**University of North Texas**

**ADTA 5900 - Advanced Data Analytics Capstone Experience**

**Deep Learning for Enhanced Trading Signal Generation:   
A Hybrid CNN-LSTM Approach to S&P 500 Technical Analysis**

Group Members – Biniam Abebe

**Introduction**

Adapting to rapidly changing trends, recognizing, and tracking trading opportunities is essential for financial investment success today. The process of technical analysis is heavily based on human psychology, being susceptible subjectivity of interpreting patterns or indicators on charts (Murphy, 2022). Recent advances in deep learning, especially hybrid CNN-LSTM architectures (Sezer et al., 2020) , provide an opportunity to improve the reliability and profitability of technical trading signals. In this project, we will build a hybrid trading signal generation system using convolutional neural networks (CNN) for pattern recognition and using LSTM (long short-term memory networks) for temporal analysis (Livieris et al., 2021), hoping to achieve better signal accuracy and better risk-adjusted returns in comparison to classical methods.

**Dataset Characteristics and Feature Descriptions**

I will analyze an extensive dataset consisting of the S&P 500 companies, which covers from January 2020 to December 2024. The dataset sourced from the Yahoo Finance API(2024), forms a solid groundwork for analysis with the following attributes:

**Table 1: Dataset Overview**

|  |  |  |
| --- | --- | --- |
| **Characteristic** | **Value** | **Definition** |
| Total Companies | 501 | Number of unique companies included in the dataset |
| Total Observations | 622,641 | Total number of daily data points across all companies |
| Date Range | 2020-02-02 to  2025-01-31 | Temporal span of the dataset |
| Number of Features | 76 | Total number of variables tracked per observation |
| Data Points per Company | 1,242.80 (avg) | The average number of trading days recorded per company |
| Missing Values | 1.90% | Percentage of data points with missing values |
| Dataset Size | 365.78 MB | Total memory usage of the dataset |

**Table 2: Feature Categories and Descriptions**

|  |  |  |
| --- | --- | --- |
| **Category** | **Features** | **Description** |
| Price Indicators | Close, Returns, Log\_Returns, Price\_Range, Price\_Range\_Pct | Basic price measurements and their derivatives, capturing daily price movements and ranges |
| Moving Averages | MA\_X, EMA\_X, Returns\_Xd | Various time-window averages (X=5,10,20,50,200 days) providing trend information |
| Volatility Metrics | Volatility\_Xd, Volume\_MA\_Xd, BB\_Width\_X | Measures of price and volume variability, including Bollinger Band indicators |
| Technical Indicators | RSI\_X, MACD, Signal\_Line, MACD\_Histogram, Momentum\_14, ROC\_14, MFI\_X, Channel\_Width\_X | Advanced technical analysis indicators measuring momentum, trend strength, and price dynamics |
| Volume Indicators | OBV, Volume\_Ratio, Volume\_StdDev | Metrics tracking trading volume patterns and anomalies |
| Fundamental Features | PE\_Ratio, PB\_Ratio, Dividend\_Yield, Profit\_Margin, Beta, Enterprise\_Value, Forward\_EPS, Trailing\_EPS | Company-specific financial and valuation metrics |
| Market Features | Market\_Return, Market\_Volatility, Rolling\_Beta, VIX, VIX\_MA\_10 | Broader market indicators and their relationship to individual securities |

\*Please note that this dataset forms a strong foundation for technical analysis based on deep learning, covering both traditional technical indicators and fundamental market metrics. The lack of missing values (just 1.90%) and large number of observations (622641) implies a high data quality suitable for training more complicated machine learning models.

**Industry Context:**

The U.S. equity market accounts for more than $7 trillion of market capitalization and algorithmic trading constitutes 60-70% of daily trading volume. In this context, advanced pattern recognition tools have grown vital to preserving competitive advantage in systematic trading businesses.

**Research Questions and Hypotheses**

Primary Research Question:

How does a hybrid CNN-LSTM deep learning model improve the reliability and profitability of technical trading signals in S&P 500 stocks compared to traditional technical analysis methods?

This broader question addresses the practical benefits of utilizing sophisticated deep learning methods to implement technical analysis that can be evaluated based on quantifiable trading-yielding results.

Specific Hypotheses:

H1: Signal Generation Quality

* Null Hypothesis (H0): Hybrid CNN-LSTM model does not provide a significantly better accuracy in signal generation accuracy over that what is obtained from traditional technical analysis
* Null Hypothesis (H0): A hybrid CNN-LSTM model does not show significant improvement in trading signal accuracy compared against traditional approaches.

H2: Performance of The Trading Strategy

* H0: CNN-LSTM based trading strategies does not provide statistically superior risk adjusted returns as compared to regular technical analysis
* H2: Trading strategies based on CNN-LSTM generated signals lead to significantly greater risk-adjusted returns (Sharpe ratio, maximum drawdown) (Sharpe, 1994)

These hypotheses will be assessed using:

1. Pattern Recognition Metrics:

* Signal accuracy (precision/recall)
* Pattern detection timing
* False signal rate comparison
* Trading Performance Metrics:

1. Returns adjusted for risk (Sharpe ratio)(Sharpe, 1994)

* Maximum drawdown
* Win/loss ratio
* Profit factor

The analysis will employ the full dataset of 501 companies & 622,641 observations, making use of price data as well as technical indicators via a DEEP learning architecture in which CNN is used for pattern recognition, while LSTM is additionally used to account for temporal dependencies.

The project will apply fundamental concepts and skills learned in the MS in Advanced Data Analytics throughout the program, specifically:

* ADTA 5550 Deep Learning with Big Data
* ADTA 5560 Recurrent Neural Networks for Sequence Data
* ADTA 5240 Harvesting, Storing, and Retrieving Data

Expected Outcomes

The research will apply deep learning techniques to provide targeted results over a few weeks period, proving that advanced pattern recognition works for stock trading:

Expected Outcomes

1. Deep Learning Model Development

- Implementation of a CNN-LSTM hybrid model for pattern recognition:

\* CNN component for spatial feature extraction from price patterns

\* LSTM component for temporal dependency analysis

\* Integration of technical indicators as additional features

- Model training and validation using the pre-processed dataset of 501 companies

- Focus on pattern recognition in daily timeframes

2. Trading Signal Evaluation

- Development of a framework to evaluate the deep learning model's effectiveness:

\* Comparison of model-generated signals vs. traditional technical analysis

\* Pattern recognition accuracy assessment

\* Signal timing precision metrics

- Statistical analysis of prediction reliability

**References**

1. Livieris, I. E., Pintelas, E., & Pintelas, P. (2021). A CNN-LSTM model for gold price time-series forecasting. Neural Computing and Applications, 33(7), 2445-2459. https://link.springer.com/article/10.1007/s00521-020-04867-x
2. Murphy, J. J. (2022). Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. New York Institute of Finance. https://archive.org/details/technicalanalysi0000murp
3. S&P Global. (2024). S&P 500 Index Methodology. Retrieved from S&P Global website. https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-us-indices.pdf
4. Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied Soft Computing, 90, 106181. https://www.sciencedirect.com/science/article/abs/pii/S1568494620301216?via%3Dihub
5. Sharpe, W. F. (1994). The Sharpe ratio. Journal of Portfolio Management, 21(1), 49-58. https://www.scirp.org/reference/referencespapers?referenceid=1451308
6. Yahoo Finance. (2024). Yahoo Finance API Documentation. Retrieved from Yahoo Finance website https://pypi.org/project/yfinance/