

A Comparative Study of Hybrid LSTM Frameworks for Volatility Forecasting in the NASDAQ-100 and S&P 500 Markets

Yue Yihua, Rong Jia, Lv Zimeng, Xiao Tongren, Li Ke

November 2, 2025

Abstract

This article presents a comparative study on a method that integrates deep learning with classical econometric models to forecast volatility across different financial markets. Recognizing the limitations of standalone models, we propose and test a hybrid framework that enhances a Long Short-Term Memory (LSTM) network with exogenous features derived from GARCH models and implied volatility indices. The study is conducted on two major, distinct indices: the tech-heavy NASDAQ-100 (NDX) and the broader S&P 500 (SPX). Our framework uses one-step-ahead GARCH volatility predictions, alongside realized volatility (RV) and the respective implied volatility indices (VNX for NDX, VIX for SPX), as inputs to the LSTM. Through a comprehensive comparison of various feature combinations for both markets, our results reveal consistent patterns. The hybrid models that combine realized volatility with GARCH-family predictions demonstrate the highest predictive accuracy for both indices. Notably, adding implied volatility as a feature consistently degrades performance across both markets, a significant finding. The study confirms the robustness of the hybrid LSTM-GARCH approach and provides valuable insights into the role of different features in volatility forecasting across markets with different compositions.

Keywords:

Deep Learning; Long Short-Term Memory Networks; GARCH; Implied Volatility; NASDAQ-100; S&P 500; Volatility Forecasting

1. Introduction

Financial risk management is crucial to maintaining market stability, and accurate volatility forecasting is a cornerstone of this effort [GMAM24]. Given

⁰Code available at: <https://github.com/easyglider/lstm-garch-for-SPX-NDX>

the complexity and nonlinear structure of financial markets, this study explores an innovative method combining deep learning algorithms with classic econometric models to solve the volatility prediction problem [PTW24]. However, the performance of such a model may vary depending on the underlying asset’s characteristics. To test the robustness and generalizability of a hybrid framework, we conduct a comparative analysis on two of the world’s most significant stock indices: the technology-focused NASDAQ-100 (NDX) and the broad-market S&P 500 (SPX).

This paper builds a comprehensive framework integrating deep learning with econometric principles and tests it on both indices. Through this dual-market fusion, we aim to provide not only solid financial theoretical explanations but also to effectively fit nonlinear characteristics in data, thereby improving forecast accuracy and drawing more generalizable conclusions.

Subsequently, this paper adopts a hybrid approach to predict volatility for both markets. Specifically, we construct and compare a series of LSTM-based models with different combinations of exogenous features. The main innovations of this paper are:

1. We design a hybrid framework that feeds GARCH_{NDX}-family volatility predictions and implied volatility directly into an LSTM network, and we apply this framework consistently to both the NDX and SPX.
2. We conduct a comparative analysis of numerous feature combinations across both markets, systematically evaluating the contribution of GARCH predictions and implied volatility (VNX and VIX).
3. Our results yield a consistent and robust conclusion: for both the NASDAQ-100 and the S&P 500, simple hybrid LSTM models enhanced with GARCH-family predictions provide the most significant performance, while the inclusion of implied volatility consistently degrades predictive accuracy.

2. Related work

In the ever-changing financial market, organically integrating classic econometric models, such as the GARCH family, with emerging deep learning methods has become a critical area of research. The use of hybrid models has shown significant promise across various domains. For instance, Kim et al. [KW18] demonstrated the effectiveness of a hybrid approach for the KOSPI 200 index by integrating an LSTM network with predictions from GARCH, EGARCH, and EWMA models. Their work found that the hybrid model produced forecasts with lower error rates compared to standalone models. Similarly, recent research [hyb24] has further validated the effectiveness of hybrid LSTM-GARCH frameworks for financial market volatility risk prediction. This principle has been extended to other markets, with Verma [Ver21] applying a GARCH-RNN hybrid to forecast crude oil volatility, and Zeng et al. [ZSB⁺22] using a complex

hybrid model to predict natural gas volatility. Our study builds upon this foundation by applying a similar hybrid philosophy to the major US stock indices and conducting a detailed comparative analysis of a wider range of GARCH variants and exogenous features.

