Prediction in Cryptocurrency Trading: A Random Forest Learning Approach

Aseel Iyad Radwan, ikhlas Mohammad Idely

***Abstract*—The purpose of this research study is to create and assess a trading strategy that makes use of technical indicators from a chosen dataset. The Random Forest algorithm is specifically used by the proposed Candle Strength technique to get higher results and gain insightful information for bitcoin trading. The method calls for the creation of trading signals as well as a profit estimate for each deal. Corrective AI allows traders to more accurately predict the chances of profitability in their upcoming deals, which enhances risk management and capital allocation procedures. Additionally, the selection process uses the Random Forest algorithm, assisted by LSTM, to solve transaction and data restrictions, reducing the danger of overfitting. The capacity to foresee future trends, remedial AI, and random forests work together to create a more effective and sophisticated method of trading bitcoin.**

I. Introduction

Bitcoin is a digital currency that was created in 2009. It is decentralized, meaning that it is not subject to government or financial institution control. Bitcoin transactions are verified by a network of computers and recorded in a public ledger called the blockchain.Bitcoin is a volatile asset, meaning that its price can fluctuate wildly. This makes it a risky investment, but it also offers the potential for high returns.The cryptocurrency market is dynamic and highly volatile, presenting both opportunities and challenges for traders. To navigate this landscape and optimize trading strategies, the integration of artificial intelligence (AI) has gained significant attention. In this article, we explore the application of AI in Bitcoin trading, specifically focusing on the implementation of Corrective AI using the Candle Strength Strategy. We discuss the use of the Random Forest algorithm, the extraction of features from Bitcoin 15-minute data, and the prediction of trade profitability. Furthermore, we address the decision to use Random Forest over LSTM due to limited trades and data to avoid overfitting. Finally, we explore the potential of Reinforcement AI for future enhancements in overall trading performance.

The candle strength approach employs the closing price of a candle to decide whether to buy or sell a security. It is a technical analysis trading strategy. A candle is considered bullish if it closes higher than the open or bearish if it closes lower than the open. The candle strength approach is a simple trading strategy, and the following describes the shape of these candles. However, they can be helpful in spotting market patterns.

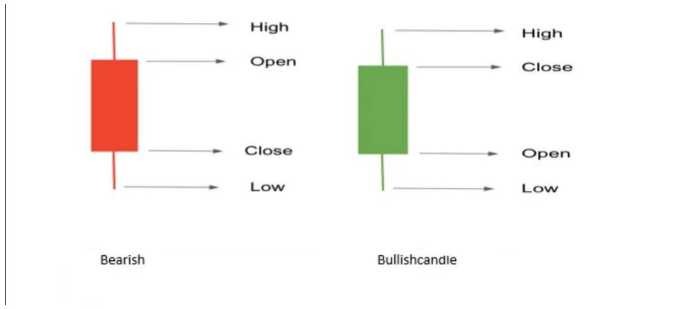

Fig. 1. candle

The Candle Strength strategy, a reliable trading method, was applied to 15-minute Bitcoin data from 2018 to 2022. As a result, the random forest algorithm's ability to adapt and prevent overfitting was limited. This method seeks to find profitable trading opportunities by examining candlestick patterns and using technical indicators. The Average Directional Index (ADX), Simple Moving Average (SMA), Exponential Moving Average (EMA), and Average True Range (ATR) are among the items retrieved from the 15-minute Bitcoin data. These characteristics provide information about market patterns, volatility, and potential price changes.

A random forest algorithm is a type of machine learning algorithm that can be used for classification or regression tasks. It works by creating a number of decision trees and then combining their predictions to get a final result. Random forest algorithms are less prone to overfitting than other machine learning algorithms, such as LSTMs. This makes them well-suited for tasks where the data set is limited.

Corrective AI is a machine learning tool that can help traders predict the probability of profit for their next trade. It uses a user's past trading record and big data to calculate the probability of profit. These probability predictions can then be used for risk-management as well as capital allocation.

Corrective AI is a promising tool that can help traders improve their trading performance. However, it is important to note that it is not a silver bullet. Traders should still use their own judgment and experience when making trading decisions.

Corrective AI, a machine learning tool, is utilized to predict the probability of profit for each trade. By leveraging a trader's historical trading records and big data, Corrective AI calculates the likelihood of a profitable outcome.

