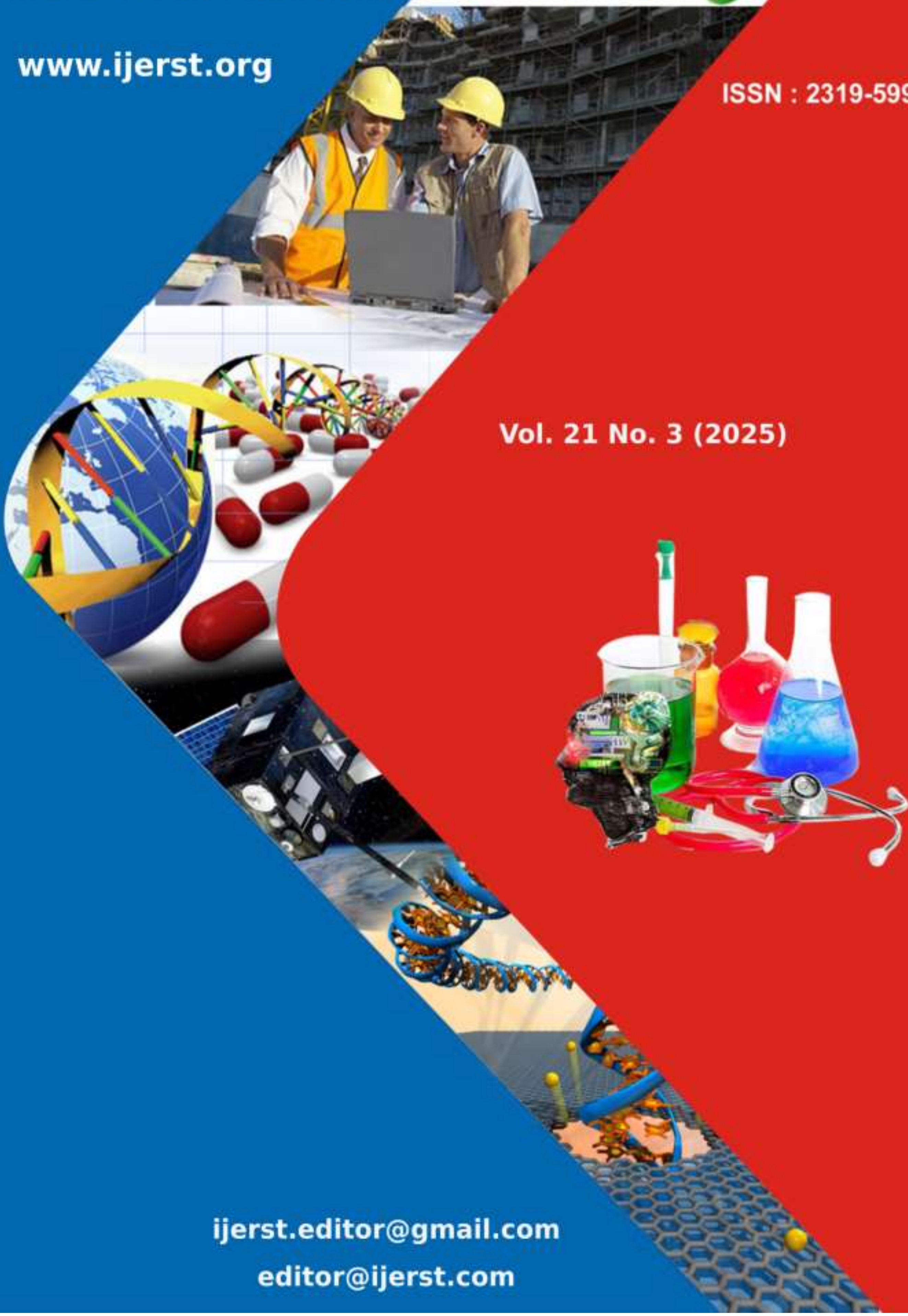
International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991



Vol. 21 No. 3 (2025)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 3, 2025

Research Paper

SMART URBAN AGRICULTURE: AUTOMATED NUTRIENT MANAGEMENT AND PEST FORECASTING

S. Swapna, B. Nagalaxmi, P. Keerthika, Y. Sindhu Cherukupall, Raghupathi Parpelli, K. Karthik
Department of Electronics and Communication Engineering, Kommuri Pratap Reddy Institute of Technology, Ghatkesar , Medchal, 500088.