3. Methodology

Our methodology is built on a hybrid framework that combines traditional econometric models with a deep learning architecture. We first define and calculate the core components—Realized Volatility, GARCH-family predictions, and the LSTM network—before detailing how they are integrated.

3.1. Realized Volatility (RV) Calculation

The target variable for our forecasting models is the daily Realized Volatility (RV). Unlike volatility measures based on daily data, RV provides a more accurate, model-free estimate of the true latent volatility by leveraging high-frequency, intra-day data. For each trading day t , the RV is calculated from 5-minute logarithmic returns as:

$$RV_t = \sqrt{\sum_{i=1}^M r_{t,i}^2}$$

where M is the number of 5-minute intervals in a trading day (78 in our case), and $r_{t,i}$ is the log-return of the i -th interval on day t , calculated as $r_{t,i} = \ln(P_{t,i}) - \ln(P_{t,i-1})$.

3.2. GARCH-Family Models for Volatility Prediction

We employ a suite of GARCH-family models to generate one-step-ahead volatility forecasts, which are then used as exogenous features for the LSTM. The standard GARCH(1,1) model defines the conditional variance σ_t^2 as:

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

In addition to the standard GARCH, we utilize five other variants to capture more complex volatility dynamics:

- **GJR-GARCH:** Accounts for the leverage effect, where negative shocks (bad news) tend to increase volatility more than positive shocks of the same magnitude. Its conditional variance is given by:

$$\sigma_t^2 = \omega + (\alpha + \gamma I_{t-1}) \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

where I_{t-1} is an indicator function that equals 1 if $\varepsilon_{t-1} < 0$ and 0 otherwise. The term γ captures the leverage effect.

- **TARCH (Threshold ARCH):** As implemented in the code with ‘power=1.0’, this model is also known as ZARCH and models the conditional standard deviation directly. It captures asymmetry using the sign of past shocks, not just their magnitude:

$$\sigma_t^2 = \omega + \alpha(|\varepsilon_{t-1}| - \gamma\varepsilon_{t-1}) + \beta\sigma_{t-1}^2$$

Here, the term $\alpha(|\varepsilon_{t-1}| - \gamma\varepsilon_{t-1})$ responds differently to positive and negative ε_{t-1} .

- **EGARCH (Exponential GARCH):** Models the logarithm of the variance, ensuring that the conditional variance is always positive without imposing non-negativity constraints on the parameters. It also captures leverage effects through the γ term:

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \alpha \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}}$$

- **AVGARCH (Absolute Value GARCH):** A variant that models the conditional standard deviation instead of the variance, based on the absolute value of past shocks. For a power of 1, as used in this study, the model is:

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

- **FIGARCH (Fractionally Integrated GARCH):** Captures long-range dependence (or "long memory") in the volatility process. Unlike standard GARCH where the influence of past shocks decays exponentially, FIGARCH allows for a slow, hyperbolic decay. The conditional variance is given by:

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + [\alpha - \beta\phi(L)](1-L)^d\varepsilon_{t-1}^2$$

where $\phi(L)$ is the lag polynomial, $(1-L)^d$ is the fractional differencing operator with parameter $d \in (0, 1)$, and L is the lag operator.

These models are fitted on a rolling window of 22 trading days to generate a continuous series of one-day-ahead volatility predictions.

3.3. Long Short-Term Memory (LSTM) Network

The Long Short Term Memory (LSTM) network is a type of Recurrent Neural Network (RNN) particularly well-suited for time series forecasting. Unlike traditional RNNs, LSTMs incorporate a "cell state" and gating mechanisms (forget, input, and output gates) that allow the network to selectively remember or forget information over long sequences, thus overcoming the vanishing gradient problem.

In our framework, the LSTM takes a sequence of historical data with a look-back window of 22 days as input. Each element in the sequence is a vector containing multiple features (e.g., historical RV, GARCH predictions, and/or implied volatility). The network’s output is a single value, which is the prediction for the next day’s RV.

