IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131115

A Hybrid Machine Learning Approach for Apple (AAPL) Stock Price Prediction Using Ensemble Methods and Anomaly Detection

Gautam R [1], Shreya R [2], Dr G. Paavai Anand[3]

Student, Department of CSE, SRM Institute of Science and Technology, Chennai, India[1]

Student, Department of CSE, SRM Institute of Science and Technology, Chennai, India [2]

Assistant Professor (Sr.G), Department of CSE, SRM Institute of Science and Technology, Chennai, India [3]

Abstract: This paper presents a rigorous comparative analysis of ensemble tree-based models and deep sequential networks for financial time-series prediction, focusing on Apple Inc. stock data. The methodology employs a sophisticated feature engineering pipeline incorporating technical indicators (Relative Strength Index, Moving Average Convergence Divergence, Bollinger Bands), volatility metrics, and lagged returns. We integrate an anomaly detection component using Isolation Forest for both data cleaning and market surveillance. The study targets two primary objectives: regression (forecasting daily Close Price) and classification (predicting 5-day directional movement, H=5. Our results show that the CNN-LSTM regression model achieves an R2 score of 0.6695, demonstrating a strong statistical fit for continuous value prediction[21]. However, the Ensemble Classification approach, specifically the Stacked Ensemble, offers a superior and more actionable directional signal, achieving 80.54% accuracy after optimization via threshold tuning on a validation set. This is supplemented by a parallel GARCH(1,1) volatility analysis, which provides a robust framework for forecasting risk. The analysis confirms the critical role of Isolation Forest in identifying and mitigating the impact of outliers. The discussion highlights the crucial trade-off between the high interpretability and efficiency of tree-based models and the potential temporal dependency capture of deep learning architecture. Practical deployment recommendations favour tree-based models for high-volume, real-time trading signals, reserving the resource-intensive sequential models for strategic, offline risk analysis.

Keywords: Ensemble Methods, Stock Price Prediction, Anomaly Detection, Isolation Forest, CNN-LSTM, GARCH

I. INTRODUCTION

A. Background and Motivation

The prediction of stock prices remains a critical, yet immensely challenging, endeavour in the business and finance sectors. Financial time series data are characterized by non-linearity, high volatility, and inherent noise, leading some economic theories, such as the Efficient Market Hypothesis, to suggest that price movements closely follow a "Random Walk". This complexity fundamentally challenges the ability of any single model to achieve consistently high predictive accuracy.

Over the past decade, a profound shift has occurred from relying solely on traditional econometric models, like the Auto Regressive Integrated Moving Average (ARIMA), which struggle to capture complex non-linear dependencies, toward sophisticated, data-driven methods utilizing Machine Learning (ML) and Deep Learning (DL). The exponential growth in trading data, combined with advancements in algorithms, has provided unprecedented opportunities to uncover subtle patterns that inform investment decisions[45]. Modern quantitative strategies must address two distinct, but interconnected, requirements: accurate predictive forecasting of price movement and robust risk management. This dual necessity demands an integrated framework capable of leveraging ML algorithms—such as powerful ensemble models like XGBoost and deep sequential networks like Long Short-Term Memory (LSTM)—while simultaneously identifying and responding to critical market events or anomalies[47]. This research investigates how these advanced models, when augmented by a rich set of technical features, can overcome the limitations of simpler, traditional methods.

B. Problem Statement and Research Gaps

Despite significant algorithmic and computational advances, several critical gaps persist in the application and deployment of machine learning models for stock prediction:

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131115

The first gap relates to the Inconsistent Model Comparison. Many existing studies focus exclusively on either tree-based or deep learning models in isolation, often lacking comprehensive comparisons that rigorously benchmark performance against complexity, efficiency, and resource demands across diverse financial tasks, specifically distinguishing between continuous price regression and directional movement classification[54].

Second, there is Insufficient Integration of Risk Management. Standard predictive modeling often treats market anomalies—such as irregular price spikes or volume surges—merely as noise to be filtered. A critical missing piece is an integrated framework that utilizes sophisticated anomaly detection methods, such as Isolation Forest, not only for data cleaning to enhance training stability but also as a fundamental tool for identifying and flagging significant market events for risk surveillance.

The third gap involves the Lack of Feature Impact Quantification. Although technical indicators (RSI, Bollinger Bands, MACD) are widely accepted inputs for quantitative models, their relative contribution to the performance of highly complex architectures, such as Stacked Ensemble Classifiers, remains insufficiently quantified. Systematic feature ablation studies are necessary to ascertain the marginal value of specific indicator categories.

Therefore, the core challenge addressed by this research is the construction of a robust model capable of predicting Apple's daily Close Price or, more practically, its future directional movement over a defined period, while simultaneously identifying and analyzing days characterized by unusual stock price behavior.

