On the Forecasting of Market Prices for Agricultural Commodities

Ross De Chi and Patrick Hosein
The University of the West Indies
St. Augustine, Trinidad
ross.dechi@my.uwi.edu, patrick.hosein@uwi.edu

***Abstract*—Accurate forecasting of agricultural market prices is essential for improving production planning, supply chain efficiency, and policy development. This study compares the performance of three forecasting models: Auto-regressive Integrated Moving Average (ARIMA), ARIMA with exogenous variables (ARIMAX) and feed-forward Artificial Neural Networks (ANN) including ANN with exogenous inputs (ANN-X). Historical monthly pricing and volume data covering 25 agricultural products over a 17-year period were used for model training and evaluation. Forecast accuracy was assessed using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Results show that ARIMA remains effective for stable, trend-dominated price patterns, while ARIMAX provides modest improvements when relevant external inputs are available. ANN models capture nonlinear fluctuations more effectively while ANN-X achieves the highest overall accuracy by integrating exogenous variables. Results show ARIMA remains effective for smoother, trend-dominated series, ARIMAX helps only when external signals are reliably predictive, and ANN/ANN-X best capture nonlinear volatility. Findings support data-driven selection: linear baselines for stable series; ANN-X where volatility and vetted exogenous signals prevail.**

***Index Terms*—Time Series Forecasting, ARIMA, ARIMAX, Artificial Neural Networks, Agricultural Markets**

I. Introduction

Forecasting agricultural market prices is challenging due to volatility from weather, seasonality, import competition, and shifting demand. Reliable forecasts support production planning, supply chains, and policy. Traditional models such as Autoregressive Integrated Moving Average (ARIMA) are widely used for their interpretability but assume linearity and struggle with irregular patterns [1], [2]. ARIMA with exogenous variables (ARIMAX) can incorporate predictors such as sales volume when relevant [3]. More recently, Artificial Neural Networks (ANNs) have been applied to capture nonlinear dependencies, with exogenous variants (ANN-X) expanding this flexibility [4]–[6].

A related line of work examines the design of the forecasting *strategy* itself. Prior research contrasts recursive (one–step–ahead models iterated over the horizon) with direct (horizon–specific) forecasting for crop prices, highlighting trade–offs between error propagation, horizon–specific bias, and implementation complexity [16]. This evidence suggests that strategy choice can materially affect evaluation outcomes and deployment practicality, reinforcing the need to pair model class (e.g., ARIMA vs. ANN) with a forecasting strategy suited to data characteristics and decision horizons.

This study systematically compares ARIMA, ARIMAX, ANN, and ANN-X in forecasting monthly prices for a diverse set of agricultural products using 17 years of historical price and volume data. It evaluates predictive performance on four common error metrics, MSE, RMSE, MAE, and MAPE, following established practice in the literature [1], [2] and investigates the role of seasonality through statistical testing. The goal is to provide clear guidance on model selection based on data characteristics, helping researchers and practitioners balance accuracy, complexity, and interpretability in agricultural price forecasting.

Other related work outside agriculture points to similar methodological themes. Deep models for stock prices capture nonlinear temporal structure beyond linear baselines [17]. In cryptocurrency forecasting, robust pipelines with rolling validation are emphasized to mitigate regime-change overfitting [18]. Studies comparing classical learners and deep architectures further underscore the importance of regularization and horizon-aware training for volatile, nonstationary series [19]. ARIMA provides a transparent, stable baseline [1], while ANN/ANN-X add nonlinear capacity at the cost of interpretability and tuning [4]–[6]; this trade-off motivates our controlled comparison on a common split.

Agricultural price forecasting is both methodologically and operationally demanding. Structural breaks from weather, pests, and policy can violate stationarity [1], [2]; seasonality is pronounced yet irregular across commodities; and exogenous signals (e.g., volume) are predictive in some markets but not others, complicating their use in ARIMAX or neural models [3], [4]. At the same time, small errors affect planting, contracts, spoilage, inventory, and affordability. A comparative evaluation under a common pipeline therefore clarifies when interpretable baselines suffice, when exogenous drivers add value, and when higher-capacity neural models justify their complexity, yielding decision-ready guidance for routine monitoring and procurement workflows.

Equally important is reproducibility. Organizations need forecasts that remain stable across refits and resilient to data anomalies, with clear fallbacks when external drivers weaken. Framing the problem as "method–data fit" rather than a one-model contest helps align technical choices with governance, risk, and operational constraints, and turns accuracy gains into actions stakeholders can trust.

