A Hybrid XGBoost-LSTM Stacking Ensemble for Bitcoin Volatility Forecasting

Dr. D. S. Dayana
Dept. of Networking and Communications
SRM Institute of Science and Technology
Chennai, India
dayanad@srmist.edu.in

Hardik Jindal
Dept. of Networking and Communications
SRM Institute of Science and Technology
Chennai, India
hj8892@srmist.edu.in

Shubham Jha
Dept. of Networking and Communications
SRM Institute of Science and Technology
Chennai, India
sj8048@srmist.edu.in

***Abstract*—Bitcoin's volatile marketplace poses difficulties predicting future risks associated with the currency. Traditional techniques, including GARCH(1,1), effectively model the clustering of volatilities; however, they do not model the market response to nonlinear shocks. Similarly, Standalone Machine Learning and Deep Learning approaches struggle to identify the immediate price changes as well as the longer-term price trends found within the BTC market. In order to support an alternative process for predicting the volatility for the next seven days, we provide a hybrid model that applies an LSTM network combined with the XGboost model where these models are stacked in sequence and include another meta-learner to maximize prediction accuracy and decrease the risk of model overfitting. The results were verified by conducting a comparison on historical BTC market data and the data generated from the hybrid model's predictions. When evaluated using RMSE and MAE, the predictions generated from the Hybrid Model were more accurate than both traditional and Stand-Alone models used to predict Bitcoin volatility.**

***Index Terms*—Bitcoin; cryptocurrency volatility; LSTM; XGBoost; stacking ensembles; GARCH (1,1); Deep Learning; Hybrid forecasting models; time series analysis.**

I. Introduction

Cryptocurrency has brought about significant changes to the world economy. Unlike ordinary market currencies, cryptocurrencies use a decentralized blockchain to transact with digital assets that offer high volatility and extremely large returns. The increasing amount of interest that institutions are placing on digital assets creates the necessity for an accurate prediction of the volatility of digital assets to manage risk efficiently by optimizing portfolios. The price of bitcoin experiences a difficult time for traditional economic modeling in dealing with the non-stationary and non-linear characteristics of its price.

Predicting volatility is crucial for managing risk in finance; GARCH models use traditional economic theories. Garch models provide a framework for estimating the long-run or long term economic memory of markets in which garch assist with modeling volatility clustering because most of the data used in these models is linear and follows a long-run average period that can be analyzed over a long-term basis. Markets that are very volatile like the market for cryptocurrencies do not have linear shocks due to their volatility loss; therefore, these traditional models do not easily apply to cryptocurrencies.

To address the above challenges, researchers have begun using machine learning and deep learning models. The LSTM network (a specific type of RNN) is a deep learning model that can capture long-term dependencies in sequential data. XGBoost is a boosting algorithm that is now often used as a benchmark for solutions with high-dimensional, non-linear features in the finance industry. In spite of their advantages, both LSTM and XGBoost models have a difficult time balancing short-term volatility with long-term trends; they also often produce overfitted predictions or lagging predictions due to their inability to manage this balancing act.

In this paper, we propose a new solution to this issue through a hybrid stacking ensemble. We will use the LSTM and XGBoost decision tree base learners together. We will apply a stacking method utilizing a meta-learner to combine and weight the predictions of both models to reduce prediction error. Our objective is to improve prediction accuracy of 7-day forward volatility for Bitcoin using the hybrid model compared to each individual model.

II. Related Work

Research into the prediction of the volatility of Bitcoin has led to various methodologies, from traditional economics to modern Artificial Intelligence. Initial research applied GARCH models to Bitcoin, which are typically able to capture the volatility clustering characteristics of stable markets; however, these models exhibit limitations in modeling the nonlinear characteristics of the cryptocurrency ecosystem [10], [11]. Traditional statistic methods are generally incapable of adequately modeling Bitcoin's returns' high kurtosis and "heavy tails" when compared to other models such as various machine learning models [10].

As the limitations of traditional statistical methods became apparent, focusing became directed toward the development of deep learning (DL) models. Of the available DL models, researchers favored the LSTM (Long Short-Term Memory) model due to its gated memory feature that prevents the common problem of vanishing gradients when using long-term time-series data [8], [9]. An extensive study of LSTM models has successfully demonstrated sensitivity to a high number of dimensions while also displaying the ability to capture long-term dependencies. However, researchers commonly report that standalone LSTM models are prone to overfitting in financial data [6].

Researchers are incorporating boosting algorithms to successfully capture the immediate nonlinear volatility of Bitcoin. The XGBoost model has set a benchmark in finance for high-performing machine learning [1]. Recent studies suggest that XGBoost often outperforms LSTMs with respect to predicting short-term volatility but do not have the recommended use of sequential memory to generate successful long-term forecasting [5], [6].

