Experimental Evaluation of Agriculture and Horticulture Commodities Price Prediction Using Histogram Based Gradient Boosting Algorithm

S. Saranya, Arul Murugan N., Bharanidharan K., Jeevanantham C. and Jeevak S.
Computer Science and Engineering, Mahendra Engineering College, Namakkal, Tamil Nadu, India

Keywords: Agriculture Price Forecasting, Horticulture, RF, Commodity Price Prediction, Histogram, Gradient Boost, HGB, Random Forest.

Abstract: Accurately predicting agricultural prices are critical to achieve the sustainable and healthy growth of agriculture, which is why agricultural price prediction is a prominent study issue in the sector. On the flip side, it's affected by a lot of things, the most important of which are the fluctuations in agricultural commodity prices. This paper dives into investigating the pricing patterns of important agricultural commodities among different producers worldwide, acknowledging the potential of Deep Learning in agricultural applications. Farmers, dealers, and lawmakers all have a critical responsibility for agricultural price prediction to enable them to make informed decisions regarding planting, pricing, and distribution. Agriculture is a notoriously complex and cluttered market, and typical price perspective models have proven themselves incapable. Recent research has shown that deep learning algorithms analyzing large volumes of historical data to identify nonlinear dependencies, significantly improves the accuracy of price predictions. We propose Histogram based Gradient Boosting (HGB), a time series dominate based deep learning model that predicts agricultural prices. The standard learning model Random Forest (RF) is used to cross-validate the effectiveness of the proposed model. Along with historical pricing, the proposed model is trained on other influencing factors such as seasonality, weather, and market demand indicators. The predictions made using the deep learning one were more accurate and robust in comparison to more traditional models. Experimental results suggest that this approach can improve on-farm decision-making and result in more efficient and stable market systems.

1 INTRODUCTION

The agricultural products market is highly sensitive to price information; large and frequent price spikes have had a devastating effect on people's ability to earn a living and on social stability. Apart from the economic conditions of individual countries or areas, Agricultural prices forecasting should also be done with a balance between food supply and demand across the globe (Girish Hegde et al., 2020) With the exponential growth of the human population, the subject of food security has become global. Accurate forecast of commodity prices should be beneficial to international organizations, governments, and agribusinesses in terms of ensuring there is enough food supply around the world, which maintains the global food system. Therefore, it is important to properly predict the prices of agricultural goods to enhance the quantity-based safety of agricultural products and promote social and economic development (Fajar Delli Wihartiko, et al., 2021). More variables lead to increased impermanence, and, thus, agricultural commodity prices are more volatile, complicated and non-stationary than general commodity prices. Food security on a national and international scale might be impacted by the regular and extreme swings in the pricing of agricultural commodities. Agrarian pricing is heavily influenced by market forces of supply and demand, according to the research. Prices are subject to change since production influences both supply and demand.

Factors including the cost of labour, the cost of growing, and the state of the global market all have an impact on the prices of agricultural commodities (Laveti Krishna Babu, 2024). The practice of using scientific methods to predict, from current and historical data, the direction and magnitude of future

Saranya, S., N., A. M., K., B., C., J. and S., J.
Experimental Evaluation of Agriculture and Horticulture Commodities Price Prediction Using Histogram Based Gradient Boosting Algorithm.
DOI: 10.5220/0013871900004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 1*, pages 730-737
ISBN: 978-989-758-777-1

price changes in agricultural products is known as agricultural product forecasting. There are two main approaches to predicting agricultural prices: qualitative and quantitative. (Peng Chen et al., 2023.) In qualitative analysis, all available market price data is considered, and an overall trend in price direction is made based on past experience; in quantitative analysis, all available market price data is compiled, and specific quantitative judgments about the number or magnitude of commodity price changes are made using specific forecasting methods.