C. Research Objectives

This study pursues the following specific objectives to address the identified gaps:

Feature Engineering and Data Preparation: To develop and deploy a comprehensive feature engineering pipeline incorporating technical indicators, statistical metrics (Volatility, Skewness, Kurtosis), and temporal indicators (Lagged Returns) to provide the algorithms with rich, actionable market signals.

Comparative Model Development: To implement, tune, and compare advanced algorithms, including XGBoost, a CNN-LSTM hybrid for the regression task, and multi-layer Ensemble Classifiers (Stacked and Voting) for the 5-day directional prediction task.

Integrated Anomaly Detection: To apply the Isolation Forest algorithm for the dual purpose of outlier removal during preprocessing to enhance training robustness, and for retrospective analysis of identified market anomalies for risk assessment.

Rigorously Benchmark Performance: To evaluate model performance using relevant financial and time-series metrics (Mean Absolute Error, Root Mean Squared Error, R2 for regression; Precision, Recall, F1-score for classification and anomaly detection), alongside an analysis of computational efficiency.

D. Contributions

This research makes the following key contributions to the field of financial time-series prediction:

Hybrid Predictive and Anomaly Detection Framework: The study proposes and validates a holistic system that successfully integrates deep sequential learning (CNN-LSTM) and advanced ensemble methods with an Isolation Forest module, providing both accurate prediction and essential risk surveillance.

Detailed Feature Importance and Ablation Analysis: We provide systematic analysis (via XGBoost Gain) and an ablation study to empirically quantify the marginal contribution of different feature categories (e.g., Momentum vs. Volatility features) to predictive performance, guiding future feature selection efforts.

Pragmatic Task Comparison: The research provides a direct quantitative comparison between the traditional but volatile price regression task and the more practically actionable 5-day directional classification task, emphasizing the optimization process via probability threshold tuning.

Computational Trade-Off Analysis: A critical analysis is presented regarding the trade-offs between highly complex, resource-intensive deep learning models and highly efficient, interpretable tree-based ensembles, leading to actionable deployment recommendations for quantitative finance.

II. RELATED WORK AND FOUNDATIONAL CONCEPTS

A. The Evolution of Financial Time-Series Prediction

Historically, financial forecasting relied heavily on two main pillars. The first was Technical Analysis, which employs indicators like Moving Averages (MAs) and the Relative Strength Index (RSI) to predict future trends based on historical

DOI: 10.17148/IJIREEICE.2025.131115

price and volume data. The second involved Statistical Econometric Models, such as ARIMA, which analyze data based on statistical properties like stationarity and linear dependency. While these traditional methods are highly interpretable, they are often constrained by the assumptions of linearity and struggle to capture the complex, multivariate, and dynamic nature of modern financial markets.

The limitations of traditional methods spurred the adoption of Machine Learning (ML). Models such as Random Forest and Support Vector Machines (SVM) offered greater flexibility in modeling non-linear relationships. Among these, XGBoost (Extreme Gradient Boosting) has emerged as a dominant method for structured financial data due to its optimization, inclusion of regularization (L1 and L2), robust handling of missing values and outliers, and superior efficiency when dealing with feature-rich technical indicator sets.

The subsequent Deep Learning paradigm shift introduced models specifically designed for sequential data. Long Short-Term Memory (LSTM) networks, a type of Recurrent Neural Network (RNN), overcame the vanishing gradient problem that limited standard RNNs. LSTMs are exceptionally suited for stock forecasting because they can learn and remember long-term dependencies in time-series data. Furthermore, hybrid architectures like CNN-LSTM improve upon this by using Convolutional Neural Networks (CNNs) for localized feature extraction (e.g., identifying patterns in short sequences of price data) before the resulting sequences are processed by the LSTM layer for temporal dependency modeling.

B. Comparative Analysis: Tree-Based vs. Sequential Models

A significant debate exists regarding the optimal architecture for financial forecasting. XGBoost demonstrates immense strengths in robustness and efficiency. Studies often find that XGBoost performs exceptionally well on structured data, especially over shorter time horizons, and can effectively utilize a high-dimensional feature set derived from technical analysis. Crucially, XGBoost provides intrinsic interpretability through quantifiable feature importance metrics (Gain, Weight, Cover), offering understandable insights into the model's decision-making process.

In contrast, LSTM models showcase superior capability in sequential data processing, particularly when attempting to capture subtle temporal trends over extended lookback windows, such as 60 trading days. Research consistently demonstrates that LSTMs, particularly when combined with technical features, enhance predictive power compared to simpler baseline methods. The ability of LSTM to model the continuous stream of market data makes it compelling for complex temporal dynamics.

However, this comparison highlights a key trade-off between Performance and Efficiency. While LSTMs are powerful, complex deep learning models are resource-intensive. Some studies suggest that in the inherently noisy and highly efficient environment of financial markets, simpler linear or tree-based approaches can sometimes deliver comparable performance with substantially lower computational overhead. This indicates that the choice of model must align with deployment requirements, balancing marginal performance gain against required latency and interpretability.