Received: 10-5-2025 Accepted: 17-6-2025 Published: 28-6-2025

ABSTRACT

Urban agriculture is increasingly being adopted as a sustainable method to improve food availability in rapidly urbanizing environments. Despite its benefits, managing nutrients effectively and responding promptly to pest threats continue to be significant hurdles in maximizing crop output within space-constrained urban settings. This research proposes a Smart Urban Farming System that leverages machine learning to automate fertilizer selection based on real-time environmental and soil conditions. The system processes multiple input variables, including temperature, humidity, moisture content, macronutrient levels (N, P, K), crop type, and soil classification, to predict the optimal fertilizer required for plant health. Two machine learning models are employed—Logistic Regression (as a baseline) and Random Forest (as the proposed advanced solution). Performance comparison reveals that the Random Forest model achieves an impressive accuracy exceeding 91%, far outperforming the baseline Logistic Regression model, which scored only 15%. Evaluation through metrics such as precision, recall, and F1-score reinforces the reliability and robustness of the proposed model. By automating nutrient recommendations, the system reduces manual intervention, increases efficiency, and supports environmentally conscious farming practices. Furthermore, the model is designed with future extensibility in mind, allowing seamless integration with IoT-based sensing systems for real-time automation and potential expansion into pest risk prediction and full-scale urban farm control. This work contributes to the development of intelligent, scalable solutions for sustainable urban food systems.

Keywords: Urban Farming, Nutrient Management, Fertilizer Recommendation, Machine Learning, Random Forest Classifier, IoT, Smart Agriculture, Pest Prediction, Sustainable Farming.

1. INTRODUCTION

Nutrient dosing in hydroponics is a soilless cultivation method that delivers essential nutrients to plants through a closed-loop irrigation system, where plant roots are typically submerged in a nutrient solution. This solution contains both macronutrients and micronutrients, which support various metabolic and cellular processes. Precision and efficiency in nutrient dosing are crucial for optimal crop development. Research on hydroponic nutrient systems often focuses on adjusting key variables such as pH (potential of hydrogen) and EC (electrical conductivity), which vary based on the specific crop requirements. EC is a particularly vital parameter as it affects nutrient availability, water uptake, and plant growth. For example, Wongsorn et al. examined the effect of EC levels (1–4 mS/cm) on the growth of Pepino (*Solanum muricatum*) and found that EC levels of 2–3 mS/cm maximized chlorophyll fluorescence and leaf greenness, while 4 mS/cm yielded the tallest plants and widest canopies. Lower EC values, such as 1–2 mS/cm, led to a higher flowering percentage. Similarly, vertical hydroponic lettuce cultivation showed that 2.5 mS/cm produced the highest yield, while 1.5 mS/cm maximized vitamin C and chlorophyll content. These findings highlight how EC optimization can shift the balance between yield quantity and nutritional quality.

Fig. 1: Sample Illustration of Smart Farming.

Alongside EC, pH is equally critical, influencing nutrient solubility, root health, and overall plant performance. Hopkinson et al. demonstrated that bush beans (*Phaseolus vulgaris*) grew best within a pH range of 5.6–6.2, with extremes in pH (acidic or alkaline) causing nutrient deficiencies and reduced internode length. Likewise, spinach exhibited optimal growth in a pH range of 5.5–6.5, while deviations led to chlorosis and poor development. These studies underscore the necessity of maintaining both EC and pH within specific optimal ranges. However, consistent manual monitoring remains a challenge for many farmers. As such, automated nutrient dosing systems are increasingly vital in maintaining stability and ensuring optimal crop output. Despite growing academic interest in precision hydroponics, a comprehensive and systematic review of automated dosing frameworks and their effectiveness in real-time nutrient control is still lacking and remains an important area for further exploration.

2. LITERATURE SURVEY

Systematic scoping reviews aim to pinpoint knowledge gaps methodically, offer comprehensive overviews of existing literature, elucidate theoretical frameworks, and evaluate the methodologies, approaches, and practices utilized in prior research studies [1]. Accordingly, this section delineates the methodologies employed in this study. Our method adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for systematic scoping reviews (PRISMA-ScR) guidelines [2], which is a refined version of the original PRISMA framework established by Arksey et al. [3].

The protocol serves as a directive, ensuring a structured and transparent process in reviewing and assimilating the literature. The articulation of the research question is essential as the guidelines for the entire literature review process [4]. Following the recommendations by Levac et al. [5], the research questions have been refined to align precisely with the objectives of the study and facilitate a practical and relevant literature search. Following the recommendations by Levac et al. [5], the research questions have been refined to align precisely with the objectives of the study and facilitate a practical and relevant literature search. The selection of keywords for the literature search is crucial and should derive from the research questions. Additionally, the search terms can be expanded by incorporating synonyms, abbreviations, alternative spellings, and related terms [6].