Against this backdrop, our contribution is threefold. (i) We provide a head-to-head evaluation of ARIMA, ARIMAX, ANN, and ANN-X across twenty-five diverse agricultural series using a unified pipeline and consistent hyperparameter selection criteria. (ii) We quantify the incremental value of exogenous volume information by contrasting ARIMA vs. ARIMAX and ANN vs. ANN-X. (iii) We couple aggregate error comparisons with a commodity-group analysis to show how seasonality and volatility patterns relate to model performance, providing practical guidance on method selection for stakeholders balancing accuracy, transparency, and implementation cost.

II. Methodology

Monthly price and volume data for seventy-seven agricultural products were compiled from January 2006 to October 2023. Original records were stored in separate annual workbooks; these were aggregated using Python to minimize manual error. Dates were standardized to a uniform YYYY–MM–DD format, numeric fields were cleaned to remove formatting artifacts, and all series were transposed into a monthly time series layout. Products with missing observations were removed (no imputation). Where multiple size categories existed, one representative variant (typically "M") was retained. A random sample of 25 complete series (Jan 2006–Oct 2023; max 214 months) was selected for analysis.

For the ARIMA and ARIMAX models, the Box–Jenkins methodology guided parameter selection. Each series was tested for stationarity using the Augmented Dickey-Fuller test [8]; differencing was applied as needed to achieve stationarity. Autocorrelation and partial autocorrelation plots informed the autoregressive and moving average orders (p and q), while the Akaike Information Criterion (AIC) was used to finalize model configurations [1], [2], [7]. ARIMAX models included monthly sales volume as an exogenous regressor after confirming lag correlation to avoid spurious inclusion [3].

Neural networks were implemented as feed-forward, fully connected architectures. Input features included lagged price observations and, for ANN-X, contemporaneous and lagged volume data. Networks contained a single hidden layer, with the number of neurons tuned by grid search to balance expressiveness and overfitting risk. The rectified linear unit (ReLU) activation function was applied in hidden layers and a linear output activation was used for regression. Models were trained with the Adam optimizer and a mean squared error loss function, with early stopping applied if validation loss failed to improve for 20 epochs [4]–[6].

Performance was evaluated using four complementary error metrics: mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). These were calculated for both training and test sets to assess generalization [1], [2]. To detect seasonality, each product's price series was subjected to the Kruskal–Wallis H test; a p-value below 0.05 indicated statistically significant month-to-month differences. This seasonal information helped contextualize model performance, particularly for crops with strong periodic supply fluctuations. All models were trained and validated on identical data splits to ensure fair comparison, and results were aggregated to compare performance across commodities.

III. Results and Discussion

Table I summarizes the comparative predictive performance of ARIMA, ARIMAX, ANN, and ANN-X across the 25 agricultural commodities (metrics computed as in [1], [2]; neural model setup consistent with [4]–[6]).

TABLE I
Comparative Forecasting Performance Across Models (Averaged Over 25 Commodities)

Metric	ARIMA	ARIMAX	ANN	ANN-X
MSE	4419.439	4419.345	2285.049	**1457.186**
RMSE	26.474	26.460	18.012	**15.985**
MAE	30.559	30.549	19.840	**18.942**
MAPE (%)	19.260	19.205	12.990	**11.599**

Across all products, ARIMA remained competitive on relatively stable, trend-driven price series. Commodities such as root vegetables and plantains, which exhibited smooth long-term movements and limited seasonal shocks, were forecasted with low mean error using ARIMA. The simplicity and interpretability of the model make it attractive when patterns are linear and data are limited. However, ARIMA's accuracy dropped sharply for highly variable series, particularly where price fluctuations were abrupt or nonlinear.

ARIMAX provided modest but measurable improvements when exogenous inputs carried predictive signal. For crops where trading volume correlated with upcoming price changes, such as selected leafy greens and short-cycle vegetables, ARIMAX lowered error metrics compared to ARIMA; however, gains were inconsistent where exogenous variables carried little signal [3]. Figure 1 provides the forecasts obtained for the various models for limes.

