Received: 06 May 2023 | Revised: 14 August 2023 | Accepted: 23 August 2023 | Published: 14 Sept 2023
Volume 1(2) Pages 1-16
DOI: https://doi.org/10.56279/jicts.v1i2.9011

JICTS
Journal of ICT Systems

Investigation of a Suitable Hybrid Time Series Model for Predicting Clove Price

Abdul-malik Bakar Ali [a,b 1], **Joseph Cosmas Mushi** [b]

[a] *Department of Economic Statistics, Office of the Chief Government Statistician, P. O. Box 2321, Zanzibar, Tanzania*
[b] *Department of Computer Science and Engineering, College of Information and Communication Technologies, University of Dar es Salaam, P. O. Box 33335, Dar es Salaam, Tanzania*

[1]Corresponding author
Email: ibnhassanali@yahoo.com

Funding information

This work is self-sponsored

Abstract

The prediction of clove prices in domestic markets is affected by non-linear factors, including the monopoly market operational environment. Most of the hybrid time series models used in the prediction of crop prices do not consider the monopoly market operational environment as a nonlinear factor. This study investigated a suitable hybrid time series model for predicting clove price under the monopoly market in Zanzibar, Tanzania. The study conducted desk reviews on existing hybrid time series models and realized that ARIMA-ANN, ARIMA-SVM, ARIMAX-ANN, and SARIMA-NARNN are the most common and efficient models that have been employed to predict crop prices. Based on identified models, the study performed several experiments to investigate the accuracy of these models on predicting clove prices with the monopoly market operational environment as a nonlinear factor. The mean absolute percentage error (MAPE) was used as a performance metric. The monthly average prices of cloves from January 2007 to December 2019 were used. Results show that the ARIMA-SVM (MAPE = 0.45%) outperformed the ARIMA-ANN model (MAPE = 0.48%) in predicting clove prices under a monopoly operational market. The study recommended future research to investigate hybrid models for predicting production and planted areas of cloves.

Keywords

Clove Price
Hybrid
Monopoly market
Prediction Accuracy
Time series

1. Introduction

Clove is a major cash crop in Zanzibar and has been the principal source of foreign earnings within the agriculture sector. The clove market in Zanzibar has distinct features that vary nationally and geographically. The market consists mainly of the government control monopoly system, in which a state-owned company (Zanzibar State Trading Corporation) takes control of the production and marketing of cloves [1]. This system influences other nonlinear factors, such as market law, government policy, and development strategy, in which quantification of their effects on the market is challenging [2]. Because of these factors, the future trend of the domestic market price of clove is becoming more uncertain and its typical predictive behaviour is being affected, making it very difficult to predict clove prices accurately [2, 3]. Generally, the existence of a monopoly in the market causes experts to disagree over its influence on the trend of agricultural prices. According to Muflikh et al. [4], the market monopoly stabilizes price trends and maintains levels of self-sufficiency. However, Shebanina and Burkovska [5] believed that the monopoly boosts unjustified price increases, abuses unfair competition, and lowers producers' productivity and overhead costs. Analysis of the effects of nonlinear factors, such as monopoly markets on the price of cloves, is guided by theories, including grounded theory, phenomenology, and ethnography. These theories may not be the best for research that does not need a highly theoretical setting and aims to be true to phenomena and experiences [6]. This is because a controlled monopoly on crop prices makes nonlinear components more complex and varied while threatening the model's capacity to predict accurately [7].

Researchers have proposed time series models as a way to manage complexity and varied nonlinear factors in the prediction of the trend of crop prices [8]. The time series trend consists of linear components and nonlinear components [9]. Linear models using linear functions, such as Autoregressive Integrated Moving Average (ARIMA), have been the most popular in the prediction of linear components in time series data [10]. Later, Seasonal Autoregressive Integrated Moving Average (SARIMA) was introduced to incorporate ARIMA and seasonal variation [11], while the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) was introduced to include the long-run effect of external variables in the ARIMA model [12]. However, these linear models are insufficient to predict nonlinear components and the uncertainty and volatility of time series, leading to several errors [13]. Nonlinear time series models which use nonlinear functions in prediction, such as Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and Artificial Neural Networks (ANN), have been used to handle nonlinear components of time series data [11, 12]. However, the prediction accuracy of these models is challenged by the fact that time series data rarely consist of pure nonlinear patterns alone, sometimes comprising both linear and nonlinear components [14].

To solve these challenges, researchers have proposed combination techniques, known as hybrid, to enhance accuracy by integrating linear and nonlinear models [8, 15, 16]. Mitra and Paul [14] introduced a hybrid Autoregressive Integrated Moving Average Artificial Neural Network (ARIMA-ANN) and a hybrid Autoregressive Integrated Moving Average Generalized Autoregressive Conditionally Heteroscedastic (ARIMA-GARCH) to predict wholesale potato price in the Agra market of India. The ARIMA-ANN and ARIMA-GARCH models showed promising results compared with generalized models. In addition, Yin et al. [16] applied hybrid Attention Long-Short Term Memory Seasonal Trend Decomposition using the Loess (STL-

ATTLSTM) to predict vegetable prices. The STL-ATTLSTM model outperformed single models. Anggraeni et al. [17] applied a hybrid of Autoregressive Moving Integrated Average with exogenous variable Artificial Neural Network (ARIMAX-ANN) and ARIMA-ANN to predict the price of Indonesia's rice. The authors found that they have outstanding performance compared with generalized models. Despite the added value to existing hybrid models, they disregard the monopoly market operational environment as an important nonlinear time series factor that influences the prediction. Therefore, the investigation of a suitable hybrid time series model that can be applied in the prediction of clove price, which is highly affected by the monopoly market operational environment, remains an open-ended challenge worth addressing.

