predicted volatilities.

Profit To evaluate the practical applicability of our approach to real-world trading, we assess profitability using two metrics: Sharpe Ratio (SR) [Sharpe, 1994] and Maximum Drawdown (MDD). Note that we trade one unit of each stock independently. 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: SR $= \frac{E[R_a]}{std[R_a]}$. The Maximum Drawdown measures the maximum loss from a peak r_p to a trough r_t of a portfolio (in terms of returns), and is defined as: MDD $= \frac{r_t - r_p}{r_p} \times 100$. Larger values of MDD indicate potentially risky trades.

6 RESULTS AND DISCUSSION

6.1 PERFORMANCE COMPARISON

Table 2 shows the superior stock movement classification and volatility regression performance of THA-Net and FinCLASS over baseline and proposed methods. In general, methods that study stock affecting information across text, classify stock movements and predict stock volatility more accurately than methods that only exploit historical prices. These improvements re-validate the effectiveness of leveraging textual sources to capture stock affecting signals such as market surprises, announcements, and public sentiment. We attribute the higher performance of THA-Net + FinCLASS

FinCLASS Variant	Movement Prediction				Volatility Regression	
	US S&P 500		China & HK		US S&P 500	China & HK
	Accuracy↑	MCC↑	Accuracy↑	MCC↑	MSE↓	MSE↓
No Curriculum	57.82 ± 2e-2	0.088 ± 5e-3	55.46 ± 4e-2	0.034 ± 6e-4	0.309 ± 2e-3	2.761 ± 7e-2
Model only	56.65 ± 3e-2	0.091 ± 6e-4	55.23 ± 7e-3	0.039 ± 4e-3	*0.306* ± 5e-4	2.755 ± 3e-3
Price only	56.65 ± 1e-3	0.085 ± 2e-3	55.43 ± 6e-2	0.037 ± 7e-4	0.308 ± 3e-3	2.762 ± 4e-2
Text only	57.01 ± 4e-3	0.097 ± 7e-4	*55.55 ± 2e-3*	0.043 ± 1e-3	0.308 ± 6e-4	2.767 ± 1e-2
Price + Text	*57.92 ± 3e-3*	*0.119 ± 4e-3*	55.52 ± 3e-3	*0.047 ± 5e-4*	0.307 ± 4e-3	*2.741 ± 5e-3*
Model + Price + Text	**58.29** ± **1e-3**	**0.131** ± **4e-4**	**55.72** ± **4e-3**	**0.051** ± **2e-4**	**0.304** ± **2e-3**	**2.718** ± **6e-3**

Table 1: Ablation study over the components of FinCLASS. Model, Price and Text denote the curriculums defined using difficulty scores obtained using model-complexity (M_d), price-features, and textual-features only, respectively. Intense color indicates better performance. **Bold** and *italics* represent **best** and *second-best* results across the metric, respectively.

Baselines and Models	Movement (Accuracy↑)		Volatility (MSE↓)	
	US S&P 500	China & HK	US S&P 500	China & HK
WLSTM (P)	52.95	52.12	0.647	4.149
RandForest (T)	53.15	52.91	0.587	3.845
TSLDA (T+P)	54.14	52.95	0.553	3.624
CHRNN (T+P)	54.31	53.22	0.467	3.719
StockEmb (T+P)	55.43	53.91	0.458	3.482
SN-DA (T+P)	56.15	54.28	0.403	2.991
Chaotic (T+A)	56.16	*55.61*	*0.308*	2.994
FAST (T+A)	57.61	55.34	0.321	2.799
Adv LSTM (P+A)	57.05	55.13	0.362	2.984
SN-HFA (T+P+A)	*58.23*	55.32	0.319	*2.751*
THA-Net -t-A	57.56	54.91	0.312	2.777
THA-Net -A	57.75	55.05	0.309	2.755
THA-Net	57.82	55.46	0.309	2.761
THA-Net + FinCLASS	**58.29**	**55.72**	**0.304**	**2.718**

Table 2: FinCLASS performance comparison against baseline models and variants of THA-Net. The symbols P, T, t and A denote Price, Text, TLSTM and Attention, respectively. Values in **Bold** and *italics* denote **best** and *second-best* results, respectively. Intense color scheme is indicative of a better performance trend across models.

over all baselines to two major reasons that follow. First, through the TLSTM and the hierarchical attention, THA-Net *accurately* captures the fine-grain temporal irregularities and the diverse influence of different texts and days over stock performance, as we observe performance drops across both financial tasks across the two datasets upon removal of the TLSTM and the hierarchical attention. The drop upon removal of the TLSTM as the intra-day encoder suggests that THA-Net benefits by factoring the fine-grain time irregularities in texts to model the flow of stock-affecting signals [Kalev et al., 2004]. The drop due to removal of attention demonstrates that complementing the intra-day with the inter-day attention allows THA-Net to model the *variable influence* of texts on market, hierarchically within and across days. Second, we note that FinCLASS enhances THA-Net's learning process by training it on easier examples of stock affecting signals, before exposing it to samples having complex linguistic characteristics and larger price-based temporal variations. This eventually leads to better performance across both tasks. Consequently, FinCLASS also enables THA-Net to outperform strong baselines. Next, we probe into the improvements due to FinCLASS via a series of ablative studies on the financial curriculum.