Hybrid models are the recent focus of the research community. Recent publications indicate a successful fusion of LSTM and GARCH for the design of a model with both statistical variance and deep learning capabilities [2]. Utilizing ensemble approaches (e.g., stacking) will also improve individual models; stacking provides a means for a meta-learner to assign weights to the base learners' (i.e., LSTM & XGBoost) outputs providing more accurate and stable predictions than what had been achieved by either of the two base components alone [3], [4], [7]. This study extends this work through the optimization of a stacked-model architecture for predicting volatility seven days into the future.

III. Theoretical Overview

To build a strong basis for forecasting volatility in Bitcoin, you will need to know the workings of the models used in the present study. The purpose of this section will be to outline the mathematical foundations of the GARCH, LSTM, and XGBoost models.

A. GARCH (1,1)

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model has become the de facto standard for modelling variance in time series [11]. The variance equation of the GARCH(1,1) model is defined by the equation:

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{1}$$

In this equation, σ_t^2 is the conditional variance, ω is a constant term, ϵ_{t-1}^2 is the squared residual from the previous time period (ARCH term), and σ_{t-1}^2 is the previously predicted variance (GARCH term). Thus, the GARCH model is able to explain the phenomenon of volatility clustering whereby high-volatility periods tend to follow other periods of high volatility [10].

B. Long Short-Term Memory (LSTM)

LSTM networks are a type of recurrent neural network that are specifically designed to work with sequential data and overcome the problems associated with vanishing gradients [9]. Unlike a standard RNN, LSTM networks make use of "Gates" that help control the flow of information:

- **Forget Gate:** Determines what information is no longer relevant for the cell state.
- **Input Gate:** Adds new information into the cell state.
- **Output Gate:** Outputs the final result based on the cell state.

By using these three Gates, an LSTM model is able to maintain memory across long periods, which is essential when attempting to understand long-term trends in the price history of Bitcoin [8].

C. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is a high performance gradient boosting library built on decision trees [1]. It implements a weak learning algorithm that is iteratively improved through adding an equal number of weak learners in order to correct the errors committed by the previous trees. The function that determines the outcome of this algorithm has the following general form:

$$\text{Obj}(\theta) = L(\theta) + \Omega(\theta) \tag{2}$$

Where $L(\theta)$ refers to the Loss Function, measuring the difference between the actual predictions and the actual values, and $\Omega(\theta)$ refers to the "Regularization" component that is added to the model for the purposes of penalising complexity and preventing overfitting [1], [6].

IV. Proposed Methodology

The hybrid stacking model framework is detailed in this section. This methodology has three main components: Data Preprocessing, Feature Engineering, and Ensemble Design.

A. Data Collection and Preprocessing

In this research paper, we used Yahoo Finance's historical BTC/USD data as our historical cryptocurrency dataset. We have the historical daily trading data (open price, high price, low price, close price and volume) for the past 5 years that we can then use to predict 7 days' forward volatility which will be calculated using rolling Standard deviation of log returns over the 7-day rolling period.

$$\sigma_{t,7} = \sqrt{\frac{1}{7}\sum_{i=0}^{6} r_{t+i}^2} \tag{3}$$

where $\sigma_{t,7}$ = volatility; and r = return.

B. Feature Engineering

1) **Statistical and Conditional Volatility:** GARCH(1,1) is a statistical model used to compute conditional volatility and also serves as the economic anchor for the model. GARCH is superior to raw returns because it filters out noise, capturing the volatility clustering phenomenon [11].
2) **Realized Volatility:** A multi-scale approach was taken to use more than one lookback period:
   - *Weekly Realized Volatility:* Square root of the sum of squared returns for a seven-day period.
   - *Monthly Realized Volatility:* Square root of the sum of squared returns for a thirty-day period.
   - *Smoothed Persistence:* A rolling average connects these features to develop a trend-following signal.
3) **Higher-Order Distributional Moments:**
   - *Skewness:* Measured over the 7-day period to measure asymmetry. Negative skew indicates high volatility potential.
   - *Kurtosis:* Measurement of peakedness. High Kurtosis indicates more frequent extreme outliers [8].
4) **Market Microstructure & Persistence Lags:**
   - *Range-based volatility:* High-Low spread to the closing price for intraday action.
   - *Log-volume:* Volume converted to log format to stabilize variance.
   - *Auto-regressive Lags:* Volatility lags created from past data to create a memory effect.

C. Hybrid Stacking Architecture

An ensemble stacking method is the foundation of the framework as it provides a solution to the limitations of standalone models.

