Sensex Forecasting Using Haar Wavelet Transform and LSTM: A Comparative Study

Harshith Devraj
Manipal, India
harshikollur302@gmail.com

Abstract

Financial time series data such as the Bombay Stock Exchange (BSE) Sensex are highly volatile and non-stationary, posing challenges for traditional forecasting methods. This study proposes a comparative analysis of two deep learning models: a baseline Long Short-Term Memory (LSTM) network and a Wavelet-enhanced LSTM model using the Haar Wavelet Transform for denoising. By decomposing the closing price signal into approximation and detail components, the Haar Wavelet effectively reduces noise while retaining key temporal features. Experimental results demonstrate that the Wavelet-LSTM outperforms the standard LSTM with a 31% reduction in RMSE and a 33% reduction in MAE, showing its effectiveness for financial forecasting.

Keywords: Wavelet Transform, LSTM, Financial Forecasting, Time Series Analysis, Deep Learning, Sensex

1 Introduction

Stock market forecasting remains one of the most complex tasks in financial analysis due to the inherent volatility, non-linearity, and noise in price movements. Traditional statistical methods such as ARIMA or GARCH assume stationarity and often fail to capture long-term dependencies.

Recent advances in deep learning, particularly recurrent architectures like the Long Short-Term Memory (LSTM) network, have shown promise in modeling sequential dependencies in time series. However, LSTM networks can still be affected by high-frequency noise in financial signals.

The Haar Wavelet Transform provides a simple yet powerful denoising mechanism that preserves time and frequency information simultaneously. This study explores the integration of Haar-based denoising with LSTM to improve predictive performance for the BSE Sensex index.

2 Methodology

2.1 Dataset and Preprocessing

Daily closing prices of the BSE Sensex were obtained from Yahoo Finance from January 2000 to October 2024. Missing values were handled via linear interpolation. Features such as daily returns, log returns, and temporal breakdown (year, month, day) were also added.

2.2 Haar Wavelet Transform

The Haar Wavelet Transform decomposes a time series into approximation (A) and detail (D) coefficients as follows:

$$A_i = \frac{s_{2i-1} + s_{2i}}{\sqrt{2}}, \quad D_i = \frac{s_{2i-1} - s_{2i}}{\sqrt{2}}$$

Here, A_i captures low-frequency components (trend) while D_i captures high-frequency components (noise). The inverse transform reconstructs the signal via:

$$s_{2i-1} = \frac{A_i + D_i}{\sqrt{2}}, \quad s_{2i} = \frac{A_i - D_i}{\sqrt{2}}$$

The approximation coefficients A_i were retained as the input for the Wavelet-LSTM model to train on denoised data.

2.3 Long Short-Term Memory (LSTM) Network

The LSTM is designed to capture long-term temporal dependencies in sequential data. It consists of memory cells and three gates—input, forget, and output—regulated as:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

$$h_t = o_t \odot \tanh(C_t)$$

Where σ is the sigmoid function, $\odot$ denotes element-wise multiplication, C_t is the cell state, and h_t is the hidden state.

2.4 Proposed Architecture

Two models were trained:

1. **Baseline LSTM:** Trained on raw closing price data.

2. **Wavelet-LSTM:** Trained on denoised approximation coefficients A_i.

The models were evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 score.

3 Results and Discussion

Table 1 summarizes the comparative performance between the baseline LSTM and the Wavelet-LSTM model.

Table 1: Performance Comparison between LSTM and Wavelet-LSTM

Model	**RMSE**	**MAE**	R^2
Baseline LSTM	1200.51	985.21	0.9509
Wavelet-LSTM	**827.77**	**660.89**	**0.9940**

The Wavelet-LSTM model significantly outperformed the baseline LSTM, achieving over 30% improvement in RMSE and MAE. This confirms the role of Haar Wavelet-based denoising in enhancing the stability and accuracy of deep learning models for non-stationary time series.

