Proceedings of the Indian National Science Academy (2026) 92:1204–1225
https://doi.org/10.1007/s43538-025-00482-3

RESEARCH PAPER

FuzzyML: fuzzy based machine learning models for predicting agricultural prices

Anita Sarkar[1] · Ranjit Kumar Paul[2] · Md Yeasin[2] · Ankit Kumar Singh[1] · A. K. Paul[2]

Received: 7 April 2025 / Accepted: 11 June 2025 / Published online: 30 June 2025
© Indian National Science Academy 2025

Abstract
Prediction of agricultural commodity prices is challenging due to the intricate interaction of multiple influencing factors which poses limitations for traditional statistical tools. The machine learning (ML) techniques e.g. artificial neural network (ANN), random forest (RF), support vector regression (SVR) are widely used in the domain of time series forecasting to capture the nonlinearity. On the other hand, Fuzzy technique has proven its potential in field of forecasting for ambiguous and vague data. Traditional stochastic and ML models often fail to capture the noise, uncertainty, incompleteness, non-linearity, and complexity presence in data and thus, cannot give accurate predictions. To improve forecasting accuracy by handling uncertainty and imprecise information in real-world datasets, this study proposes a fuzzy-based machine learning model (FuzzyML). It combines fuzzy logic with machine learning techniques. The proposed models i.e. Fuzzy-ANN (F-ANN), Fuzzy-RF (F-RF) and Fuzzy-SVR (F-SVR) have been developed by incorporating advantages of fuzzy relations in ML models by using membership and non-membership values through intuitionistic fuzzy c-means technique. To examine the predictive performance of FuzzyML model, India's monthly wholesale prices (Rs/q) of 18 agricultural commodities namely Rice, Wheat, Maize, Gram, Urad, Lentil, Mung, Arhar, Potato, Onion, Tomato, Soyabean, Sunflower, Groundnut, Mustard, Turmeric, Coriander, and Cumin have been used during the period January 2010 to December 2022. The accuracy of prediction of proposed FuzzyML model is evaluated in comparison with other benchmark models i.e. Autoregressive Integrated Moving Average (ARIMA), ANN, SVR, RF and also Fuzzy ARIMA (F-ARIMA) using different accuracy measures. To gain a comprehensive understanding of all accuracy measures, the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) has been applied. The study revealed that incorporating fuzzy logic to ML model increases the predictive performance of the models.

Keywords Agriculture · Intuitionistic fuzzy logic · Machine learning · Volatility

Introduction

India is one of the key players in the global agriculture sector. Accurate and reliable forecast of agricultural production as well as prices play a crucial role to ensure equilibrium between the supply and demand of agricultural commodities, boosting the income of farmers, and forming appropriate policies (Zhang and Na 2018). In light of the intricate influences of factors such as spatio-temporal dynamics, price correlations, and market diversity, there is a growing imperative to enhance the precision of agricultural price prediction for the sake of sustainable agricultural development. Consequently, the successful forecasting of agricultural product prices plays a pivotal role in government market regulation and the maintenance of national food security stability. In the last decade, numerous studies (Sankaran 2014; Mitra and Paul 2017; Paul and Garai 2021) have been undertaken to predict the prices of agricultural commodities. Brandt & Bessler (Brandt and Bessler 1983) outlined various approaches, including exponential smoothing, autoregressive integrated moving average (ARIMA) model (Box et al. 2015), econometric models, expert judgment, and a composite forecasting approach for predicting

✉ Md Yeasin
yeasin.iasri@gmail.com

[1] The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi, India

[2] ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India

time series data. They conducted a comparative analysis of the effectiveness of these methods in predicting the heightened volatility in prices of agricultural commodities over the preceding decade. Other time series forecasting models like exponential models (Goodwin 2010), Trend and Seasonal components model (Hyndman et al. 2008), BSTS: Bayesian Structural Time Series model (Scott and Varian 2014), etc. are treated as the alternatives of ARIMA model. Sometimes the linearity assumption restricts the practical application of these traditional statistical models to real world time series data and face challenges in capturing the intricate patterns, non-linear relationships, and high-dimensional nature.

In contrast, machine learning algorithms excel at adapting to dynamic patterns and automatically extract relevant features. At present times, there has been an increasing tendency in employing machine learning (ML) models (Ahmed et al. 2010; Vui et al. 2013; Palomares-Salas et al. 2014; Liakos et al. 2018; Yaslan and Bican 2017; Chaitra and Meena 2023) for predicting the large amount of non-linear time series data. Sinta et al. (Sinta et al. 2014) explored ensemble K-Nearest neighbors (KNN) method for forecasting of prices of rice. Shengwei et al. (Shengwei et al. 2017) investigated the factors influencing fluctuations in agricultural prices and formulated a prediction model for wholesale agricultural prices using least square Support Vector Regression (SVR). Paul et al. (Paul et al. 2022) investigated the effectiveness of various ML algorithms, such as Generalized Neural Network (GRNN), SVR, Random Forest (RF), and Gradient Boosting Machine (GBM), in forecasting the wholesale prices of Brinjal and shown that the ML technique performed better than the other models.

Though the ML models have good accuracy than the traditional statistical models, Agricultural commodity prices are impacted by various factors, such as supply and demand dynamics; weather patterns; global economic conditions; government policies; exchange rates; technology and innovation; and geopolitical events etc., which make it difficult to forecast the prices due to the ambiguity, instability, complexity, and uncertainty inherent in price data. In these situations, fuzzy logic can more appropriately handle the uncertainty, volatility, complexity, and non-linearity present in the data through membership function, signifying the extent of relevance of that element. Fuzzy logic can be a great inclusion for intelligent prediction methods to address the dynamic features of market price fluctuations in agricultural commodities. Ghosh et al. (Ghosh et al. 2016) developed the fuzzy time-series model by allocating values of membership functions that are asymmetric to the basic intervals set, resembling the concept of L–R fuzzy sets (Dubois and Prade 1981) and superiority of their proposed model over the existing fuzzy time series models had been established (Song and Chissom 1993; Huarng 2001). In recent times, many researchers (Tseng et al. 2001; Torbat et al. 2018; Jajarmi et al. 2019; Nasiri and Ebadzadeh 2022; Zhang and Xiang 2024) have attempted to incorporate the fuzzy logic in their model to forecast the time series data. But this model fails to give satisfactory result when the data involves higher degree of fuzziness. Atanassov (Atanassov 1999) developed an improved version of fuzzy logic called intuitionistic fuzzy logic (IFL), where non-membership values and hesitation degrees were considered along with the membership values for capturing the intensive uncertainty or randomness and indecision that might be present in real-world situations. Atanassov's work has contributed to the development of IFL and its uses in various fields, including decision-making (Da Costa et al. 2013; Atanassov 2015) and system modelling (Parvathi et al. 2013; Andreev et al. 2014; Mavrov et al. 2022). Adnan et al. (2021) explored the feasibility of a new hybrid neuro-fuzzy approach, combining a hybrid adaptive fuzzy system with innovative heuristic algorithms known as Water Cycle-Moth Flame Optimization (WCAMFO), referred to as ANFIS-WCAMFO, for modeling monthly reference evapotranspiration at Dhaka and Mymensing stations in Bangladesh. Dwivedi et al. (2023) defined an intuitionistic fuzzified time series and proposed a comprehensive time series prediction framework, which involves both intuitionistic fuzzification and defuzzification. In contrast with the partly intuitionistic fuzzified designs suggested by Kumar and Gangwar (2015), where only fuzzification incorporated hesitation to handle non-determinism in time series data. Recent studies (Ikram et al. 2023; Adnan et al. 2023) have consistently demonstrated the effectiveness of hybrid fuzzy-based models in various forecasting applications. Sarkar et al. (2024a) introduced WaveFLSTM, a wavelet-based fuzzy LSTM model that effectively combines wavelet transformation with fuzzy logic and deep learning to capture both temporal and frequency domain patterns in complex datasets. This approach demonstrated superior performance in modeling nonlinear and non-stationary time series. In another study, Sarkar et al. (2024b) developed IFNN, an intuitionistic fuzzy logic-based neural network model to handle uncertainty and imprecision inherent in time series data. Agricultural price data experiences significant volatility and uncertainty due to various factors, including erratic weather patterns, pest outbreaks, and shifting market demand. Furthermore, changes in the global economy and trade policies can also add to the instability of prices in the agricultural sector. This uncertainty present in data poses significant challenges to apply existing time series

forecasting models across various domains. It often leads to inaccurate predictions, increasing the risk associated with decision-making processes and potentially resulting in suboptimal outcomes. Furthermore, uncertainty complicates model selection, contributes to volatile market behavior, and may lead to misinterpretation of forecasting results. Due to these difficulties the traditional stochastic time series models and ML models often fail to predict the actual behaviors of the data. Dealing with uncertainty requires robust methodologies, careful consideration of underlying factors, and effective risk management strategies to improve forecasting accuracy and mitigate its impact on decision-making. Keeping these in mind, the primary goal of this study is to integrate intuitionistic fuzzy logic with the existing well-known time series forecasting models, especially machine learning model by leveraging expertise and domain-specific knowledge that can aid in pinpointing potential sources of uncertainty and fine-tuning forecasting models accordingly in such a way that it can effectively tackle the inherent uncertainty or high degree of fuzziness present in agricultural commodity prices. Thus, an innovative approach is proposed to strengthen the predicting capacity of machine learning models by integrating the intuitionistic fuzzy logic. This fuzzy logic in machine learning provides distinct advantages, particularly in improving the capacity to model and reason in uncertain situations. Unlike traditional models, this approach combines data-driven strengths of machine learnings with the human-like reasoning of fuzzy logic, enabling more precise management of ambiguity and vagueness in time series data. This method captures complex, non-linear patterns while preserving interpretability through fuzzy rules, which helps explain the model's decisions. Additionally, deep fuzzy logic systems are resilient to noisy or incomplete data and can integrate expert knowledge, making them particularly advantageous for time series forecasting. This approach provides more precise and reliable results by handling uncertain or imprecise time series data using membership and non-membership values, along with the degree of ambiguousness within each fuzzy cluster. Consequently, the predictions will be more robust, offering a closer reflection of real-world conditions and enhanced forecasting performance, particularly in cases with non-linear patterns or incomplete data. This study introduced intuitionistic fuzzy based four novel models namely, F-ARIMA, F-SVR, F-RF, and F-ANN which combine higher order deep fuzzy logic with the stochastic ARIMA model and ML models like SVR, RF, and ANN. This paper examines predictive capabilities of efficient fuzzy-based forecasting model with traditional time series model for forecasting prices of agricultural commodities in India.

Model construction

Intuitionistic fuzzy information theory

Let A be an intuitionistic fuzzy set (IFS) defined in the universal set U. Then the A can be represented as $A = \{\langle x, m_A(x), n_A(x)\rangle : x \epsilon U\}$, where $m_A : U \rightarrow [0,1]$ and $n_A : U \rightarrow [0,1]$ are the membership and non-membership function of the element $x \epsilon U$ respectively, and for all $x \epsilon U$ in A, $0 \leq m_A(x) + n_A(x) \leq 1$ satisfies. The intuitionistic fuzzy index, denoted as $h_A(x)$, is a measure associated with a fuzzy subset A on the universal set U and defined as $h_A(x) = 1 - m_A(x) - n_A(x)$, alternatively known as the degree of hesitancy or uncertainty of x in the fuzzy subset A. It is important to note that $0 \leq h_A(x) \leq 1;\ \forall x \in U$.When employing the intuitionistic fuzzy logic for time series forecasting, the process primarily involves three key steps: converting the given time series data into the deep intuitionistic fuzzy set by employing the membership and non-membership values to each observations, this is known as fuzzification. Then employ the deep fuzzy logic to find out the fuzzy relation in the process. After that the results have been converted into the crisp set. This process is known as defuzzification. Architecture of intuitionistic fuzzy rule has been depicted in Fig. 1.

