

# Opening the Black Box of Financial AI with CLEAR-Trade: A Class-Enhanced Attentive Response Approach for Explaining and Visualizing Deep Learning-Driven Stock Market Prediction

Devinder Kumar  
Graham W. Taylor  
Alexander Wong

University of Waterloo, ON, Canada  
University of Guelph, CIFAR and Vector Institute, ON, Canada  
University of Waterloo & Waterloo AI Institute, ON, Canada

## Abstract

Deep learning has been shown to outperform traditional machine learning algorithms across a wide range of problem domains. However, current deep learning algorithms have been criticized as uninterpretable "black-boxes" which cannot explain their decision making processes. This is a major shortcoming that prevents the widespread application of deep learning to domains with regulatory processes such as finance. As such, industries such as finance have to rely on traditional models like decision trees that are much more interpretable but less effective than deep learning for complex problems. In this paper, we propose CLEAR-Trade, a novel financial AI visualization framework for deep learning-driven stock market prediction that mitigates the interpretability issue of deep learning methods. In particular, CLEAR-Trade provides an effective way to visualize and explain decisions made by deep stock market prediction models. We show the efficacy of CLEAR-Trade in enhancing the interpretability of stock market prediction by conducting experiments based on S&P 500 stock index prediction. The results demonstrate that CLEAR-Trade can provide significant insight into the decision-making process of deep learning-driven financial models, particularly for regulatory processes, thus improving their potential uptake in the financial industry.

## 1 Introduction

Do machine learning algorithms need to be explainable? This is an important question in today's world, where machine learning algorithms, especially those based on deep learning, are being used at a wide range of tasks and have shown tremendous efficacy in performing these tasks. Deep learning has been touted as potentially disruptive to multiple sectors, including finance. However, deep learning models have also been criticized as uninterpretable "black boxes", with no clear explanation as to how they reach particular decisions [7]. This is a major hindrance to the widespread adoption of deep learning in industries like finance, where regulations are very tight. In such industries with strict regulatory requirements, AI technology must be transparent, interpretable, and explainable. Many experts in these sectors believe that relying on such "black box" methods is a growing problem that is already very relevant due to regulatory processes in these sectors, and it is going to be increasingly more relevant in the future. For example, in some jurisdictions, companies are legally bound to explain certain financial decisions that impact their customers [4]. As such, current approaches for leveraging deep learning are not feasible under these scenarios.

The limitation of deep learning in terms of transparency and interpretability have forced industries dealing with regulatory processes to use comparatively simple machine learning algorithms such as linear or logistic regression, decision trees, or ensemble methods such as random forests which are significantly more explainable and quite effective in simple cases. However, as the complexity of the problem increases, which is very true in finance, deep learning algorithms have been shown to outperform such traditional algorithms by a wide margin across a wide range of problem domains [3]. As such, strategies for explaining the decisions made by deep learning algorithms are highly desired to enable their widespread use in sectors that have strong regulatory processes.

More recently, a number of methods were proposed to mitigate this issue of interpretability and transparency in deep learning. For example, Zeiler & Fergus [8] proposed the formation of a parallel deconvolutional network to interpret the features extracted at different levels of a convolutional neural network. Ribeiro [5] introduced a method to build trust in models that are locally accurate, i.e., correct near the input data sample. Selvaraju et al. [6] proposed a method called Grad-CAM that enables users to discern "strong" networks from weaker ones by providing a combination of visualization of two other techniques. While promising, all of the aforementioned approaches are restricted to identifying features or