Agricultural price forecasting primarily makes use of quantitative analysis, which can be further subdivided into several methods based on the variables being considered, such as univariate and multivariate forecasting, regression analysis (also known as causal analysis), machine learning, time series analysis, and combined models (M. Durga Sai Sandeep, et al.,2025). Supply and demand, weather patterns, government actions, market rivalry, foreign commerce, etc. all have an impact on the pricing of agricultural products. (Dian Dharmayanti, et Al.,2024) It is challenging to represent and quantify prices and the interactions between elements using simple mathematical models since they are typically nonlinear, dynamic, and unpredictable. While conventional methods are simple and quick to use, they need more a priori knowledge and assumptions and have poor prediction impacts when dealing with high-dimensional, nonlinear, and non-smooth data. There are a number of drawbacks to intelligent approaches, including their inability to handle complicated data reliably and consistently, their high data and computing resource requirements, and their lack of interpretability and stability (Sourav Kumar Purohit, et al.,2021).

A significant portion of the overall expenditure on food production in developing countries is attributed to inefficient supply networks, as reported by the World Food Programme. Predictions are crucial to the handling of these challenges. Everyone from farmers to lawmakers may benefit from reliable predictions of global food prices when it comes to strategic planning and making educated decisions. An accurate forecast may reduce price volatility by 20%, says the International Food Policy Research Institute. Progress in AI and ML has led to an uptick in the precision of predicting models in recent years (Luana Gonçalves Guindani, et al.,2024). By 2030, the global agriculture sector could reach into a $2.3 trillion market made possible by digital technologies like AI and ML, according to the World Economic Forum. This study covers the history of agricultural commodity price prediction with the simple solution methods, hybrid combination models, conventional and intelligent prediction approaches. (Manas Kumar Mohanty, et al., 2023) This article has anover view of the methods that are used for forecasting the prices of the agricultural products, it states each with their advantages and disadvantages along with real-life examples, and then describes the overall pathway of growth within this area. The present research first discusses the history of agricultural price forecasting tools and then analyzes their currently underlying development status.

2 RELATED WORKS

In this paper, a hybrid forecasting model based on VMD, EEMD, and LSTM is proposed to address this problem and solve the large prediction errors caused by the non-linear features and the large price fluctuations of agricultural products (Changxia Sun, et al., 2024). This model is known by the acronym "VMD-EEMD-LSTM". Starting from the initial time series of agricultural commodity prices, which is decomposed using VMD, we see a residual component that is more complex. So, what you end up getting from this process is called the VMFs, or variationally mode functions. After that, the remaining part is decomposed again using EEMD. Each component's predictions are then derived by training an LSTM model with all of the components. Lastly, the most accurate price prediction is obtained by linearly combining the forecasts for all components. We conducted empirical experiments to assess the VMD-EEMD-LSTM model's efficiency for one-step and multi-step predictions using weekly pricing data from China's wholesale agricultural marketplaces for shiitake mushrooms, cauliflower, Chinese chives, and pork. This study's composite model improved predicting accuracy, as shown by the findings.

Accomplishing Sustainable Development Goal 2, "Zero Hunger" (Anket Patil, et al.,2023), and enhancing human health and social well-being relies on achieving food security globally. However, the volatility of agricultural commodity prices is just one of several factors that influence food insecurity. This paper dives into investigating the pricing patterns of important agricultural commodities among different producers worldwide, acknowledging the potential of Machine Learning in agricultural applications. This paper presents a Hybrid SARIMA-LSTM (HySALS) to predict the worldwide values of agricultural commodities based on extensive testing and performance comparison of appropriate Machine Learning algorithms. This study examines the suggested method by looking at five key commodities' price histories: wheat, millet, sorghum, maize, and rice. Developing nations that produce a major portion of the world's these crops or are among the top producers globally are the primary focus of the study, along with the average production share worldwide. From 2005 to 2017, we use training data. From 2018 to 2022, we test the model. From 2023 to 2030, we predict the worldwide prices of key commodities. The goal of making these forecasts is to provide farmers and policymakers with more information they can use to make informed decisions that will help in the global struggle for food security.