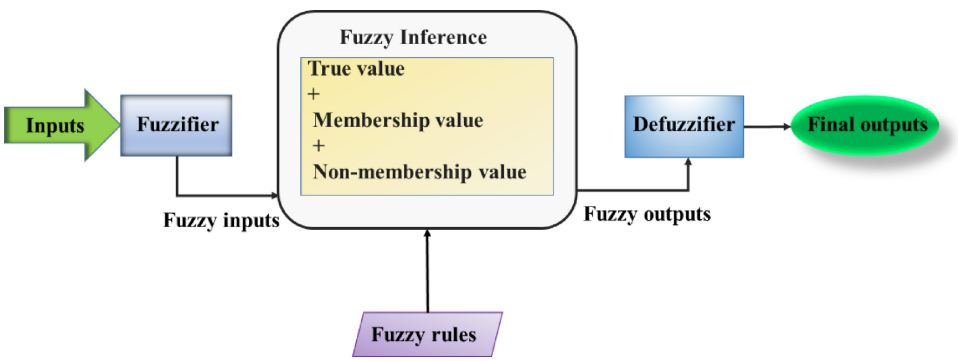


Fig. 1 The Intuitionistic fuzzy inference architecture

ARIMA model

The ARIMA model, widely used for time series forecasting and analysis, integrates autoregressive (AR) and moving average (MA) parts with differencing to enhance predictability in non-stationary time series data. Represented as ARIMA (p, d, q), where p is the AR order, d denotes the degree of differencing, and q signifies the MA order, the model is especially effective for managing time series data featuring trends and seasonality. Determining model parameters involves statistical techniques like autocorrelation and partial autocorrelation analysis. The form of ARIMA (p,d,q) model is given by

$$\Phi(B)(1-B)^{d}X_t = \Theta(B)\varepsilon_t \tag{1}$$

where, $\Phi(B)$ and $\Theta(B)$ are respectively the AR and MA polynomials defined as:

$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p \tag{2}$$

and

$$\Theta(B) = 1 - \Theta_1 B - \Theta_2 B^2 - \cdots - \Theta_p B^p \tag{3}$$

B is the backshift operator, defined as

$$BX_t = X_{t-1} \tag{4}$$

and ε_t is a white noise process.

Support vector regression (SVR) model

In time series analysis, the SVR is a machine learning technique that leverages Support Vector Machines principles to forecast future values in a time series by learning the underlying patterns and relationships in historical data. SVR aims to minimize the prediction error while considering the temporal sequence of the data. The SVR model (Vapnik 1999) can be defined as:

$$f(\boldsymbol{x}) = \boldsymbol{w} \cdot \Phi(\boldsymbol{x}) + b \tag{5}$$

where, $\Phi(\boldsymbol{x})$ is a mapping function that employs a non-linear transformation to convert non-linear input into a linear form within a higher-dimensional feature space such as

$$\Phi(\cdot) = R^n \rightarrow R^{n_h},$$

$\boldsymbol{x}$ is the input vector, $\boldsymbol{w}$ is the vector of weights associated with the model, and b is the bias which have been calculated by minimizing the regularized risk function given as:

$$R(\lambda) = \frac{1}{2}\|\boldsymbol{w}\|^2 + \lambda\frac{1}{n}\sum_{i=1}^{n} L_\varepsilon\left(f(x_i), y_i\right) \tag{6}$$

where, $\frac{1}{2}\|\boldsymbol{w}\|^2$ is the regularised term which gauges the smoothness or flatness of the function and $\frac{1}{n}\sum_{i=1}^{n} L_\varepsilon\left(f(x_i), y_i\right)$ is the empirical error calculated by the Vapnik's ε-insensitivity loss function (L_ε), formulated as:

$$L_\varepsilon\left(f(\boldsymbol{x}), \boldsymbol{y}\right) = \begin{cases} |f(\boldsymbol{x}) - \boldsymbol{y}| - \varepsilon & if\ |f(\boldsymbol{x}) - \boldsymbol{y}| \geq \varepsilon \\ 0 & otherwise \end{cases} \tag{7}$$

where, $\boldsymbol{y}$ is the output vector. Here λ and ε are the user defined hyperparameters. Architecture of SVR algorithm has been depicted in Fig. 2.

Random forest (RF) model

RF is based on the bagging technique, which involves applying bootstrap aggregation to decision trees. This approach reduces the variance of the base algorithms, particularly when they exhibit weak correlation. This versatile and user-friendly supervised machine learning algorithm is widely used due to its simplicity and versatility. RF constructs multiple decision trees and combines them to yield more precise and robust predictions. To mitigate tree correlation, RF utilizes randomization in two manners: initially, each tree is trained on a bootstrapped subset, and second, the feature used for splitting in each node is chosen from a random subset of size m rather than from all possible features. The independence of the N trees in the RF algorithm facilitates easy parallelization, and each tree forms a complete binary tree with a maximum extent of depth, contributing to the effectiveness of RF performance.

Let $r(\bullet)$ be the represent the function derived from the training of RF, which is utilized for predicting y_t $(t \in N)$ based on the historical values $(y_1, y_2, \ldots, y_{t-1})$. When employing k lagged variables, the forecasted value $(\widehat{y}_t)$ is determined by the following equation:

$$\widehat{y}_l = r\left(y_{l-1}, y_{l-2}, \ldots, y_{l-k}\right), l = k+1, \ldots, t \tag{8}$$

Architecture of random forest has been presented in Fig. 3.

Artificial neural network (ANN) model

The Artificial Neural Network (ANN) model (Haykin 1998) is a computational framework inspired by the

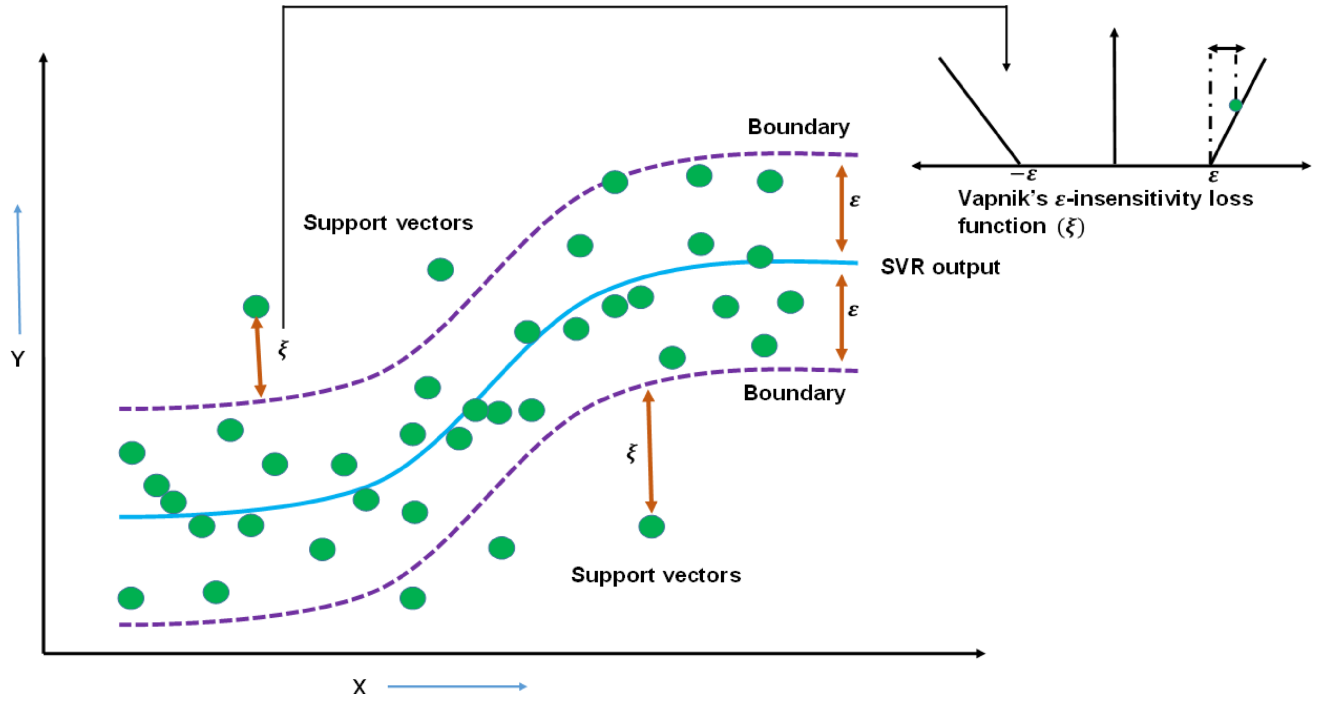


Fig. 2 Diagrammatic representation of SVR algorithm

structure of the human brain. It consists of interconnected neurons arranged in layers. It includes an input layer for receiving data, one or more hidden layers for processing through mathematical operations, and an output layer for producing results. The layers are connected through non-circular arcs (Khashei and Bijari 2010). The learning is achieved by adjusting connection weights to minimize prediction errors using the backpropagation algorithm.

The predominant choice for modeling and predicting time-series data is a single hidden layer feed-forward network which is structured with three layers of computing units interconnected by non-circular arcs. The relationship between one output (y_t) and the k inputs $(X_1, X_2, \ldots, X_k)$ is given as

$$y = \omega_\circ + \sum_{i=1}^{r} \omega_i f\left(\omega_{\circ i} + \sum_{j=1}^{k} \omega_{ji} X_j\right) + \varepsilon \tag{9}$$

where, $\omega_i\,(i = 0,1,\ldots,r)$ is the weight associated with the i^{th} hidden node to the output and $\omega_{ji}is$ the weight associated with the connection between j^{th} input node and i^{th} hidden layer node; k is the input nodes number and r is the hidden layer nodes number and f is the activation function. Architecture of ANN model has been depicted in Fig. 4.

The neural network system emulates the learning process of the human brain and can accomplish of handling nonlinear data, a characteristic often found in agricultural datasets. This makes ANNs particularly suitable for accurately modeling the agricultural commodity prices data.

Proposed fuzzy based machine learning models

A fuzzy-based model is a computational approach that leverages fuzzy logic to handle uncertainty, ambiguity, vagueness and imprecision in decision-making, allowing for more flexible and precise representation of relationships between variables. Fuzzy based model improves the model's ability to capture intricate patterns of time series data and make predictions in real-world scenarios with prevalent uncertainty. Steps for fuzzy based algorithm can be summarized as below:

Step 1. Specify the parameters like total number of the intuitionistic fuzzy set, lag of the time series data, total number of train and test set observation etc. employed in the fuzzy based model.

Step 2. Fuzzy logic representation where each observation is linked to multiple membership and non-membership functions, assigning linguistic labels to each observation. The fuzzy membership function assesses the extent to which an observation is associated with a specific fuzzy set. Here, the input variable is a one-dimensional input vector derived from the given time series observations.

Step 3. Prepare input for the time series model by combining membership, non-membership, and lag values.