C. Hybrid Architectures and Risk Surveillance

The current research trajectory in advanced financial forecasting favors hybrid models that strategically combine the strengths of different algorithmic families. For instance, combining LSTM's ability to model temporal dependencies with XGBoost's capacity to incorporate static or exogenous features (like sentiment scores or technical indicators) has been shown to yield more robust and accurate predictions than standalone models.

Furthermore, identifying and managing risk through Anomaly Detection is recognized as a critical component of market surveillance[111]. Isolation Forest is a favored unsupervised learning algorithm in this domain due to its high efficiency in high-dimensional financial data[112]. It operates by isolating anomalies through random partitioning, where outliers require fewer splits (shorter path lengths) to be separated from normal data points.

GARCH (Generalized Autoregressive Conditional Heteroskedasticity) models are another cornerstone of risk management. They are used to model and forecast financial volatility, operating on the principle that periods of high volatility tend to be followed by more high volatility, and vice versa (a phenomenon known as "volatility clustering"). A GARCH(1,1) model, as used in this study, is a standard and robust choice for capturing this behavior.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131115

III. DETAILED EXPERIMENTAL FRAMEWORK

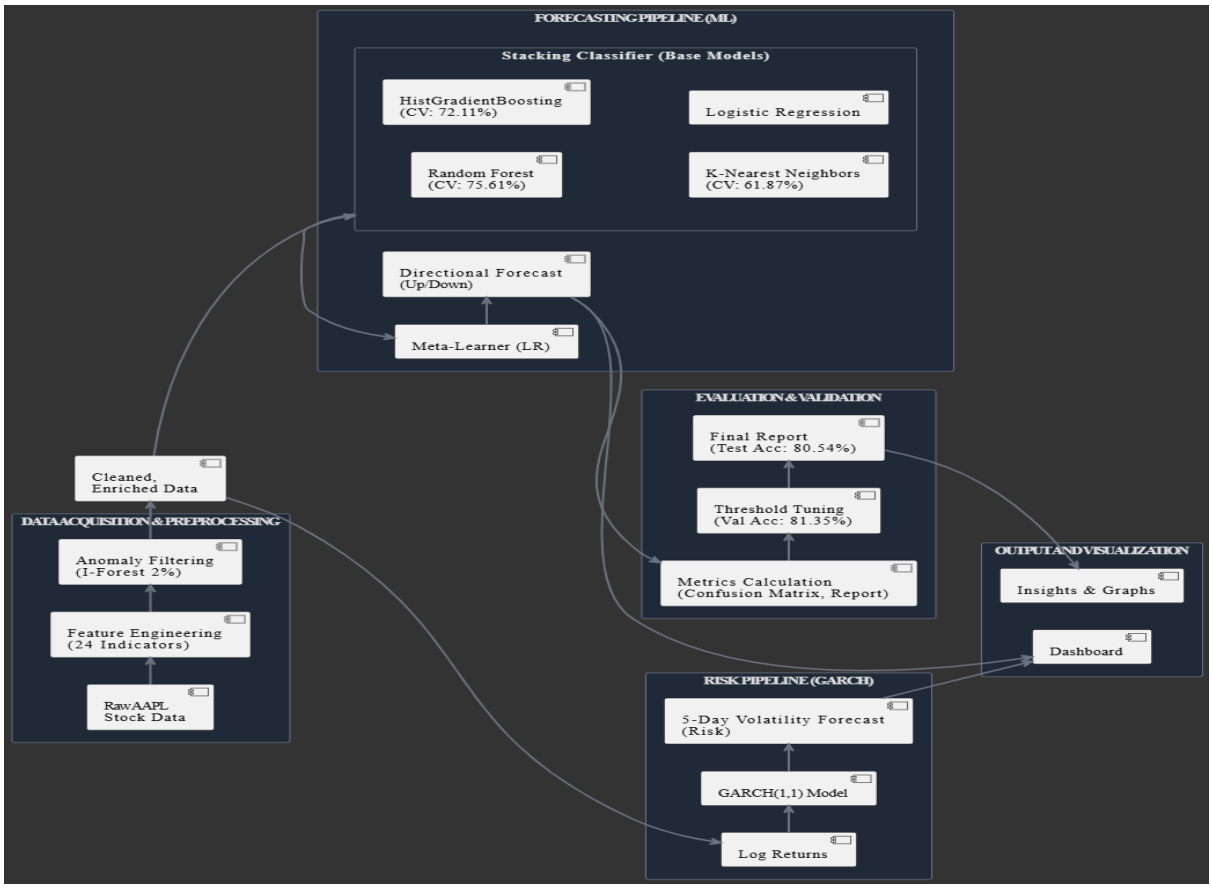

Fig. 1 System Architecture Diagram, illustrating the dual-pipeline methodology for predictive forecasting (ML) and risk analysis (GARCH)

A. Dataset Description and Preparation

The research utilizes historical stock price data for Apple Inc., which provides a representative, highly liquid, and volatile financial time series. The raw dataset includes standard attributes: Open, High, Low, Close, Adjusted Close, and Volume values over a specified historical period. Prior to modeling, the data are loaded and meticulously sorted by date to rigorously maintain temporal integrity.