In support of the viability of the economy and agricultural development, there needs to be stability in the agricultural futures market (Tingting Zhang, et al.,2023). Due to the complexity of changes in agricultural futures prices, it is not easy to overcome the restrictions that the current data preprocessing techniques put in place for improvement of the models' ability to forecast. In this study, we propose a novel VMD-SGMD-LSTM model that combines state-of-the-art quadratic decomposition with an AI framework. First, we use VMD to clean up the raw futures price data, and then we let SGMD deal with the rest of the components. Secondly, several modal components are predicted using the LSTM model, and then the result is achieved using the predicted values from the different components. In addition, using data from the Chinese agricultural futures market for wheat, maize, and sugar, this study offers empirical analysis in one-step, two-step, and four-step forward forecasting scenarios, respectively. By outperforming other benchmarked models in terms of predictive power and resilience across several agricultural futures, the results show that the VMD-SGMD-LSTM hybrid model suggested here overcomes the constraints of earlier research.

Predicting agricultural prices accurately is critical to achieving the agricultural sector's sustainable and healthy development, making it a popular study issue in the sector (Feihu Sun, et al.,2023). It delves into the many ways of forecasting, including classic, intelligent, and combination model approaches, and discusses the difficulties that researchers have when trying to estimate the prices of agricultural commodities. The findings of the study propose the following: (1) the ARIMA and exponential smoothing price forecasting of agricultural products will be a developing trend for the future, and understanding the reasons for the combination will help improve accurate forecasting; (2) future forecasting models will continue to incorporate structured, unstructured data, and variables; and (3) when forecasting these agricultural product price estimates, accuracy of values in addition to trend forecasting accuracy will be advantageous. This manuscript serves to progress a future durability research agenda, as it reviews and analyses price forecasting agricultural product methods.

Predicting agricultural commodity prices with any degree of accuracy is difficult because of how complicated and unpredictable these markets are (Kapil Choudhary, et al., 2025). Predictions made using current models are generally inaccurate because they do not account for non-stationary and nonlinear trends in pricing data. A new hybrid VMD-LSTM model is introduced to address these challenges; it combines genetic algorithm, variationally mode decomposition and long short-term memory (LSTM) to enhance prediction accuracy. The proposed model uses GA-optimized VMD, a technique for breaking down price series into intrinsic mode functions (IMFs) with the desirable property of sparsity, to speed up the convergence process. Next, model and forecast each of these IMFs separately using LSTM models that have been optimized using GA. The final step in generating the actual price series is to combine the predictions of all IMFs. The VMD-LSTM is put to the test in comparison to three other LSTM and decomposition-based models using monthly pricing data for maize, palm oil, and soybean oil (EEMD-LSTM, CEEMDAN-LSTM). It is possible to measure the efficacy using directional prediction statistics, root mean square error and mean absolute percentage error. As compared to the next best CEEMDAN-LSTM, VMD-LSTM decreases RMSE by 56.93%, MAPE by 44%, and palm oil by 21.67% and soybean oil by 25.85%, respectively. The improved prediction accuracy of VMD-LSTM is further supported by TOPSIS and the Diebold-Mariano test. Farmers, dealers, and policymakers might all benefit from the proposed model's improved agricultural price predicting capabilities.