These probability predictions are essential for effective risk management and capital allocation. Traders can make informed decisions based on the predicted probabilities, enhancing their overall trading strategies and performance.

The Candle Strength Strategy, coupled with Corrective AI, is executed on Bitcoin data from 2022-2023. This application yields promising results, including decreased drawdown and increased yearly returns. The integration of Corrective AI aids in minimizing losses and optimizing profitability, demonstrating the effectiveness of AI-driven approaches in cryptocurrency trading.
The decrease in drawdown is significant because it means that the trader was able to make more money on their winning trades and lose less money on their losing trades. The increase in yearly return is also significant because it means that the trader was able to make more money overall.

II. RELATED WORK

Find Bitcoin price prediction using the machine Learning algorithms specifically lack. Implemented a The latent source model as developed by for price prediction of Bitcoin indicates 89from 4.1. There was also work using text data from social media Media platforms and other sources of bitcoin price prediction. . Research on the relationship between bitcoin price, tweets, and bitcoin opinions on Google Trends.

implemented a similar methodology except instead of Predict the Bitcoin price by predicting the trading volume using it google trends views. However, there is one limitation to such studies Often the sample size is small, and the tendency for misinformation to spread through various (social) media channels such as Twitter or on message boards like Reddit, which artificially price inflation/deflation .

Analyze the Bitcoin Blockchain for price prediction of Bitcoin using Support Vector Machines (SVM) and synthetics Neural networks (ANNs) report price direction accuracy 55Limited predictability in Blockchain data alone. also used Blockchain Data, SVM Implementation, Random Forest and Binomial GLM (Generalized Linear Model) with prediction note Over 97Models limit the generalizability of their results. wavelets; They have also been used to predict bitcoin prices, with , Noting positive correlations between search engine views, Network hash rate and mining difficulty at bitcoin price.

Another form of RNN is long-term memory Network (LSTM). It differs from Elman RNN in that In addition to having memory, they can choose the data they want Remember and data to forget based on weight and the importance of this feature. implemented LSTM's LSTM time series prediction task Plus an RNN for the job. This kind of model is implemented here too. One limitation in both training RNN and LSTM is the important computation needed. to For example, a grid of 50 days can be compared to training 50 individual MLP models. Since developing CUDA framework by NVIDIA in 2006, developed An application that takes advantage of parallelism too GPU capabilities have grown exponentially including in the field of machine learning. reported three times faster training and testing of their ANN model when implemented on a GPU instead of the CPU. Similarly, recorded an increase in velocity In rated time to the size of eighty times when Execute SVM on the GPU on an alternate SVM Algorithm running on the CPU. In addition, the training time was nine Greater CPU execution times. also received Speeds that were forty times faster in Training A when training Deep neural networks for image recognition on a GPU vs. CPU. Due to the obvious benefits of using a GPU, RNN and LSTM models are implemented on both CPU and GPU.

III. SYSTEM ARCHITECTURE :

The website of the cryptocurrency exchange Binance was used to collect the data for this study. The dataset included a number of parameters, including volume, date, open price, high price, low price, and open price. The information was split into two independent groups: training data for the years 2018 through 2022 and testing data for the years 2022 through 2023. Any duplicate or incorrect records were deleted from both parts before analysis.

Variable name	Attribute Descriptio
Date	Trading Date
Open	Bitcoin Open price for particular time
High	Bitcoin High price achieved for particular time
Low	Bitcoin Low price achieved for particular time
Close	Bitcoin Close price for particular time
Volume	Coin volume traded

TABLE I
DATA SET VARIABLE DESCRIPTION

Traders utilize a range of indicators to assess market conditions and make well-informed trading decisions. These analytical tools provide insights into price fluctuations, patterns, and potential opportunities. Among the commonly used indicators are moving averages, such as the Simple Moving Average (SMA) and Exponential Moving Average (EMA), which help identify trends. Oscillators like the Relative Strength Index (RSI) indicate overbought or oversold conditions.Additionally, the Average Directional Index (ADX) offers information about market trends, and the Average True Range (ATR) measures volatility and potential price changes. By analyzing these indicators, traders aim to enhance their market understanding and optimize their trading strategies.

A. *Average True Range (ATR)*

Technical analysis uses the Average True Range (ATR) indicator to quantify volatility. The ATR does not serve as a price direction indication, unlike many of the widely utilized indicators of today. It is instead a metric that is only used to quantify volatility, particularly volatility brought on by price gaps or limit changes.