This study set three hypothetical questions to achieve the intended objective. The first question intended to examine which of the existing hybrid time-series models have been efficiently used for the prediction of crop prices in general. The second question examined the performance of efficient models under complex and variably monopoly market operational environments. The third question intended to evaluate the suitability of the model that outperformed others in experiments.

2. Methods

Desk reviews were conducted on various literature to identify the hybrid time series models that have been employed to predict crop prices in general. Then, experiments were performed to investigate suitable hybrid models in the prediction of clove prices under complex and variably nonlinear factors in the operational environment. The suitable hybrid model is the one that demonstrates high performance in terms of prediction accuracy [18, 19]. To measure the accuracy of hybrid models, this study used the mean absolute percentage error (MAPE). This is because MAPE has scale independence and interpretability advantages that can be used to measure the performance of the models regardless of the units of the variables used [20]. The model with MAPE value less than or equal to 10% was considered a suitable model due to the high accuracy attained, 11–20% was considered a good model, 21–50% was considered a rational model, and above 50% was considered an inaccurate model [21].

In general, this study adopted an exploratory mixed (qualitative-quantitative) method research design. A qualitative part of the research design was applied to analyze the data from the Literature search (Narrative Literature Review). The quantitative design has been used to provide a setting for investigating the hybrid time series model's capacity for prediction and assessing the model's accuracy. The secondary data collected from the Office of the Chief Government Statistician Zanzibar (OCGS) has been used to support the training of the models along with the experiments.

2.1 Narrative Review

A narrative review is a qualitative analysis designed to collect quantitative information from studies that employ a variety of methodologies or theoretical conceptualizations without focusing on the statistical significance of the results [22]. This study employed five (5) stages of the narrative review to retrieve the hybrid time series model used to predict crop prices. In the first stage, the study identified the search engines to retrieve relevant study reports in the field of prediction of crop prices. Based on criteria elaborated by Siddaway et al. [22], the study included the following search engines:

Google Scholar, Science Direct, Bielefeld Academic Search Engine (BASE), USDA Department of Agriculture Repository (PubAg), *and Institute of Electrical and Electronics Engineers (IEEE)*

In the second stage, the study sets up the inclusion and exclusion criteria to enhance the search process. The inclusion criteria were as follows: English language publications between January 2012 and July 2022. Articles whose titles included keywords, such as "hybrid", "combined", "mixed", "integrated ", or showing a combination structure of the model (e.g., ARIMA-ANN); articles where the term "hybrid" refers to a combination of prediction models (eg. ARIMA-ANN); and articles whose data represent patterns in crop pricing time series. On the other hand, the exclusion criteria included articles published without abstracts, commentaries, conference abstracts, editorials, opinions, perspectives, peer-reviewed, grey literature and unpublished theses. It also excluded dissertations, books, periodicals, or newspapers that are available offline.

In the third stage, the study dealt with the development of searching code that is relevant to the study under investigation. The study applied a search rule to retrieve the hybrid time series models that have been used to predict crops in general:

"hybrid time series" OR "mixed time series" OR "combined time series" OR "integrated time series" AND (crop forecasting OR crop prediction OR crop modeling).

In the fourth stage, the study recorded the information from each search engine from a large scope (Number of Hits) to a narrow scope (exact match of hybrid time series model to crop prediction). Table 1 shows the results of various search engines as of December 16, 2022, along with pre-inclusion and exclusion criteria.

Table 1: Number of Searched Articles

Name of Search Engine	Number of Hits	Filter[1] 1	Filter[2] 2	Filter[3] 3
Google Scholar	570	13	18	14
Science Direct	113	2	3	1
IEEE	6	2	3	0
BASE	190	56	5	4
PubAg	1	1	1	0
Total	880	74	30	19

The initial database search yielded 880 articles. After applying three filters to these articles, nineteen (19) articles were extracted as perfect matches because of their relevance in the title and abstract to hybrid time series and crop prediction. Five (5) of these articles were judged to be unrelated to the issue of crop prediction, while three (3) articles were found to be duplicates. As tabulated in Table 2, eleven (11) articles were found to be relevant for the identification of the hybrid time series models used to predict crops in general.

Table 2: Relevant Articles with Hybrid Models

Number of articles	Author (Years)
1	Jong et al.,(2020) [23]
2	Devi et al., (2021) [24]
3	Alam et al., (2018) [25]
4	Sujjaviriyasup, (2019) [26]
5	Naveena & Subedar, (2017) [27]
6	Rathod et al., (2017) [28]
7	Anggraeni et al., (2019) [17]
8	Alam, et al., (2018) [12]
9	Neog et al., (2022) [8]
10	Sanjeev & Bhardwaj, (2022) [29]
11	Chi, (2021) [30]

In the fifth stage, the study created a dataset (Appendix A) for hybrid time series models to

[1] Title Matching (hybrid time series not in crops)
[2] Title Matching (hybrid time series in crops)
[3] Abstract Matching (hybrid time series in crops)

predict crops. The dataset was created based on seven variables, namely "Hybrid model", "Type of crops", "Country", "Performance", "Author (year)", "Dataset size", and "Prediction Variable". Table 3 gives a brief description of these variables.

Table 3: Variable Description of the Dataset

Variable Name	Value	Description
Hybrid Model	String	Type of hybrid model applied
Type of Crop	String	Type of crop or crops used
Country	String	Location/Area where the experiment took place
Performance	Float	Value of MAPE
Author (year)	String (Numerical)	Author(s) of study and respective year of publication
Dataset size	Numerical	Size of Dataset used
Prediction Variable	String	Time series variable for prediction

2.2 Secondary Data

The secondary data used to support the experiments of this study came from the Statistical Abstract Reports for the periods from January 2007 to December 2019 published by the Office of the Chief Government Statistician of Zanzibar (OCGS). The data included in the experiments were the monthly average price per kilogram of clove (TZS/Kg), Minimum Temperature in Celsius (°C), Maximum Temperature in Celsius (°C), Relative Humidity (percent), Sunshine (hours), and Rainfall (Millimeters).