6.2 ABLATION STUDY

We now study how THA-Net's stock classification and volatility regression performance benefits via curriculums defined using model-based and multi-modal entropy-based complexity heuristics. Table 1 presents the influence of factoring price-based, text-based, and both multi-modal (price and text-based) stock complexities (S_d), and the model complexity (M_d) to compute the difficulty scores χ for arranging a curriculum to train THA-Net. The difficulty score as an intrinsic property of a training example is sometimes best decided by the neural model itself [Xu et al., 2020], as is indicated by the performance improvement observed for the task of volatility regression in the case of US S&P 500. However, as stock prices and textual signals are highly stochastic in nature, model training benefits from a potent blend of hand-crafted heuristics to better analyze sample difficulty. THA-Net enjoys more significant performance gains across all tasks and datasets when trained over price and text-based curriculums compared to model-based curriculum.

However, we observe that the price curriculum typically leads to lower relative performance gains over the base model compared to the text curriculum. This difference arises as textual stock affecting signals reflect more complex linguistic and temporal variations, especially around events such as the release of quarterly earnings calls, mergers and acquisitions, breaking news, etc. Such variations are likely to be more apparent in the case of the China & Hong Kong data, which comprises chaotic stock-relevant news from the period of 2015-16 China Stock Market Turbulence [Liu et al., 2016]. Consequently, a text curriculum is more likely to introduce the difficult textual samples to THA-Net at later stages of training. We observe optimum performance when using difficulty scores obtained via a blend of price, text, and model complexities to define the curriculum. We attribute these improvements to the entropy-based multi-

Model	US S&P 500		China & HK	
	SR↑	MDD↓	SR↑	MDD↓
THA-Net	-0.924	54.62	0.019	14.07
THA-Net + FinCLASS	**0.633**	**43.31**	**0.104**	**13.67**

Table 3: Real-world trading analysis using THA-Net across stocks in S&P 500 and China A-Shares indexes. We observe significantly high profit improvements and risk reduction when we use FinCLASS to train THA-Net for trading.

modal stock-complexity (S_d) that jointly helps measure the linguistic and price-based temporal variations across the lookback in a time-aware, hierarchical fashion.

6.3 PROFIT ANALYSIS

We examine the practical applicability of FinCLASS to real-world stock trading by analyzing the risk-adjusted returns (Sharpe ratio) and the maximum risk (Maximum Drawdown) associated with the trades executed using THA-Net across stocks in China A-shares and the S&P 500 indexes. We first train THA-Net for stock trading without curriculum learning, and observe poor performance in terms of profits and a higher risk over both indexes as shown in Table 3. This observation indicates that THA-Net takes riskier trading decisions and often experiences losses of large magnitude. However, when we train THA-Net using FinCLASS, we observe significant improvements in risk-adjusted returns (US S&P 500: 168.5%, China A-Shares: 447.3%) and a strong reduction in maximum losses (US S&P 500: 20.7%, China A-Shares: 2.8%). Such improvements indicate the efficacy of FinCLASS in enhancing the real-world applicability of neural stock prediction methods. We further elucidate on the benefits of FinCLASS via a qualitative study on the stocks in the China & Hong Kong dataset.

6.4 ANALYZING STOCK COMPLEXITY

We now study the performance improvements obtained via FinCLASS over non-curriculum THA-Net against samples of varying difficulty levels χ. In Table 4 we divide the dataset into three buckets of low, medium, and high sample difficulty according to the stock complexity S_d. We observe significant improvements over all three difficulty levels, demonstrating that FinCLASS improves performance across both financial tasks over data having varying levels of complexity. Interestingly, we see a corresponding increase in relative improvement for volatility regression as the stock complexity increases from low to medium to high, demonstrating that FinCLASS incurs greater performance gains on increasingly difficult samples, which are otherwise harder to learn for THA-Net. In Figure 3, we divide our training data into 10 fine-grained groups according to the stock complexity S_d, and show a heat-map of relative improvement via

Difficulty Range (S_d)	Relative Gains (%)
Volatility Regression	
0.01-0.41	22.94
0.41-0.47	25.95
0.47-0.93	30.40
Movement Prediction	
0.01-0.41	3.30
0.41-0.47	4.43
0.47-0.93	2.92

Table 4: Performance improvements for stock movement prediction and volatility regression tasks obtained in China & HK via THA-Net+FinCLASS over non-CL THA-Net across samples having varying stock complexity scores (S_d).