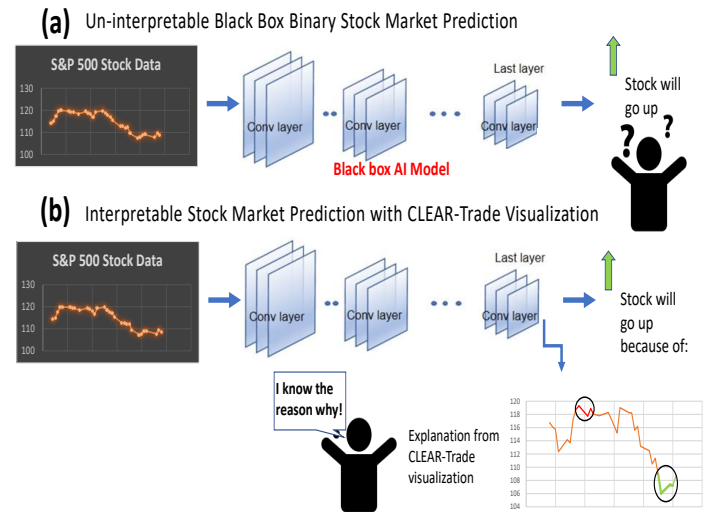


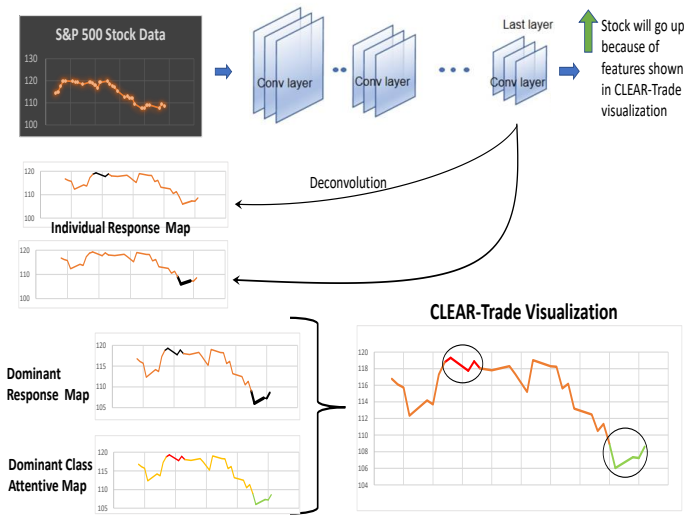
Fig. 1: Two different scenarios for stock market prediction using deep financial models: a) Stock market prediction without interpretability, b) interpretable stock market prediction via CLEAR-Trade. The proposed CLEAR-Trade visualization framework improves financial model interpretability by providing effective visual interpretations of the decision-making process. CLEAR-Trade allows for the visualization of i) the attentive time windows responsible for stock market prediction decisions based on the financial AI model (marked in red and green), ii) their level of contribution to the stock market prediction decision (in this case, stock market index rise (green) or stock market index fall (red)), as well as iii) the dominant class (rise or fall) associated with each attentive time window. This visualization enables financial analysts to better understand the rationale behind the stock market prediction decisions made by the model.

regions of interest and their influence in the decision made by the deep neural network only, thus restricting their utility for gaining a more detailed understanding of the decision process. To address this issue, Kumar et al. [2] recently proposed a **Class Enhanced Attentive Response** (CLEAR) approach that not only identifies attentive regions of interest and their influence on the decision made, but more importantly provides the dominant classes associated with the attentive regions of interest. This additional information about the dominant classes and their influence on the decision making process leads to a higher degree of human interpretability, which makes it very well suited for scenarios that necessitate regulatory processes such as in finance.

Motivated by this, in this paper, we propose CLEAR-Trade, a **Class Enhanced Attentive Response** approach to explaining and visualizing deep learning-driven stock market prediction. In particular, CLEAR-Trade is designed in this paper to provide detailed explanations for the prediction decisions made by a deep learning-driven binary stock market prediction network, as shown in Fig. 1. Our aim is to create a powerful tool for peering into the minds of these otherwise uninterpretable financial AI models to better visualize and understand why they are making the decisions the way they do. We hope this will have a significant impact on day-to-day work of financial analysts in helping them better understand these deep financial models, and facilitating their adoption.