As elaborated in Appendix B, the monthly average price per kilogram of clove (TZS/Kg) was obtained by dividing the Value of Cloves (TZS) by the Quantity of clove (Kg). The exogenous variables (Minimum Temperature in Celsius, Maximum Temperature in Celsius, Relative Humidity, Sunshine, and Rainfall) were used to support the prediction of the monthly average price of cloves.

2.3 Hybrid Time Series Procedure

The approach used in the study considers time series y_t as a function of linear and nonlinear components [10]. Firstly, the linear model was used to capture the linearity and the nonlinear model to capture the uncertainties (seasonality of clove harvesting, government policies, climatic changes, market law, clove development strategy, and monopoly market system) in the form of residuals. To predict the nonlinear components of time series data, the residuals are then used as input for other nonlinear time-series models (Figure 1).

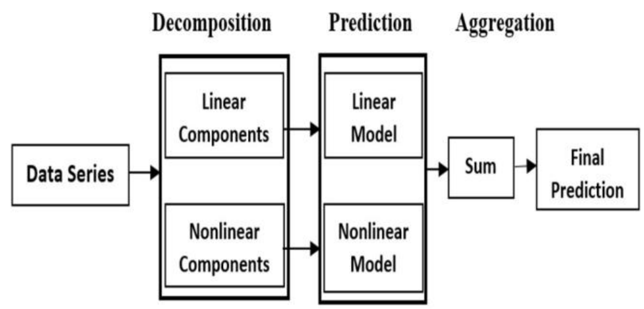

Figure 1: Hybrid Time-series Procedures **[10]**

2.4 The Performance Evaluation Metrics

This study used the Mean Absolute Percentage Error (MAPE) to evaluate hybrid time series models in the prediction of clove prices. MAPE is the sum of the absolute differences between the actual and predicted values divided by the number of observations, mathematically defined as

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100 \quad (1).$$

The y_t is the actual price of cloves, $\hat{y}_t$ is the predicted price of cloves, n is the number of observations, and t is the period in a month.

3. Experiments

The chosen hybrid time-series models for clove price prediction were tested using experimental analysis. During experiments, data were split into 80% for training and 20% for testing. The Python and Sigma Excel tools were used to carry out experiments. The MinMaxScaler from the Sklean library in Python was used to transform the clove price data before being splitted into training and test sets.

3.1 Hybrid ARIMA-ANN Model

In this experiment, the ARIMA (0,1,1) model was experimented to predict the price of cloves, and an additional series (called residual series) was created by deducting the ARIMA prediction output from the actual price (Appendix B). Then, the ANN model received the residual dataset as input. The study used Python programming's optimizer parameter search to find the ANN model's most optimal parameter for training. The ANN configuration included an optimizer (Adam), two hidden layers, 25 neurons, a batch size of 32, 500 epochs, 0.2 dropouts, activation (ReLU), and a loss function (linear). Finally, the hybrid output was created by adding the results of the ANN model with ARIMA outputs.

3.2 Hybrid ARIMA-SVM Model

In this experiment, the hybrid time-series model was formed by predicting the clove price using the ARIMA (0,1,1) model and then creating a new series (residual series) by subtracting the ARIMA prediction output from the actual price (Appendix B). The SVM model was fed with the residual dataset as an input. To find the most ideal parameter for training the SVM model, the study used Python's optimizer parameter search function. The SVM configuration used throughout the experiments with the model was Regularization factor (C) (6388.5), Kernel (rbf), Gamma (0.001), and epsilon (0.001). Finally, the hybrid output was created by adding the results of the SVM model with ARIMA outputs.

3.3 Hybrid ARIMAX-ANN Models

The experiments for hybrid ARIMAX-ANN failed because the inclusion of exogenous variables (X) into the ARIMA model increased the Akaike Information Criteria (AIC) values while all the exogenous variables were found insignificant during analysis. This means that the exogenous variables (X) do not increase the regression power to the ARIMA model in the prediction of clove prices. After grid search, the summary results of the ARIMAX model were ARIMA (0, 1, 1) with (p, d, q) parameters, respectively, five (5) predictors, seasonal frequency = 1, and AIC = 2500.82. The Seasonal frequency = 1 indicates the absence of seasonality in the trend of clove prices. The experiment also involved testing the effect of exogenous variables (X) on the model. The parameter estimates test was carried out to measure the degree of confidence that can be accredited to the estimate. The null hypothesis being tested is that the coefficient (b_i) is statistically zero against the alternative that it is different from zero.

$$H_0:\ b_i = 0 \quad \text{against } H_a:\ b_i \neq 0$$

Failure to reject the null hypothesis implies that the independent (explanatory) variable X_i to attached coefficient b_i has no impact on the dependent variable, making it a useless variable in the model.

Table 4 indicates that only MA_1 has p-values < 0.05 (0.000), while all exogenous variables (predictors) have p-values > 0.05, which indicates the insignificance of these variables in the ARIMAX model. Therefore, the exogenous variables (X) needed to establish the ARIMAX model have failed the test, and the remaining parameters only allow us to build the ARIMA model. In addition to that, the AIC value of ARIMAX was much higher (AIC = 2500.82) compared with the ARIMA (AIC = 906.20) model. This indicates that the inclusion of exogenous variables (X) decreases the accuracy of the model in the prediction of clove prices, instead of adding explanatory value to the regression [12].