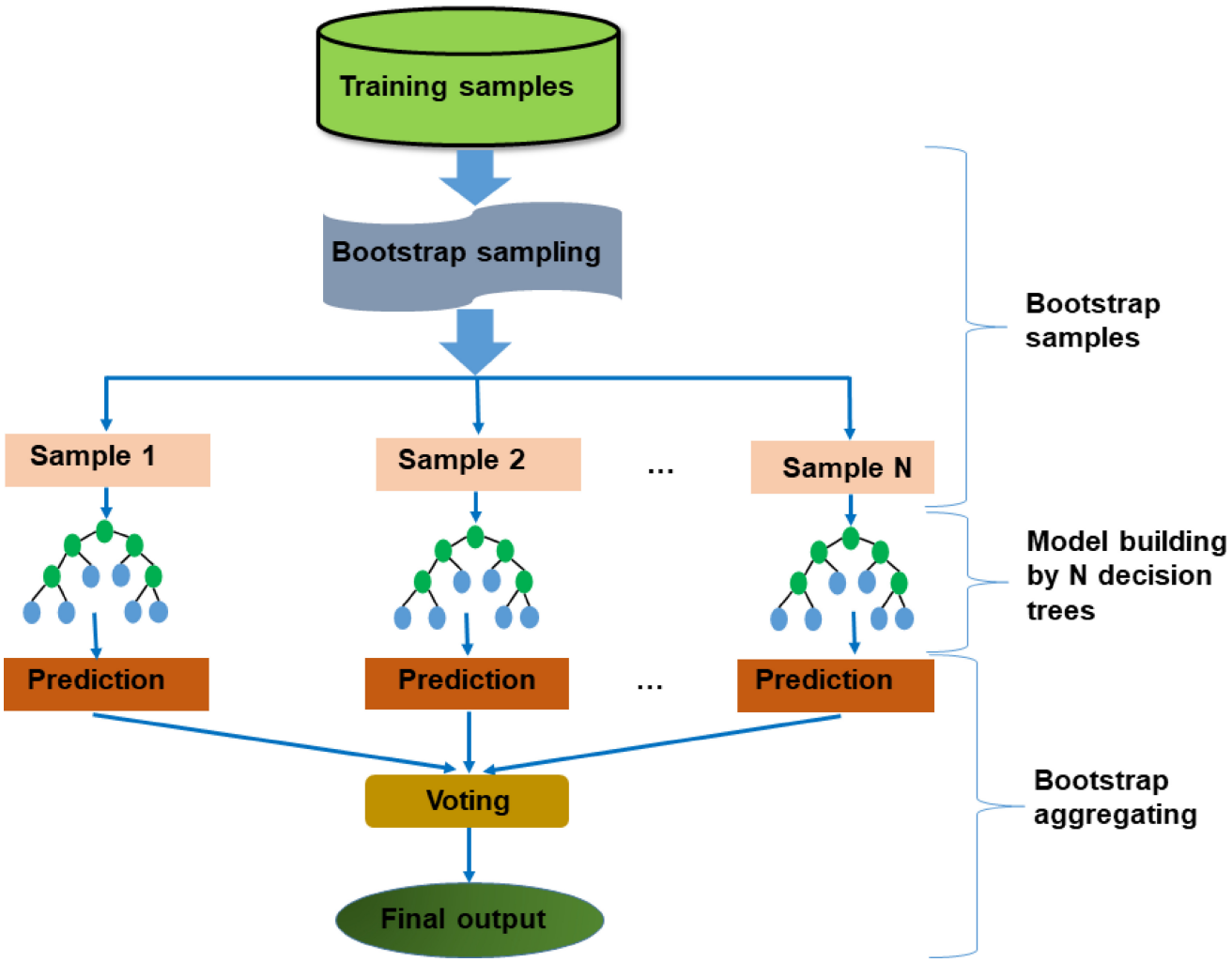


Fig. 3 Schematic diagram of random forest

Step 4. Developed the fuzzy based model using the input series prepared in previous step.

In contrast to conventional methods, this approach computes fuzzy membership and non-membership values using the intuitionistic fuzzy c-means (IFCM) clustering algorithm instead of relying on specific membership functions like triangular, trapezoidal, Gaussian, or others. The IFCM algorithm extends the traditional fuzzy clustering method, particularly the Fuzzy C-Means (FCM) algorithm. Atanassov and Gargov (Atanassov and Gargov 1989) introduced IFCM, introducing an extra level of uncertainty called hesitation alongside the membership and non-membership degrees found in fuzzy clustering. The algorithm iteratively refines cluster centers and membership values by minimizing an objective function that incorporates the parameters of the intuitionistic fuzzy set. In IFCM, each data point is defined by three values: membership degree, non-membership degree, and hesitation degree, collectively defining the intuitionistic fuzzy set membership which can be structured as.

$$\boldsymbol{M} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \cdots & \mu_{1c} \\ \mu_{21} & \cdots & & \cdots & \mu_{2c} \\ \vdots & & \ddots & & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mu_{n1} & \mu_{n2} & & \cdots & \mu_{nc} \end{bmatrix} \text{ and}$$

$$\boldsymbol{N} = \begin{bmatrix} \vartheta_{11} & \vartheta_{12} & \cdots & \cdots & \vartheta_{1c} \\ \vartheta_{21} & \cdots & & \cdots & \vartheta_{2c} \\ \vdots & & \ddots & & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \vartheta_{n1} & \vartheta_{n2} & & \cdots & \vartheta_{nc} \end{bmatrix}$$

where, $\{\mu_{ij}\}$ and $\{\vartheta_{ij}\}$ are the membership and non-membership values with respect to the i^{th} $(i = 1,2,\ldots,n)$ observation and the j^{th} $(j = 1,2,\ldots,c)$ cluster respectively. $\boldsymbol{M}$ and $\boldsymbol{N}$ are the $n \times c$ order matrix of membership and non-membership values respectively, and c is the number of

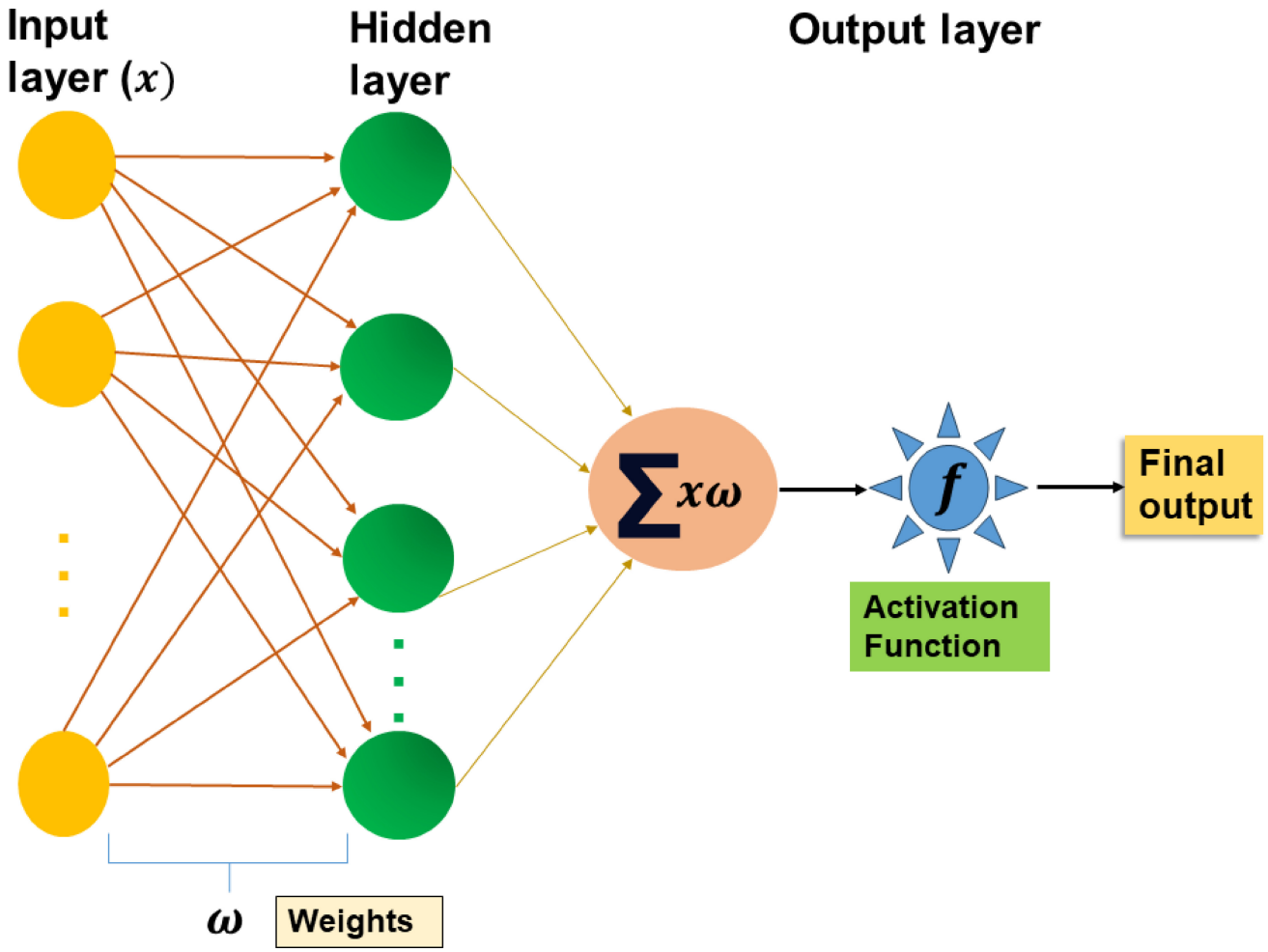


Fig. 4 Schematic diagram of ANN

clusters in IFCM method and n is the total number of available observations which can be represented as $n \times 1$ order vector $\boldsymbol{x}^T = (\quad X_1 \quad X_2 \quad \ldots \quad X_n \quad)$. Here, membership values μ_{ij}'s are obtained by

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(a-1)}} \tag{10}$$

where, a is the index of fuzziness, d_{ij} is the distance of each i^{th} observation (x_{ij}) from the j^{th} cluster centre (c_j) computed as

$$d_{ij} = \sqrt{(x_{ij} - c_j)^2} \tag{11}$$

Non-membership values ϑ_{ij}'s are computed by

$$\vartheta_{ij} = 1 - \mu_{ij} - h_{ij} \tag{12}$$

where, h_{ij} is the hesitation degree obtained as

$$h_{ij} = 1 - \mu_{ij} - \left(\left(1 - {\mu_{ij}}^b\right)\right)^{1/b}, b > 0 \tag{13}$$

Architecture of fuzzy based model has been depicted in Fig. 5.

Note that as this study produces actual forecasts, there is no requirement for a defuzzification phase in the procedure.

Using the above steps, different fuzzy based models have been made to predict the prices of agricultural commodity. For constructing the fuzzy based ARIMA model (F-ARIMA), ARIMA-X (Auto Regressive Integrated Moving Average with Exogenous variables) is used with the membership and non-membership values as the exogenous variable. The model can be written as

$$\begin{aligned} \boldsymbol{X}_t =& \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} \\ &+ \Theta_1 \varepsilon_{t-1} + \ldots + \Theta_q \varepsilon_{t-q} + \boldsymbol{\beta F_t} + \boldsymbol{\varepsilon_t} \end{aligned} \tag{14}$$

where, $\boldsymbol{F_t}$ is the matrix of membership and non-membership values, represented as

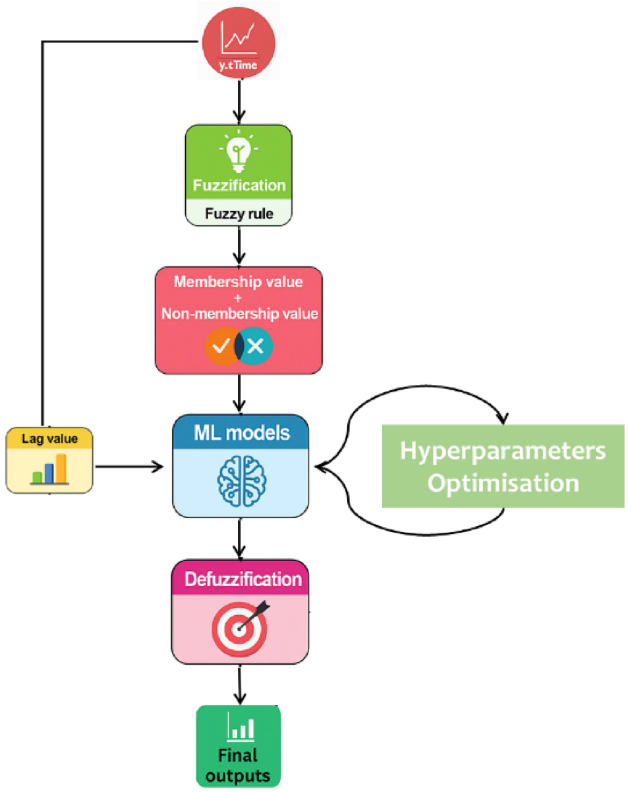


Fig. 5 Architecture of fuzzy based models

Algorithm Proposed fuzzy-based ML model

Input: Time series data (y_t), lag (L)
Output: Final predicted value(s) $\hat{y}_{t+1}$, $\hat{y}_{t+2}$, ...