## 2 Methodology

With the goal of enabling transparent and interpretable deep learning-driven stock market prediction, the proposed CLEAR-Trade visualization framework presents the financial analyst with the following



**Fig. 2:** An overview of the proposed CLEAR-Trade visualization framework for explaining and visualizing deep learning-driven stock market prediction. As an illustrative case in this paper, for predicting whether a stock market index will rise or fall (two classes), the index's past 30 days of trade data is fed into the deep learning-driven financial AI model and individual attentive response maps are computed for each class (stock market index rise or stock market index fall: here black represents attention regions for individual classes and the thickness of line represents the strength of attention) based on the last layer of the deep financial model. Based on this set of attentive response maps, two different maps are computed: 1) a dominant attentive response map, which shows the level of contribution of each time point to the decision-making process, and 2) a dominant class attentive map, which shows the dominant class associated with each time point influencing the decision-making process. Finally, the dominant attentive response map and the dominant attentive class map are combined to produce the final CLEAR-Trade visualization, thus enabling the financial analyst to visualize the factors leveraged by the deep financial model in predicting whether the stock market index will rise (green) or fall (red).

information pertaining to the decision-making process:

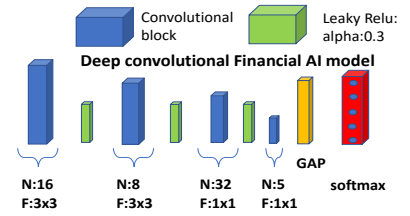
1. the attentive time windows responsible for the decision made by the financial AI model;
2. the attentive levels at these time windows so that their level of influence over the decision made by the financial AI model can be understood; and
3. the dominant class (in this paper, stock market index rise or fall) associated with these attentive time windows so that we can better understand why a decision was made.

The procedure for obtaining the CLEAR-Trade visualization for stock market prediction is shown in Fig. 2 and can be explained as follows.

First, a forward pass with a time-series input of historical trade information about a particular stock market index (in this case, an index's 30 days worth of open, close, highs, lows, and trade volumes) is performed through the deep financial model and a stock market prediction decision output is obtained. To create a CLEAR-Trade visualization associated with this particular stock market prediction decision, we first compute a set of individual response maps  $\{R(\underline{x}|c) | 1 \leq c \leq K\}$ , where  $K$  is the total number of classes present for stock market prediction (in this case, there are two classes: stock market index rise and fall). The deep financial model is set up such that it contains similar number of kernels in the last layer as the number of classes. To elaborate the process, first consider the response  $\hat{h}_l$  for all the kernels  $w$ , at the last layer  $l$  of the financial AI model which can be calculated as:

$$\hat{h}_l = \sum_{k=1}^K z_{k,l} * w_{k,l}. \quad (1)$$

where  $*$  denotes the convolution operation and  $z$  are the feature maps. To calculate the response of last layer in the input domain,



**Fig. 3:** Architecture of the deep convolutional financial model used for stock market prediction process. The financial AI model is embedded in the CLEAR-Trade visualization process, which augments a set of fully convolutional layers, a leaky rectified linear unit layer, global average pooling (GAP) and a softmax layer at the end of the model for the learning process.

we can extend this formulation for response of the specific kernel  $s \in \{1 \dots K\}$  of the deep financial AI with Un-pooling layer  $P'$  as:

$$R(\underline{x}|c) = H_1 P'_1 H_2 P'_2 \dots H_{L-1} P'_{L-1} H_L^c z_L. \quad (2)$$

where  $H$  denotes the combined operation of convolutional and summation, for notational brevity.  $H_L^c$  represents the convolution matrix operation in which the kernel weights  $w_L$  are all zero except that at the  $c^{\text{th}}$  time point.