Table 4: ARIMAX Parameter Estimates

Term	Coefficient	T-test	p-value
MA_1	-0.301	3.815	0.000
Minimum Temperature	77.843	0.681	0.497
Maximum Temperature	-27.357	0.312	0.755
Rainfall	0.158	0.371	0.712
Relative Humidity	-22.331	1.252	0.213
Sunshine	-24.716	0.398	0.691

3.4 Hybrid SARIMA-NARNN Models

The experiment for the SARIMA-NARNN model failed because the clove price did not show seasonal (S) variation, which is needed to establish the SARIMA model. Table 5 indicates that, after grid search, the summary results of the SARIMA model were ARIMA (0, 1, 1) with corresponding (p, d, q) parameters, seasonal frequency = 1, and AIC = 906.21.

Table 5: SARIMA Parameter Summary

SARIMA Model	Value
Autoregressive (AR) Order (p)	0
Integrated Order (d)	1
Moving Average (MA) Order (q)	1
Seasonal Frequency (S)	1

The Seasonal frequency = 1 indicates the absence of seasonality in the trend of clove prices, and that's why the SARIMA-NARNN model failed.

4. Results and Discussions

4.1 Hybrid Models from Literature

The narrative review has revealed eleven (11) articles that have used the hybrid models to predict crops in general. Out of these articles, the study found seven (7) hybrid models, namely ARIMA-ANN, ARIMA-SVM, ARIMA-TDNN, ARIMA-NLSVR, ARIMAX-SVM, ARIMAX-ANN, and SARIMA-NARNN. Table 6 indicates that five (5) hybrid models, out of the seven models, have been used to predict the production of crops. Examining the data from the narrative review, the study found that ARIMA-ANN and ARIMA-SVM have been mostly applied in the prediction of the production of crops. However, in terms of performance, the ARIMAX-SVM is the most accurate model with a MAPE of 0.37% followed by the ARIMAX-ANN with a MAPE of 1.11% in the prediction of production of crops.

Table 6: Hybrid Models for Prediction of Production of Crops

Hybrid Model	Number of Occurrence	Minimum MAPE (%)
ARIMA-ANN	5	2.16
ARIMA-SVM	5	1.48
ARIMAX-ANN	2	1.11
ARIMAX-SVM	2	0.37
ARIMA-NLSVR	1	6.78
ARIMA-TDNN	1	12.49

As indicated in Table 7, The MAPE analysis on the four (4) hybrid models to predict the prices of crops from the narrative review revealed that although ARIMA-ANN is the most used model for the prediction of the price of crops, the ARIMAX-ANN is the most accurate hybrid model in prediction followed by ARIMA-ANN.

Table 7: Hybrid Models for Prediction of Price of Crops

Hybrid Model	Number of Occurrence	Minimum MAPE (%)
ARIMA-ANN	2	1.21
ARIMA-SVM	1	6.31
ARIMAX-ANN	1	0.23
SARIMA-NARNN	1	3.72

4.2 Performance of the Hybrid Time Series Models in Clove Prices Prediction

In general, all four hybrid models in Table 7 have shown outstanding performance in the prediction of the price of crops. The study investigated the suitability of these hybrid models in the prediction of clove price under a monopoly operational environment. The summary statistics from Table 8 show that the average price of TZS 13,959.85 per Kg was predicted by the ARIMA-ANN model, the minimum average price of cloves was TZS.13,716.68 per Kg and the maximum average price of cloves was TZS.14,205.44 per Kg. While the average price predicted by the ARIMA-SVM model was TZS.13,957.93 per Kg, the minimum average price of cloves was TZS.13,716.94 per kg and the maximum average price was TZS.14,205.99 per kg.

Overall statistics indicated that the ARIMA-SVM performed better than ARIMA-ANN in the prediction of clove price (Table 7). However, this analysis used the assumption that the residual from the linear component solely contained nonlinear relationships and that there was an additive relationship between linear and nonlinear components [10]. This does not necessarily mean that any significant nonlinear patterns will be present in the linear component's residuals.

Table 8: ARIMA-ANN and ARIMA-SVM Model Summary

Statistics	Value in TZS	
	ARIMA-ANN	ARIMA-SVM
Mean (Average Price)	13959.85	13959.80
Range	488.76	488.06
Minimum	13716.68	13716.94
Maximum	14205.44	14204.99

Clove price fluctuates over time according to both ARIMA-ANN and ARIMA-SVM models (Figures 2 and 3). Both ARIMA-ANN and ARIMA-SVM models show that the price of clove fluctuates significantly between the periods of 2 to 10 and 20 to 24.