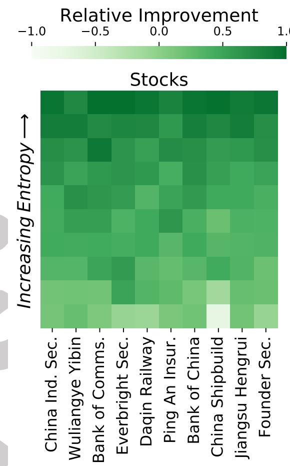

Figure 3: Heatmap depicts performance gains using FinCLASS over THA-Net, for stock data samples in increasing order of entropy.

FinCLASS over non-curriculum THA-Net across 10 stocks. We observe improvements across a diverse range of stocks where FinCLASS improves training performance across a wide range of sample complexities. The heat-map shows some interesting trends; for instance, there is a significant performance increase for high entropy samples compared to low entropy ones. Such trends may arise as non-curriculum approaches may better learn easy samples only, but FinCLASS also enables THA-Net to better learn samples having complex linguistic and price-based temporal variations along with easy ones to bolster stock prediction.

6.5 PARAMETER SENSITIVITY

In Figure 4, we study the influence of the parameter B on the performance of FinCLASS. The value of B directly impacts the number of meta-datasets formed, granularity of the model difficulty score, and the number of learning stages in the curriculum. We gradually vary B from smaller to larger values in the range $B \in [2, 20]$, as shown in Figure 4. Initially, we observe poor performance across both tasks as for lower values of B, the copy-models are likely to inherit a bias towards only one type of difficulty [Xu et al., 2020]. This bias occurs as the sample distribution in meta-datasets would remain identical to the original dataset for lower values of B. On the other hand, larger values of B lead to smaller meta-datasets, which may individually not possess enough information (variations in stock-affecting signals) to train robust copy-models. Further, we observe that middle-ranged values ($B = 10$) give the best performance across both tasks and datasets. Note that FinCLASS is robust to the value of B, owing to the competitive stock prediction

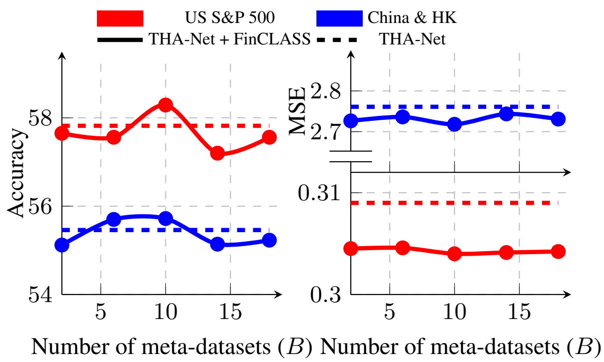

Figure 4: The performance of THA-Net + FinCLASS for stock movement prediction and volatility regression tasks across both the datasets, as the number of meta-datasets (B) is varied in the range of 2 to 20.

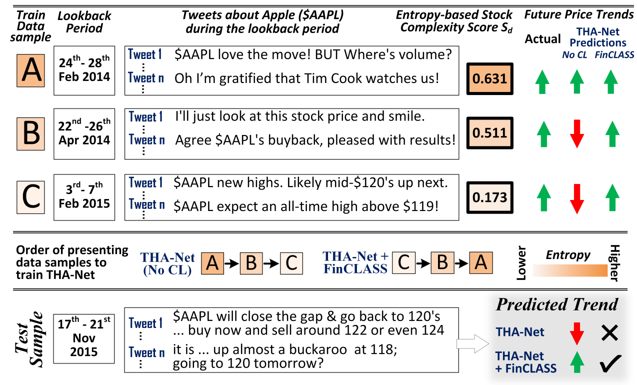

Figure 5: Case study on the US S&P 500 dataset showing how FinCLASS allows THA-Net to learn across easier examples of stock affecting information before exposing the model to linguistically complex samples. This enhances model training and stock prediction performance.

performance against non-curriculum THA-Net for a breath of variations in the value of B.