To calculate the ATR, the True Range first needs to be discovered. True Range takes into account the most current period high/low range as well as the previous period close if necessary. There are three calculation which need to be completed and then compared against each other. The True Range is the largest of the following: The Current Period High minus (-) Current Period Low The Absolute Value (abs) of the Current Period High minus (-) The Previous Period Close The Absolute Value (abs) of the Current Period Low minus (-) The Previous Period Close

true range = max[(high - low), abs(high - previous close), abs (low - previous close)]

*Absolute Value is used because the ATR does not measure price direction, only volatility. Therefore there should be no negative numbers. *Once you have the True Range, the Average True Range can be plotted. By default on TradingView the ATR is a Relative Moving Average (RMA) of the True Range, but the smoothing type can be changed to SMA, EMA or WMA in the settings.

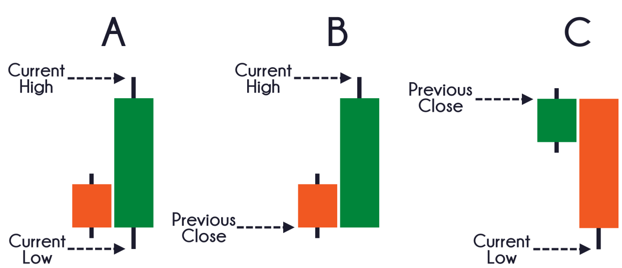

Fig. 2. Average-True-Range

average True Range(ATR) is a volatility indicator that shows us how much an asset may move during a given timeframe. We can use the ATR indicator to fine-tune our exits, by also taking the asset's volatility into account using the ATR method of the talib library.

$$ATR = \left(\frac{1}{n}\right) \sum_{(i=1)}^{(n)} TR_i$$

where:

TR_i = A particular true range

n = The time period employed

Fig. 3. Formula ATR

B. *Simple Moving Average (SMA)*

Simple moving averages, often known as arithmetic moving averages, are created by adding up the components of a time series and dividing the result by the total number of periods. The simple moving average is the most basic kind of moving average, as its name suggests. It is perhaps the most often utilized trading tool for technical analysis. The SMA's components are all equally weighted.

The SMA is usually used to identify trend direction, but it can also be used to generate potential trading signals. The formula for calculating the SMA is straightforward:

SMA = (Sum of data points in the moving average period)/(Total number of periods)

C. *Exponential moving averages*

When the price of a security surges, the simple moving averages are occasionally too simplistic and are ineffective. The most recent periods are given more weight by exponential moving averages. They can therefore be utilized to develop a superior moving average strategy because they are more trustworthy than the SMA and a better depiction of the security's recent performance.

The EMA is calculated as shown below: Weighting multiplier = 2 / (moving average period +1) EMA = (Closing price - EMA of previous day/bar) x multiplier) + EMA of previous day/bar Rewritten as: EMA = (Closing price) x multiplier + (EMA of previous day/bar) x (1 - multiplier) The weightage to the most recent data is greater for a shorter period EMA than for a longer period EMA. For example, a 10 period EMA applies a weightage of 18.18The name exponential moving average is because each term in the moving average period has an exponentially greater weightage than its preceding term. The exponential moving average is faster to react than the simple moving average which can be seen in the chart below. In the chart below, blue line represents the daily closing price, red line represents the 30 day SMA and the green line represents the 30 day EMA.

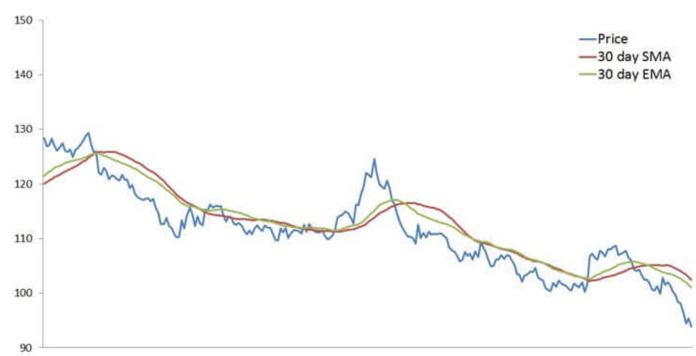

Fig. 4. Exponential-moving-average

IV. Diagonal correlation matrix

The diagonal correlation matrix is a statistical tool used to analyze the relationships between variables in a dataset. It is a square matrix that displays the correlation coefficients between pairs of variables. Each element on the diagonal represents the correlation of a variable with itself, which is always equal to 1. The off-diagonal elements represent the correlation between different pairs of variables. By examining the values in the diagonal correlation matrix, researchers can gain insights into the strength and direction of relationships among variables. This matrix is particularly useful in identifying patterns, detecting multicollinearity, and understanding the dependencies between variables. It serves as a valuable tool in various fields, including finance, economics, social sciences, and data analysis.