Step 1: Initialize parameters
- Number of clusters (c)
- Fuzziness index (a)
- Prediction window (P_t)
- Lag window size (L)

Step 2: Implement IFCM
- for i = 1 to c do
  - Compute cluster center c_i
- for j = 1 to n do
  - Compute distance $d_{ij} = \sqrt{((x_{ij} - c_j)^2)}$
  - Compute membership degree: $\mu_{ij} = [\sum_{k=1}^{c} (d_{ij} / d_{kj})^{\wedge}(2/(a-1))]^{-1}$
  - Compute hesitation degree: $h_{ij} = 1 - \mu_{ij} - [(1 - \mu_{ij}^b)]^{\wedge}(1/b)$
  - Compute non-membership degree: $\theta_{ij} = 1 - \mu_{ij} - h_{ij}$

Step 3: Construct input matrix
- Combine lagged time series values μ_{ij}, and θ_{ij}

Step 4: Train fuzzy-ML hybrid model
- Choose model: SVR, RF, ANN
- • Perform optimization of model parameters
- Repeat until loss function is minimized:
  - Train model
  - Evaluate loss function
- if loss is minimized then proceed to Step 5

Step 5: Defuzzification
- Convert output fuzzy value to crisp predicted value(s)

Step 6: Output predicted values
- $\hat{y}_{t+1}$, $\hat{y}_{t+2}$, ...

$$\boldsymbol{F_t} = [\ \boldsymbol{M}\quad \boldsymbol{N}\ ] = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \cdots & \mu_{1c} & \vartheta_{11} & \vartheta_{12} & \cdots & \cdots & \vartheta_{1c} \\ \mu_{21} & \cdots & & \cdots & \mu_{2c} & \vartheta_{21} & \cdots & & \cdots & \vartheta_{2c} \\ \vdots & & \ddots & \vdots & & \vdots & & \ddots & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mu_{n1} & \mu_{n2} & & \cdots & \mu_{nc} & \vartheta_{n1} & \vartheta_{n2} & \cdots & \cdots & \vartheta_{nc} \end{bmatrix} \tag{15}$$

and $\boldsymbol{\beta}$ is the vector of coefficients for membership and non-membership values.

For fuzzy based ANN (F-ANN), support vector regression (F-SVR), and random forest (F-RF) model, the inputs are given with membership and non-membership values as exogenous variable which are calculated by using IFCM algorithm. Here the inputs for these models are the values of time series data and membership and non-membership values which can be written as

$$\boldsymbol{X_{input}} = [\ \boldsymbol{x}\quad \boldsymbol{M}\quad \boldsymbol{N}\ ] = \begin{bmatrix} & \mu_{11} & \mu_{12} & \cdots & \cdots & \mu_{1c} & \vartheta_{11} & \vartheta_{12} & \cdots & \cdots & \vartheta_{1c} \\ X_1 & \mu_{21} & \cdots & & \cdots & \mu_{2c} & \vartheta_{21} & \cdots & & \cdots & \vartheta_{2c} \\ X_2 & & \vdots & \ddots & \vdots & & & \vdots & \ddots & \vdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ X_n & \mu_{n1} & \mu_{n2} & & \cdots & \mu_{nc} & \vartheta_{n1} & \vartheta_{n2} & \cdots & \cdots & \vartheta_{nc} \end{bmatrix} \tag{16}$$

where, $\boldsymbol{X_{input}}$ is the $n \times (1 + 2c)$ order input matrix, $\boldsymbol{x}$ is the $n \times 1$ order vector of all the n observations.

The algorithm of proposed fuzzy based ML model is described as follows:

Model validation

The fuzzy based models are empirically evaluated using some statistical criteria, specifically Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Coefficient of determination (R^2) which are formulated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_k - \widehat{y}_k)^2} \tag{17}$$

$$\text{MAPE} = \frac{1}{\text{n}} \sum_{\text{k}=1}^{n} \left| \frac{y_k - \widehat{y}_k}{y_k} \right| \times 100 \tag{18}$$

$$\text{MAE} = \frac{1}{n} \sum_{\text{k}=1}^{\text{n}} |y_k - \widehat{y}_k| \tag{19}$$

$$R^2 = 1 - \frac{\sum_{k=1}^{n} (y_k - \widehat{y}_k)^2}{\sum_{k=1}^{n} (y_k - \overline{y_k})^2} \tag{20}$$

where, y_k, $\overline{y_k}$, and $\widehat{y}_k$ are the true, mean and forecasted values of the study variable, respectively. Lower values of RMSE, MAPE, and MAE; and higher value of R^2 indicates the greater accuracy of the model.

Empirical analysis

Data description

Data on the monthly wholesale prices (Rs/q) and arrival information for 18 agricultural commodities namely Rice, Wheat, Maize, Gram, Urad, Lentil, Mung, Arhar, Potato, Onion, Tomato, Soyabean, Sunflower, Groundnut, Mustard, Turmeric, Coriander, and Cumin were collected from the AGMARKNET portal (https://agmarknet.gov.in/) spanning from 1st January 2010 to 31st December 2022. This portal is managed by the Directorate of Marketing and Inspection, Government of India. These 18 commodities were selected based on their economic importance, availability of consistent historical data and relevance in regional agricultural markets. These commodities represent a diverse range of Indian diet such as food grains, pulses, vegetables, spices and oilseeds. These commodities are not only staples in food security but also play a pivotal role in agricultural policy, trade, and industry. Rice, Wheat, and Maize, for example, are key food crops with substantial contributions to the global food supply and the livelihoods of millions of farmers. Other crops like Onion and Tomato are highly volatile in price, impacting both producers and consumers. Oilseeds such as Soyabean, Sunflower, and Groundnut are vital in the global edible oil market. Furthermore, pulses such as Gram, Urad, Lentil, and Mung are crucial for food security in many developing countries due to their nutritional value and low cost. The inclusion of these crops is aligned with addressing challenges in agricultural markets and enhancing food security. The total 154 markets have been selected based on their representation in the respective commodity and their share in arrivals. For each of the commodities, national level aggregated price has been computed weighted by arrival in the market. The selected markets of each commodity have been depicted Figure 6.

Descriptive statistics of the aggregated data series has been presented in Table 1. A close examination of Table 1 reveals that the variation measured in terms of CV in the price data is less for cereal and pulse crops (Rice-19.33%, Wheat- 18.36, Maize- 23.66%, Mung- 23%, Groundnut-24.19%, and Cumin- 20.78%). The prices for vegetables exhibit the greatest variability (Potato- 52.61%, Onion-63.43%, and Tomato-57.10%). Rest of the food crop prices have moderate variation as CV values ranges from 27 to 37%. It is also noticed that the price dataset for all commodities is positively skewed except Mung.

Model implementation

The stochastic models i.e. ARIMA and ML models such as SVR, RF, and ANN along with proposed models as detailed in the Sect. "Model construction" have been fitted. The agricultural commodity prices dataset comprises 156 data points, and the series is partitioned into training and testing sets with 90:10 ratio. The training set encompasses the initial 140 months of observations, utilized for model construction, while the remaining 16 months of observations are reserved for assessing the precision of the fuzzy based ML models. The dependency on the current price extends up to 12 lags for all commodity. The selection of the best-fitted ARIMA model was based on information criteria such as Akaike Information Criterion (AIC), and Schwartz Bayesian Criterion (SBC). Machine learning techniques were trained by optimizing its important hyperparameters. The grid search method is used which systematically optimizes hyperparameters by examining a predetermined range of values. This algorithm explores various combinations of hyperparameters for a specific model to find the optimal configuration. The range of the parameters/ hyperparameters for all agricultural commodities are given in the below Table 2.

To implement the intuitionistic fuzzy logic, the five clusters has been selected using the Elbow method. This method identifies the optimal number of clusters by examine the rate of decrease in within-cluster variance. Membership and non-memberships values of prices are treated as exogenous variables in machine learning techniques. The precision of the fitted models on the validation set is presented in Table 3.

A perusal of Table 3, indicates that across all agricultural commodities, fuzzy-based ML models exhibit significantly superior performance compared to conventional models. The optimal model for Rice is determined by F-ARIMA, identified through the lowest RMSE value,