Given the set of individual attentive response maps, we then compute the dominant attentive class map,  $\hat{C}(\underline{x})$ , by finding the class that maximizes the attentive response level,  $R(\underline{x}|c)$ , across all classes:

$$\hat{C}(\underline{x}) = \underset{c}{\operatorname{argmax}} R(\underline{x}|c). \quad (3)$$

Given the dominant attentive class map,  $\hat{C}(\underline{x})$ , we can now compute the dominant attentive response map,  $D_{\hat{C}}(\underline{x})$ , by selecting the attentive response level at a particular time point based on the identified dominant class, which can be expressed as follows:

$$D_{\hat{C}}(\underline{x}) = R(\underline{x}|\hat{C}). \quad (4)$$

To form the final CLEAR-Trade visualization, we map the dominant class attentive map and the dominant attentive response map in the HSV color space as follows:

$$\begin{aligned} H &= F(\hat{C}(\underline{x})), \\ S' &= 1, \\ V &= D_{\hat{C}}(\underline{x}). \end{aligned} \quad (5)$$

where  $F(\cdot)$  is the color map dictionary that assigns an individual color to each dominant attentive class,  $c$ . Fig. 2 shows an example of the CLEAR-Trade visualization.

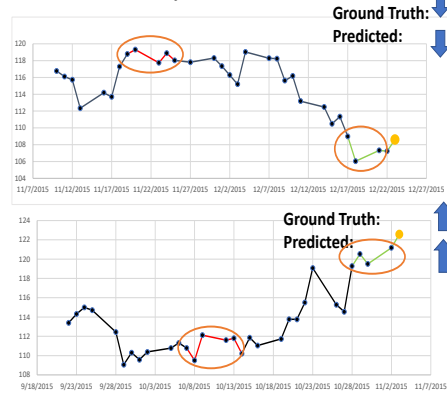
### 3 Experiments and Results

This section explains the experimental setup, the deep financial model built for performing binary stock market prediction on the S&P 500 stock market index, and summarizes our results.

#### 3.1 Experimental Setup

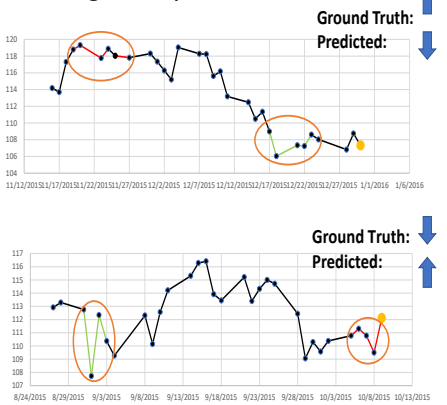
For training purposes, we selected the last three years worth of trade data of the S&P 500 stock market index to train a deep convolutional neural network, shown in Fig. 3, as the deep financial model used in this study. For preparing this data for training the financial model, we divided the data into 30-day time segments and treated the class (index rise or index fall) on the 31st day as '1' if the index was higher than previous day or '0' if the index was lower [1]. We used 90% of the data as training set and considered 10% for evaluation purposes. The trained deep financial model achieved a prediction accuracy of 61.22%, though it is important to note that the focus of this paper is on the ability to visualize and understand the decision-making process of the financial AI model, not in attaining the best possible accuracy, and therefore improving the accuracy of the reference model are plans for future work.

