ISSN: 2525 - 2186
Tạp chí Khoa học Lạc Hồng, **2026**, 26, 025-029

MÔ HÌNH HỌC MÁY TỔ HỢP CÓ KHẢ NĂNG GIẢI THÍCH CHO DỰ BÁO CHÍNH XÁC TỒN KHO VÀ DOANH SỐ TRONG NÔNG NGHIỆP THỦY CANH

Đặng Đăng Khoa[1], Phạm Thị Hồng Diễm[2], Phạm Minh Triết[3], Phan Văn Nam[4*]

[1] *Trường Đại học Lạc Hồng, Phường Trấn Biên, Tỉnh Đồng Nai, Việt Nam*
[2] *Sở Nội vụ tỉnh Vĩnh Long, Phường Long Châu, Tỉnh Vĩnh Long, Việt Nam*
[3] *Phòng Hậu cần Công an Thành phố Hồ Chí Minh, Phường Cầu Ông Lãnh, Thành phố Hồ Chí Minh, Việt Nam*
[4] *Trường Đại học Nông Lâm, Phường Linh Xuân, Thành phố Hồ Chí Minh, Việt Nam*
* Tác giả liên hệ: *phanvannambd@gmail.com*

THÔNG TIN BÀI BÁO

Ngày nhận:	05/03/2026
Ngày hoàn thiện:	27/03/2026
Ngày chấp nhận:	03/04/2026
Ngày đăng:	15/04/2026

TỪ KHÓA

linear regression, random forest, XGBoost;
Model-agnostic local explanation method;
Sustainable Development Goals;
Stock and Sales predicted.

TÓM TẮT

Nhu cầu lương thực toàn cầu ngày càng gia tăng đòi hỏi các phương thức nông nghiệp bền vững hơn. Mặc dù canh tác thủy canh sử dụng tài nguyên hiệu quả, việc dự đoán mức tồn kho và doanh số bán hàng vẫn gặp nhiều khó khăn do các điều kiện biến động. Nghiên cứu này đề xuất một khung mô hình tổ hợp được tối ưu hóa, tích hợp hồi quy tuyến tính (Linear Regression), rừng ngẫu nhiên (Random Forest) và XGBoost, với trọng số được điều chỉnh bằng thuật toán tiến hóa. Mô hình được đánh giá bằng sai số căn phương trung bình (RMSE) và sai số tuyệt đối trung bình (MAE), cho thấy hiệu quả vượt trội so với các mô hình riêng lẻ, với mức giảm RMSE 41,62% cho dự báo tồn kho và 56,10% cho dự báo doanh số. Ngoài ra, một phương pháp giải thích cục bộ độc lập với mô hình được sử dụng để diễn giải kết quả, qua đó xác định sản lượng thu hoạch và dữ liệu bán hàng trong quá khứ là những yếu tố dự báo quan trọng. Trong tương lai, nghiên cứu sẽ tập trung tích hợp dữ liệu thời gian thực nhằm nâng cao khả năng thích ứng và hiệu quả dự báo của mô hình.

EXPLAINABLE ENSEMBLE LEARNING FOR ACCURATE HYDROPONIC STOCK AND SALES PREDICTION

Dang Dang Khoa[1], Pham Thi Hong Diem[2], Pham Minh Triet[3], Phan Van Nam[4*]

[1] *Lac Hong University, Tran Bien Ward, Dong Nai Province, Vietnam*
[2] *Department of Home Affairs of Vinh Long Province, Long Chau Ward, Vinh Long Province, Vietnam*
[3] *Logistics Division, Ho Chi Minh City Police, Cau Ong Lanh Ward, Ho Chi Minh City, Vietnam*
[4] *Nong Lam University, Linh Xuan Ward, Ho Chi Minh City, Vietnam*
*Corresponding Author: *phanvannambd@gmail.com*

ARTICLE INFO

Received:	Mar 5th, 2026
Revised:	Mar 27th, 2026
Accepted:	Apr 3rd, 2026
Published:	Apr 15th, 2026

KEYWORDS

Linear regression, random forest, XGBoost;
Model-agnostic local explanation method;
Sustainable Development Goals;
Stock and Sales predicted.