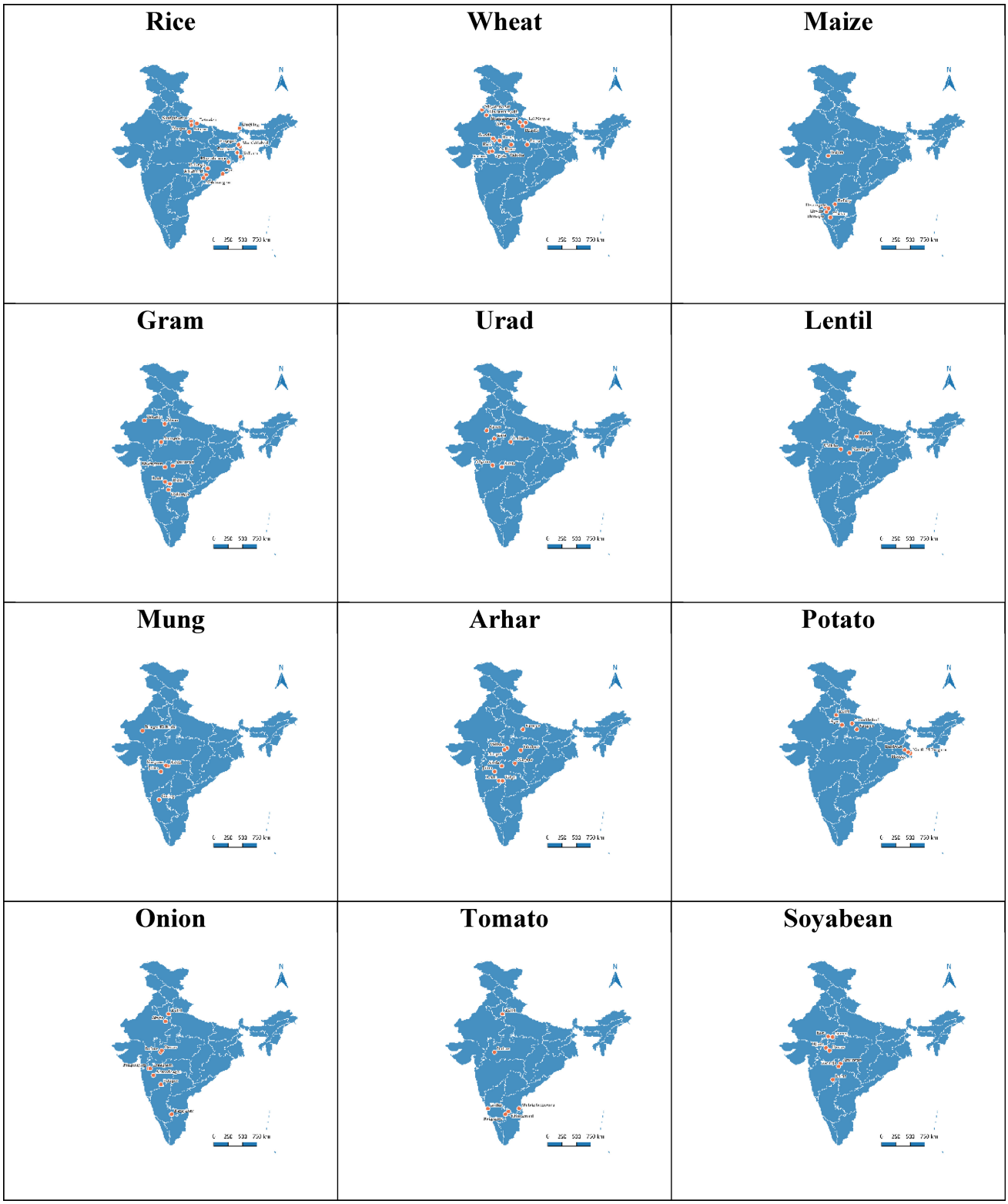


Fig. 6 Selected markets of different food commodity

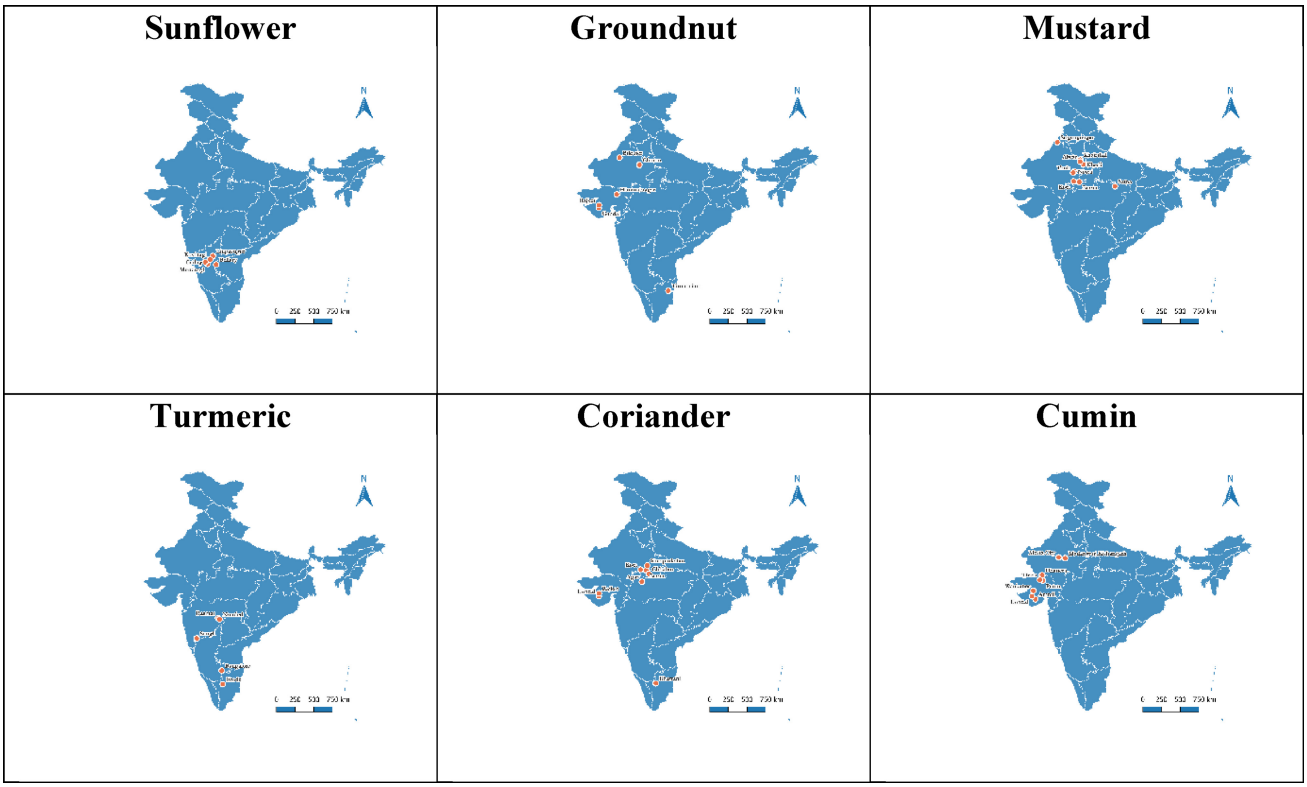


Fig. 6 (continued)

Table 1 Summary statistics of the monthly prices for the agricultural commodity

Commodity	Mean	S.E	Median	S.D	Kurtosis	Skewness	Minimum	Maximum	CV (%)
Rice	1990.41	30.80	1960.57	384.70	− 0.54	0.21	1268.15	3032.47	19.33
Wheat	1639.15	24.23	1636.30	302.68	− 0.03	0.08	1062.36	2551.30	18.47
Maize	1390.49	26.34	1351.21	328.95	0.06	0.67	841.17	2380.58	23.66
Gram	3941.55	101.79	4040.56	1271.39	2.75	1.14	1934.44	8934.84	32.26
Urad	4842.81	143.06	4277.02	1786.86	0.62	1.02	2594.74	10,277.83	36.90
Lentil	4295.64	96.25	3998.69	1202.16	− 0.93	0.45	2517.07	6975.86	27.99
Mung	5032.22	92.65	4828.91	1157.23	− 0.76	− 0.03	2282.64	7868.73	23.00
Arhar	4940.26	112.87	4418.61	1409.74	0.14	0.87	2832.87	9264.84	28.54
Potato	933.75	39.33	861.02	491.27	2.77	1.42	273.54	3052.76	52.61
Onion	1317.25	66.89	1080.17	835.52	3.89	1.76	417.20	5378.38	63.43
Tomato	1180.75	53.98	978.73	674.23	3.11	1.61	392.69	4237.91	57.10
Soyabean	3639.92	105.06	3479.55	1312.24	2.86	1.59	1839.51	8839.24	36.05
Sunflower	3750.84	90.86	3344.17	1134.85	1.51	1.57	2294.58	7648.71	30.26
Groundnut	4579.64	88.70	4448.05	1107.82	0.13	0.59	2692.93	8170.03	24.19
Mustard	3969.73	102.04	3665.28	1274.54	0.91	1.18	2175.33	7848.13	32.11
Turmeric	8094.28	198.18	7730.10	2475.21	0.73	0.99	3990.49	16,697.14	30.58
Coriander	5754.33	168.25	5716.17	2101.45	− 0.40	0.47	2244.70	10,588.47	36.52
Cumin	13,911.98	231.42	13,102.04	2890.44	1.35	1.04	9001.38	24,838.12	20.78

while F-ANN stands out for its lowest MAPE and MAE values, and F-RF has highest R^2 value. Conversely, for Wheat, F-ARIMA exhibits the lowest RMSE value, while F-RF is favored for its lowest MAPE and MAE values and highest R^2 value. In the case of Maize, Gram, Mung, Potato, Onion, Groundnut, and Turmeric, F-RF emerges as the superior model across all metrics, including lowest RMSE, MAPE, and MAE values and highest R^2 value. For Urad and Arhar, F-RF leads in terms of the lowest RMSE, MAPE, and MAE values whereas F-SVR is preferred for its highest R^2 value. For Lentil and Soyabean, F-RF is identified as the top model, achieving the lowest RMSE, MAPE and MAE values, while RF excels with the highest R^2 value. Regarding Tomato, F-ARIMA is identified

Table 2 Hyperparameters for the different models

Models	Hyperparameters	Candidates/ Ranges
ARIMA	p, d, q	Iterated with significant lags
SVR	#SV	410 to 598
	Cost	0.5 to 1.5
	Gamma	0.1 to 0.4
	Kernal	Radial basis
RF	Trees	458 to 599
	#MSR	0.0002 to 0.0007
	#VE	92.56 to 95.41
ANN	#HN	32 to 52
	#DR	0.25 to 0.5
	#LR	0.01to 0.08
	sigma2	0.005 to 0.01

#*HN* Number of nodes in hidden layer; *DR* Dropout rate; *#LR* Learning rate; *#SV* Number of support vectors; *#MSR* Mean of squared residuals; *#VE* % of variation explained

as the top model, achieving the lowest RMSE and MAE values, while F-RF excels with the lowest MAPE value and highest R^2 value. For Sunflower, F-RF leads in terms of the lowest RMSE, whereas F-ARIMA is preferred for its lowest MAPE and MAE values and highest R^2 value. For Mustard and Coriander, F-ANN emerges as the top-performing model, showcasing superiority across RMSE, MAPE, and MAE values but F-SVR emerges as superior based on R^2 value compared to other models. In the case of Cumin, F-SVR is identified as the optimal model based on the lowest RMSE, while F-RF excels with the lowest MAPE and MAE values and highest R^2 value. Additionally, it's worth noting that the F-RF model consistently demonstrates superior performance across different commodities. As it is cleared that for almost all commodities fuzzy-based models have lower RMSE, MAPE, and MAE values and higher R^2 value compare to their counter traditional models. The Diebold Mariano (DM) has also been implemented for check the supremacy of the proposed model and it has been found that our proposed model outperformed other benchmark models. Therefore, it can be inferred that fuzzy-based machine learning models outperformed traditional machine learning models and stochastic models across all commodities. Additionally, the percentage change in accuracy of ARIMA, SVR, RF, and ANN models compared to their fuzzy versions has been calculated in Table 4 for enhanced comprehension.

Upon closer examination of the predictive performance of fuzzy and non-fuzzy techniques, it becomes evident that models employing fuzzy algorithms consistently surpass traditional models across all three accuracy measures utilized in this study. It's noted that F-ARIMA demonstrates over 50% enhancement in accuracy compared to the traditional ARIMA model across most commodities, excluding Rice, Groundnut, and Mustard (where it exceeds almost 30%). Similarly, F-SVR displays over 30% improvement in accuracy compared to the traditional SVR model for most commodities, except Gram, Urad, Potato, Tomato, Soyabean, Mustard, and Turmeric (where it surpasses 6%). Additionally, F-RF showcases over 20% improvement in accuracy compared to the traditional RF model for most commodities, except Lentil (where it exceeds 13%). Finally, F-ANN exhibits over 40% improvement in accuracy compared to the traditional ANN model across most commodities, except Wheat, Groundnut, and Mustard (where it exceeds 14 Overall improvement of proposed fuzzy-based models over corresponding traditional for cereals, pulses, vegetables, oilseeds, and spices crop are almost 41%, 52%, 56%, 42%, and 64% respectively. The findings revealed that the fuzzy based model outperformed both traditional stochastic and machine learning models in terms of accuracy across all metrics. This underscores the effectiveness of employing a fuzzy based approach for improving price forecasting accuracy. In summary, F-ARIMA and F-ANN exhibited over 60% higher accuracy compared to ARIMA and ANN models, while F-SVR and F-RF showed approximately 35% and 40% greater accuracy than SVR and RF models, respectively. Table 4 demonstrates that the proposed fuzzy-based machine learning models consistently surpass traditional models in situations where datasets contain substantial noise or outliers. These models perform well when dealing with data that is uncertain, imprecise, incomplete, imbalanced, skewed, or characterized by non-linear and complex relationships. Deep fuzzy logic integrates degrees of membership, non-membership, and hesitation, enabling a more refined representation of uncertainty. In contrast, traditional models often struggle in environments with ambiguity or vagueness, as they rely on precise, crisp inputs. Also, proposed fuzzy systems excel in handling noise, capturing uncertainty, leading to better generalization and more stable predictions, whereas traditional models may overfit or fail to handle noise effectively. Additionally, fuzzy logic flexibly models complex decision boundaries by managing uncertainty and partial truths, while traditional machine learning models may require more intricate tuning, such as deep architectures, to achieve similar results.

To enhance understanding of the accuracy and variability across different models, violin plots are presented, providing a more detailed evaluation. Violin plots are used to visualize the distribution of error metrics such as RMSE, MAPE, MAE, and the coefficient of determination (R^2) across different models, providing a clearer picture of model performance. Each plot illustrates how these error values vary across multiple models, revealing the spread and distribution of errors metrics for each

Table 3 Validation of different models for predicting agricultural commodity prices

Model Accuracy	ARIMA	SVR	RF	ANN	F-ARIMA	F-SVR	F-RF	F-ANN
Rice								
RMSE	340.96	527.55	364.38	385.60	266.43	435.36	320.35	312.52
MAPE	10.61	16.01	11.34	12.70	7.30	10.13	7.62	7.30
MAE	293.74	443.16	313.98	347.99	205.03	277.17	207.35	199.43
R^2	0.52	0.34	0.54	0.29	0.71	0.52	0.59	0.50
Wheat								
RMSE	306.93	333.85	245.54	268.96	216.05	274.23	228.91	256.65
MAPE	10.62	9.79	7.82	9.60	6.39	6.86	5.48	9.02
MAE	243.08	228.89	181.12	218.51	148.94	154.93	124.54	204.18
R^2	0.19	0.35	0.60	0.53	0.37	0.56	0.75	0.32
Maize								
RMSE	399.04	123.72	86.64	678.23	41.80	72.77	33.43	130.05
MAPE	19.89	5.42	3.83	24.56	1.87	3.32	1.27	5.25
MAE	373.11	103.51	71.74	444.80	34.18	60.76	24.29	96.60
R^2	0.54	0.71	0.86	0.41	0.82	0.83	0.95	0.61
Gram								
RMSE	221.85	125.17	147.53	1565.51	104.32	93.52	47.21	335.79
MAPE	4.42	2.05	2.88	26.34	2.23	1.89	0.90	6.25
MAE	192.20	93.23	128.53	1175.26	98.96	83.98	40.21	273.09
R^2	0.52	0.92	0.85	0.23	0.87	0.93	0.96	0.44
Urad								
RMSE	1389.96	212.48	296.30	3679.11	134.49	154.10	127.21	303.46
MAPE	21.90	2.56	4.07	50.23	1.86	2.09	1.53	4.20
MAE	1356.60	156.16	249.42	3059.88	112.98	126.03	92.47	256.17
R^2	0.32	0.70	0.66	0.24	0.55	0.84	0.82	0.33
Lentil								
RMSE	573.51	1078.41	206.13	1663.21	336.18	696.64	171.97	383.58
MAPE	8.49	16.78	2.43	24.15	3.28	10.48	2.14	4.58
MAE	529.40	1060.24	157.66	1516.61	207.75	662.85	135.05	289.26
R^2	0.31	0.40	0.89	0.42	0.57	0.49	0.79	0.78
Mung								
RMSE	386.75	167.65	120.98	1393.89	139.90	94.09	61.23	274.36
MAPE	5.95	2.37	1.53	21.68	1.97	1.38	0.67	3.98
MAE	343.37	136.84	86.21	1250.04	116.16	82.02	39.44	233.17
R^2	0.44	0.97	0.94	0.56	0.70	0.95	0.97	0.81
Arhar								
RMSE	608.65	265.04	188.02	1982.62	128.82	180.17	126.30	375.64
MAPE	7.01	3.66	1.79	24.57	1.98	2.23	1.07	5.15
MAE	465.97	231.29	116.83	1604.13	123.68	145.17	71.03	318.00
R^2	0.31	0.83	0.81	0.33	0.61	0.96	0.94	0.28
Potato								
RMSE	367.89	115.26	113.76	1000.77	59.11	65.35	45.32	277.36
MAPE	27.29	4.96	7.36	73.54	3.76	4.10	2.24	20.77
MAE	333.62	60.77	83.75	843.19	47.34	48.53	29.34	201.03
R^2	0.44	0.88	0.91	0.50	0.82	0.97	0.98	0.51
Onion								
RMSE	925.93	700.42	121.64	577.57	143.19	112.72	57.89	177.09
MAPE	32.21	19.35	7.57	36.33	9.93	7.69	2.33	9.31
MAE	648.36	431.06	99.67	457.43	124.23	92.66	39.16	144.57
R^2	0.57	0.94	0.94	0.49	0.91	0.95	0.99	0.61
Tomato								
RMSE	960.10	693.38	421.31	1334.46	166.40	514.37	238.91	223.80
MAPE	36.80	26.64	15.66	52.00	6.79	20.94	5.76	14.91
MAE	728.25	506.01	295.81	911.07	123.82	390.67	133.19	188.20
R^2	0.49	0.63	0.93	0.47	0.77	0.89	0.97	0.84