## Correct Binary Stock Prediction



Date	Open	High	Low	Close	Volume
11/27/2015	107.27	108.85	107.2	108.61	3.3e+07
11/28/2015	107.4	107.72	106.48	107.23	3.3e+07
11/29/2015	107.28	107.7	105.77	107.23	4.8e+07
11/30/2015	108.91	109.52	105.88	106.03	9.6e+07
12/1/2015	112.02	112.25	108.98	108.98	4.5e+07
12/2/2015	111.07	111.99	108.8	111.34	5.6e+07
12/3/2015	111.94	112.8	110.35	110.49	5.3e+07
12/4/2015	112.18	112.48	109.79	112.48	6.4e+07
12/5/2015	115.58	115.39	112.85	113.18	4.7e+07
12/6/2015	116.64	116.84	115.51	116.17	2.9e+07
12/7/2015	117.64	117.69	115.08	115.62	4.6e+07
12/8/2015	117.52	118.4	116.68	118.23	4.8e+07
12/9/2015	118.98	119.86	117.81	118.28	3.2e+07
12/10/2015	115.29	119.25	115.11	119.03	5.8e+07
12/11/2015	116.55	116.79	114.22	115.2	4.2e+07
12/12/2015	117.34	118.11	116.08	116.28	3.3e+07
12/13/2015	118.79	119.41	117.75	118.1	3.9e+07
12/14/2015	118.29	118.41	117.6	117.81	1.3e+07
12/15/2015	119.21	119.25	117.6	118.03	2.1e+07
12/16/2015	117.1	119.35	117.11	118.88	4.3e+07
12/17/2015	118.19	119.27	117.72	117.75	3.2e+07
12/18/2015	117.64	119.75	116.76	118.78	3.8e+07
12/19/2015	117.64	119.75	116.76	118.78	3.8e+07
12/20/2015	115.76	117.49	115.5	117.28	4.7e+07
12/21/2015	114.52	115.05	113.32	113.69	2.8e+07
12/22/2015	111.38	114.24	111	114.18	3.8e+07
12/23/2015	115.2	115.07	112.77	113.28	4.6e+07
12/24/2015	116.26	116.82	115.65	115.72	3.3e+07
12/25/2015	116.37	117.42	115.21	116.11	4.5e+07
12/26/2015	116.9	118.07	116.06	116.77	5.9e+07

## Wrong Binary Stock Prediction



Date	Open	High	Low	Close	Volume
11/27/2015	111	112.14	108.21	111.12	2.9e+07
11/28/2015	110.74	111.7	108.21	109.21	3.4e+07
11/29/2015	111.74	112.7	109.43	110.76	5.9e+07
11/30/2015	113.63	114.13	109.77	111.12	4.6e+07
12/1/2015	109.08	111.37	107.08	107.08	5.2e+07
12/2/2015	108.01	111.01	107.55	110.98	5.8e+07
12/3/2015	109.07	109.63	107.91	108.56	6.4e+07
12/4/2015	110.17	111.54	108.73	110.3	6.6e+07
12/5/2015	111.83	113.51	107.86	109.06	7.3e+07
12/6/2015	113.85	114.57	112.44	112.44	5.2e+07
12/7/2015	116.44	116.89	114.02	114.71	5.6e+07
12/8/2015	113.25	115.5	112.37	115	5e+07
12/9/2015	113.63	114.72	113.3	114.33	3.6e+07
12/10/2015	113.38	114.38	112.52	113.4	5e+07
12/11/2015	113.67	115.97	113.66	115.21	3e+07
12/12/2015	112.21	114.3	111.87	113.45	7.4e+07
12/13/2015	115.66	116.49	113.72	113.92	6.4e+07
12/14/2015	116.25	116.54	115.44	116.41	3.7e+07
12/15/2015	115.93	116.53	114.43	116.28	4.3e+07
12/16/2015	116.58	116.89	114.86	115.31	5.8e+07
12/17/2015	111.79	114.21	111.76	114.21	5e+07
12/18/2015	110.27	113.28	109.9	112.57	6.3e+07
12/19/2015	113.76	114.09	109.77	110.53	8.5e+07
12/20/2015	111.75	112.56	110.32	112.31	5.5e+07
12/21/2015	108.97	110.45	108.53	109.27	5e+07
12/22/2015	112.89	112.79	110.04	110.37	5.3e+07
12/23/2015	110.28	112.34	108.13	112.34	6.2e+07
12/24/2015	110.15	111.88	107.36	107.72	7.7e+07
12/25/2015	112.03	114.53	112	112.76	5.6e+07
12/26/2015	112.17	113.31	111.54	113.29	5.3e+07
12/27/2015	112.23	113.24	110.02	112.93	8.5e+07

Fig. 4: Correctly and mis-classified binary stock market prediction for predicting the S&P 500 stock market index as visualized and explained using CLEAR-Trade for a deep financial model trained for experimental purposes. For both cases, the CLEAR-Trade visualizations (left) pinpoint the attentive time windows that are responsible for the particular stock market prediction decision made by the financial AI model (green for index rise and red for index fall). In the CLEAR-Trade visualizations, the thickness of color lines indicate the influence of the attentive time windows on the prediction decision output. On the right hand side, the time-wise stock data information sheet (open, high, low, close, and trade volume) of one of the graphs is shown in reverse chronological order from the day on which the prediction is made. Here also, the attentive levels (transparency of color) and the dominant class (green or red) associated with these attentive time windows is shown so that we can better understand why a decision was made.