ABSTRACT

Growing global food demand calls for more sustainable agricultural practices. While hydroponic farming uses resources efficiently, predicting stock levels and sales remains difficult due to fluctuating conditions. This study proposes an optimized ensemble framework that integrates linear regression, random forest, and XGBoost, with weights refined through an evolutionary algorithm. Evaluated using root mean square error (RMSE) and mean absolute error (MAE), the model outperforms individual approaches, reducing RMSE by 43.82% for stock prediction and 55.3% for sales forecasting. A model-agnostic local explanation method is used to interpret the results, highlighting harvested crop volume and historical sales as key predictors. Future work will focus on integrating real-time data to improve adaptability and forecasting performance.

Doi: https://doi.org/10.61591/jslhu.26.1085
Available online at: https://lhj.vn

1. INTRODUCTION

The growing demand for sustainable food production has drawn increasing attention to hydroponic farming, a cultivation method known for its efficient use of water, space, and nutrients. Despite these advantages, managing inventory in hydroponic systems remains a persistent operational challenge. Farms frequently encounter issues such as overstocking, product shortages, and unnecessary food waste. Earlier studies suggest that weak forecasting practices may result in up to 30% of harvested produce being lost in urban hydroponic operations, largely due to mismatches between supply and market demand [1]. These inefficiencies highlight the need for more reliable predictive approaches to support production planning and reduce losses. Forecasting in hydroponic environments is not straightforward. Crop growth cycles can vary, environmental conditions often shift, and resource availability may fluctuate over time [2]. Such variability introduces complex and nonlinear relationships within operational data, making prediction tasks difficult. Traditional statistical techniques, as well as single machine-learning models, often struggle to capture these dynamics, which can lead to unstable forecasts and suboptimal decision-making in stock management [3].

To address these limitations, this study proposes an optimized ensemble framework that integrates three complementary models: linear regression (LR) [4], random forest (RF) [5], and XGBoost [6]. Each contributes a distinct analytical capability. Linear regression captures broader linear trends in stock and sales movements, while random forest improves robustness by modeling nonlinear interactions and highlighting feature importance [7]. XGBoost further strengthens predictive performance through gradient boosting, enabling the model to learn complex patterns more effectively [8]. By combining these approaches, the framework aims to produce more stable and accurate forecasts than any individual model alone. To refine the contribution of each model, an evolutionary algorithm (EA) is employed to optimize the weighting structure within the ensemble. Rather than assigning fixed weights, the EA iteratively adjusts the influence of each component model, allowing the framework to better adapt to the characteristics of the dataset and minimize forecasting error [9]. This optimization process improves both predictive accuracy and model adaptability in dynamic production environments.

Interpretability is also an important consideration for practical adoption. Predictive models are most useful when decision-makers can understand the reasoning behind their outputs. For this reason, a model-agnostic local explanation method is applied to reveal the variables that most strongly influence prediction outcomes [10]. By highlighting key factors, such as harvest volume or historical sales patterns. The model provides actionable insights that can support more informed decisions in inventory control and demand forecasting.

In addition to its methodological contribution, this work also aligns with broader sustainability objectives. More accurate forecasting helps reduce mismatches between production and demand, thereby lowering the risk of food waste and improving resource efficiency. In this way, the proposed framework contributes to several Sustainable Development Goals (SDGs), particularly SDG 12.3, which focuses on reducing food loss and waste, and SDG 9.3, which emphasizes improving market access and productivity for smaller-scale agricultural producers. By combining improved predictive performance with transparent model interpretation, the framework offers a practical tool for supporting more efficient and sustainable hydroponic farming operations.

2. MATERIALS AND METHODS

The study follows four main stages: data acquisition, model development, model evaluation, and model interpretability. An overview of the workflow is illustrated in Figure 1, with each step described in the following subsections.