3 METHODOLOGY

The research on the topic of agricultural product price forecasting is one that is all-encompassing, interdisciplinary, and constantly evolving. Data sources, data types, data quality, data processing techniques, model design techniques, and model evaluation techniques are always evolving, which means that the methods used to predict the values of agricultural goods will also be improved and updated. The employment of combination optimization methods by predictive models has been demonstrated to outperform models utilizing a single optimization methodology by researchers over an extended period of time. Combination parameter optimization techniques have several benefits over single parameter optimization algorithms, including their capacity to tackle complicated optimization problems including discrete, nonlinear, and multi-modal functions. While the latter have difficulty with issues like differentiability and convexity, the former usually necessitate them. One advantage of combination parameter optimization approaches over single parameter optimization algorithms is their ability to avoid local optima, which can be caused by initial value influence and lead to slow convergence or inferior solutions. In addition, different problems have different needs and combination parameter optimization techniques may respond to that. For example, the are free to apply neighborhood structures, multiple fitness functions and crossover mute strategies. On the other hand, if optimization algorithms only consider one parameter, they are usually less flexible and reach more stable results, which are harder to adapt or improve. Combination parameter optimization strategies have notable downsides including their computational complexity, theoretical analytical reliance, and sensitivity to parameter selections. So, in real-world applications, one of the critical challenges is to select the right optimization algorithms to tackle a certain problem under a particular set of objectives to make the required adjustments and to enhance the solution. The introduction and widespread use of social media has amplified the influence of public opinion on consumers and farmers. This influence is complex and leads to at least unreasonable planting or purchasing behavior.

For the purpose of price forecasting and study of price dynamics, Figure 1 shows the proposed method flow diagram. The suggested method, Histogram based Gradient Boosting (HGB), improves upon the present approach, Random Forest (RF), by making it more practical and increasing the accuracy of forecasts. Figure 2 of the system architecture above shows the suggested method for price forecasting and price dynamics analysis. We consider two types of data in order to tackle the worldwide food security problem that this study aims to solve: One is the Crop data, which includes detailed information on the commodity, nation, price (in the currency of the country in which the crop is grown), and quantity (for each month and year) among other factors that impact the crop's cost on a worldwide scale. The price data is another important part of the process since it includes the nation, month, year, and standard currency value. These pieces of information are utilized to convert currency units to US dollars.

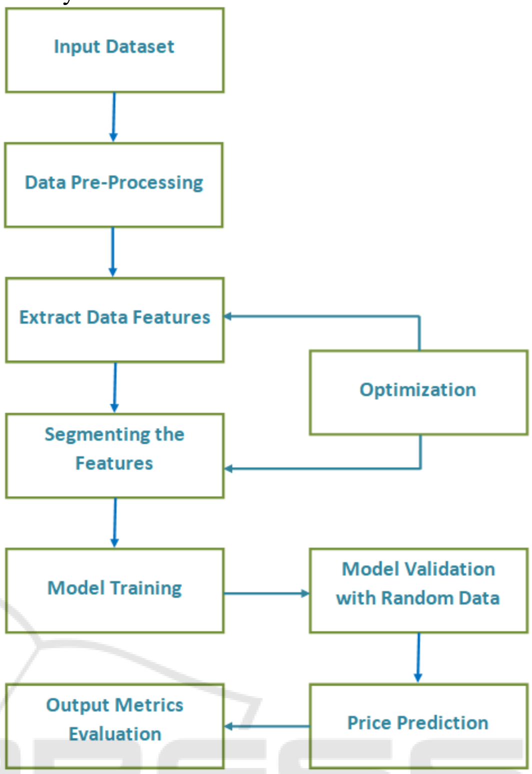

Figure 1: System Flow Diagram.

(i) Data Acquisition: To address the issue of food security on a worldwide scale, this study draws on two sets of data: The Crop data is one such resource; it provides comprehensive information on the crop's worldwide cost by month and year as well as by product, nation, price (in currency unique to that country), and quantity. An additional source is the price data, which includes the following fields: nation, month, year, and standard currency value. These fields are utilized for currency unit conversions to US dollars.
(ii) Data Mapping: Common characteristics, such as nation, year, and price, are used to map the two sets of data. More accurate and insightful analysis is made possible by ensuring thorough and integrated data generation.