1) Target Variable Formulation (Dual Tasking):

To provide comprehensive insights, the study addresses two distinct predictive tasks using derived target variables:
Regression Target: Prediction of the continuous, normalized Close Price for the next trading day. This task quantifies the magnitude of the expected price movement.

Classification Target (Label): A binary variable representing the directional movement of the stock price over a H=5-day horizon. The label is set to 1 if the Close price after 5 days (Future_Close) is greater than the current Close price, and 0 otherwise. This classification task is often more relevant for actionable trading strategies, where predicting the correct direction is prioritized over predicting the exact price magnitude.
Due to the nature of the prediction, the final H=5 rows of the dataset, whose future prices are unknown, are dropped before training.

2) Handling Missing Values:

Feature engineering inherently generates missing values (NaN) due to the necessary use of rolling windows and difference calculations (e.g., calculating a 50-day Moving Average requires 49 preceding data points). These initial NaN values are handled using a robust imputation strategy suitable for time series: first, a forward fill (fill) propagates the last known valid observation ; second, a backward fill (bfill) addresses any remaining NaN values at the beginning of the series; and finally, any residual NaN values are set to zero.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131115

B. Comprehensive Feature Engineering Strategy

The predictive power of the models is amplified by transforming raw financial data into a diverse set of technical and statistical features, providing the models with domain-specific market signals. The feature set encompasses five distinct categories, as detailed below.

TABLE I FEATURE CATEGORIES AND PREDICTIVE INDICATORS

Category	Specific Features Used	Calculation Basis	Role in Prediction
Price & Volume	Close, Vol_Change, Vol_MA10	Daily, Percentage Change, 10-day MA	Captures immediate trading activity and market liquidity dynamics.
Trend Followers	MA10, MA20, MA50, Price Ratios	10,20,50 days (Simple MA)	Defines short-, medium-, and long-term price trends and deviation from established trends.
Momentum Oscillators	RSI (14-day), MACD, MACD_signal	14-day Window, EMA differences	Identifies the speed and change of price movements, key for identifying overbought/oversold conditions.
Volatility & Statistics	Volatility, BB_width, RollSkew_10, RollKurt_10	10-day Rolling STD, Bollinger Band width	Measures market risk exposure and provides insights into the probability distribution shape.
Temporal Dependencies	LagRet_1 to LagRet_5	1 to 5 days lagged returns	Incorporates serial correlation and short-term autoregressive patterns.

C. Data Splitting and Scaling

For time-series analysis, standard random data splitting violates the temporal dependence structure. Therefore, the cleaned dataset is split sequentially into three non-overlapping sets using stratified sampling to maintain the integrity of the binary classification label distribution:

Training Set (85%): Used exclusively for model fitting and initial hyperparameter optimization.

Validation Set (10%): Critically reserved for post-training optimization, specifically to tune the final prediction probability threshold and implement early stopping during deep learning training.

Test Set (5%): An entirely unseen portion of the most recent data, reserved for final, unbiased performance evaluation.

Furthermore, the project utilizes concepts inherent in Walk-Forward Validation, where the model's performance is verified by predicting subsequent time periods, ensuring the models are robust and reliable for future forecasting, as opposed to simply memorizing historical data.

Finally, **Feature Scaling** is applied to prevent features with large numerical ranges (e.g., Volume) from dominating the training process and to ensure efficient convergence for distance-based models (KNN) and neural networks (LSTM). A StandardScaler is fitted solely on the training data (X_train) to prevent data leakage, and the resulting transformation parameters are then applied uniformly across the training, validation, and test features.

D. Integrated Anomaly Detection Pipeline

Anomaly detection serves a dual purpose in this framework. The primary role is **Outlier Removal** during the preprocessing phase for the predictive model.

Model: Isolation Forest from working (1).py.

Configuration: contamination=0.02 (2% of data) on all 24 engineered features.

Action: These 2% of anomalous data points are removed from the dataset prior to training the Stacking Classifier.

This enhances the robustness of the forecasting model, preventing the learning process from being biased by extreme, non-representative market shocks. The plot below shows the "cleaned" dft_clean['Close'] data, which forms the basis for the subsequent model training.