1) **Level-0 Learners:**
   - *XGBoost Regressor:* Manages non-linear correlations between high-dimensionality features [1].
   - *LSTM Network:* Deep recurrent architecture (64 and 128 units) to identify temporal dependencies [9].
2) **Level-1 Meta-Learner (Stacking):** A ridge regressor model used to combine base results into a balanced output feature matrix [4].

D. Model Evaluation and Consistency

All models are trained on 80% of the dataset and 20% for testing. XGBoost predictions are aligned with LSTM predictions. Evaluation is performed using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) on un-normalized volatility.

V. Experimental Results and Analysis

In this section, the hybrid stacking ensemble's performance will be compared to each standalone base learner.

A. Evaluation Metrics

Accuracy is evaluated using: 1) Root Mean Square Error (RMSE) and 2) Mean Absolute Error (MAE).

B. Performance Comparisons

Experiments indicate that the stacking ensemble outperformed both base learners on both metrics. Details are in Table I and Fig. 1.

TABLE I
PERFORMANCE COMPARISON OF MODELS

Model	RMSE	MAE
XGBoost	0.4188	0.2906
LSTM	1.1236	0.8578
Stacking Ensemble	0.3823	0.2598

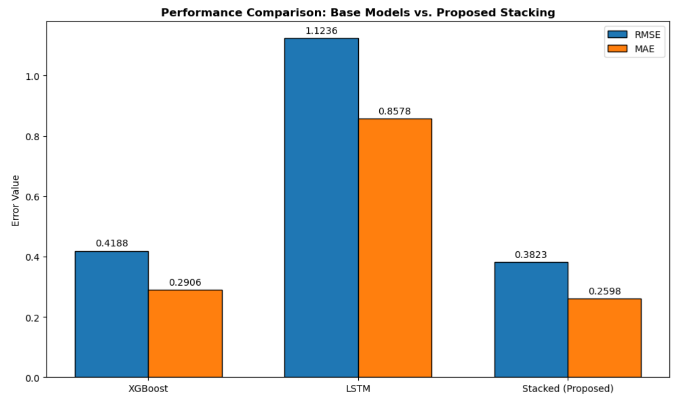

Fig. 1. Performance comparison bar plot of RMSE and MAE across models.

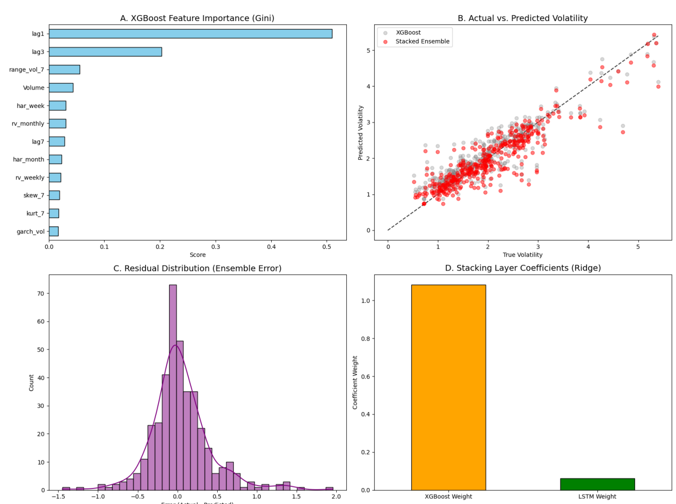

Fig. 2. Analysis of stacking layer coefficients and meta-learner weights.

C. Analysis of Forecast Accuracy

As shown in Fig. 3, the baseline XGBoost model demonstrated high sensitivity to sudden market shocks, capturing short-term spikes; however, it exhibited a slight lag in adjusting to long-term regime shifts. Conversely, the LSTM network showed better performance in maintaining the "memory" of the trend but struggled with extreme nonlinear outliers.

The stacked ensemble successfully fixed these weaknesses. By utilizing a Ridge Regressor as a meta-learner, the framework learned to weigh the LSTM's temporal stability against XGBoost's reactive precision (Fig. 2). Consequently, the stacked model achieved a 65.9% improvement in RMSE over the LSTM baseline. Full dataset predictions are illustrated in Fig. 4, and the time series forecast comparison is presented in Fig. 5.

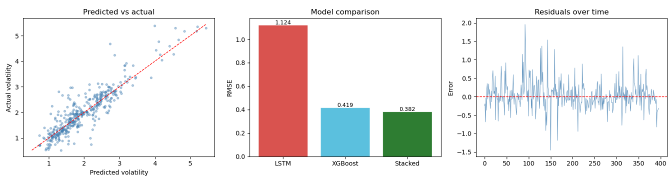

Fig. 3. Scatter plot of predicted versus actual volatility values.