- **Data Pre-Processing:** The procedures outlined in the next section are used to pre-process the data.
- **Missing Values Handling:** The dataset is filled up with the average price for each unique set of country and year to replace any missing values.

(iii) Data Grouping: Relevant data points are grouped from the dataset by filtering and grouping appropriate nations and commodities within a range.

(iv) Weight Standardization: Weight standardization is a method for transforming different weights into a universally accepted amount, in this case 1 kilogram. This guarantees that all measurements taken within the system are consistent and uniform.

(v) Data Validation: In order to verify the accuracy of the recorded information and the logical coherence of the values across different features, data consistency tests were conducted by comparing the precision of the crop data with the accessible data.

Grey models, regression analysis, and time series forecasting are some of the most conventional methods for projecting agricultural prices. These methods are effective in situations when the variables are independent, the data is normally distributed, and the link is either linear or simple nonlinear. Nevertheless, these requirements are never satisfied by practical agricultural price forecasting, which frequently poses complicated issues such nonlinearity, high dimensionality, and short sample size. Intelligent forecasting methods are able to successfully represent nonlinear interactions in price series and have less modelling constraints and assumptions than econometric and mathematical-statistical approaches. Decision trees, support vector machines, as well as plain Bayes are examples of conventional machine learning approaches; however, despite their simplicity, speed to train, and robustness, these methods struggle with complicated nonlinear interactions, require human involvement for feature selection and extraction, and provide poor generalization.

Feature engineering is unnecessary when deep learning models are properly supervised, given sufficient data quantity and quality, and allowed to extract feature information from the original sequence. Additionally, they excel at handling sequences with nonlinear dependencies. Constraints of the deep learning approach include an inability to easily alter parameters, a high data volume required, the risk of overfitting, and a lack of interpretability. In real-world forecasting scenarios, many forecasting approaches like RF can be used to solve the same problem because of diverse modelling systems and starting points. Various forecasting approaches offer varying insights, each with their own set of pros and cons. Rather than being incompatible, they work hand in hand and complement one another. Agricultural commodity price forecasting models sometimes have less-than-ideal characteristics when they are first established. To get a better prediction model in these situations, optimizing the parameters is essential. Parameter optimization commonly employs techniques such as simulated annealing, particle swarm optimization, genetic algorithms, grid search, and cross validation.

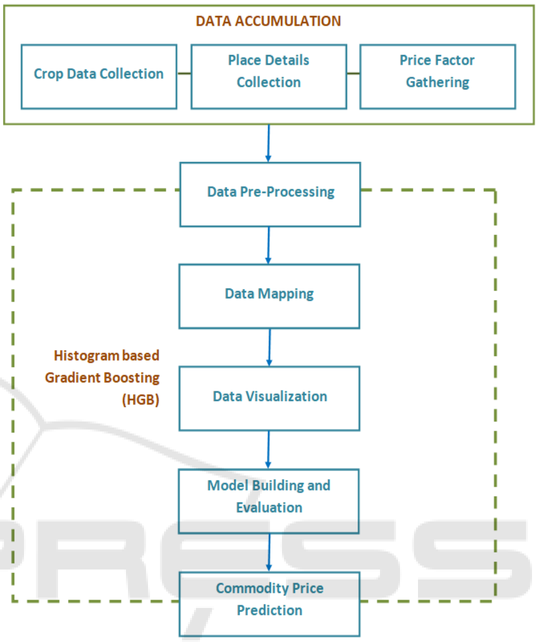

Figure 2: Architecture Diagram.

A more scientific way would be to integrate many valid forecasting methods into one, which is known as the combined forecasting technique. Combination forecasting models combine two or more models to predict variables. They are more accurate, comprehensive, and make better use of sample data than individual models. This helps with synthesizing useful information from different methods and improves forecasting accuracy. Since the goal of the project is to forecast food costs throughout the world, it is necessary to convert all rates into the same currency. The prices are converted to USD using the currency rates of the other nations included in the data at the specified timestamps. The following figure, Figure 2 represents the system architecture diagram of the proposed approach.