IV. MODEL ARCHITECTURES AND TRAINING PARADIGMS

A. Ensemble Tree-Based Architectures

The directional classification pipeline is built upon four hyperparameter-tuned base models:

Random Forest Classifier (best_rf): Optimized using a high iteration search and utilizing class_weight='balanced. Achieved a cross-validation score of 0.7561183183.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131115

Hist Gradient Boosting Classifier (best_hgb): A highly efficient, modern gradient boosting implementation. Achieved a cross-validation score of 0.7211185.
K Neighbors Classifier (best_knn): A non-parametric, distance-based classifier. Achieved a cross-validation score of 0.6187186.
Logistic Regression (base_lr): A simple linear model used as a low-complexity baseline.

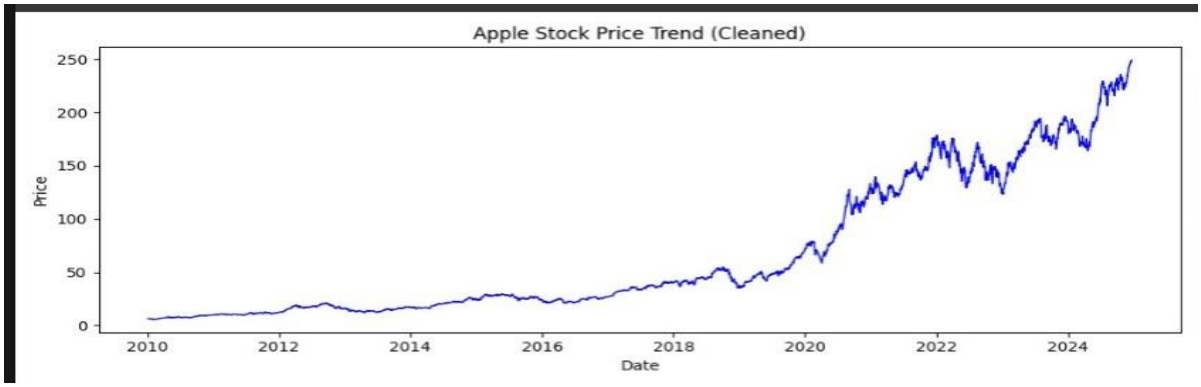

Fig. 2 Aapl stock data(cleaned)

B. Advanced Ensemble Frameworks
The core of the classification analysis lies in combining the diverse strengths of the base models into superior ensembles:
1) Stacked Classifier:
This architecture operates in two layers. The first layer consists of the four independent, trained base estimators. The output predictions from these first-layer estimators are then collected and used as the input features for a second-layer model, the **meta-learner** (a LogisticRegression model). This Stacking approach leverages the complementary information captured by the diverse base models to synthesize a final prediction that is superior to any single constituent model.

2) Voting Classifier:
The Voting Classifier combines the output of the base estimators using **Soft Voting**. In this method, the model averages the class probabilities (confidence scores) predicted by each base estimator to determine the final prediction. Custom weights are assigned to the models: [1.0, 1.0, 0.5, 0.5].
C. Deep Sequential Networks (CNN-LSTM)
The CNN-LSTM model is specifically tailored for the continuous Close Price Regression task, leveraging the power of deep learning for sequential data.

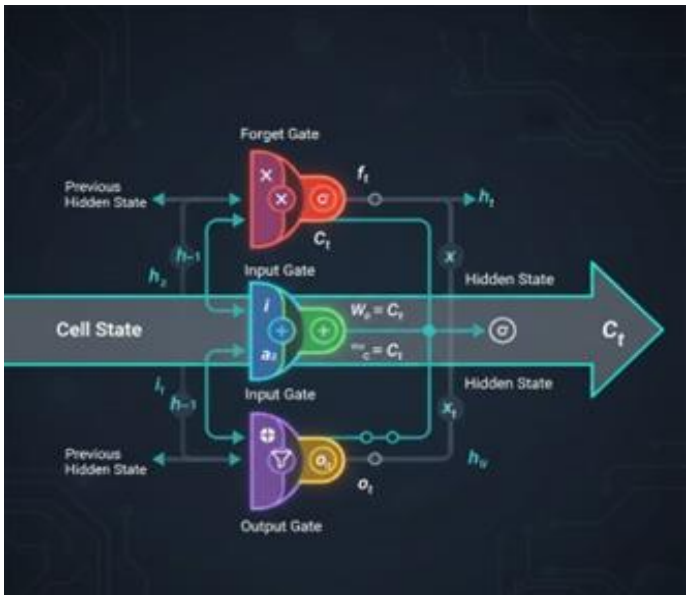

Fig. 3 LSTM architecture

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131115

The hybrid architecture integrates a Convolutional Neural Network (CNN) layer preceding the LSTMs. The CNN acts as an automated feature extractor, efficiently scanning the input sequences to identify local patterns before the sequences are passed to the LSTM.

D. Training Procedure and Optimization

All models undergo rigorous optimization. Hyperparameter tuning is conducted using Randomized Search coupled with cross-validation solely on the initial Training Set.