4 Visualization

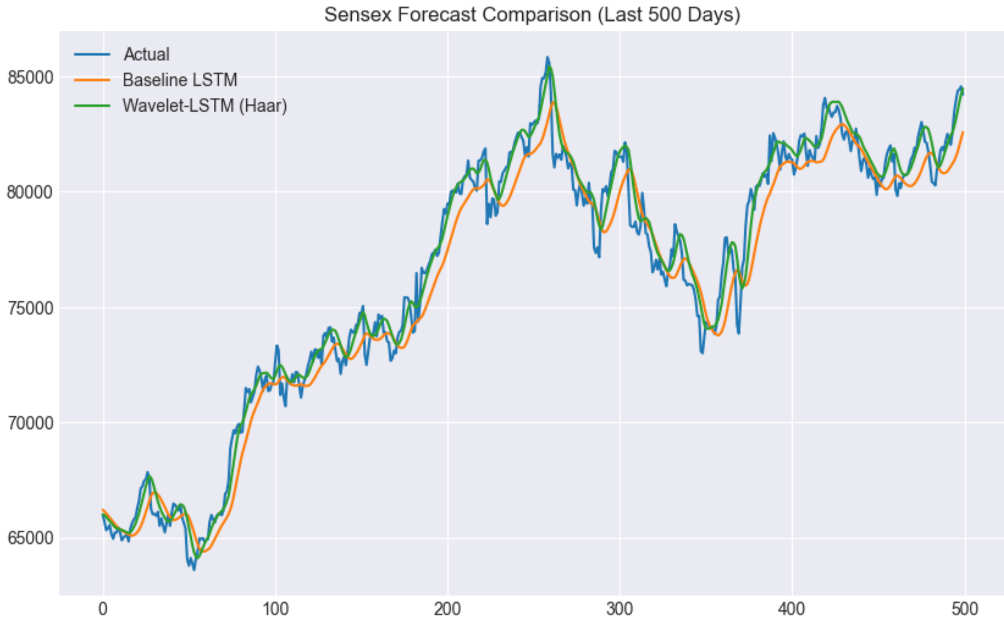

Figure 1: Comparison of Predicted vs Actual Sensex Values using LSTM and Wavelet-LSTM

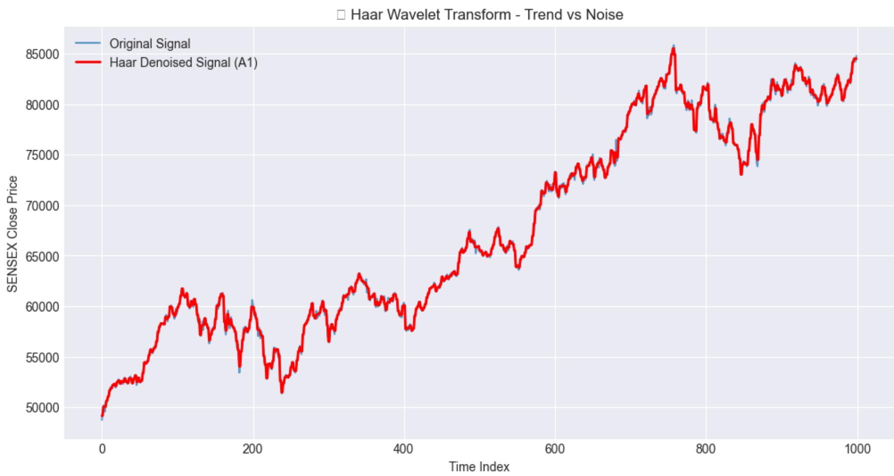

Figure 2: Haar Wavelet Decomposition showing Approximation and Detail Components

5 Conclusion

This research demonstrates that incorporating Haar Wavelet-based denoising prior to LSTM training can substantially improve forecasting performance for financial time series. The results validate the hypothesis that removing high-frequency noise enhances the model's ability to learn meaningful temporal patterns. Future work can extend this approach using hybrid CNN-LSTM architectures and adaptive wavelet families for improved generalization.

References

1. Mallat, S. (1999). *A Wavelet Tour of Signal Processing.* Academic Press.

2. Hochreiter, S., & Schmidhuber, J. (1997). *Long Short-Term Memory.* Neural Computation, 9(8), 1735–1780.

3. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175.