The annual price trends of many commodities may be examined through data visualization. To get a better picture of when crop prices were highest and lowest, we may sort the price data by year and assign each one an equal amount of weight.

4 RESULTS AND DISCUSSION

The study applies ML capabilities to address the fundamental issue of global food security that is affected by the price of agricultural commodities. Two significant results are achieved as a result of the work: In the first place, an examination of the price fluctuations of important agricultural commodities is carried out, with a special focus on Wheat, Millet, Sorghum, Maize, and Rice. The study sheds insight on the patterns and shifts that have occurred in the pricing on a worldwide basis. Particular attention is paid to emerging countries that are either the most prolific producers of these crops or the ones that attain the maximum output of this crop in comparison to other producing countries. The price dynamics study and anticipated prices provide valuable information to policymakers, farmers, researchers, as well as other stakeholders in order to ensure global food security by mitigating the effects of price-impacting variables, fostering sustainable agriculture, and making informed decisions. For stakeholders including farmers, merchants, and legislators, precise forecasting of agricultural commodity prices is essential.

Conventional forecasting methods such as Random Forest and the like often struggle with the nonlinear and highly volatile nature of agricultural price data. Deep learning algorithms have demonstrated potential for tackling these challenges due to their ability to capture complex patterns on the time series. This approach introduced HGF (Histogram based Gradient Boosting), a new deep learning method that is particularly useful for temporal sequence modelling and has been successfully implemented for price predictions. Four of the recommended models include external variables such temperature and precipitation. When these models were employed on potato, onion, and tomato's pricing in major Indian markets, they outstripped traditional statistical and machine learning approaches, thereby underscoring the importance of including external entities into prediction models by realizing diminished error metrics.

Machine learning techniques, particularly HGB algorithm and its variants, have shown significant potential in improving the accuracy of agricultural commodity price forecasts. Which not only contain external factors but also cover hybrid models helps to better accuracy of integration forecasting and thus can provide useful input for agriculture sector actors to make reasonably informed decisions. For various groups including farmers, merchants, and legislators, accurate forecasting of agricultural prices is critically important. Agricultural pricing data is nonlinear and volatile, making it sometimes difficult to handle with traditional statistical methods. Due to their ability to capture complex patterns in time series data, deep learning algorithms have shown promise in addressing these challenges. Deep learning models have exhibited potential in addressing these challenges as a result of their ability to capture complex regularities within time series data. Figure 3 shows the user registration page of the proposed method. 3, requires the user or farmer to verify their identification before visiting to the login page.

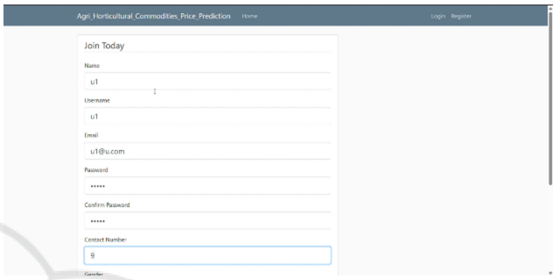

Figure 3: User Registration.

The results of the proposed method's user login, commodity details registration, and commodity price prediction pages are shown in Figures 4, 5, and 6, respectively.

The accuracy ratio of the proposed scheme is evaluated by cross-validating it with the standard learning model called Random Forest (RF). The accompanying figure, Figure 7, shows the output prediction accuracy of the proposed technique HGB. Table 1 provides a descriptive representation of the same.