2.1 Model Selection

The dataset was collected from a hydroponic farm over 32 weeks. It records several operational variables, including the number of crops planted, the number harvested, and weekly sales. During preprocessing, missing values were handled by estimating the remaining stock based on harvest and sales data and carrying this value forward to the following week. The dataset was then divided into training and testing sets using an 80:20 split to ensure reliable model evaluation [11].

To improve prediction performance, the study employs an EA to determine the optimal weighting of three models: LR, RF, and XGBoost. Instead of assigning equal influence to each model, the EA adjusts their weights based on predictive performance. The objective is to minimize forecasting errors measured by root mean square error (RMSE) and mean absolute error (MAE). In this way, models that perform better contribute more strongly to the final prediction while still benefiting from the stability of the ensemble.

The optimization begins with a population of 100 candidate solutions, each representing a different set of model weights. Their performance is evaluated using RMSE and MAE to identify the most effective combinations. The top 50% of solutions are retained for the next generation, while crossover and mutation operations introduce variation and help avoid premature convergence. Crossover is applied with a probability of 50%, allowing weight combinations to exchange information, while mutation occurs with a 20% probability to explore new candidate solutions. The process runs for 50 generations, gradually refining the weight distribution until the solution stabilizes. Once the optimal weights are determined, the ensemble model is evaluated using standard performance metrics. To improve transparency, a model-agnostic local explanation method is also applied to interpret the predictions and highlight the factors that most influence stock and sales forecasts, providing practical insights for decision-makers [12].

To improve transparency, the study also examines how different variables influence the model's predictions. For this purpose, a model-agnostic local explanation method was applied to interpret the contribution of key features, including the number of harvested crops, sales data, remaining stock, and the number of plants cultivated. By highlighting how these variables shape prediction outcomes, the approach offers clearer insight into the model's decision process and helps build greater confidence among stakeholders using the forecasting results [13].

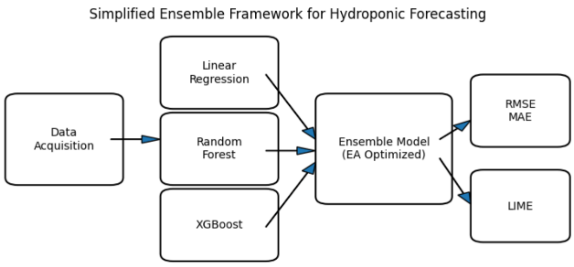

Figure 1. *The overview framework*

3. RESULTS AND DISCUSSION

This section reports the performance of the proposed ensemble framework, along with its interpretability and broader implications for hydroponic farm management. The results suggest that optimizing model weights through an evolutionary algorithm noticeably improves forecasting performance for both stock levels and sales. In addition, incorporating a model-agnostic local explanation method makes the model's predictions easier to interpret, which is important for practical decision-making.

3.1 Model performance comparison

Model performance was assessed using RMSE and MAE. The results presented in Table 1 highlight clear differences in predictive performance among the tested models. The LR produces relatively stable results for remaining stock prediction, with an RMSE of 1.82 and an MAE of 1.07. This suggests that part of the stock variation follows a reasonably linear pattern. However, LR performs less effectively when predicting sales, where the RMSE increases to 5.41, indicating that sales behavior is influenced by more complex factors that cannot be fully captured by a simple linear model.

Table 2. *Case Model performance comparison*

Model	RMSE		MAE	
	stock	sales	stock	sales
LR	1.82	5.41	1.07	2.18
RF	2.91	6.54	1.47	2.10
XGBoost	4.28	15.71	1.15	3.34
Ensemble model	1.0	2.43	0.91	1.67

The RF is designed to capture nonlinear relationships, shows mixed results. While RF is expected to handle complex interactions more effectively, its performance in this dataset appears less stable, particularly in remaining stock prediction where the RMSE rises to 2.91. This may be due to the relatively small dataset, which can limit the model's ability to generalize patterns effectively. For sales forecasting, RF performs slightly better than LR in terms of MAE, but its RMSE remains higher, suggesting occasional larger prediction errors. XGBoost demonstrates the largest prediction errors, particularly for sales forecasting where the RMSE reaches 15.71. Although gradient boosting models are generally powerful for complex datasets, they can become sensitive to limited training data or noise. In this case, the relatively short observation period of 32 weeks may restrict the model's ability to learn stable boosting patterns, leading to overfitting or unstable predictions.