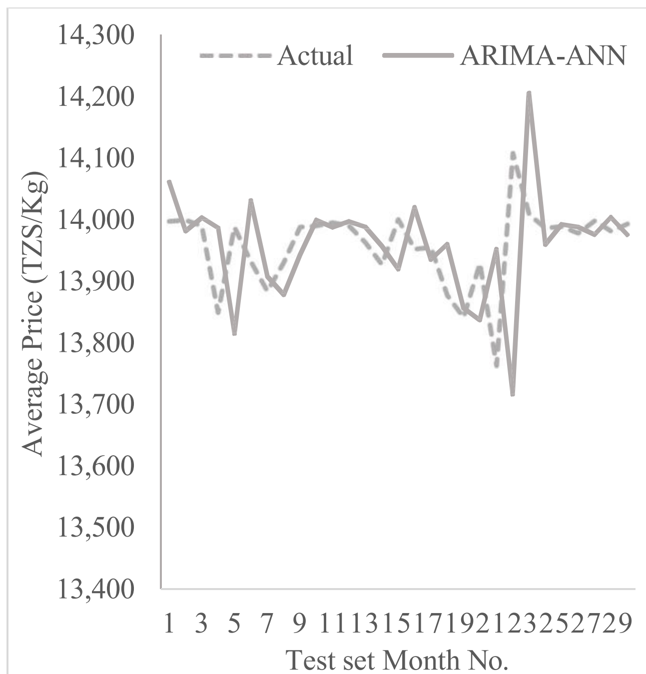

Figure 2: Clove Price Prediction from the ARIMA-ANN Model

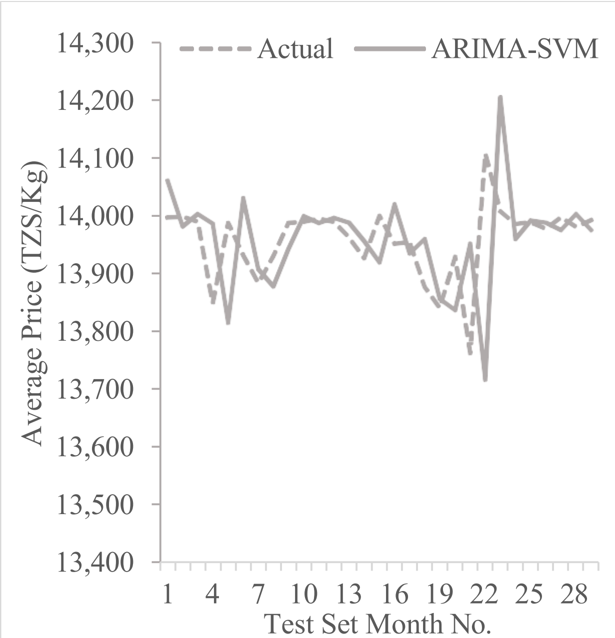

Figure 3: Clove Price Prediction from the ARIMA-SVM Model

4.3 Hybrid Time Series Model for Clove Prices Prediction

Results from experiments show that the prediction accuracy of the ARIMA-SVM model is 99.55% (MAPE=0.45%) compared with 99.52% (MAPE = 0.48%) of ARIMA-ANN model in the prediction of clove prices (Table 8). This means that the average percentage between the price predicted by ARIMA-SVM and ARIMA-ANN models and the actual price of cloves differ by 0.45% and 0.48%, respectively. The trend of clove prices is correctly predicted by the ARIMA-SVM model when compared with the ARIMA-ANN model under a monopoly market operational environment. Based on Lewis [21], both ARIMA-ANN and ARIMA-SVM models were suitable for the prediction of clove prices under a monopoly environment. Although the hybrid ARIMA-SVM is more suitable to handle the nonlinear factors affecting clove prices during prediction, the two hybrid models (ARIMA-ANN and ARIMA-SVM) have exhibited outstanding performance in the prediction of clove price compared to the price of other crops. For instance, Anggraeni et al. [17] applied ARIMA-ANN to predict the price of Indonesia's rice; the result showed a MAPE of 1.21%. In addition, Naveena and Subedar [27] applied the ARIMA-ANN model to predict the price of washed coffee in India; the model exhibited a MAPE of 1.49%. Likewise, the ARIMA-SVM model applied by Jong et al. [23] to predict the price of Rubber achieved a MAPE of 6.31%. These findings exhibited the fact that, even though hybrid models have extensive prediction accuracy, none of the hybrid models provides the best accuracy in all crops over time [31]. Therefore, the investigation of hybrid time series models to predict the price of cloves under a monopoly market operational environment was a necessary exercise.

The results have also shown that the experiments to investigate the performance of ARIMAX-ANN and SARIMA-NARNN in the prediction of clove price were not successful. The exogenous variables (X) used to create the ARIMAX model were found insignificant. In addition, the inclusion of exogenous variables in the ARIMA model increased the Akaike Information Criteria (AIC) values and decreased the explanatory power of models [8]. In the case of SARIMA-NARNN, the seasonal variation (S) parameter which was needed to establish the SARIMA model was not observed in clove price data. The clove prices trend showed Autoregressive (AR), Integrated (I), and Moving Average (MA) of values 0, 1, and 1, respectively. These parameters only allow the ARIMA (0,1,1) model. However, the ARIMAX-ANN showed outstanding results (MAPE = 0.23%) in the prediction of the price of rice (Table 7).

Table 8: Accuracy Measures of Hybrid Models in the Prediction of Clove Price

SN	The average price of cloves (TZS/Kg)	Prediction		MAPE (%)	
		ARIMA-ANN	ARIMA-SVM	ARIMA-ANN	ARIMA-SVM
1	13997	14060.99	14060.68	0.00457	0.00455
2	13999	13981.40	13981.46	0.00126	0.00125
3	13990	14003.48	14003.40	0.00096	0.00096
4	13849	13986.82	13886.80	0.00995	0.00273
5	13988	13815.14	13815.35	0.01236	0.01234
6	13932	14030.93	14030.66	0.00710	0.00708
7	13884	13907.77	13907.90	0.00171	0.00172
8	13930	13878.23	13878.22	0.00372	0.00372
9	13988	13942.96	13942.82	0.00322	0.00323
10	13990	13999.33	13999.20	0.00067	0.00066
11	13995	13987.86	13987.83	0.00051	0.00051
12	13990	13996.91	13996.85	0.00049	0.00049
13	13963	13988.44	13988.40	0.00182	0.00182
14	13927	13956.86	13956.85	0.00214	0.00214
15	14000	13919.76	13919.76	0.00573	0.00573
16	13952	14020.02	14019.83	0.00488	0.00486
17	13955	13935.38	13935.45	0.00141	0.00140
18	13877	13959.98	13959.89	0.00598	0.00597
19	13841	13856.66	13856.76	0.00113	0.00114
20	13929	13837.24	13837.22	0.00659	0.00659
21	13763	13951.88	13951.68	0.01372	0.01371
22	14108	13716.68	13716.94	0.02774	0.02772
23	14008	14205.44	14204.99	0.01409	0.01406
24	13986	13959.64	13959.90	0.00188	0.00187
25	13989	13992.61	13992.50	0.00026	0.00025
26	13978	13988.27	13988.23	0.00073	0.00073
27	13998	13975.61	13975.58	0.00160	0.00160
28	13981	14003.69	14003.60	0.00162	0.00162
29	13993	13975.55	13975.54	0.00125	0.00125
Total Error				0.13910	0.13171
MAPE (%)				0.47966	0.45415