6.6 QUALITATIVE ANALYSIS

We now conduct an extended study to elucidate the benefits of FinCLASS for stock prediction on the US S&P 500, as shown in Figure 5. Tweets about Apple in training samples A and B possess a sarcastic tone, making it hard to analyze their plausible influence on the stock. Without a curriculum, such linguistically challenging samples would appear *before* the easier sample (C) due to the chronology of the training dataset, likely making it harder for THA-Net to learn stock affecting signals across text to predict movements accurately. We observe that FinCLASS correctly assigns a higher difficulty score to samples A and B while defining the learning curriculum and presents them *after* the less complicated sample C to train THA-Net, allowing it to learn stock affecting trends via easier samples first. This paradigm later makes it easy for THA-Net to learn stock-affecting information across difficult samples, thus maximizing performance. Note that predictions improve across both training and testing datasets when using the curriculum generated via FinCLASS to train THA-Net. Lastly, we show that for a moderately complex test-data sample, movement trend is wrongly classified when training THA-Net without the curriculum, but when trained using FinCLASS, the trend is classified accurately. We attribute THA-Net+FinCLASS's overall improved performance to the generated curriculum that ameliorates the efficiency of the learning process.

7 CONCLUSION

We present FinCLASS, a curriculum learning framework to enhance the training of price-based and language-based neural models for financial tasks. FinCLASS defines an optimum learning curriculum using model-based and hand-crafted entropy-based multi-modal sample difficulty heuristics that reflect the linguistic and price-based temporal complexities of training samples in datasets. FinCLASS is generalizable to a variety of language-based tasks involving a sequence of textual and multi-modal data, embeddings, or feature representations. Experiments on benchmark English tweets and Chinese financial news headlines related to stocks in the S&P 500 and the China A-Shares indexes demonstrate the efficacy of FinCLASS to enhance the performance of neural quantitative trading methods. FinCLASS helps our neural trading model THA-Net outperform state-of-the-art baselines and increases its practical applicability across multiple real-world financial settings.

References

Klaus Adam, Albert Marcet, and Juan Pablo Nicolini. Stock market volatility and learning. *The Journal of Finance*, 71(1):33–82, 2016.

Leif BG Andersen. Efficient simulation of the heston stochastic volatility model. *Available at SSRN 946405*, 2007.

Dogu Araci. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*, 2019.

Wei Bao, Jun Yue, and Yulei Rao. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLOS ONE*, July 2017.

Brad M. Barber and Terrance Odean. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *The Review of Financial Studies*, 21(2), 12 2007. ISSN 0893-9454.

Inci M. Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K. Jain, and Jiayu Zhou. Patient subtyping via time-aware lstm networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, page 65–74, New York, NY, USA, 2017. Association for Computing Machinery.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48, 2009.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *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)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

Xin Du and Kumiko Tanaka-Ishii. Stock embeddings acquired from news articles and price history, and an application to portfolio optimization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3353–3363, Online, July 2020. Association for Computational Linguistics.

Fuli Feng, Huimin Chen, Xiangnan He, Ji Ding, Maosong Sun, and Tat-Seng Chua. Enhancing stock movement prediction with adversarial training, 2019.

William Forbes. *Behavioural finance*. John Wiley & Sons, 2009.

Ehsan Hoseinzade, Saman Haratizadeh, and Arash Khoeini. U-cnnpred: A universal cnn-based predictor for stock markets, 2019.

Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan 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*, WSDM '18, page 261–269, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450355810.

Jieyun Huang, Yunjia Zhang, Jialai Zhang, and Xi Zhang. A tensor-based sub-mode coordinate algorithm for stock prediction, 2018.

Weiwei Jiang. Applications of deep learning in stock market prediction: recent progress, 2020.

Petko S Kalev, Wai-Man Liu, Peter K Pham, and Elvis Jarnecic. Public information arrival and volatility of intraday stock returns. *Journal of Banking & Finance*, 28 (6):1441–1467, 2004.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

Allison Koenecke and Amita Gajewar. Curriculum learning in deep neural networks for financial forecasting. In *Workshop on Mining Data for Financial Applications*, pages 16–31. Springer, 2019.

Kazuhiro Kohara, Tsutomu Ishikawa, Yoshimi Fukuhara, and Yukihiro Nakamura. Stock price prediction using prior knowledge and neural networks. *Intelligent Systems in Accounting, Finance & Management*, 6(1):11–22, 1997.

Kai A Krueger and Peter Dayan. Flexible shaping: How learning in small steps helps. *Cognition*, 110(3):380–394, 2009.

Xiaowei Lin, Zehong Yang, and Yixu Song. Short-term stock price prediction based on echo state networks. *Expert systems with applications*, 36(3):7313–7317, 2009.