In contrast, the proposed ensemble model consistently achieves the lowest RMSE and MAE values across both prediction tasks. For remaining stock, the ensemble reduces RMSE to 1.00 and MAE to 0.91, indicating a notable improvement in prediction accuracy compared with the individual models. A similar pattern appears in sales forecasting, where the RMSE drops to 2.43 and the MAE to 1.67. These improvements suggest that combining LR, RF, and XGBoost allows the framework to balance their individual strengths while compensating for their weaknesses. Another important observation from Table 1 is the magnitude of improvement achieved by the ensemble approach. Compared with the best-performing individual model, the ensemble significantly reduces prediction errors, particularly for sales forecasting where variability is higher. This result supports the idea that ensemble learning can provide a more robust solution when dealing with dynamic agricultural datasets, where both linear trends and nonlinear relationships coexist. Overall, the findings indicate that relying on a single predictive model may not be sufficient for hydroponic stock and sales forecasting. The optimized ensemble framework offers a more reliable alternative by integrating multiple modeling perspectives and adjusting their contributions through evolutionary optimization. This combination ultimately leads to more accurate and stable predictions, which can support better operational planning in hydroponic farming systems.

The experimental results show that the proposed ensemble model significantly improves forecasting performance compared with individual models. Using RMSE and MAE as evaluation metrics, the optimized ensemble consistently achieves lower error values. In particular, the model reduces RMSE by 41.62% for remaining stock prediction and 56.10% for sales forecasting compared with the best-performing single model. These results indicate that the evolutionary optimization effectively balances the contribution of each base model, leading to more accurate and stable predictions in hydroponic farm operations.

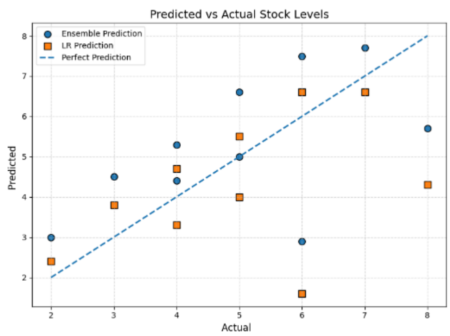

Figure 2. *The predicted vs actual stock levels*

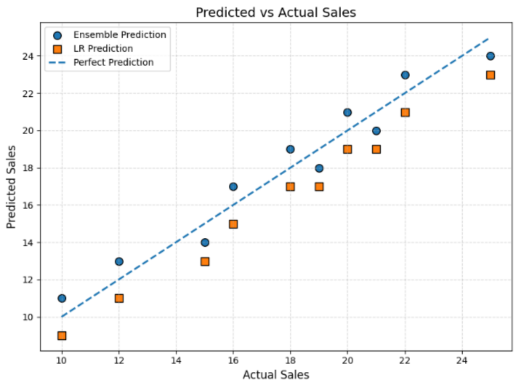

***Figure 3.** The predicted vs actual sales*

Figures 2 and 3 further illustrate the predictive behavior of the models by comparing the predicted and actual values for remaining stock and sales. In both figures, the ensemble predictions tend to follow the observed data more closely than those produced by the individual model. This indicates that the ensemble framework is better able to capture the underlying patterns in the dataset. The distribution of points around the perfect prediction line also suggests lower prediction error and improved consistency. In particular, the ensemble model shows fewer large deviations from the actual values, reflecting stronger stability when dealing with fluctuations in production and demand. Such stability is especially important in hydroponic farming, where operational conditions and market demand can vary considerably from week to week. Overall, these visual results reinforce the numerical findings reported earlier, confirming that the optimized ensemble approach provides more reliable forecasting performance. By producing predictions that align more closely with real observations, the framework offers a practical tool for improving stock planning and sales management in hydroponic farm operations.