5. Conclusion

This study investigated a suitable hybrid time series model for the prediction of clove price under a nonlinear factor of a monopoly market environment, which has never been used before. To achieve this objective, the study developed an approach that involves the narrative review as elaborated in the way of science [22] and experimentation using a dataset collected from the Office of the Chief Government Statistician of Zanzibar. Our approach is more detailed, simple to follow and, to the best of our knowledge, the first narrative review on the hybrid time series models in crop prediction. The narrative review revealed seven (7) hybrid models that have commonly been used to predict crops in general. Four (4) among these seven models (ARIMA-ANN, ARIMA-SVM, ARIMAX-ANN, and SARIMA-NARNN) have been applied in the prediction of the prices of crops. In the experimentation methodology, the study developed the dataset containing hybrid time series models that have been used to predict crops with respective seven (7) variables (Hybrid model, Type of crops, Country, Performance, Author (year), Dataset size, and Prediction variable). This dataset serves as a benchmark for further studies on the hybrid time series models on crop prediction. To the best of our knowledge, a dataset containing hybrid time series models to predict crops has never been generated before.

The experiments were done to investigate which of the four models is suitable for the prediction of clove prices under a monopoly operational environment. The study finally found that ARIMA-SVM with MAPE of 0.45% is the most suitable hybrid time series model to predict clove prices. This model helps policymakers to determine the future trend of the price of cloves for sustainable growth of the clove industry in Zanzibar. Accurate prediction of the prices of cloves helps the government regulate post-harvest storage and management of the production of cloves to stabilise price volatility throughout the year and helps farmers decide when to sell cloves for the most profit. The model results also align with the government strategic direction 2.2 of Zanzibar Vision 2050, which aims to orient scientific and socio-economic research to address the development needs of the nation. Future studies should look into the investigation of the hybrid time series models on production and planted areas of cloves under the monopoly.

CONTRIBUTIONS OF CO-AUTHORS

Abdul-malik Bakar Ali		Conceptualized the idea, conducted experiments, data analysis, and report writing.
Joseph Cosmas Mushi	[ORCID: ORCID: 0000-0003-4095-0433]	Reviewed the paper and provided technical support on procedures and resources

APPENDIX

Appendix A: Dataset from Narrative Search

No	Hybrid Model	Type of Crop	Country	Performance (MAPE %)	Author (Year)	Dataset_size	Prediction variable
1	ARIMA-SVM	Rubber	Bulk Latex	6.31	Jong et al., (2018)	488	Price
2	ARIMA-ANN	Wheat	Haryana	4.02	Devi et al., (2021)	38	Production
3	ARIMA-ANN	Rice	Aligarh district, India	3.59	Alam et al., (2018)	456	Production
4	ARIMA-ANN	Washed Coffee	India	1.49	Naveena & Subedar, (2017)	230	Price
5	ARIMA-ANN	Rice	Indonesia	1.21	Anggraeni et al., (2019)	108	Price
6	ARIMA-SVM	Rice	Aligarh district, India	2.07	Alam et al., (2018)	38	Production
7	ARIMA-SVM	Rice	Thailand	13.11	Sujjaviriyasup, (2019)	2559	Production
8	ARIMA-SVM	Rice	India	17.47	Sujjaviriyasup, (2019)	2559	Production
9	ARIMA-SVM	Rice	Viet Nam	23.25	Sujjaviriyasup, (2019)	2559	Production
10	ARIMA-TDNN	Banana	Karnataka State, India	12.49	Rathod et al., (2017)	60	Production
11	ARIMA-NLSVR	Banana	Karnataka State, India	6.78	Rathod et al., (2017)	60	Production
12	ARIMAX-ANN	Rice	Indonesia	0.23	Anggraeni et al., (2019)	108	Price
13	ARIMAX-ANN	Rice	Aligarh district, India	1.11	Alam et al., (2018)	38	Production
14	ARIMAX-SVM	Rice	Aligarh district, India	0.37	Alam et al., (2018)	38	Production
15	SARIMA-NARNN	Soybeans	Global market	3.72	Chi, (2021)	361	Price
16	ARIMA-ANN	Sugarcane	Panipat district	5.46	Sanjeev & Bhardwaj, (2022)	49	Production
17	ARIMA-ANN	Sugarcane	Yamunanagar district	2.16	Sanjeev & Bhardwaj, (2022)	49	Production
18	ARIMAX-ANN	Rice	Assam	2.82	Neog et al., (2022)	37	Production
19	ARIMAX-SVM	Rice	Assam	0.94	Neog et al., (2022)	37	Production
20	ARIMA-ANN	Rice	Meerut District	2.94	Alam et al., (2018)	456	Production
21	ARIMA-SVM	Rice	Meerut District	1.48	Alam et al., (2018)	38	Production