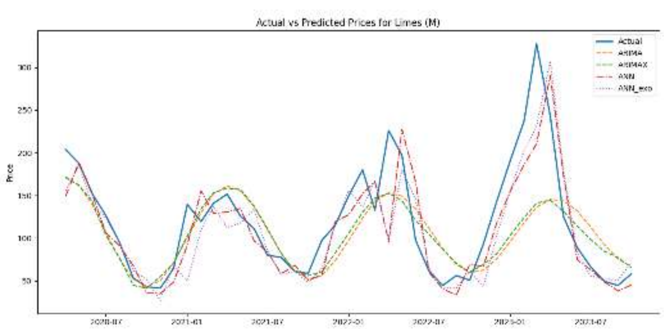

Fig. 1. Limes (M): actual prices vs. one-step-ahead forecasts on the test window. ANN-X tracks peaks and reversals more closely than ARIMA.

Neural network models delivered clear benefits for complex series. ANN models captured nonlinear dependencies and volatility better than both ARIMA-based approaches, and ANN-X extended this advantage by combining neural flexibility with exogenous features [4]–[6]. When meaningful volume data existed and seasonality was strong, ANN-X achieved the lowest overall error measures across all metrics, confirming the value of integrating additional signals into a flexible nonlinear model. We also grouped commodities (Table II) to determine the performance of each group. As quantified in Table IV, adding exogenous inputs yields the largest Test MAPE reductions in Other Vegetables (-3.32 points, -18.63%) and Root Crops (-1.35 points, -22.31%), while effects are negligible for Fruits and Leafy Greens.

TABLE II
COMMODITY GROUPS USED IN ANALYSIS

Group	Commodities
Root crops	Cassava; Dasheen (Local); Sweet Potatoes (Local)
Leafy greens	Lettuce (M); Patchoi; *Amarantus* spp. (Spinach); Cabbage (White)
Other vegetables	Celery; Chive (L); Hot Peppers (40 lb); Cucumber; Melongene (M); Seim beans; Sweet Pepper (M); Pumpkin; Christophene; Caraillie (M); Ginger; Tomato (M)
Fruits	Pineapple; Paw Paw; Plantain (Green); Oranges (L); Limes (M)

Table III presents category-wise MAPE, further illustrating differential benefits by commodity group (metric computation as in [1], [2]; neural setup per [4]–[6]).

TABLE III
CATEGORY-WISE AVERAGE TEST MAPE (%) BY MODEL

Category	ARIMA	ARIMAX	ANN	ANN-X
Fruits	29.30	29.30	**15.66**	16.59
Leafy Greens	11.24	11.24	**7.61**	7.71
Root Crops	9.07	9.35	6.05	**4.70**
Other Vegetables	23.21	23.21	17.82	**14.50**

Seasonality analysis provided further insight: models able to flexibly capture recurring patterns, particularly ANN-X, performed best on commodities with strong cyclical behaviour, whereas ARIMAX only partially benefited when exogenous inputs were aligned with seasonal drivers. Despite their superior accuracy, neural networks required careful hyperparameter tuning, more computation, and greater technical expertise than ARIMA-based methods, and their opacity can be a limitation for stakeholders who value interpretability [4]. Hybrid approaches that couple linear ARIMA structure with nonlinear neural components provide an alternative route to capturing trend and residual nonlinearities [14]. Figure 2 summarizes the distribution of Test MAPE across commodities, highlighting median and dispersion by model.

Overall, these findings reinforce that model choice should be data-driven rather than one-size-fits-all. ARIMA remains a strong baseline for stable, data-limited contexts [1], [2].

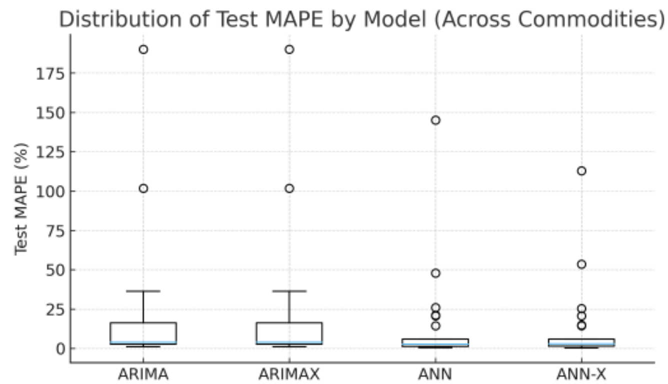

Fig. 2. Distribution of Test MAPE across commodities by model. Boxes show IQR with medians; whiskers denote 1.5×IQR.