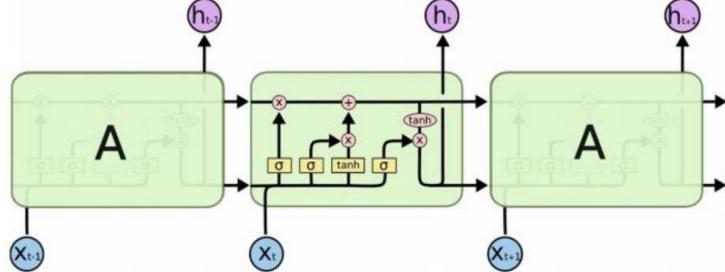


Figure 1: LSTM Network Architecture and Information Flow

3.4. Hybrid Framework and Model Training

This article builds a new fusion model by introducing GARCH-family volatility predictions and market implied volatility as exogenous input variables to an LSTM neural network. The core idea is that while LSTM can capture temporal patterns in the target variable (Realized Volatility), its predictive power can be significantly enhanced by providing it with forward-looking information from other well-established models.

This study systematically tests various combinations of these features across both the NDX and SPX markets to identify the most effective and generalizable model. The main model configurations are described in Table 1. For clarity, in all subsequent tables and discussions, ‘GARCH’ refers to the standard GARCH(1,1) model, ‘GJR-GARCH’ refers to the GJR-GARCH(1,1) model, and so on.

Table 1: Description of Key Hybrid Model Configurations

Model Abbreviation	Input Features
LSTM-RV (Baseline)	Historical Realized Volatility (RV)
LSTM-RV-IV	Historical RV, Implied Volatility (VNX or VIX)
LSTM-RV-GARCH	Historical RV, GARCH Prediction
LSTM-RV-GJR-GARCH	Historical RV, GJR-GARCH Prediction
LSTM-RV-TARCH	Historical RV, TARCH Prediction
LSTM-RV-EGARCH	Historical RV, EGARCH Prediction
LSTM-RV-AVGARCH	Historical RV, AVGARCH Prediction
LSTM-RV-FIGARCH	Historical RV, FIGARCH Prediction
LSTM-RV-IV-GARCH	Historical RV, IV, GARCH Prediction
LSTM-RV-IV-GJR-GARCH	Historical RV, IV, GJR-GARCH Prediction
LSTM-RV-IV-TARCH	Historical RV, IV, TARCH Prediction
LSTM-RV-IV-EGARCH	Historical RV, IV, EGARCH Prediction
LSTM-RV-IV-AVGARCH	Historical RV, IV, AVGARCH Prediction
LSTM-RV-IV-FIGARCH	Historical RV, IV, FIGARCH Prediction
LSTM-RV-IV-All-GARCH	Historical RV, IV, All 6 GARCH Predictions

The prepared sequences of features are used to train the LSTM network. To ensure a fair comparison, a consistent set of optimal hyperparameters, determined through experimentation, is used for all LSTM models. These parameters are detailed in Table 2. The model is trained using the Adam optimizer and Mean Squared Error (MSE) as the loss function, and its performance is evaluated on a hold-out test set for each market separately.

Table 2: LSTM Hyperparameter Settings

Epochs	Batch Size	Optimizer	Learning Rate	Hidden Size	Layers	Dropout
500	32	Adam	0.001	50	2	0.2

4. Experiment

4.1. Evaluation indicators

Since a single error evaluation index has its limitations, if only one index is used to evaluate the model effect, it will produce a large deviation and it is difficult to make a comprehensive evaluation of the model's prediction results. Therefore, we use 6 evaluation indicators to evaluate the model, which are shown in Table 3.

Table 3: Different evaluation indicators

Evaluation	Notes
MSE	$\frac{1}{T} \sum_{t=1}^T (RV_t - \sigma_t^2)^2$
MAE	$\frac{1}{T} \sum_{t=1}^T RV_t - \sigma_t^2 $
RMSE	$\sqrt{\frac{1}{T} \sum_{t=1}^T (RV_t - \sigma_t^2)^2}$
HMSE	$\frac{1}{T} \sum_{t=1}^T \left(1 - \frac{\sigma_t^2}{RV_t}\right)^2$
HMAE	$\frac{1}{T} \sum_{t=1}^T \left 1 - \frac{\sigma_t^2}{RV_t}\right $
QLIKE	$\frac{1}{T} \sum_{t=1}^T \left(\ln \sigma_t^2 + \frac{RV_t}{\sigma_t^2}\right)$

4.2. Datasets and Results

The datasets used are the 5-minute historical price data for the NASDAQ 100 (NDX) and the S&P 500 (SPX). From these, daily Realized Volatility (RV) is calculated. We also incorporate the respective CBOE implied volatility indices: the VNX for the NASDAQ-100 and the VIX for the S&P 500. This dual-market setup allows for a robust test of the model’s performance across different market structures.