In our project, we created a correlation matrix heatmap visualization for a given dataset. The code first sets up the drawing theme using the seaborn library. Then, it generates a random dataset or loads a dataset from the ”train dataset” directory. Computes the correlation matrix for the dataset using the ‘corr()‘ function. Next, it creates a mask to display only the upper triangle of the correlation matrix. The code prepares the matplotlib shape and defines a custom color scheme for the heatmap. Finally, it uses the heatmap seaborn() function to generate the heatmap visualization, apply the mask and set the color scheme.Then the heatmap is saved, as shown in the following image.

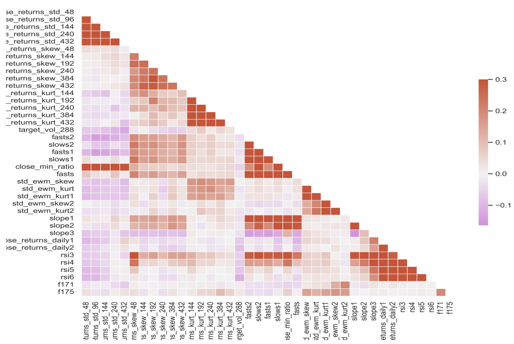

Fig. 5. Diagonal correlation matrix

V. Algorithem:Random Forest

we explored several machine learning algorithms, including the Support Vector Classifier (SVC), Decision Tree Classifier, and Random Forest Classifier, to develop a trading strategy. Each algorithm was tested and evaluated based on its performance in predicting market outcomes and generating profitable trading decisions. After thorough analysis and comparison, we found that the Random Forest Classifier algorithm consistently produced the most satisfactory results in terms of accuracy and profitability. Therefore, we concluded that the Random Forest Classifier algorithm is the most suitable choice for our trading strategy, considering its superior performance and predictive capabilities.
Random forest is a machine learning algorithm that is widely used for classification and regression tasks. It operates by creating an ensemble of decision trees and combining their predictions to make more accurate predictions. Each decision tree in the random forest is trained on a different subset of the data and considers only a random subset of features at each split, hence the name ”random forest.”

In the context of trading, ”random forest trading” could refer to using the random forest algorithm to make trading decisions. Here’s a generalized explanation of how it could work:

1. Data collection: Relevant financial and market data is gathered, such as historical price data, company fundamentals, technical indicators and other relevant information.

2. Feature selection: From the collected data, a set of features is chosen as input variables for the random forest model. These features could include indicators like moving averages, trading volume, volatility measures, or any other data that might be useful for predicting future price movements.

3. Training: The random forest model is trained using data, where the input features are associated with known outcomes, such as whether the price increased or decreased in the next time period. The model learns patterns and relationships in the data to make predictions.

4. Prediction: Once the random forest model is trained, it can be used to make predictions on new, unseen data. For example, it can be used to predict whether the price of a stock will increase or decrease in the next time period based on the current market conditions and input features.

5. Trading decisions: Based on the predictions made by the random forest model, trading decisions can be executed. For instance, if the model predicts a price increase, a trader might decide to buy the stock or take a long position. Conversely, if the model predicts a price decrease, a trader might decide to sell the stock or take a short position.

It’s important to note that successful trading strategies involve many considerations beyond the choice of algorithm, such as risk management, portfolio diversification, transaction costs, and market conditions. While machine learning algorithms like random forest can provide insights and aid decision-making, they are not guaranteed to generate profitable trading strategies on their own. Professional traders often use a combination of techniques, including statistical analysis, fundamental analysis, and market expertise, to make informed trading decisions.