### 3.2 Stock Market Prediction Results

To present the effectiveness of the CLEAR-Trade visualization (as explained in Section 2) to enable interpretable deep learning-driven financial AI, we used the trained stock market predictive model and obtained the CLEAR-Trade visualization results as shown in Fig. 4. In Fig. 4, for both cases (correct and wrong predictions), it can be clearly observed from the CLEAR-Trade visualizations which time windows are most crucial to the decision-making process of the model for reaching a particular stock market prediction.

Specifically, in the correctly predicted cases, it can be observed that the deep financial model primarily leveraged the past four days of trade data for correctly predicting whether the S&P 500 stock market index will rise or fall. This is intuitive as the past few days are more likely to have a major effect on the index's behavior compared to data from a couple of weeks back. In the case whether the financial AI model predicts incorrectly, we can observe that the model primarily leverages trade data for nearly 3 weeks ago for making its decision. Another observation that can be made in both cases is that in the cases where the stock market prediction is correct, the deep financial model leverages only open, high and low values to make a decision. This trend if leveraging only open, high, and low values is observed across the majority of correct predic-

tions made by the model. This is again intuitive as unless there is a significant change in trade volume, knowledge of trade volume generally does not have a significant impact on either index rise or fall. Conversely, it can be seen that when making incorrect predictions, the financial AI model strongly takes into account the trade volume as well, which is not a strong predictive feature as indicated above and as such can incorrectly influence its decisions. Finally, it can be observed that the confidence of the stock market prediction in this case is low when it makes incorrect predictions, as indicated by the color transparency in the data-sheet, while in correct predictions the confidence of the stock market prediction is high.

Hence, based on the above mentioned observations, it is evident that CLEAR-Trade visualization not only provides a justification for particular stock market prediction decision output, it can also provide considerable insights that financial analysts can take into account while making trading decisions.

## 4 Conclusion

In this paper, we proposed CLEAR-Trade, a visualization framework that provides insight into the decision processes of deep financial models used for stock market prediction. Experiments predicting the rise or fall of the S&P 500 index showed that CLEAR-Trade visualization leads to a higher degree of human interpretability and transparency in predictions made using deep financial models, hence paving a way for their use in regulatory settings. We believe that the approach has potential to create an industry-wide effect by facilitating the use of state-of-the-art deep learning models for areas in finance that are under significant regulatory pressure.

## Acknowledgments

This research has been supported by Canada Research Chairs programs, Natural Sciences Engineering Research Council of Canada (NSERC), and Canada Foundation for Innovation (CFI).

## References

- [1] Alex Honchar. Neural networks for algorithmic trading. Simple time series forecasting. <http://bit.ly/2gvmJCK>, 2016. [Online; accessed 28-08-2017].
- [2] Devinder Kumar, Alexander Wong, and Graham W Taylor. Explaining the unexplained: A class-enhanced attentive response (clear) approach to understanding deep neural networks. *IEEE Computer Vision and Pattern Recognition (CVPR) Workshop*, 2017.
- [3] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [4] MIT Technology Review. The Financial World Wants to Open AI's Black Boxes. <https://www.technologyreview.com/s/604122/the-financial-world-wants-to-open-ais-black-boxes/>, 2015. [Online; accessed 28-08-2017].
- [5] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144. ACM, 2016.
- [6] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. *IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [7] Daniel LK Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, 19(3):356, 2016.
- [8] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.