After calculating RV and generating GARCH predictions, the data for each market is divided into training and test sets in a ratio of 6 : 4.

The experimental results for the NASDAQ-100 and S&P 500 are presented visually in Figures 2 and 3, with detailed metrics in Tables 4 and 5.

All LSTM Models: Volatility Predictions vs Actual RV on Test Set

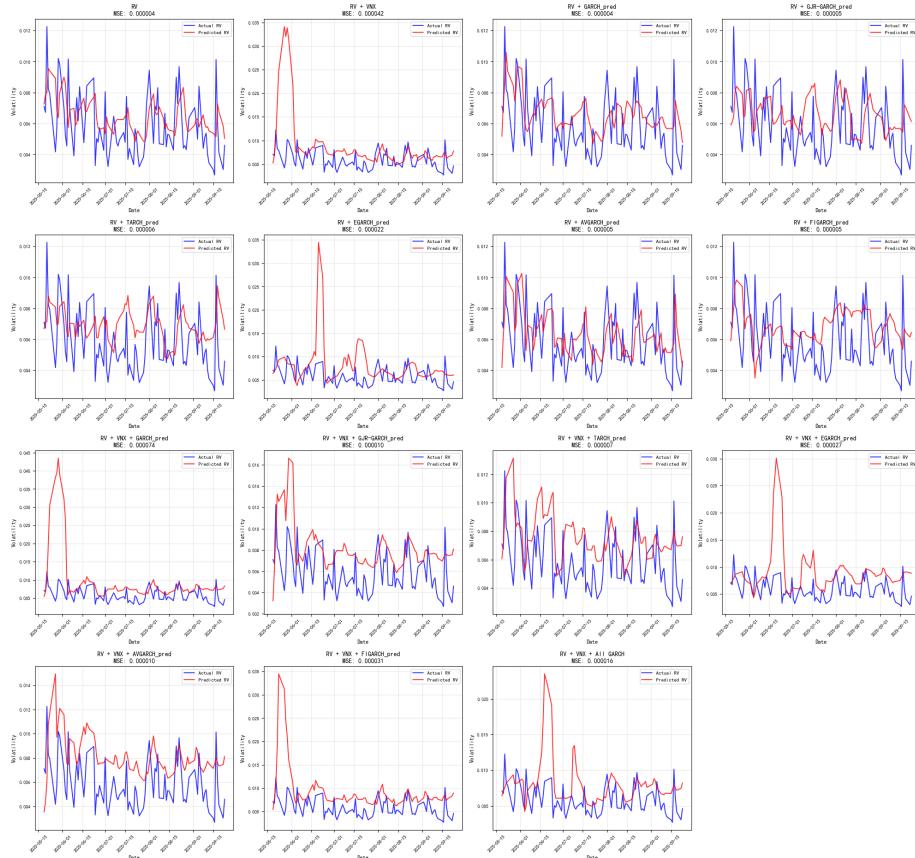


Figure 2: Volatility Predictions vs Actual RV on NASDAQ-100 Test Set

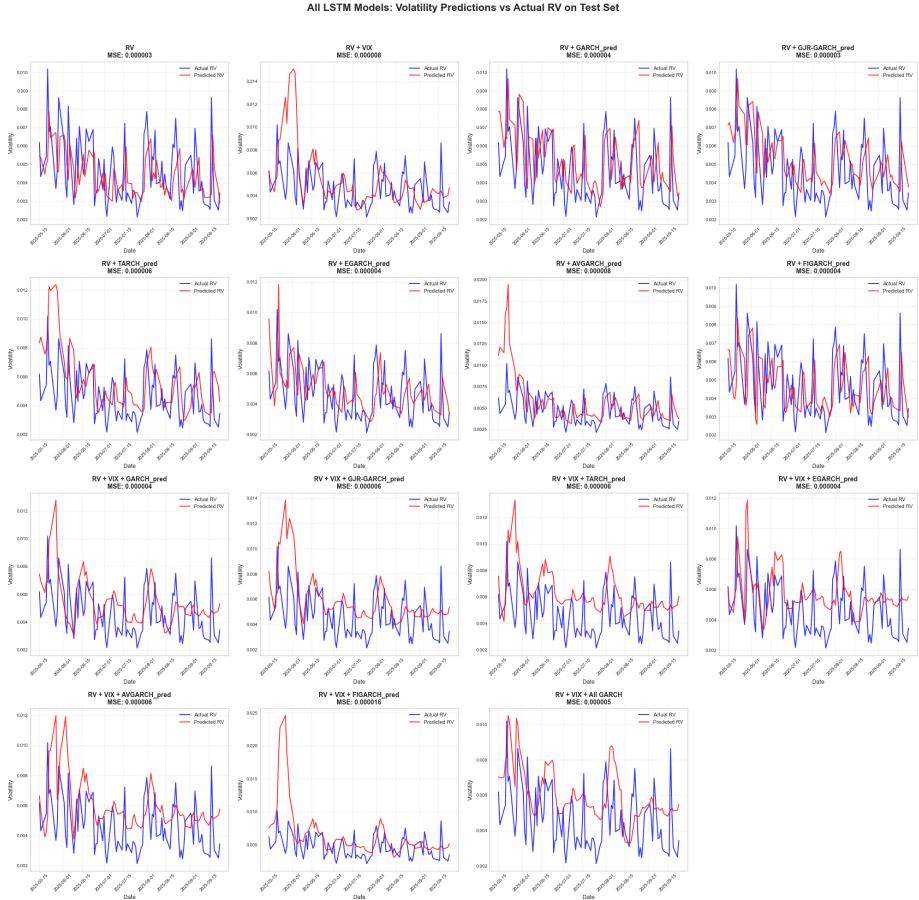


Figure 3: Volatility Predictions vs Actual RV on S&P 500 Test Set

Table 4: Full Experimental Results for NASDAQ-100 (NDX) on Test Set

Model	MSE	MAE	RMSE	HMAE	HMSE	QLIKE
RV	0.000004	0.001728	0.002085	0.333830	0.189105	-4.121149
RV + VNX	0.000042	0.003484	0.006461	0.643581	1.361482	-4.064308
RV + GARCH	0.000004	0.001763	0.002043	0.345066	0.189642	-4.122119
RV + GJR-GARCH	0.000005	0.001844	0.002265	0.362703	0.228146	-4.111549
RV + TARCH	0.000006	0.001952	0.002353	0.401267	0.288057	-4.108184
RV + EGARCH	0.000022	0.002929	0.004684	0.564699	0.732345	-4.066076
RV + AVGARCH	0.000005	0.001753	0.002142	0.330436	0.184342	-4.118387
RV + FIGARCH	0.000005	0.001957	0.002333	0.390539	0.252976	-4.104031
RV + VNX + GARCH	0.000074	0.004361	0.008619	0.793911	2.182858	-4.034769
RV + VNX + GJR-GARCH	0.000010	0.002577	0.003148	0.525713	0.466104	-4.083498