Table 3 (continued)

Model Accuracy	ARIMA	SVR	RF	ANN	F-ARIMA	F-SVR	F-RF	F-ANN
Soyabean								
RMSE	2623.38	1767.06	655.86	2330.28	1219.77	1619.99	417.76	945.09
MAPE	35.60	25.65	10.27	32.37	16.44	22.83	5.51	14.76
MAE	2255.02	1603.02	570.02	1999.00	1004.23	1437.15	321.59	862.87
R^2	0.28	0.21	0.82	0.29	0.38	0.03	0.79	0.50
Sunflower								
RMSE	2206.09	2593.70	597.77	818.51	540.63	1473.72	525.37	546.49
MAPE	32.86	39.66	6.91	11.79	4.88	19.14	5.39	5.45
MAE	2021.60	2426.06	446.83	700.59	320.81	1251.88	353.99	349.72
R^2	0.22	0.23	0.39	0.32	0.76	0.47	0.68	0.41
Groundnut								
RMSE	935.39	1513.58	871.91	1005.48	636.71	1222.71	596.85	898.28
MAPE	10.00	18.73	7.90	12.26	6.14	14.08	5.64	8.06
MAE	687.61	1270.06	543.01	778.35	429.74	969.49	394.67	553.92
R^2	0.23	0.24	0.53	0.29	0.42	0.43	0.69	0.36
Mustard								
RMSE	1006.48	1589.83	726.80	589.62	724.32	1370.96	560.96	532.10
MAPE	14.49	24.17	10.58	7.78	9.96	20.96	7.95	6.11
MAE	906.15	1569.62	663.94	485.81	631.58	1351.21	499.21	398.70
R^2	0.24	0.73	0.74	0.40	0.39	0.86	0.48	0.66
Turmeric								
RMSE	2051.84	841.53	736.09	2860.27	350.19	767.14	345.39	497.16
MAPE	16.63	5.49	4.00	30.68	4.08	5.28	1.80	4.48
MAE	1455.66	496.83	377.52	2291.90	290.53	466.53	175.29	366.60
R^2	0.42	0.91	0.89	0.51	0.78	0.96	0.99	0.88
Coriander								
RMSE	3277.02	1767.59	1313.26	4271.16	959.96	1023.34	1220.09	606.73
MAPE	30.68	14.18	10.47	38.84	6.56	7.28	6.90	4.02
MAE	3013.45	1423.13	1067.92	3825.90	676.24	675.73	685.68	367.55
R^2	0.43	0.47	0.55	0.39	0.74	0.69	0.81	0.74
Cumin								
RMSE	6958.52	5220.41	3930.87	7357.08	3232.21	278.43	349.38	2467.38
MAPE	30.71	20.12	16.44	33.35	11.17	1.46	1.40	8.29
MAE	6233.78	4242.90	3392.27	6748.97	2394.55	193.13	185.72	1785.06
R^2	0.27	0.33	0.48	0.35	0.47	0.62	0.77	0.65

model. The violin plots of different accuracy measures of our proposed models have been depicted in Fig. 7.

The violin plots support our results, as they reveal that the F-RF model has the lowest mean values for RMSE, MAPE, and MAE, while showing the highest mean value for R^2. This is consistent with the accuracy results in the table. To get the overall insight from all accuracy measures, the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) has been implemented. This technique has been introduced by Hwang and Yoon (Hwang and Yoon 1981) to assess the performance of alternatives based on their similarity to the ideal solution. As preprocessing step, the measures were inverted, and min–max normalization was applied to ensure equal weighting across all measures. The findings of the TOPSIS method are presented in Table 5.

The rank in the Table 5, represent the performance of the model based on the all four accuracy metrics. The results of the TOPSIS methods are also aligned with our findings. Therefore, it can be concluded that integrating fuzzy logic into machine learning techniques can enhance the prediction accuracy of various forecasting models. A major factor contributing to improved forecasting accuracy is the model's capacity to manage uncertainty and ambiguity. Traditional models such as ARIMA, SVR, RF, and ANN rely on precise datasets, often falling short in accounting for the inherent uncertainty in agricultural data. On the other hand, the F-ARIMA, F-SVR, F-RF, and F-ANN models incorporate intuitionistic fuzzy logic, enabling it to handle uncertain or imprecise information more efficiently. This is particularly crucial in agriculture, where variables like climate, soil conditions, and pest activity introduce significant unpredictability in

Table 4 Percentage improvement in accuracy for predicting agricultural commodity prices

Model Accuracy	ARIMA	SVR	RF	ANN
Rice				
RMSE	21.86	17.47	12.08	18.95
MAPE	31.2	36.73	32.8	42.52
MAE	30.2	37.46	33.96	42.69
R^2	36.54	50.32	9.18	72.41
Wheat				
RMSE	29.61	17.86	6.77	4.58
MAPE	39.83	29.93	29.92	6.04
MAE	38.73	32.31	31.24	6.56
R^2	93.95	58.69	25.61	40.29
Maize				
RMSE	89.52	41.18	61.42	80.83
MAPE	90.6	38.75	66.84	78.62
MAE	90.84	41.31	66.15	78.28
R^2	51.85	16.90	10.47	48.78
Gram				
RMSE	52.98	25.29	68	78.55
MAPE	49.55	7.8	68.75	76.27
MAE	48.51	9.93	68.72	76.76
R^2	66.69	1.08	12.81	2.78
Urad				
RMSE	90.32	27.47	57.07	91.75
MAPE	91.51	18.36	62.41	91.64
MAE	91.67	19.3	62.93	91.63
R^2	73.14	20.00	24.24	36.93
Lentil				
RMSE	41.38	35.4	16.57	76.94
MAPE	61.37	37.54	11.93	81.04
MAE	60.76	37.48	14.34	80.93
R^2	82.69	23.20	11.61	83.99
Mung				
RMSE	63.83	43.87	49.39	80.32
MAPE	66.89	41.77	56.21	81.64
MAE	66.17	40.06	54.25	81.35
R^2	61.25	2.35	3.21	44.72
Arhar				
RMSE	78.84	32.02	32.83	81.05
MAPE	71.75	39.07	40.22	79.04
MAE	73.46	37.23	39.2	80.18
R^2	96.62	14.51	16.23	16.02
Potato				
RMSE	83.93	43.3	60.16	72.29
MAPE	86.22	17.34	69.57	71.76
MAE	85.81	20.14	64.97	76.16
R^2	87.45	9.61	8.44	2.42
Onion				
RMSE	84.54	83.91	52.41	69.34
MAPE	69.17	60.26	69.22	74.37
MAE	80.84	78.5	60.71	68.39
R^2	59.65	1.06	5.32	24.49
Tomato				
RMSE	82.67	25.82	43.29	83.23
MAPE	81.55	21.4	63.22	71.33
MAE	83	22.8	54.97	79.34
R^2	57.95	39.73	3.81	79.47

Table 4 (continued)

Model Accuracy	ARIMA	SVR	RF	ANN
Soyabean				
RMSE	53.5	8.32	36.3	59.44
MAPE	53.82	10.99	46.35	54.4
MAE	55.47	10.35	43.58	56.83
R^2	35.4	87.22	3.51	74.37
Sunflower				
RMSE	75.49	43.18	12.11	33.23
MAPE	85.15	51.74	22	53.77
MAE	84.13	48.4	20.78	50.08
R^2	47.79	99.87	72.67	74.37
Groundnut				
RMSE	31.93	19.22	31.55	10.66
MAPE	38.6	24.83	28.61	34.26
MAE	37.5	23.67	27.32	28.83
R^2	77.56	77.53	30.40	21.49
Mustard				
RMSE	28.03	13.77	22.82	9.76
MAPE	31.26	13.28	24.86	21.47
MAE	30.3	13.92	24.81	17.93
R^2	66.19	18.77	35.91	67.15
Turmeric				
RMSE	82.93	8.84	53.08	82.62
MAPE	75.47	3.83	55	85.4
MAE	80.04	6.1	53.57	84
R^2	86.78	5.66	10.86	73.21
Coriander				
RMSE	70.71	42.11	7.09	85.79
MAPE	78.62	48.66	34.1	89.65
MAE	77.56	52.52	35.79	90.39
R^2	74.16	45.71	48.02	87.98
Cumin				
RMSE	53.55	94.67	91.11	66.46
MAPE	63.63	92.74	91.48	75.14
MAE	61.59	95.45	94.53	73.55
R^2	75.35	90.36	59.48	86.67

yield forecasting. The line plots of actual and the predicted values of our proposed models have been depicted in Fig. 8.

The proposed FuzzyML models offer a novel advancement over existing hybrid approaches by using intuitionistic fuzzy logic instead of conventional fuzzy sets. This extension enables a richer representation of uncertainty through the inclusion of membership, non-membership and hesitation degrees. As a result, the models are better equipped to handle the ambiguity, non-linearity, and irregularities present in agricultural price data. This enhanced capability makes the proposed framework more effective for forecasting in real-world applications where data uncertainty is a major challenge.