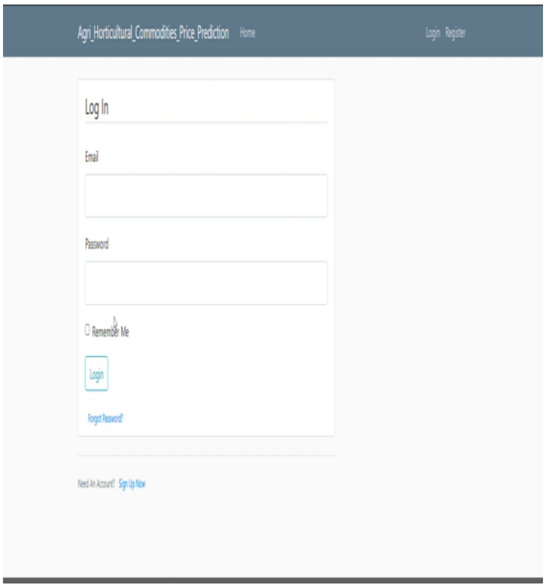

Figure 4: Login Page.

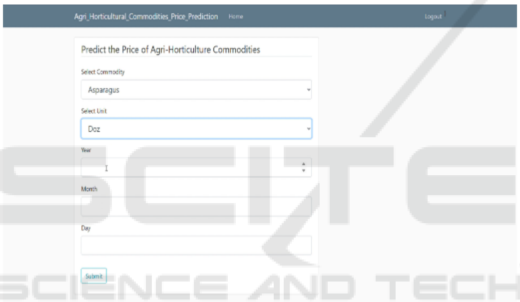

Figure 5: Commodity Details.

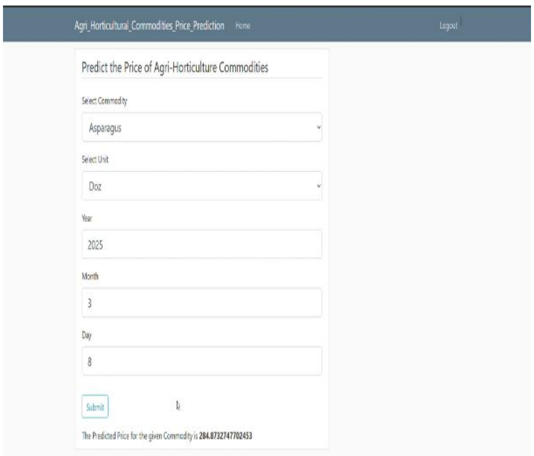

Figure 6: Commodity Price Prediction.

Table 1: Analysis of Commodity Price Prediction Accuracy Between Rf and Hgb.

Epochs	RF (%)	HGB (%)
50	86.36	96.73
75	87.39	97.46
100	88.54	96.59
125	86.67	96.48
150	87.76	96.32
175	87.87	97.64
200	86.98	97.44
225	87.08	97.71
250	86.19	97.28

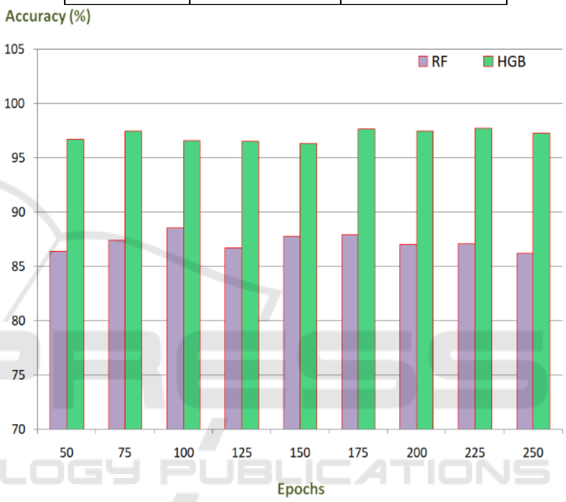

Figure 7: Commodity Price Prediction Accuracy.