RV + VNX + TARCH	0.000007	0.002189	0.002700	0.458591	0.410872	-4.100649
RV + VNX + EGARCH	0.000027	0.003599	0.005241	0.712556	1.142888	-4.042722
RV + VNX + AVGARCH	0.000010	0.002644	0.003106	0.548409	0.508732	-4.078449
RV + VNX + FIGARCH	0.000031	0.003476	0.005579	0.689190	1.215970	-4.055682

Table 5: Full Experimental Results for S&P 500 (SPX) on Test Set

Model	MSE	MAE	RMSE	HMAE	HMSE	QLIKE
RV	0.000003	0.001371	0.001786	0.297524	0.147231	-4.356314
RV + VIX	0.000008	0.001807	0.002745	0.412834	0.421819	-4.340616
RV + GARCH	0.000004	0.001519	0.001919	0.366709	0.246728	-4.354821
RV + GJR-GARCH	0.000003	0.001535	0.001866	0.373294	0.235907	-4.358079
RV + TARCH	0.000006	0.001816	0.002374	0.445140	0.380586	-4.343857
RV + EGARCH	0.000004	0.001562	0.001978	0.355456	0.208423	-4.351091
RV + AVGARCH	0.000008	0.001940	0.002883	0.430813	0.364749	-4.343037
RV + FIGARCH	0.000004	0.001424	0.001884	0.317488	0.178129	-4.346674
RV + VIX + GARCH	0.000004	0.001629	0.002068	0.400642	0.295012	-4.357120
RV + VIX + GJR-GARCH	0.000006	0.001922	0.002460	0.483620	0.422610	-4.343364
RV + VIX + TARCH	0.000006	0.002007	0.002441	0.532983	0.485529	-4.333721
RV + VIX + EGARCH	0.000004	0.001791	0.002117	0.458701	0.345931	-4.347024
RV + VIX + AVGARCH	0.000006	0.001916	0.002424	0.497191	0.450871	-4.340194
RV + VIX + FIGARCH	0.000016	0.002341	0.003971	0.546921	0.825420	-4.330617