Conclusion

The consistent and unpredictable fluctuations in the prices of agricultural commodities pose a significant challenge for producers, consumers, researchers, and policymakers. To address the challenges posed by highly uncertain and volatile time series data, a fuzzy-based model has been proposed in this study. The study examined the performance of fuzzy based machine learning model using 18 different commodities that are commonly consumed in various states of India. Most of these datasets exhibit nonlinear, non-normal, and non-stationary behaviour which makes it difficult to use traditional

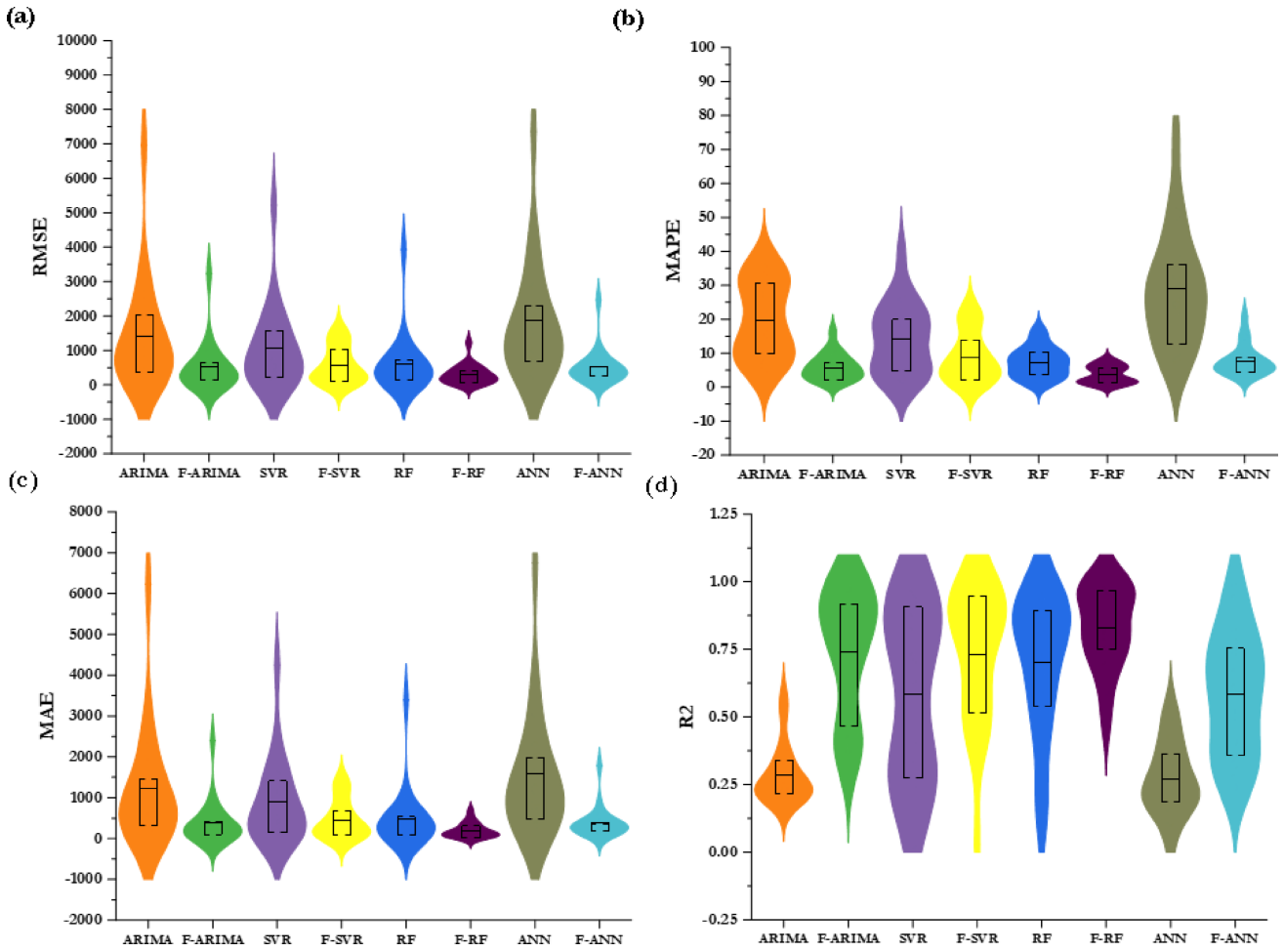


Fig. 7 Violin plots of different models for different accuracy metrices: **a** RMSE **b** MAPE **c** MAE **d** Coeeficient of determination (R^2)

models. The comparison between the fuzzy-based ML models and traditional models is conducted using performance measures such as RMSE, MAPE, and MAE. The findings indicate that the fuzzy-based ML models consistently outperform their corresponding non-fuzzy models. For most of the commodities, Random Forest (F-RF) model consistently shows better performance than other models. This study empirically concluded that incorporating fuzzy logic to machine learning model increases the predictive performance of the models. F-ARIMA and F-ANN demonstrated a substantial improvement in accuracy, exceeding 60% compared to ARIMA and ANN models. Similarly, F-SVR and F-RF exhibited approximately 35% and 45% higher accuracy, respectively, when compared to SVR and RF models. Despite giving outstanding improvements there might the possible threats to validity of the experimental study such as fuzzy-based models necessitate careful selection and tuning of parameters, which can be a time-consuming process that often demands domain expertise. Additionally, the incorporation of expert knowledge into fuzzy rules, the need for scalability with large datasets, and the management of the computational complexity of these models can pose threats to the validity of the experimental study. In future study, the accuracy of FuzzyML model can be further improved by incorporating exogenous information to the models. Also, the proposed model could be enhanced by incorporating optimization algorithms and advanced techniques like attention mechanisms or ensemble methods. Additionally, integrating external factors such as weather patterns, geopolitical events, and market policies into the forecasting framework may improve the model's accuracy and robustness in capturing the complex influences on agricultural prices. In the future, the proposed fuzzy-based model can be applied to multivariate time series analysis by simultaneously considering multiple

Table 5 Score and rank of TOPSIS method of implemented models for all commodities

Commodity	TOPSIS	ARIMA	SVR	RF	ANN	F-ARIMA	F-SVR	F-RF	F-ANN
Rice	Score	0.64	0.15	0.59	0.32	0.99	0.60	0.84	0.92
	Rank	4	8	6	7	1	5	3	2
Wheat	Score	0.06	0.23	0.66	0.47	0.53	0.67	0.97	0.31
	Rank	8	7	3	5	4	2	1	6
Maize	Score	0.29	0.84	0.90	0.00	0.98	0.91	1.00	0.80
	Rank	7	5	4	8	2	3	1	6
Gram	Score	0.77	0.95	0.92	0.00	0.95	0.96	1.00	0.73
	Rank	6	4	5	8	3	2	1	7
Urad	Score	0.57	0.94	0.92	0.01	0.99	0.99	0.99	0.83
	Rank	7	4	5	8	1	3	2	6
Lentil	Score	0.62	0.35	0.98	0.09	0.91	0.61	0.95	0.87
	Rank	5	7	1	8	3	6	2	4
Mung	Score	0.70	0.92	0.96	0.03	0.94	0.97	1.00	0.84
	Rank	7	5	3	8	4	2	1	6
Arhar	Score	0.68	0.90	0.95	0.04	0.97	0.96	1.00	0.76
	Rank	7	5	4	8	2	3	1	6
Potato	Score	0.62	0.94	0.93	0.04	0.98	0.98	1.00	0.73
	Rank	7	4	5	8	2	3	1	6
Onion	Score	0.14	0.43	0.89	0.26	0.85	0.90	1.00	0.79
	Rank	8	6	3	7	4	2	1	5
Tomato	Score	0.28	0.54	0.79	0.00	0.99	0.69	0.97	0.88
	Rank	7	6	4	8	1	5	2	3
Soyabean	Score	0.22	0.32	0.89	0.15	0.57	0.33	0.98	0.68
	Rank	7	6	2	8	4	5	1	3
Sunflower	Score	0.21	0.13	0.78	0.58	1.00	0.58	0.95	0.79
	Rank	7	8	4	5	1	6	2	3
Groundnut	Score	0.50	0.01	0.75	0.44	0.74	0.36	1.00	0.63
	Rank	5	8	2	6	3	7	1	4
Mustard	Score	0.45	0.28	0.78	0.71	0.66	0.39	0.75	0.87
	Rank	6	8	2	4	5	7	3	1
Turmeric	Score	0.39	0.85	0.89	0.00	0.94	0.87	1.00	0.92
	Rank	7	6	4	8	2	5	1	3
Coriander	Score	0.23	0.61	0.65	0.02	0.91	0.89	0.89	0.96
	Rank	7	6	5	8	2	4	3	1
Cumin	Score	0.15	0.31	0.44	0.07	0.63	1.00	0.99	0.77
	Rank	7	6	5	8	4	1	2	3

input variables that influence the output, capturing their interdependencies and interactions. This is vital in areas like finance for risk assessment and credit scoring by analyzing economic factors, demand forecasting by considering seasonality, market trends, and inventory levels, healthcare for predicting patient outcomes using health indicators, environmental modeling for forecasting climate conditions based on meteorological data, and quality control by analysing multiple production metrics. The proposed FuzzyML framework shows strong potential for real-time forecasting and broader applicability. In future work, real-time prediction can be obtained by incorporating continuous data updates and stream-based learning techniques. This will help in making timely decisions in dynamic agricultural markets. Future work will also explore the sensitivity of the model to variations in the number of clusters and fuzzification parameters, examining their impact on robustness and predictive accuracy.

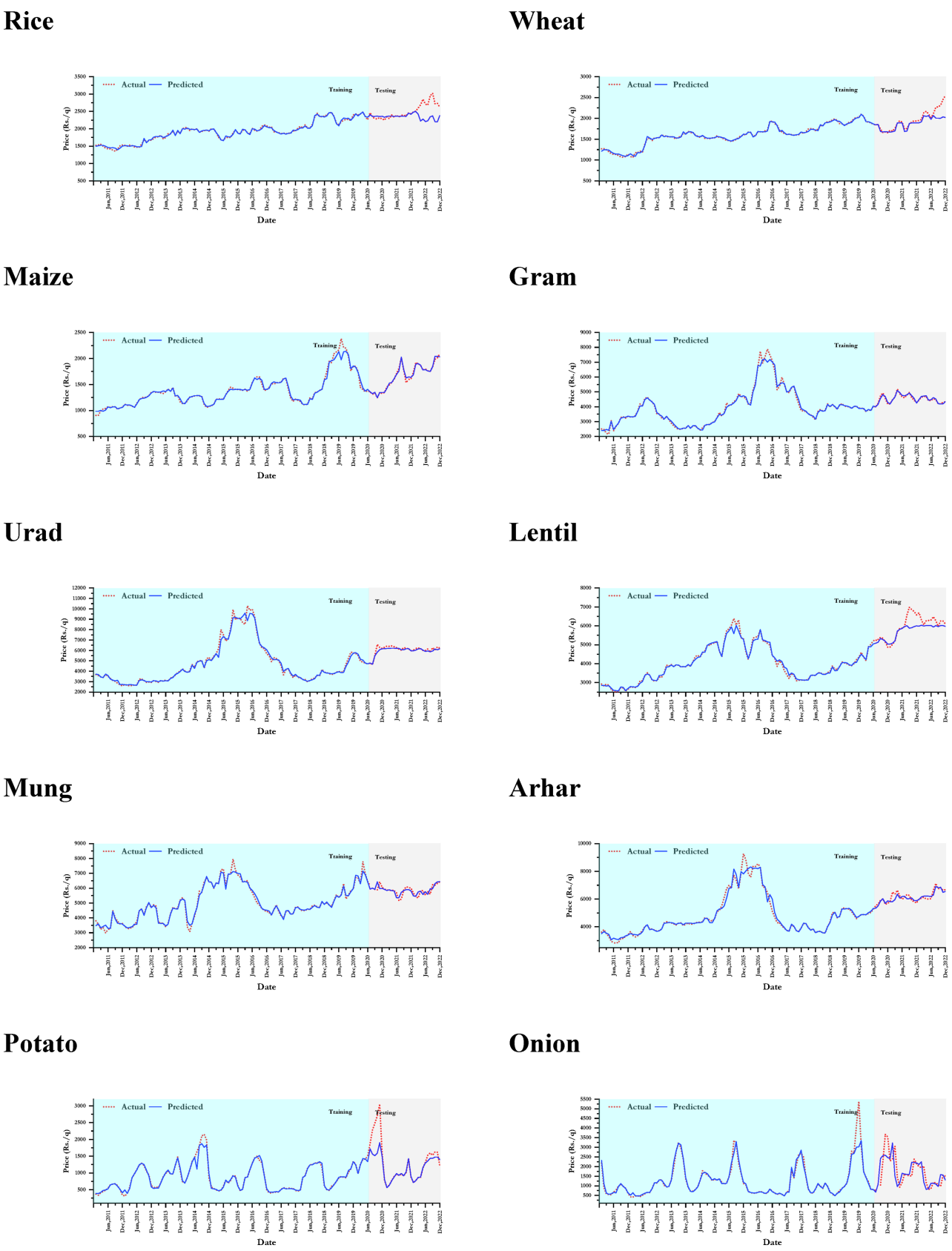
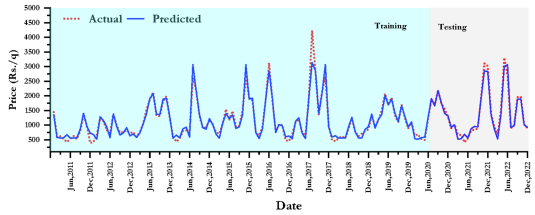
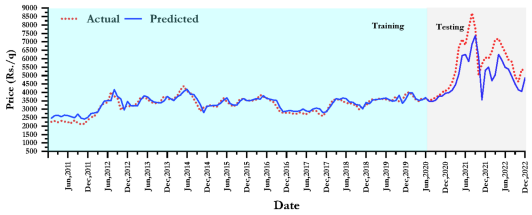