VI. Corrective AI

Artificial intelligence (AI) has gained significant traction in the field of trading due to its potential to enhance decision-making and improve trading outcomes. By leveraging advanced algorithms and machine learning techniques, AI enables traders to analyze vast amounts of data, identify patterns, and make data-driven predictions. AI-based trading systems can incorporate various components, such as natural language processing, sentiment analysis, and predictive modeling, to gain insights from news articles, social media, and market data. These systems can automatically generate trading signals, execute trades, and manage risk based on predefined strategies. AI algorithms can adapt and learn from market conditions, continuously improving their performance over time. The use of AI in trading offers benefits like increased efficiency, reduced human biases, enhanced speed, and the ability to process vast amounts of data in real-time. However, it’s important to note that AI-based trading systems are not without challenges, such as model interpretability, potential algorithmic biases, and the need for robust risk management. Nevertheless, the integration of artificial intelligence in trading holds considerable promise for increasing profitability and optimizing trading strategies in today’s complex and dynamic financial markets.

VII. Result

A. *Test data without ai*

All models had difficulty extracting useful information from the data, as evidenced by the results of the validation data. The trading period for the test data began on November 1, 2021, at 00:00:00, ended on January 20, 2023, at 18:45:00, and lasted for 445 days, 18 hours, and 45 minutes without the use of the intelligence model with a risk ratio equal to 0.2. The total equity at the end of the time period was $156,966.17. The next image serves as another example of it.

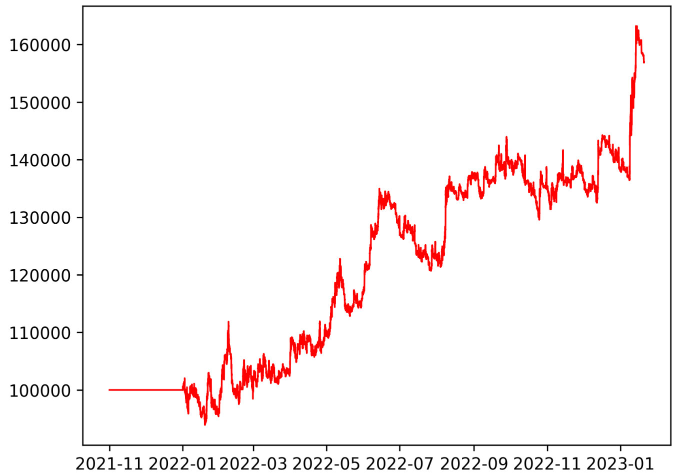

Fig. 6. equity curve

and the annualized volatility of the trading strategy was 34.332119%. 443 trades have been executed, and the winning percentage is 34.988713%. The worst trade resulted in a loss of -6.115072%, while the best trade made a profit of 22.810937%. This approach had an average turnover of 0.203791%. The link between total profit and total loss is represented by the profit factor which has been calculated as 1.317608. These results provide a summary of how strategy Trading Strategy performed over the testing period.

B. test data with ai

The trading strategy was able to calculate the percentage of risk for each trade after utilizing the machine intelligence model, which resulted in a higher profit rate than in the absence of machine intelligence. The trading period, which lasted 445 days, 18 hours, and 45 minutes, started on November 1, 2021, at 00:00:00, and finished on January 20, 2023, at 18:45:00. The amount of time the technique was actively used, or exposure time, was 80.589307%. According to the graph, the total capital at the end of the time period was $210,166.67.

The strategy's annualized volatility increased to 69.752542%, which is significant compared to market volatility. 440 deals were completed with a 35.0% victory rate. The worst trade resulted in a loss of -6.115072%, while the best trade produced a profit of 22.810937%. Each trade produced a meager return of 0.204692% on average. The strategy's profitability is gauged by the profit factor, which was calculated to be 1.318231. These outcomes show how strategy Trading Strategy performed better when the

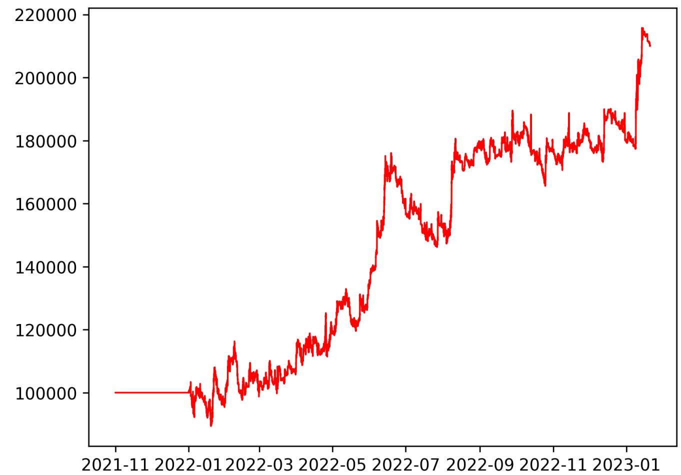

Fig. 7. Caption

artificial intelligence model was used, leading to greater risk management and total profitability.