ARIMAX is beneficial when reliable external predictors exist [3]. Neural networks excel for volatile, nonlinear markets but require greater technical resources and may be less interpretable to non-specialists [4]–[6].

The comparative errors in Table I and Table III reveal several cross-commodity patterns that help explain when each method excels. The linear structure and well-understood diagnostics of ARIMA make it resilient for series dominated by smooth trends and modest seasonal movement as these conditions minimize the need for nonlinear function classes while preserving interpretability [1], [2]. Introducing exogenous signals via ARIMAX helps primarily when those signals are causally proximate to supply or demand (e.g., trading volume, shipment counts); otherwise, the linear augmentation may add noise without improving generalization [2], [13]. Also, ANN-based models consistently reduce MAPE in the more volatile groups (e.g., fruits and other vegetables), supporting the view that flexible learners absorb nonlinearities and interactions that classical models struggle to capture [11], [12].

Figure 3 visualizes the category-wise MAPE comparison referenced in Table III. It makes clear that ANN and ANN-X dominate in the two most volatile categories (Fruits, Other Vegetables), while the performance gap narrows—and can occasionally reverse—for more regular series (Root Crops, some Leafy Greens). These patterns align with prior evidence that (i) model capacity should scale with series complexity, and (ii) exogenous variables add value when they carry genuine predictive signal with correctly specified lags [2], [5]. In our pipeline, lags for ARIMAX were constrained by cross-correlation analysis, and neural models used lagged prices and volumes with early stopping to manage overfitting, consistent with best practice [4], [10].

Two forms of seasonality further shape these outcomes. The first is calendar seasonality (e.g., harvest cycles) that differs by commodity; the second is logistics-driven quasi-seasonality (e.g., import cycles, holidays) that can be irregular across years. Our seasonality screening used the Kruskal–Wallis test to detect month-to-month differences at α=0.05; several commodities returned significant effects (e.g., ginger,

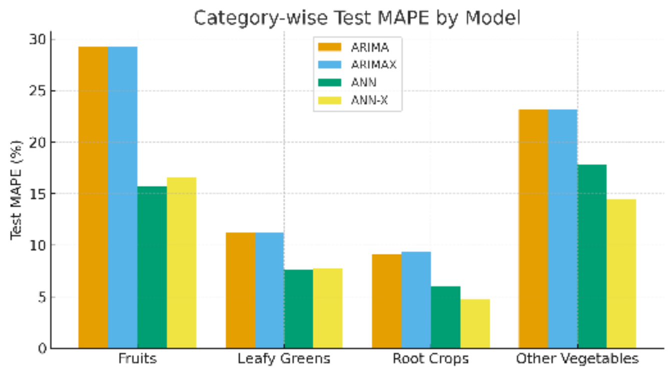

Fig. 3. Category-wise Test MAPE by model (values as in Table III).

celery, hot peppers, patchoi), consistent with the finer-grained analyses in the thesis body [1], [2], [9]. When seasonality was strong and aligned with external signals (volume), ANN-X tended to yield the lowest test errors, while ARIMAX showed mixed gains relative to ARIMA [11].

Beyond aggregate metrics, two diagnostics are informative for practice. (i) *Rolling-origin backtests* (not shown) help verify that gains are not tied to a single split; future work should adopt rolling evaluation to quantify variance of relative performance [2], [15]. (ii) *Error asymmetry* matters for pricing and inventory; quantile or pinball losses could be used to produce probabilistic forecasts that respect stakeholder risk tolerances [15]. From a modelling standpoint, ARIMA's residuals (post Box–Jenkins checks) remained approximately white for many root crops, supporting its suitability in low-volatility settings [1]. Conversely, residual autocorrelation under ARIMAX for some fruits suggested unmodeled nonlinear effects, which ANNs partially absorbed [11], [12].

The incremental value of volume varied across commodities. Where lagged volume correlated with subsequent price changes, ARIMAX and ANN-X improved accuracy; when weakly correlated, ARIMAX often resembled ARIMA, whereas ANN-X preserved much of the ANN gain due to nonlinear interactions [2], [5]. This echoes the literature on feature selection and lag specification in time series with exogenous drivers [5], [13]. Practically, we recommend (a) screening exogenous candidates with cross-correlation and domain judgment, (b) testing a minimal set of lags to limit multicollinearity, and (c) using validation loss/early stopping for neural models [4], [10].