Fig. 8 Actual and predicted values of best fitted model for different commodity

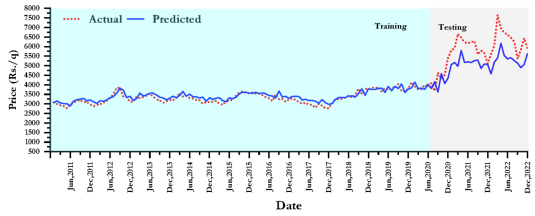
Tomato



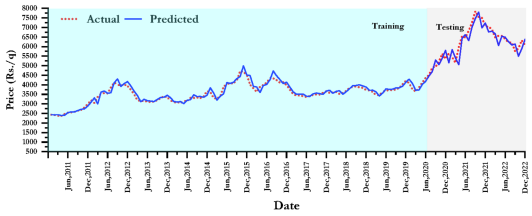
Soyabean



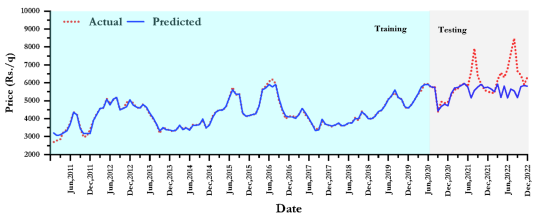
Sunflower



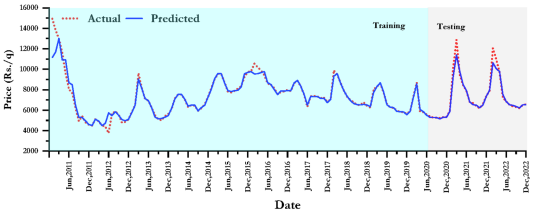
Mustard



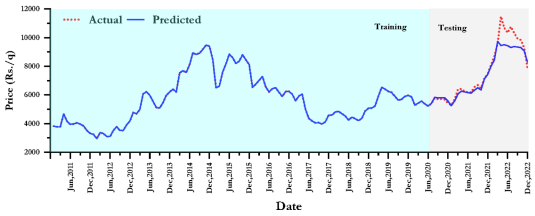
Groundnut



Turmeric



Coriander



Cumin

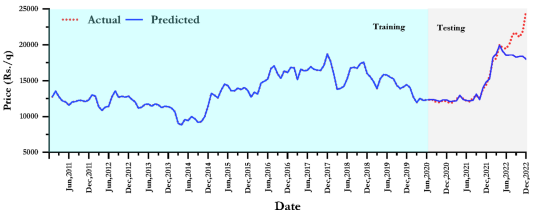


Fig. 8 (continued)

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43538-025-00482-3.

Acknowledgements The authors are thankful to the anonymous reviewers and the editor for their valuable suggestions.

Funding This research received no external funding.

Data and code availability Data and code will be available on reasonable request.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Adnan, R.M., Mostafa, R.R., Islam, A.R.M.T., Kisi, O., Kuriqi, A., Heddam, S.: Estimating reference evapotranspiration using hybrid adaptive fuzzy inferencing coupled with heuristic algorithms. Comput. Electron. Agric. **191**, 106541 (2021)

Adnan, R.M., Dai, H.L., Kisi, O., Heddam, S., Kim, S., Kulls, C., Zounemat-Kermani, M.: Modelling biochemical oxygen demand using improved neuro-fuzzy approach by marine predators algorithm. Environ. Sci. Pollut. Res. **30**(41), 94312–94333 (2023)

Ahmed, N.K., Atiya, A.F., El Gayar, N., El-Shishiny, H.: An Empirical Comparison of Machine Learning Models for Time Series Forecasting. Econom Rev **29**, 594–621 (2010). https://doi.org/10.1080/07474938.2010.481556

Andreev, S., Atanassov, K., Sotirov, S.: Generalized net model of a social network with intuitionistic fuzzy estimation. Notes on Intuitionistic Fuzzy Sets **20**, 72–83 (2014)

Atanassov, K.T.: Intuitionistic Fuzzy Sets. Physica-Verlag HD, Heidelberg (1999)

Atanassov, K.: Intuitionistic fuzzy logics as tools for evaluation of Data Mining processes. Knowl Based Syst **80**, 122–130 (2015). https://doi.org/10.1016/J.KNOSYS.2015.01.015

Atanassov, K., Gargov, G.: Interval valued intuitionistic fuzzy sets. Fuzzy Sets Syst. **31**, 343–349 (1989). https://doi.org/10.1016/0165-0114(89)90205-4

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time series analysis: forecasting and control. John Wiley & Sons (2015)

Brandt, J.A., Bessler, D.A.: Price forecasting and evaluation: An application in agriculture. J. Forecast. **2**, 237–248 (1983). https://doi.org/10.1002/FOR.3980020306

Chaitra, B., Meena, K.: Forecasting Crop Price using various approaches of Machine Learning. Proceedings - 2023 International Conference on Innovations in Engineering and Technology, ICIET 2023. (2023) https://doi.org/10.1109/ICIET57285.2023.10220616

Da Costa, C.G., Bedregal, B., Dória Neto, A.D.: Atanassov's intuitionistic fuzzy probability and Markov chains. Knowl Based Syst **43**, 52–62 (2013). https://doi.org/10.1016/J.KNOSYS.2013.01.015

Dubois, D., Prade, H.: Additions of interactive fuzzy numbers. IEEE Trans Automat Contr **26**, 926–936 (1981). https://doi.org/10.1109/TAC.1981.1102744

Dwivedi, A.K., Subramanian, U.K., Kuruvilla, J., Thomas, A., Shanthi, D., Haldorai, A.: Time-series data prediction problem analysis through multilayered intuitionistic fuzzy sets. Soft. Comput. **27**(3), 1663 (2023)

Ghosh, H., Chowdhury, S., Prajneshu,: An improved fuzzy time-series method of forecasting based on L-R fuzzy sets and its application. J. Appl. Stat. **43**, 1128–1139 (2016)

Goodwin, P.: The Holt-Winters Approach to Exponential Smoothing: 50 Years Old and Going Strong. Foresight: The International Journal of Applied Forecasting (2010)

Haykin, S.: Neural networks: a comprehensive foundation, 2nd edn. Prentice Hall PTR (1998)

Huarng, K.: Heuristic models of fuzzy time series for forecasting. Fuzzy Sets Syst. **123**, 369–386 (2001). https://doi.org/10.1016/S0165-0114(00)00093-2

Hwang, C.-L., Yoon, K.: Multiple attribute decision making. Springer, Berlin Heidelberg (1981)

Hyndman R, Koehler A, Ord J, Snyder R (2008) Forecasting with exponential smoothing: the state space approach

Ikram, R.M.A., Cao, X., Sadeghifar, T., Kuriqi, A., Kisi, O., Shahid, S.: Improving significant wave height prediction using a neuro-fuzzy approach and marine predators algorithm. J. Marine Sci. Eng. **11**(6), 1163 (2023)

Jajarmi, E., Sajjadi, S.A., Mohebbi, J.: Predicting the relative density and hardness of 3YPSZ/316L composites using adaptive neuro-fuzzy inference system and support vector regression models. Measurement **145**, 472–479 (2019)

Khashei, M., Bijari, M.: An artificial neural network (p, d, q) model for timeseries forecasting. Expert Syst. Appl. **37**, 479–489 (2010). https://doi.org/10.1016/J.ESWA.2009.05.044

Kumar, S., Gangwar, S.S.: A fuzzy time series forecasting method induced by intuitionistic fuzzy sets. Int. J. Modeling, Simulation, Scient Comput. **6**(04), 1550041 (2015)

Liakos, K.G., Busato, P., Moshou, D., et al.: Machine learning in agriculture: a review. Sensors (2018). https://doi.org/10.3390/S18082674

Mavrov, D., Atanassova, V., Bureva, V., et al.: Application of game method for modelling and temporal intuitionistic fuzzy pairs to the forest fire spread in the presence of strong wind. Mathematics (2022). https://doi.org/10.3390/MATH10081280

Mitra, D., Paul, R.K.: Hybrid time-series models for forecasting agricultural commodity prices. Model. Assist. Stat. Appl. **12**, 255–264 (2017). https://doi.org/10.3233/MAS-170400

Nasiri, H., Ebadzadeh, M.M.: MFRFNN: multi-functional recurrent fuzzy neural network for chaotic time series prediction. Neurocomputing **507**, 292–310 (2022). https://doi.org/10.1016/J.NEUCOM.2022.08.032

Palomares-Salas, J.C., Agüera-Pérez, A., de la Rosa, J.J.G., Moreno-Muñoz, A.: A novel neural network method for wind speed forecasting using exogenous measurements from agriculture stations. Measurement **55**, 295–304 (2014)

Parvathi, R., Malathi, C., Akram, M., Atanassov, K.T.: Intuitionistic fuzzy linear regression analysis. Fuzzy Optim. Decis. Making **12**, 215–229 (2013). https://doi.org/10.1007/S10700-012-9150-9/METRICS

Paul, R.K., Garai, S.: Performance comparison of wavelets-based machine learning technique for forecasting agricultural commodity prices. Soft. Comput. **25**, 12857–12873 (2021). https://doi.org/10.1007/S00500-021-06087-4/METRICS

Paul, R.K., Yeasin, M., Kumar, P., et al.: Machine learning techniques for forecasting agricultural prices: a case of brinjal in Odisha. India. Plos One **17**, e0270553 (2022). https://doi.org/10.1371/JOURNAL.PONE.0270553

Sankaran, S.: Demand forecasting of fresh vegetable product by seasonal ARIMA model. Int. J. Operational Res. **20**, 315–330 (2014). https://doi.org/10.1504/IJOR.2014.062453

Sarkar, A., Yeasin, M., Paul, R.K., Paul, A.K., Singh, A.K.: WaveFL-STM: Wavelet-based fuzzy LSTM model for forecasting complex time series data. Neural Comput. Appl. (2024a). https://doi.org/10.1007/s00521-024-10622-3

Sarkar, A., Yeasin, M., Paul, R.K., Singh, A.K. and Paul, A.K.: IFNN: intuitionistic fuzzy logic based neural network model for time series forecasting. National Academy Science Letters, pp.1–6. (2024) https://doi.org/10.1007/s40009-024-01522-0

Scott, S.L., Varian, H.R.: Predicting the present with Bayesian structural time series. Int. J. Math. Model. Numer. Optimisation **5**, 4–23 (2014). https://doi.org/10.1504/IJMMNO.2014.059942

Shengwei, W., Yanni, L., Jiayu, Z., Jiajia, L.: Agricultural price fluctuation model based on SVR. In: Proceedings of 2017 9th International Conference On Modelling, Identification and Control, ICMIC 2017. Institute of Electrical and Electronics Engineers Inc., pp 545–550 (2017)

Sinta, D., Wijayanto, H.: Ensemble K-Nearest neighbors method to predict rice price in Indonesia. Appl. Math. Sci. **8**, 7993–8005 (2014)

Song, Q., Chissom, B.S.: Fuzzy time series and its models. Fuzzy Sets Syst. **54**, 269–277 (1993). https://doi.org/10.1016/0165-0114(93)90372-O

Torbat, S., Khashei, M., Bijari, M.: A hybrid probabilistic fuzzy ARIMA model for consumption forecasting in commodity markets. Econ Anal Policy **58**, 22–31 (2018). https://doi.org/10.1016/J.EAP.2017.12.003

Tseng, F.M., Tzeng, G.H., Yu, H.C., Yuan, B.J.C.: Fuzzy ARIMA model for forecasting the foreign exchange market. Fuzzy Sets Syst. **118**, 9–19 (2001). https://doi.org/10.1016/S0165-0114(98)00286-3

Vapnik, V.: The nature of statistical learning theory. Springer science & business media (1999)

Vui, C.S, Soon. G.K, On. C.K., et al.: A review of stock market prediction with artificial neural network (ANN). In: Proceedings - 2013 IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2013. pp 477–482 (2013)

Yaslan, Y., Bican, B.: Empirical mode decomposition based denoising method with support vector regression for time series prediction: a case study for electricity load forecasting. Measurement **103**, 52–61 (2017)

Zhang, Y., Na, S.: A novel agricultural commodity price forecasting model based on fuzzy information granulation and MEA-SVM model. Math. Probl. Eng. (2018). https://doi.org/10.1155/2018/2540681

Zhang, J., Lyu, D., Xiang, J.: A model-data-fusion method for real-time continuous remaining useful life prediction of lithium batteries. Measurement **238**, 115312 (2024)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.