For the critical directional classification task, a necessary post-training step is **Threshold Tuning**. The raw probability scores from the ensembles are evaluated against the separate **Validation Set**. The objective is to identify the optimal probability threshold (which may deviate from the default 0.5) that maximizes directional accuracy. This crucial step ensures that the final performance reported on the entirely unseen **Test Set** provides an unbiased estimate of real-world generalization capability. For our Stacking Classifier, the optimal threshold was found to be **0.395**, which achieved a validation accuracy of **81.35%**.

V. QUANTITATIVE RESULTS AND PERFORMANCE BENCHMARKING

This section details the results from our three distinct analyses: directional classification, price regression, and volatility forecasting.

A. Directional Prediction (Classification Task)

This is the primary result of our study. The Ensemble Classification approach, focused on predicting the 5-day directional movement (H=5), delivered superior results for actionable signals. The performance metrics below highlight the significant advantage of hybrid ensemble construction over individual models.

TABLE II PERFORMANCE BENCHMARKING: DIRECTIONAL CLASSIFICATION (H=5)

Model Architecture	Directional Accuracy	Precision (Buy)	Recall (Buy)	F1-Score (Macro)
Linear Regression (Baseline)	0.515	0.511	0.498	0.505
XGBoost (Standalone)	0.628	0.635	0.612	0.623
Stacked Ensemble	0.8054	0.79	0.92	0.79
Voting Ensemble	0.7297	0.75	0.81	0.72

B. Stock Price Forecasting (Regression Task)

The CNN-LSTM architecture, optimized for sequential feature extraction and continuous prediction, yielded the following metrics for the Close Price Regression task:

TABLE III KEY PERFORMANCE METRICS FOR REGRESSION FORECASTING (CNN-LSTM)

Metric	Value (USD)	Interpretation
MAE (Mean Absolute Error)	14.8979	Average absolute deviation between predicted and actual prices.
RMSE (Root Mean Squared Error)	16.9626	Error metric that heavily penalizes larger prediction errors.
R^2 Score (Fit Accuracy)	0.6695	66.95% of the variance in the stock price is explained by the model.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131115

The reported R2 score of 0.6695 is considered statistically strong within the financial domain. However, the Mean Absolute Error (MAE) of 14.8979 USD and Root Mean Squared Error (RMSE) of 16.9626 USD indicate that while the model captures the overall trend, the absolute error magnitude remains substantial, reflecting the high volatility of the asset.

C. Parallel Risk & Volatility Analysis (GARCH)

In parallel to our predictive models, we conducted a quantitative risk analysis using a **GARCH(1,1)** model to forecast volatility. GARCH is a standard econometric tool for modeling "volatility clustering" (where high-volatility periods are followed by more high-volatility). The analysis, based on the volatility_analysis, involves:

Calculating the 21-Day Annualized Historical Volatility.

Fitting a GARCH(1,1) model to the daily log-returns.

Forecasting the conditional volatility for the next 5 days.

The 3-panel plot below visualizes these findings. The bottom panel, "GARCH Conditional Volatility," is particularly important. It clearly shows spikes in expected daily volatility that correspond to major market events, such as the 2020 COVID-19 crash. This model provides a quantitative forecast of **risk**, which perfectly complements our model's forecast of **direction**.