3.2 Future directions and challenges

Earlier studies on hydroponic forecasting have largely relied on single machine learning models or conventional statistical approaches. While these methods can capture certain patterns in production or sales data, they often struggle to adapt to the variability typical of hydroponic systems, where environmental conditions, plant growth cycles, and market demand frequently change. The traditional machine learning models tend to show limited robustness when applied to agricultural datasets, suggesting the need for more flexible and adaptive forecasting frameworks [14]. The results of this study indicate that the proposed EA-optimized ensemble approach addresses some of these limitations. By dynamically adjusting the contribution of individual models, the framework improves predictive stability and overall accuracy. The lower RMSE and MAE values observed in this study suggest that combining multiple learning algorithms can capture both linear and nonlinear relationships more effectively than relying on a single model [15]. This integration allows the system to adapt more readily to the diverse conditions found in real hydroponic farm operations.

Another aspect that distinguishes this work from much of the existing literature is the emphasis on model interpretability. Previous studies have typically focused on improving prediction accuracy, with less attention given to explaining how the models reach their conclusions. In contrast, this study incorporates a model-agnostic local explanation method to provide local explanations of prediction outcomes. As highlighted in [16], explainable artificial intelligence (XAI) plays an important role in agricultural decision-making because it allows practitioners to understand the factors influencing model predictions. By identifying the variables that most strongly affect stock and sales forecasts, the proposed framework offers insights that are more practical for real-world farm management. Beyond predictive performance, the framework also carries implications for sustainable agricultural management. More reliable forecasting enables better stock planning, which can reduce both overproduction and shortages in hydroponic systems. In practice, this means farmers can align harvesting schedules more closely with market demand, helping to minimize waste and improve resource efficiency.

The framework also supports broader sustainability objectives reflected in the SDGs. For instance, improved stock forecasting contributes to SDG 12.3, which aims to reduce food waste. By anticipating demand more accurately, farmers can avoid unnecessary production and reduce post-harvest losses. At the same time, the approach supports SDG 9.3, which focuses on strengthening the productivity and market participation of small-scale producers. Access to data-driven forecasting tools can help hydroponic farmers make more informed decisions regarding production planning, pricing, and inventory management, ultimately improving their competitiveness in local markets. Finally, the framework contributes to SDG 12.C by encouraging more efficient use of resources. Hydroponic farming already emphasizes controlled resource use, but predictive analytics can further optimize the allocation of water, nutrients, and energy throughout the production cycle. By integrating machine learning with evolutionary optimization, the proposed system promotes a more adaptive and data-informed approach to agricultural management. Overall, the findings suggest that combining predictive accuracy with interpretability can make advanced forecasting tools more practical for everyday agricultural use. Rather than functioning solely as a technical model, the framework has the potential to support more sustainable, efficient, and transparent hydroponic farming practices.

Real-time data integration also represents an important direction for future work. In practice, hydroponic systems increasingly rely on IoT sensors and automated monitoring tools. Connecting the forecasting framework with real-time sensor data could allow the model to adapt dynamically to changing environmental conditions, enabling more responsive decision-making. Finally, although the use of explainability techniques improves model transparency, interpreting machine learning outputs in practical farm settings remains a challenge. Farmers and stakeholders may require user-friendly visualization tools or decision-support interfaces to translate model insights into actionable strategies. Future research should therefore explore ways to integrate predictive models with intuitive dashboards or farm management systems. Addressing these challenges would strengthen the applicability of the proposed framework and further support the adoption of data-driven decision-making in hydroponic agriculture.

4. CONCLUSION

This study shows that an optimized ensemble framework combining LR, RF, and XGBoost, with weights refined through an evolutionary algorithm, can improve the forecasting of hydroponic stock and sales. The results indicate that integrating multiple models helps capture both linear and nonlinear patterns in the data, leading to more stable and accurate predictions. In addition, the use of model-agnostic local explanation method provides clearer insights into the factors influencing the model's outputs, making the system more transparent for practical decision-making. These improvements can support small-scale hydroponic farmers in managing inventory more effectively, reducing waste, and using resources more efficiently. In this sense, the framework contributes to broader sustainability goals by promoting data-driven agricultural practices that align with the SDGs. Future research should explore the integration of real-time data and more adaptive modeling approaches to further enhance forecasting reliability in dynamic farming environments.