Based on the full experimental results, we observe remarkably consistent patterns across both the NDX and SPX markets.

For both indices, the baseline LSTM model using only historical RV ('RV') performs exceptionally well and serves as a strong benchmark. The primary finding is that hybrid models that do not include implied volatility ('RV + GARCH') are the best performers. For the NDX, 'RV + GARCH' and 'RV +

AVGARCH’ have the strongest results, while for the SPX, ‘RV + GJR-GARCH’ and ‘RV + FIGARCH’ show top-tier performance. This confirms that adding GARCH-family predictions as features provides a powerful and complementary signal to the LSTM.

The most striking and consistent result across both markets is the negative impact of implied volatility. For both the NDX and the SPX, adding the respective implied volatility index (‘RV + VNX’, ‘RV + VIX’) significantly degrades performance across almost all metrics compared to the baseline ‘RV’ model. Even more telling is that the models including all three feature types (‘RV + IV + GARCH’) are consistently outperformed by their simpler ‘RV + GARCH’ counterparts. This strongly suggests that, within this hybrid LSTM framework, the information contained in implied volatility is either redundant to what the LSTM and GARCH features already capture, or it introduces noise that the model cannot effectively parse. This is a counter-intuitive but robust finding.

5. Conclusion

This comparative study on the NASDAQ-100 and S&P 500 indices yields two robust conclusions for hybrid volatility forecasting. First, integrating one-step-ahead GARCH-family predictions as exogenous features into an LSTM network consistently and significantly enhances predictive accuracy across both markets. Second, the inclusion of market implied volatility (VNX and VIX) consistently **degrades** model performance, suggesting it provides redundant or noisy information within this framework. These findings underscore the value of combining econometric forecasts with deep learning and offer a clear, generalizable principle for feature selection in financial forecasting. This work provides not only a refined tool for risk management but also affirms that thoughtfully engineered features from established econometric models are critical for building powerful and generalizable forecasting systems.

References

- [GMAM24] Andrés García-Medina and Ester Aguayo-Moreno. Lstm-garch hybrid model for the prediction of volatility in cryptocurrency portfolios. *Computational Economics*, 63(4):1511–1542, 2024. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC10013303/pdf/10614_2023_Article_10373.pdf.
- [hyb24] Hybrid lstm-garch framework for financial market volatility risk prediction. *Journal of Computer Science and Software Applications*, 4(5):22–29, 2024. URL: <https://www.mfacademia.org/index.php/jcssa/article/download/158/137>.
- [KW18] Hyoungjae Kim and Changhyeon Won. Forecasting the volatility of stock price index: A hybrid model integrating lstm with multiple

- garch-type models. *Expert Systems with Applications*, 103:25–37, 2018. URL: <https://arxiv.org/pdf/2502.15813.pdf>.
- [PTW24] Haojun Pan, Yuxiang Tang, and Guoqiang Wang. A stock index futures price prediction approach based on the multi-garchlstm mixed model. *Mathematics*, 12(11):1677, 2024. URL: <https://www.mdpi.com/2227-7390/12/11/1677>.
- [Ver21] S. Verma. Forecasting volatility of crude oil futures using a garch-rnn hybrid approach. *Intelligent Systems in Accounting, Finance and Management*, 28(2), 2021. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/isaf.1489>.
- [ZSB⁺22] Haoran Zeng, Bolin Shao, Guang Bian, Hongzhe Dai, and Fan Zhou. A hybrid deep learning approach by integrating extreme gradient boosting-long short-term memory with generalized autoregressive conditional heteroskedasticity family models for natural gas load volatility prediction. *Energy Science & Engineering*, 10(7), 2022. URL: <https://scijournals.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/ese3.1122>.