VIII. Conclusion:

The Candle Strength Strategy benefits from the incorporation of AI, more specifically the Random Forest algorithm, since it produces better results and insightful information for Bitcoin trading. Utilizing Corrective AI, traders may forecast their next trade's likelihood of success, improving risk management and capital allocation procedures.

Reinforcement learning will be used in this study's subsequent phase to enhance trading performance. is a kind of machine learning algorithm that can discover how to decide in a situation by making mistakes. This might be applied to enhance the corrective AI algorithm's trade sizing and risk management capabilities. It is crucial to keep in mind that reinforcement learning is a sophisticated technique, making it difficult to quickly establish a profitable trading strategy.

References

[1] https://www.tradingview.com/support/solutions/43000501823-average-true-range-atr/
[2] https://www.tradingwithrayner.com/atr-indicator/
[3] https://www.tradingwithrayner.com/wp-content/uploads/2018/05/Average-True-Range.png
[4] https://blog.quantinsti.com/moving-average-trading-strategies/
[5] Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8
[6] McNally, Sean; Roche, Jason; Caton, Simon (2018). [IEEE 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP) - Cambridge, United Kingdom (2018.3.21-2018.3.23)] 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP) - Predicting the Price of Bitcoin Using Machine Learning. , (), 339–343. doi:10.1109/PDP2018.2018.00060
[7] M. Papadonikolakis, C.-S. Bouganis, and G. Constantinides, "Performance comparison of gpu and fpga architectures for the svm training problem," in Field-Programmable Technology, 2009. FPT 2009. International Conference on. IEEE, 2009, pp. 388–391.
[8] . Shah and K. Zhang, "Bayesian regression and bitcoin," in Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on. IEEE, 2014, pp. 409–414

[9] G. H. Chen, S. Nikolov, and D. Shah, “A latent source model for non-parametric time series classification,” in Advances in Neural Information Processing Systems, 2013, pp. 1088–1096
[10] M. Matta, I. Lunesu, and M. Marchesi, “Bitcoin spread prediction using social and web search media,” Proceedings of DeCAT, 2015.
[11] ——, “The predictor impact of web search media on bitcoin trading volumes.”
[12] B. Gu, P. Konana, A. Liu, B. Rajagopalan, and J. Ghosh, “Identifying information in stock message boards and its implications for stock market efficiency,” in Workshop on Information Systems and Economics, Los Angeles, CA, 2006.
[13] A. Greaves and B. Au, “Using the bitcoin transaction graph to predict the price of bitcoin,” 2015.
[14] I. Madan, S. Saluja, and A. Zhao, “Automated bitcoin trading via machine learning algorithms,” 2015.
[15] R. Delfin Vidal, “The fractal nature of bitcoin: Evidence from wavelet power spectra,” The Fractal Nature of Bitcoin: Evidence from Wavelet Power Spectra (December 4, 2014), 2014.
[16] L. Kristoufek, “What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis,” PloS one, vol. 10, no. 4, p. e0123923, 2015.
[17] F. A. Gers, D. Eck, and J. Schmidhuber, “Applying lstm to time series predictable through time-window approaches,” pp. 669–676, 2001.
[18] D. Steinkrau, P. Y. Simard, and I. Buck, “Using gpus for machine learning algorithms,” in Proceedings of the Eighth International Conference on Document Analysis and Recognition. IEEE Computer Society, 2005, pp. 1115–1119.
[19] B. Catanzaro, N. Sundaram, and K. Keutzer, “Fast support vector machine training and classification on graphics processors,” in Proceedings of the 25th international conference on Machine learning. ACM, 2008, pp. 104–111.
[20] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, “Deep, big, simple neural nets for handwritten digit recognition,” Neural computation, vol. 22, no. 12, pp. 3207–3220, 2010