Appendix B: Clove Prices Dataset

Month Number	Price of cloves	ARIMA(0,1,1) Residuals	Month Number	Price of cloves	ARIMA(0,1,1) Residuals	Month Number	Price of cloves	ARIMA(0,1,1) Residuals
1	2284	N/A	53	4987	0.10192	105	13993	0.18752
2	2730	8.91379	54	5058	0.9768	106	14001	0.02154
3	2746	-1.7603	55	11964	76.2809	107	13990	-0.0983
4	2786	1.19161	56	14805	5.84859	108	14022	0.29454
5	2786	-0.2927	57	15003	0.18482	109	14023	-0.0639
6	2709	-1.3971	58	14978	-0.2496	110	13958	-0.5338
7	2255	-8.7791	59	14986	0.12669	111	13979	0.30885
8	2774	12.521	60	15201	1.71891	112	13959	-0.2451
9	2769	-3.1715	61	14989	-2.1479	113	13973	0.17869
10	2792	1.21544	62	14980	0.45423	114	13981	0.02376
11	2798	-0.1852	63	14989	-0.0381	115	13997	0.12944
12	2799	0.0644	64	14994	0.05019	116	14010	0.07805
13	2854	1.01871	65	14992	-0.0287	117	13977	-0.2981
14	3389	9.33425	66	14992	0.00704	118	14127	1.33865
15	3492	-0.5374	67	14999	0.05543	119	13996	-1.4336
16	3499	0.25045	68	9993	-45.024	120	14190	1.98645
17	3492	-0.1799	69	9992	11.0529	121	14194	-0.4545
18	3498	0.1457	70	11656	13.2901	122	13937	-2.0553
19	3498	-0.0358	71	12499	4.40644	123	13910	0.27619
20	3499	0.0257	72	12500	-1.0738	124	13882	-0.3054
21	3500	0.01059	73	12499	0.25489	125	14000	1.07443
22	3000	-8.7797	74	12496	-0.0895	126	13813	-1.8497
23	2999	2.13899	75	12495	0.01304	127	14001	2.0487
24	3000	-0.5073	76	12496	0.00574	128	13997	-0.5372
25	2499	-9.4399	77	12487	-0.0819	129	13999	0.1489
26	2486	2.05906	78	12491	0.05593	130	13990	-0.1127
27	2497	-0.2856	79	13993	13.044	131	13849	-1.1674
28	2495	0.03013	80	13983	-3.2896	132	13988	1.46505
29	2495	-0.0074	81	13997	0.92664	133	13932	-0.8339
30	2498	0.06186	82	13998	-0.2192	134	13884	-0.2021
31	3000	9.56932	83	13981	-0.0899	135	13930	0.43973
32	3000	-2.3513	84	13992	0.11509	136	13988	0.38286
33	2999	0.55947	85	13982	-0.1128	137	13990	-0.0772

Month Number	Price of cloves	ARIMA(0,1,1) Residuals	Month Number	Price of cloves	ARIMA(0,1,1) Residuals	Month Number	Price of cloves	ARIMA(0,1,1) Residuals
34	3000	-0.1192	86	13957	-0.1838	138	13995	0.06123
35	3021	0.41203	87	14554	5.04559	139	13990	-0.0573
36	3282	4.549	88	14936	1.90619	140	13963	-0.2143
37	3272	-1.2924	89	13921	-8.9197	141	13927	-0.2522
38	3286	0.56205	90	13992	2.79263	142	14000	0.67974
39	3273	-0.3651	91	13992	-0.6862	143	13952	-0.573
40	3277	0.15961	92	13997	0.21087	144	13955	0.1662
41	3248	-0.5469	93	13998	-0.0434	145	13877	-0.702
42	3293	0.92127	94	13994	-0.0232	146	13841	-0.1333
43	3487	3.10595	95	13992	-0.0112	147	13929	0.77956
44	3493	-0.6616	96	13978	-0.1156	148	13763	-1.6023
45	3496	0.21331	97	13933	-0.3525	149	14108	3.31627
46	3498	-0.0186	98	13937	0.1205	150	14008	-1.6582
47	3770	4.51744	99	14014	0.62173	151	13986	0.22149
48	4004	2.64371	100	14036	0.033	152	13989	-0.0291
49	4989	14.0618	101	13973	-0.5405	153	13978	-0.0859
50	4993	-3.3985	102	13857	-0.8506	154	13998	0.19021
51	4996	0.87749	103	13993	1.3615	155	13981	-0.1905
52	4986	-0.3572	104	13983	-0.4191	156	13993	0.14827

REFERENCES

[1] S. A. Bakar, Contribution of the Zanzibar State Trading Corporation (ZSTC) to clove production in Zanzibar, Doctoral dissertation, The University of Dodoma, 2015, doi: http://hdl.handle.net/20.500.12661/677.

[2] S. J. Kabote and J. Tunguhole, Determinants of clove exports in Zanzibar: Implications for policy, *International Journal of Business and Economic Studies*, **4(2)**: p. 127–147, 2022.

[3] I. S. Mohd, The prospects of Muzaraah financing and supply chain model in Zanzibar clove industry, Doctoral dissertation, Institute of Islamic Banking and Finance, 2018, doi: http://studentrepo.iium.edu.my/jspui/handle/123456789/2217.

[4] Y. N. Muflikh, C. Smith, and A. A. Aziz, A systematic review of the contribution of system dynamics to value chain analysis in agricultural development, *Agricultural Systems*, **189**: p. 103044, 2021.

[5] O. Shebanina and A. Burkovska, Price regulation of agrarian markets in the context of food security in Ukraine, *SHS Web Conf.*, **65**: p. 03002, 2019, doi: 10.1051/shsconf/20196503002.

[6] L. Doyle, C. McCabe, B. Keogh, A. Brady, and M. McCann, An overview of the qualitative descriptive design within nursing research, *Journal of Research in Nursing*, **25(5)**: p. 443–455, 2020, doi: 10.1177/1744987119880234.

[7] R. Sari, K. Kusrini, T. Hidayat, and T. Orphanoudakis, Improved LSTM Method of Predicting Cryptocurrency Price Using Short-Term Data, *Indonesian Journal of Computing and Cybernetics Systems*, **17(1)**: p. 33-34, 2023, doi: 10.22146/ijccs.80776.

[8] B. Neog, B. Gogoi, and A. N. Patowary, Development of Hybrid Time Series Models for Forecasting Autumn Rice Using ARIMAX- ANN and ARIMAX-SVM, *Annals of Forest Research*, **65(1)**: p. 9119-9133, 2022.

[9] D. Borges and M. C. Nascimento, COVID-19 ICU demand forecasting: A two-stage Prophet-LSTM approach, *Applied Soft Computing*, **125**: p. 109-181, 2022.