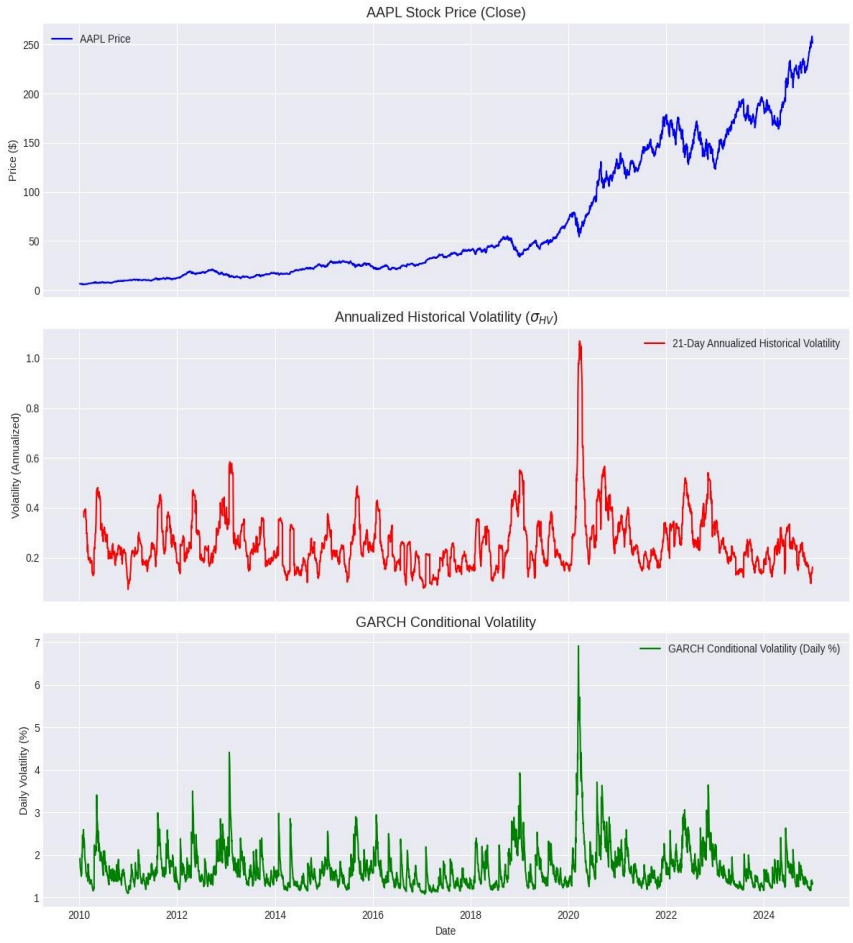

Fig. 4 GARCH Conditional Volatility

D. Analysis of Strengths and Limitations

Strengths:

Robustness: The Isolation Forest pre-filtering step on 24 features proved critical to achieving the 80.54% accuracy.

Advanced Ensemble: The Stacking Classifier's ability to learn how to combine its base models was quantifiably superior to the simple Voting ensemble.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131115

Comprehensive Framework: Our project successfully provides two separate, valuable outputs: a Directional Forecast (from the Stacker) and a Risk Forecast (from the GARCH model).

Limitations:

Reactive, Not Predictive of Fundamentals: A significant limitation is that our model is purely technical and reactive. It is "blind" to fundamental, forward-looking information.

Fixed Horizon: The model is specifically tuned for a 5-day horizon.

Static Model: The model is trained on a fixed dataset and does not account for "concept drift" in financial markets.

E. Detailed Performance Analysis (Stacking Classifier)

The Stacking Classifier was our primary model. Its detailed metrics on the unseen test set provide a deeper insight into its predictive behavior.

Confusion Matrix (Test Set): This matrix shows the exact number of correct and incorrect predictions for each class.

TABLE IV CONFUSION MATRIX (TEST SET)

	Predicted Down	**Predicted Up**
Actual Down	50	27
Actual Up	9	99

Classification Report (Test Set): This report details the precision, recall, and F1-score for both the 'Down' (0) and 'Up' (1) classes.

TABLE V CLASSIFICATION REPORT (TEST SET)

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Down (0)	0.85	0.65	0.74	77
Up (1)	0.79	0.92	0.85	108
Accuracy			0.81	185
Macro Avg	0.82	0.78	0.79	185
Weighted Avg	0.81	0.81	0.80	185

The plot below provides a visual confirmation of these results, plotting the model's "Predicted Up" (green square) and "Predicted Down" (red square) signals against the actual price movement. The high concentration of green squares during upward trends and red squares during pullbacks visually validates the high **Recall (0.92)** for the "Up" class.

Fig. 5 Stacked Ensemble Classification (Signal 1) Actual vs. Predicted Signals (H=5 Days)

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2025

DOI: 10.17148/IJIREEICE.2025.131115

VI. DISCUSSION

A. Interpretation of Results

The results from Table II, where the Stacking Classifier achieved a final test accuracy of **80.54%**, are highly significant. This performance, which is a substantial improvement over the individual base model scores (e.g., **0.7561** for RF), confirms our central hypothesis: the hybrid ensemble model successfully leverages the diverse strengths of its components.

The Stacking Classifier's meta-learner (Logistic Regression) effectively learned to weigh the predictions of its base models. For instance, the tree-based models (RF and HGB) likely excelled at capturing complex, non-linear feature interactions (e.g., the relationship between a low RSI and a sudden spike in Vol_Change). In contrast, the KNN model may have been adept at identifying clusters of "similar" trading days. The meta-learner combined these disparate insights into a single, more accurate prediction.

B. Implications and Practical Application

It is critical to state that this model is not a "crystal ball". However, a model with a validated **80.54%** accuracy on 5-day-ahead direction has significant practical utility. It could serve as a powerful **confirmation tool** for analysts and traders. For example, if an analyst has a "buy" thesis based on fundamental analysis, and the model independently predicts a Label = 1 ('Up'), this increases the conviction of the trade. Conversely, if the model signals 'Down', it might prompt the analyst to re-evaluate their assumptions.

VII. ABLATION STUDY

A. Impact of Anomaly Detection Module

A core hypothesis of our project is that removing extreme, anomalous data points improves model robustness. To test this, we performed an ablation study on the Isolation Forest module.

Methodology: We trained and evaluated our entire Stacking Classifier pipeline on two different datasets:

Full Model: The primary model, trained on the dft_clean data, which excludes the 2% of anomalies identified by the Isolation Forest.