For operations, we suggest reporting both point forecasts and uncertainty bands (e.g., bootstrapped intervals) and monitoring calibration alongside accuracy. Stakeholders (growers, wholesalers, policymakers) benefit from transparent error decomposition: trend/seasonal vs. residual contributions, plus sensitivity to lagged volume. This supports actionable guidance on planting, storage, and import timing, while clarifying the interpretability/accuracy trade-off [12], [15].

IV. Conclusions and Future Work

While the present comparison controls for data splits and evaluation metrics, several extensions could deepen the analysis and practical impact. First, richer exogenous information could be explored beyond volume (for example, weather indices, input and fuel costs, import volumes) to better capture supply and logistics shocks; model variants with targeted feature selection may help avoid overfitting when signals are weak [5]. Second, advanced neural architectures (e.g., recurrent or attention-based sequence models) and automated hyperparameter search may further improve accuracy in highly nonlinear series, complementing the feed-forward designs used here [4]. Third, hybrid strategies that combine linear components for trend/seasonal structure with nonlinear residual learners (e.g., ARIMA + ANN-X) are a promising avenue when transparency and flexibility are both desired [1], [4].

TABLE IV
Effect of Exogenous Inputs by Category (Test MAPE)

Category	**△ AX–A**	**%**	**△ ANNX–ANN**	**%**
Fruits	0.00	+0.00	+0.93	+5.94
Leafy Greens	0.00	+0.00	+0.10	+1.31
Root Crops	+0.28	+3.09	−1.35	−22.31
Other Vegetables	0.00	+0.00	−3.32	−18.63

Note: Negative values indicate improvement (lower MAPE).

Evaluation metrics can also be broadened. Beyond MSE, RMSE, MAE and MAPE, employing scale-free metrics (e.g. MASE) and rolling-origin backtests would yield a fuller picture [2]. Statistical comparisons (e.g. forecast dominance tests) would clarify when observed gains are significant rather than sample-specific. For decision support, probabilistic forecasts (for example, quantile forecasts trained with asymmetric losses) and distributional calibration checks are important when stakeholders manage inventory or price risk. Finally, model interpretability remains critical for adoption: post-hoc explanations for neural forecasts (e.g., sensitivity analyses over lags and exogenous signals) can help bridge the gap between accuracy and actionable insight [4], [5].

This study compared four forecasting approaches, ARIMA, ARIMAX, ANN, and ANN-X, for predicting agricultural market prices across twenty-five commodities. Results show that ARIMA remains a strong baseline for stable and trend-driven series [1], [2], while ARIMAX provides moderate improvements when reliable exogenous signals, such as sales volume, are available [3]. Neural networks demonstrated superior ability to capture nonlinear dynamics and volatility, with ANN-X achieving the lowest error measures overall by combining neural flexibility with external drivers [4]–[6]. Seasonality analysis further revealed that models capable of integrating recurring patterns performed best for commodities with strong cyclical price behaviour.

These findings highlight the importance of matching forecasting technique to data characteristics and practical constraints. ARIMA is suitable for interpretable, low-resource applications, while neural networks, particularly ANN-X, offer improved accuracy where data richness and computational capacity allow [4]. Future research could examine hybrid models that combine the interpretability of classical approaches with the flexibility of neural networks, explore deep learning architectures such as LSTM and GRU, and expand the range of exogenous factors to improve generalizability across markets and regions.

Taken together, the evidence suggests a pragmatic path for organizations that must balance accuracy, transparency, and operating cost. First, establish an interpretable baseline with ARIMA for continuous monitoring and routine reporting; this provides stable forecasts and clear diagnostics at low computational expense. Second, add exogenous drivers only when they are demonstrably predictive and reliably available in production. In our data, volume improved performance in select commodities but did not guarantee gains across the board, underscoring the need for disciplined feature screening and lag validation. Third, when volatility and non-linear dynamics dominate, as in many fruit and short-cycle vegetable markets, ANN models, especially with exogenous inputs, produce a materially lower error. Here, simple, repeatable training protocols (fixed splits, early stopping, light regularization) are as important as the model choice itself for dependable deployment.