[10] G. P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing*, **50**: p. 159–175, 2003.

[11] O. Adedire, Model for Retail Price Forecast of Thailand Rice In Plateau State Nigeria, *Journal of Multidisciplinary Modeling and Optimization*, **4(1)**: p. 1–10, 2021.

[12] W. Alam, M. Ray, R. R. Kumar, K. Sinha, S. Rathod, and K. N. Singh, Improved ARIMAX modal based on ANN and SVM approaches for forecasting rice yield using weather variables, *Indian J. Agric. Sci*, **88(12)**: p. 1909–1913, 2018.

[13] Abhishekh, S. S. Gautam, and S. R. Singh, A new type 2 fuzzy time series forecasting model based on three-factors fuzzy logical relationships, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **27(2)**: p. 251–276, 2019.

[14] D. Mitra and R. K. Paul, Hybrid time-series models for forecasting agricultural commodity prices, *Model Assisted Statistics and Applications*, **12(3)**: p. 255–264, 2017.

[15] M. W. Alam, K. Sinha, R. K. Ranjan, M. Roy, S. Rathod, and K. N. Singh, Development of Hybrid Time Series Model using Machine Learning Techniques for Forecasting Crop Yield with Covariates, Indian Agricultural Statistics Research Institute, ICAR-IASRI, New Delhi, Project Report AGENIASRISI: 201502000057, 2018, doi: http://krishi.icar.gov.in/jspui/handle/123456789/6472.

[16] H. Yin, D. Jin, Y. H. Gu, C. J. Park, S. K. Han, and S. J. Yoo, STL-ATTLSTM: vegetable price forecasting using STL and attention mechanism-based LSTM, *Agriculture*, **10(12)**: p. 612, 2020.

[17] W. Anggraeni, F. Mahananto, A. Q. Sari, Z. Zaini, K. B. Andri, and Sumaryanto, Forecasting the Price of Indonesia's Rice Using Hybrid Artificial Neural Network and Autoregressive Integrated Moving Average (Hybrid NNs-ARIMAX) with Exogenous Variables, *Procedia Computer Science*, **161**: p. 677–686, 2019, doi: 10.1016/j.procs.2019.11.171.

[18] S. M. R. Haj Seyed Javady, R. Heydari, and F. Abbasi, Forecasting the future price of pistachio in agricultural commodity exchange using of the hybrid model of Wavelet-XGBoost, *Agricultural Economics*, **17(1)**: p. 79–108, 2023, doi: 10.22034/iaes.2023.1974208.1959.

[19] Z. Jinliang, T. Zhongfu, and W. Yiming, An adaptive hybrid model for short term electricity price forecasting, *Applied Energy*, **258**: p. 114087, 2020, doi: 10.1016/j.apenergy.2019.114087.
[20] S. Kim and H. Kim, A new metric of absolute percentage error for intermittent demand forecasts, *International Journal of Forecasting*, **32(3)**: p. 669–679, 2016.
[21] C. D. Lewis, *Industrial and business forecasting methods: A Practical Guide to Exponential Smoothing and Curve Fitting*. Landon, UK: Butter-worth-Heinemann, 1982.
[22] A. P. Siddaway, A. M. Wood, and L. V. Hedges, How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses, *Annual Review of Psychology*, **70(1)**: p. 747–770, 2019, doi: 10.1146/annurev-psych-010418-102803.
[23] L. J. Jong, S. Ismail, A. Mustapha, M. H. A. Wahab, and S. Z. S. Idrus, The Combination of Autoregressive Integrated Moving Average (ARIMA) and Support Vector Machines (SVM) for Daily Rubber Price Forecasting, *IOP Conf. Ser.: Mater. Sci. Eng.*, **917(1)**: p. 012–044, 2020, doi: 10.1088/1757-899X/917/1/012044.
[24] M. Devi, J. Kumar, D. P. Malik, and P. Mishra, Forecasting of wheat production in Haryana using hybrid time series model, *Journal of Agriculture and Food Research*, **5**: p. 100–175, 2021.
[25] W. Alam *et al.*, Hybrid linear time series approach for long term forecasting of crop yield, *Indian J. Agric. Sci*, **88**: p. 1275–1279, 2018.
[26] T. Sujjaviriyasup, Hybrid Model of Linear and Nonlinear Models for Forecasting Annual Molumes of International Rice Exports, *Burapha Science Journal*, **24(2)**: p. 517–531, 2019.
[27] K. Naveena and S. Subedar, Hybrid time series modelling for forecasting the price of washed coffee (Arabica Plantation Coffee) in India, *International Journal of Agriculture Sciences,* **9(10)**: p. 4004–4007, 2017.
[28] S. Rathod, G. C. Mishra, and K. N. Singh, Hybrid time series models for forecasting banana production in Karnataka State, India, *Journal of the Indian Society of Agricultural Statistics*, **71(3)**: p. 193–200, 2017.
[29] Sanjeev and N. Bhardwaj, Time Series Prediction Using Hybrid ARIMA-ANN Models for Sugarcane, **34(23)**: p. 772–782, 2022, doi: 10.9734/ijpss/2022/v34i232488.
[30] Y. N. Chi, Time Series Forecasting of Global Price of Soybeans using a Hybrid SARIMA and NARNN Model: Time Series Forecasting of Global Price of Soybeans, *Data Science: Journal of Computing and Applied Informatics*, **5(2)**: p. 85–101, 2021.
[31] S. K. Purohit, S. Panigrahi, P. K. Sethy, and S. K. Behera, Time Series Forecasting of Price of Agricultural Products Using Hybrid Methods, *Applied Artificial Intelligence*, **35(15)**: p. 1388–1406, 2021, doi: 10.1080/08839514.2021.1981659.