5 CONCLUSION AND FUTURE SCOPE

The unpredictable and multi-factorial character of agricultural markets makes agricultural price prediction a difficult task. In this unit, we saw how two state-of-the-art deep learning algorithms, Random Forests (RF) and Histogram-based Gradient Boosting (HGB), non-traditional methods of agricultural price forecasting, might better than more traditional methods. Agricultural prices are affected by complex non-linear interactions among multiple variables such as weather patterns, supply and demand shocks, market trends, etc.: such interactions were captured using ensemble learning techniques. The results showed that compared with traditional methods, both Random Forests and Histogram-based Gradient Boosting models were more accurate and differential. Due to its relatively simple interpretation through feature importance scores (rankings), RF was particularly successful at identifying non-linear relationships. In contrast, HGB\u2019s prediction performance was closer to CTB, which required more time in terms of prediction, while HGB greatly improved the efficiency of the training speed in handling large datasets. Given the complexity of the data with some structured features such as historical pricing data to unstructured features like weather and market demand indicators, the suggested model performed well. This showcases the algorithms' capacity of handling the variability of continuous information throughout different land points for predicting prices in an agricultural setting. The evaluations further proved that meticulous hyperparameter tuning is necessary to achieve high performance and the disasters showed that the models can make accurate predictions for agricultural prices.

In the future, research could focus on price forecasting by incorporating climate change, quantity of warehousing, supply to demand ratios per country, and population dynamics as inputs to Machine Learning models. Those elements have a large influence on price dynamics and ultimately on food security.

REFERENCES

Anket Patil, et al., "Forecasting Prices of Agricultural Commodities using Machine Learning for Global Food Security: Towards Sustainable Development Goal 2", International Journal of Engineering Trends and Technology, 2023.

Changxia Sun, et al., "A Study on Agricultural Commodity Price Prediction Model Based on Secondary Decomposition and Long Short-Term Memory Network", Agriculture, 2024.

Dian Dharmayanti, Et Al., "Application Of Data Mining For Predicting Horticultural Commodities Price", Journal of Engineering Science and Technology, 2024.

Fajar Delli Wihartiko, et al., "Agricultural Price Prediction Models: A Systematic Literature Review", Proceeding s of the 11th Annual International Conference on Industrial Engineering and Operations Management, 2021.

Feihu Sun, et al., "Agricultural Product Price Forecasting Methods: A Review", Agriculture, 2023.

Girish Hegde, et al., "A Study On Agriculture Commodities Price Prediction and Forecasting", International Conference on Smart Technologies in Computing, Electrical and Electronics, 2020.

Kapil Choudhary, et al., "A genetic algorithm optimized hybrid model for agricultural price forecasting based on VMD and LSTM network", Scientific Reports, 2025.

Laveti Krishna Babu, "Analysis of AI-ML Models for Prices Forecasting of AgricultureandHorticultural Commodities", International Journal of Research Publication and Reviews, 2024.

Luana Gonçalves Guindani, et al., "Exploring current trends in agricultural commodities forecasting methods through text mining: Developments in statistical and artificial intelligence methods", Heliyon, 2024.

M. Durga Sai Sandeep, et al., "AI-ML Based Price Prediction Model for Agri-Horticultural Commodities", International Journal of Research Publication and Reviews, 2025.

Manas Kumar Mohanty, et al., "Agricultural commodity price prediction model: a machine learning framework", Neural Computing and Applications, 2023.

Peng Chen, et al., "Short-term Forecast of Agricultural Prices Using CNN+LSTM", Proceedings of the 7th International Conference on Intelligent Information Processing, 2023.

Prashantha S, et al., "Agricultural Crop Commodities Price Prediction Using Machine Learning Techniques", International Research Journal of Innovations in Engineering and Technology, 2020.

Sourav Kumar Purohit, et al., "Time Series Forecasting of Price of Agricultural Products Using Hybrid Methods", Applied Artificial Intelligence, 2021.

Tingting Zhang, et al., "Agricultural commodity futures prices prediction based on a new hybrid forecasting model combining quadratic decomposition technology and LSTM model", Frontiers in Sustainable Food Systems, 2024.