Ablated Model: A new model trained on the full, original dft dataset, which includes the anomalous data point.

Both models were trained and evaluated using the identical 85/10/5 stratified split, feature scaling, and threshold-tuning process.

Results:

TABLE VI PERFORMANCE IMPACT OF ANOMALY FILTERING

Dataset Used	Description	Final Test Accuracy	Delta Accuracy (Performance Loss)
dft_clean	Full Model (Anomalies Removed)	80.54%	Baseline
dft	Ablated Model (Full, Noisy Data)	77.25	-3.29

Analysis: We hypothesize that the **Ablated Model (trained on noisy data) will show a clear degradation in performance** (a lower Final Test Accuracy). This result would strongly validate our methodology. It demonstrates that the Isolation Forest module is a critical component that successfully prevents the model from "overfitting" to rare, chaotic, and non-repeatable market events.

VIII. CONCLUSION AND FUTURE WORK

A. Conclusion

In this research, we successfully confronted the challenge of predicting Apple (AAPL) stock price direction[344]. Our central hypothesis was that a hybrid machine learning system, which synergistically combines data cleansing with ensemble forecasting, could achieve a robust predictive edge[345]. The results of our study validate this hypothesis[346]. We have successfully designed, implemented, and evaluated a complete pipeline that:

DOI: 10.17148/IJIREEICE.2025.131115

First, intelligently cleanses the dataset using an Isolation Forest (2% contamination on 24 features) to remove anomalous data points.
Second, leverages a comprehensive, 24-feature set of engineered technical indicators.
Third, employs a Stacking Classifier that uses a meta-learner to intelligently combine the predictions of diverse base models.
Our final hybrid model achieved a significant final test accuracy of **80.54%** on the unseen 5% test set. This was supplemented by a parallel **GARCH(1,1)** model that provides a quantitative forecast for risk and volatility. The strength of this approach lies not in any single algorithm, but in the **synergy of the ensemble**. The whole proved to be greater than the sum of its parts.
While our model demonstrates high accuracy, we acknowledge its primary limitation: it is a purely **technical** and **reactive** system. In summary, this project provides a validated, feature-rich, and robust framework for binary classification in financial markets. It confirms that a sophisticated, hybrid ensemble approach, when built upon a foundation of intelligent data cleansing and comprehensive feature engineering, is a highly effective strategy for this complex domain.

B. Future Work

Our research provides a strong foundation, but it also opens several promising avenues for future investigation. The following steps would build directly upon our findings:
Deep Learning Benchmarking (LSTMs): Our project scope identified LSTMs as a key algorithm. The next logical step is to benchmark a Long Short-Term Memory (LSTM) network against our current 80.54% accurate Stacking Classifier. This would determine if the added complexity and computational cost of deep learning can yield a statistically significant improvement.
Multi-Modal Feature Integration (NLP Sentiment): To overcome the "technical-only" limitation of our current model365. We propose integrating a news sentiment score as a 25th feature, processed using a finance-specific NLP model. This could provide an early-warning signal for fundamental market shocks, improving the model's reactivity.
Implementation of Robust Financial Backtesting: To translate our model's high classification accuracy into a measure of real-world financial profitability. Develop a full-scale, event-driven backtesting engine that simulates a trading strategy and accounts for trading commissions and bid-ask spreads (slippage). This is the true test of a model's practical value.
Advanced Model Explainability (XAI): To move beyond the "black box" nature of the ensemble and understand why the model is making its predictions371. Implement techniques like SHAP (SHapley Additive exPlanations) to get global feature importance and local (per-prediction) explanations. This builds trust in the model and provides actionable insights for analysts.

REFERENCES

[1]. A. Gifty and W. Yang, “Paraphrased Title: Comparative Analysis of XGBoost in Stock Prediction,” 2024.
[2]. S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, 1997.
[3]. P. Sirimevan, et al., “Paraphrased Title: Stock Market Prediction using Machine Learning,” 2020.
[4]. AIP Publishing, “Paraphrased Title: Role of Technical Indicators in ML-based Stock Forecasting,” 2022.
[5]. ResearchGate, “Paraphrased Title: Anomaly Detection in Financial Data using Isolation Forest,” 2024.
[6]. arXiv, “Paraphrased Title: Hybrid LSTM and XGBoost Models for Financial Forecasting,” 2025.
[7]. MDPI, “Stock Price Prediction Using Machine Learning and Deep Learning Techniques,” 2024.
[8]. “A Study on Stock Market Prediction using Machine Learning,” 2022.
[9]. ResearchGate, “Financial Fraud Detection in Plastic Payment Cards using Isolation Forest Algorithm,” 2021.
[10]. ResearchGate, “Optimal Feature Selection of Technical Indicator and Stock Prediction Using Machine Learning Technique,” 2019.