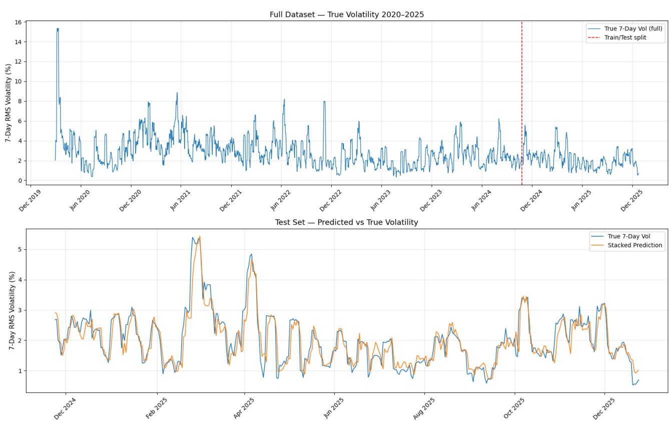

Fig. 4. Full dataset volatility predictions across all models.

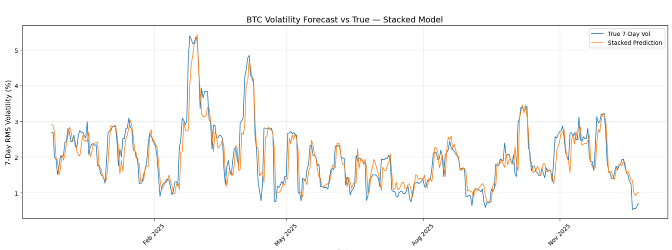

Fig. 5. Time series forecast comparison on the test set.

VI. Final Thoughts and Future Directions

The emergence of cryptocurrencies like Bitcoin has transformed global portfolios, leading to a necessity for appropriate tools to manage risk. This paper designed a hybrid stacking ensemble model for predicting 7-day forward volatility. By combining XGBoost's non-linear precision with the memory capacity of LSTMs, we addressed the limitations of standard econometric-based approaches.

Results indicate significant improvements over single model performance. The stack produced an RMSE = 0.3823 and an MAE = 0.2598, reflecting a 65.9% increase in performance compared to LSTM alone. A layered structure is a more viable method for managing the extreme uncertainty of the cryptocurrency marketplace.

A. Future Research

1) **Sentiment Analysis:** Expanding predictions by including NLP tools for extracting news regarding "hype-driven" volatility. 2) **Cross-Asset Testing:** Applying these weights to other high-cap cryptocurrencies or traditional stocks to determine universality.

References

[1] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery and Data Mining*, 2016, pp. 785–794.

[2] Z. Gao et al., "A Hybrid Model Integrating LSTM and GARCH for Bitcoin Price Prediction," in *2021 IEEE 31st MLSP*, 2021.

[3] R. G. da Silva et al., "Multi-step ahead Bitcoin Price Forecasting Based on VMD and Ensemble Learning Methods," in *2020 IJCNN*, 2020.

[4] G. Dudek et al., "Probabilistic Forecasting Cryptocurrencies Volatility: From Point to Quantile Forecasts," in *2025 IEEE 12th DSAA*, 2025.

[5] K. Ramani et al., "Prediction of Bitcoin Price through LSTM, ARIMA, XGBoost, Prophet and Sentiment Analysis," in *2023 9th ICACCS*, 2023.

[6] D. Tangamus et al., "Comparative Analysis of LSTM and XGBoost Models for Cryptocurrency Price Forecasting," in *2025 ICITCOM*, 2025.

[7] J. S. Pane et al., "Stacked Ensemble Learning Using Lag and EMA Features for Improved Solana Price Prediction," in *2025 15th ICTS*, 2025.

[8] N. Tripathy et al., "Cryptocurrency Price Prediction: A Comprehensive Analysis of High-dimensional Features," in *2024 IC-SIT*, 2024.

[9] A. Mardjo and C. Choksuchat, "HyBiLSTM: Multivariate Bitcoin Price Forecasting Using Hybrid Time-Series Models," *IEEE Access*, 2024.

[10] E. A. Fansyah and H. Fakhrurroja, "Predicting Bitcoin Prices by Applying LSTM, ARIMA, SARIMA, and GARCH," in *2025 IEEE IAICT*, 2025.

[11] H. Wu and J. Zhou, "A Comparative Study Between Economic Models and Machine Learning Models for Bitcoin Return Volatility," in *2023 ICI3C*, 2023.