5. REFERENCES

[1] Bezner Kerr, R., Liebert, J., Kansanga, M., & Kpienbaareh, D. (2022). Human and social values in agroecology: A review. Elem Sci Anth, 10(1), 00090.

[2] Choudhary, S., Boruah, A., Ram, N., Gulaiya, S., Choudhary, C. S., & Verma, L. K. (2023). Millet's role in sustainable agriculture: a comprehensive review. International Journal of Plant & Soil Science, 35(22), pp. 556-568.

[3] Galt, R. E., Pinzón, N., Robinson, N. I., & Coronil, M. B. B. (2024). Agroecology and the social sciences: A half-century systematic review. Agricultural systems, 216, 103881.

[4] Cravero, A., Pardo, S., Sepúlveda, S., & Muñoz, L. (2022). Challenges to use machine learning in agricultural big data: a systematic literature review. Agronomy, 12(3), 748.

[5] Wetmore, S., Truong, K., Belmar, J. A., Fritz, T., Amer, S., & Nara, P. S. C. (2025). Agri-Virtual Assistant-A Tool for Farmers. International Journal of Computer Applications, 186(76), pp. 52-58.

[6] Mariadass, D. A., Moung, E. G., Sufian, M. M., & Farzamnia, A. (2022, November). EXtreme gradient boosting (XGBoost) regressor and shapley additive explanation for crop yield prediction in agriculture. In 2022 12th International Conference on Computer and Knowledge Engineering (ICCKE), pp. 219-224.

[7] Yuan, G. N., Marquez, G. P. B., Deng, H., Iu, A., Fabella, M., Salonga, R. B., ... & Cartagena, J. A. (2022). A review on urban agriculture: technology, socio-economy, and policy. Heliyon, 8(11).

[8] Hernández-Martínez, N. R., Blanchard, C., Wells, D., & Salazar-Gutiérrez, M. R. (2023). Current state and future perspectives of commercial strawberry production: A review. Scientia Horticulturae, 312, 111893.

[9] Tang, K., Zhao, X., Xu, Z., & Sun, H. (2024). A stacking ensemble model for predicting soil organic carbon content based on visible and near-infrared spectroscopy. Infrared Physics & Technology, 140, 105404.

[10] Mana, A. A., Allouhi, A., Hamrani, A., Rehman, S., El Jamaoui, I., & Jayachandran, K. (2024). Sustainable AI-based production agriculture: Exploring AI applications and implications in agricultural practices. Smart Agricultural Technology, 7, 100416.

[11] Soltanifar, M., & Tavana, M. (2024). A novel pairwise comparison method with linear programming for multi-attribute decision-making. EURO Journal on Decision Processes, 12, 100051.

[12] Rubo, S., & Zinkernagel, J. (2025). Enhancing the prediction of irrigation demand for open field vegetable crops in Germany through neural networks, transfer learning, and ensemble models. Agricultural Water Management, 312, 109402.

[13] Miller, C., Portlock, T., Nyaga, D. M., & O'Sullivan, J. M. (2024). A review of model evaluation metrics for machine learning in genetics and genomics. Frontiers in bioinformatics, 4, 1457619.

[14] Rahman, M. A., Chakraborty, N. R., Sufiun, A., Banshal, S. K., & Tajnin, F. R. (2024). An AIoT-based hydroponic system for crop recommendation and nutrient parameter monitorization. Smart Agricultural Technology, 8, 100472.

[15] Borges, A. F., Laurindo, F. J., Spínola, M. M., Gonçalves, R. F., & Mattos, C. A. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. International journal of information management, 57, 102225.

[16] Kalasampath, K., Spoorthi, K. N., Sajeev, S., Kuppa, S. S., Ajay, K., & Angulakshmi, M. (2025). A literature review on applications of explainable artificial intelligence (XAI). IEEE access.