From an implementation standpoint, we recommend tiered decision rules. Use the linear baseline for weekly operational updates and price bands communicated to stakeholders, while promoting the ANN-X model to primary use when (i) recent backtests show a persistent accuracy margin over ARIMA, (ii) exogenous signals remain well-calibrated, and (iii) the forecast is high-stakes (inventory buys, contract pricing, or import timing). Operational guardrails—such as drift checks, periodic refits on rolling windows, and fallbacks to the baseline during data anomalies—help maintain service continuity. Finally, complement point forecasts with uncertainty bands and short, commodity-specific notes (e.g., seasonality flags, sensitivity to volume), so end users can translate statistical improvement into actionable, risk-aware decisions.

One avenue for future work is to pair ANN/ANN-X forecasts with post-hoc interpretability tools (e.g., SHAP or LIME) to provide feature-attribution summaries without retraining the models [20], [21]. For time-series use, explanations would be computed on the test window using only training-distribution backgrounds to avoid leakage, attributing each prediction to lagged prices and exogenous volume. Aggregating attributions by commodity and month could expose stable patterns (e.g., volume shocks preceding peaks), while rolling-window explanations would reveal regime shifts. Practically, we would report per-commodity "top contributors," sanity-check sign/lag consistency across splits, and include cautionary notes where attributions are unstable. This interpretability layer could improve stakeholder trust and decision rationale in procurement and pricing workflows, with modest additional compute.

This closing perspective emphasizes method–data fit and governance, enabling measurable gains in forecast quality without sacrificing maintainability or trust.

References

[1] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: Wiley, 2015.

[2] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. OTexts, 2018.

[3] M. Khashei, M. Bijari, and G. A. R. Ardali, "Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks," *Neurocomputing*, vol. 72, no. 4–6, pp. 956–967, 2009.

[4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[5] S. F. Crone and N. Kourentzes, "Feature selection for time series prediction—A combined filter and wrapper approach for neural networks," *Neurocomputing*, vol. 73, no. 10–12, pp. 1923–1936, 2010.

[6] C. Hamzacebi, "Improving artificial neural networks' performance in seasonal time series forecasting," *Information Sciences*, vol. 178, no. 23, pp. 4550–4559, 2008.

[7] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, 1974.

[8] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *J. Amer. Stat. Assoc.*, vol. 74, no. 366, pp. 427–431, 1979.

[9] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *J. Amer. Stat. Assoc.*, vol. 47, no. 260, pp. 583–621, 1952.

[10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2015.

[11] K. Benidis *et al.*, "Deep learning for time series forecasting: Tutorial and literature survey," *ACM Comput. Surveys*, vol. 55, no. 6, pp. 1–36, 2022. [arXiv:2004.10240]

[12] T. Ahmed, I. El Adab, and A. E. Hassanien, "Time series forecasting using artificial neural networks methodologies: A systematic review," *Future Comput. Informatics J.*, vol. 3, no. 2, pp. 334–340, 2018.

[13] H. Choi and H. Varian, "Predicting the present with Google Trends," *Econ. Record*, vol. 88, pp. 2–9, 2012.

[14] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[15] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 competition: Results, findings, conclusion and way forward," *Int. J. Forecast.*, vol. 34, no. 4, pp. 802–808, 2018.

[16] K. Harrykissoon and P. Hosein, "Recursive vs. Direct Forecasting of Crop Prices," in *Proc. IEEE Int. Conf. on Technology Management, Operations and Decisions (ICTMOD)*, 2023, pp. 1–6, doi: 10.1109/ICTMOD59086.2023.10438140.

[17] K. V and P. E. Anusree, "Stock Market Analysis and Forecasting Using Deep Learning," in *Proc. 2025 Int. Conf. on Machine Learning and Autonomous Systems (ICMLAS)*, Mar. 2025, pp. 87–95, doi: 10.1109/icmlas64557.2025.10968208.

[18] S. Hameed, D. Javaheri, and W. Naseem, "Machine Learning-Driven Cryptocurrency Price Forecasting: Advanced Predictive Analytics for Market Trend Modeling," *International Journal of Computer Applications*, vol. 187, no. 31, pp. 44–59, Aug. 2025, doi: 10.5120/ijca2025925547.

[19] H. M. Fadhil and N. Q. Makhool, "Forecasting Cryptocurrency Market Trends with Machine Learning and Deep Learning," *BIO Web of Conferences*, vol. 97, p. 00053, 2024, doi: 10.1051/bioconf/20249700053.

[20] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD)*, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[21] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, pp. 4765–4774.