Dehong Liu, Hongmei Gu, and Tiancai Xing. The meltdown of the chinese equity market in the summer of 2015. *International Review of Economics & Finance*, 45:504 – 517, 2016. ISSN 1059-0560.

Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation, 2015.

Burton G Malkiel. Efficient market hypothesis. In *Finance*, pages 127–134. Springer, 1989.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

Francesco Carlo Morabito, Domenico Labate, Fabio La Foresta, Alessia Bramanti, Giuseppe Morabito, and Isabella Palamara. Multivariate multi-scale permutation entropy for complexity analysis of alzheimer's disease eeg. *Entropy*, 14(7):1186–1202, 2012. ISSN 1099-4300.

Meinard Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

Sanmit Narvekar, Bei Peng, Matteo Leonetti, Jivko Sinapov, Matthew E Taylor, and Peter Stone. Curriculum learning for reinforcement learning domains: A framework and survey. *Journal of Machine Learning Research*, 21(181): 1–50, 2020.

Thien Hai Nguyen and Kiyoaki Shirai. 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, Beijing, China, July 2015. Association for Computational Linguistics.

Maureen O'Hara. High frequency market microstructure. *Journal of Financial Economics*, 116(2):257 – 270, 2015.

Anastasia Pentina, Viktoriia Sharmanska, and Christoph H Lampert. Curriculum learning of multiple tasks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5492–5500, 2015.

Steve Pincus and Rudolf E Kalman. Irregularity, volatility, risk, and financial market time series. *Proceedings of the National Academy of Sciences*, 101(38):13709–13714, 2004.

Yu Qin and Yi Yang. What you say and how you say it matters: Predicting stock volatility using verbal and vocal cues. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 390–401, Florence, Italy, July 2019. Association for Computational Linguistics.

Dongning Rao, Sihong Huang, Zhihua Jiang, Ganesh Gopal Deverajan, and Rizwan Patan. A dual deep neural network with phrase structure and attention mechanism for sentiment analysis. *Neural Computing and Applications*, pages 1–12, 2021.

Ramit Sawhney, Shivam Agarwal, Arnav Wadhwa, and Rajiv Ratn Shah. Deep attentive learning for stock movement prediction from social media text and company correlations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8415–8426, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.676. URL https://www.aclweb.org/anthology/2020.emnlp-main.676.

Ramit Sawhney, Arnav Wadhwa, Shivam Agarwal, and Rajiv Ratn Shah. FAST: Financial news and tweet based time aware network for stock trading. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2164–2175, Online, April 2021a. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/2021.eacl-main.185.

Ramit Sawhney, Arnav Wadhwa, Shivam Agarwal, and Rajiv Ratn Shah. Quantitative day trading from natural language using reinforcement learning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4018–4030, Online, June 2021b. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/2021.naacl-main.316.

Robert P. Schumaker and Hsinchun Chen. Textual analysis of stock market prediction using breaking financial news: The azfin text system. *ACM Trans. Inf. Syst.*, 27(2), March 2009. ISSN 1046-8188. doi: 10.1145/1462198.1462204.

Benjamin Segal and Dan Segal. Are managers strategic in reporting non-earnings news? evidence on timing and news bundling. *Review of Accounting Studies*, 21(4): 1203–1244, Dec 2016. ISSN 1573-7136.

William F Sharpe. The sharpe ratio. *Journal of portfolio management*, 21(1):49–58, 1994.

Yauheniya Shynkevich, T.M. McGinnity, Sonya A. Coleman, Ammar Belatreche, and Yuhua Li. Forecasting price movements using technical indicators: Investigating the impact of varying input window length. *Neurocomputing*, 264:71 – 88, 2017. ISSN 0925-2312.

Huizhe Wu, Wei Zhang, Weiwei Shen, and Jun Wang. Hybrid deep sequential modeling for social text-driven stock prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, page 1627–1630, New York, NY, USA, 2018. Association for Computing Machinery.

Benfeng Xu, Licheng Zhang, Zhendong Mao, Quan Wang, Hongtao Xie, and Yongdong Zhang. Curriculum learning for natural language understanding. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6095–6104, Online, July 2020. Association for Computational Linguistics.

Yumo Xu and Shay B. Cohen. 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, July 2018. Association for Computational Linguistics.

Dawei Zhou, Lecheng Zheng, Yada Zhu, Jianbo Li, and Jingrui He. Domain adaptive multi-modality neural attention network for financial forecasting. In *Proceedings of The Web Conference 2020*, WWW '20, page 2230–2240, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450370233. doi: 10.1145/3366423.3380288. URL https://doi.org/10.1145/3366423.3380288.