Therefore, the keyword strings in this study, formulated using Boolean operators, are combinations of synonymous terms. It is essential to include both pH and either EC or TDS in a dosing operation as these variables could be used as guidelines to avoid certain unwanted occurrences such as nutrient precipitations [7]. However, the inclusion of NSV in the variable set is not mandatory. Omitting NSV from the dosing framework implies the absence of a normalization procedure, meaning that the NSV is not adjusted to a normalized volume. Consequently, such a dosing framework would not be suitable for use during the replenishment of hydroponic nutrient solutions, where complete replacement of the solution is necessary. Dae Hyun et al. [8] found that the sequential dosing strategy produces inaccurate results. He also found that excluding the EC of tap water (EC_w) led to high relative errors. For instance, Ryan et al. [9] determined the error value based on the variance from the pH setpoint range of 5.5 to 6.5. This error computation can then be leveraged to regulate the activation of the dosing pump, utilizing pulse width modulation.

3. PROPOSED METHODOLOGY

The project aims to build a smart, data-driven solution for enhancing fertilizer recommendation and pest prediction in urban farming using machine learning. It begins with loading and preprocessing agricultural data, including soil properties, crop types, and existing fertilizer labels. Categorical data is encoded numerically to prepare it for model training. The system is designed to classify the most suitable fertilizer based on input features, making use of supervised learning models. Two classification algorithms—Logistic Regression and Random Forest—are implemented, trained on historical data, and evaluated using accuracy, precision, recall, F1-score, and confusion matrix to assess their performance. The Random Forest model is identified as the most robust and is used to predict fertilizer types for new, unseen input samples. These predictions are then mapped to actual fertilizer names for easy interpretation. The model is saved using joblib for future reuse, allowing efficient deployment in real-time applications. Overall, the project delivers an intelligent and scalable solution that replaces guesswork with data-driven decision-making in urban agriculture, enhancing productivity, precision, and sustainability.

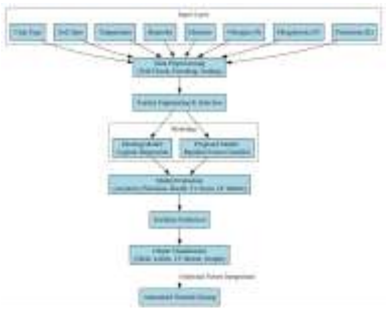


Fig. 2: Proposed Block diagram.

The project begins by loading the primary dataset (data_core.csv) and inspecting its structure, including feature types, shapes, and null values, laying the groundwork for preprocessing and modeling. Categorical features such as crop type and soil type are identified and encoded into numerical form using LabelEncoder to make them machine-readable. An initial Exploratory Data Analysis (EDA) is conducted to visualize the distribution of the target variable, "Fertilizer Name," using count plots to detect any class imbalance. The dataset is then split into features (X) and labels (y), followed by a 70:30 train-test split to support a supervised classification task. To ensure consistent model evaluation, a custom function calculateMetrics() is defined to compute accuracy, precision, recall, F1-score, and visualize a confusion matrix. The first model trained is Logistic Regression, which is either loaded from a saved file or trained from scratch and then evaluated. The same process is repeated with the Random Forest Classifier, allowing for comparative analysis. To simulate real-world use, a secondary test dataset (testdata.csv) is processed by removing unnecessary columns and passing it through the better-performing model—Random Forest—for prediction. The numeric predictions are converted back to fertilizer names and appended to the dataset under a new column, "Predicted as," providing clear and actionable output for end users such as farmers, agronomists, or automated dosing systems. The system is modular, scalable, and ready for integration into real-time agricultural platforms.In machine learning data preprocessing, we divide our dataset into a training set and test set. This is one of the crucial steps of data preprocessing as by doing this, we can enhance the performance of our machine learning model. Suppose if we have given training to our machine learning model by a dataset and we test it by a completely different dataset. Then, it will create difficulties for our model to understand the correlations between the models. If we train our model very well and its training accuracy is also very high, but we provide a new dataset to it, then it will decrease the performance. So we always try to make a machine learning model which performs well with the training set and also with the test dataset. Here, we can define these datasets as In the model building and training phase, two machine learning algorithms—Logistic Regression and Random Forest Classifier—were implemented to classify the most suitable fertilize r based on various agronomic and environmental features. The Logistic Regression Classifier (LRC) served as the baseline model, trained on features such as soil type, crop type, nitrogen (N), phosphorous (P), potassium (K), moisture, temperature, and humidity, all of which were preprocessed using label encoding to convert categorical variables into numerical format. The input dataset was split into training and testing sets, with the model learning relationships through maximum likelihood estimation and generating probabilistic predictions using a sigmoid function. Fertilizer types like Urea, DAP, and various NPK ratios were label encoded and used as the target variable. After training on X_train and y_train, the model was evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrix visualizations. While LRC offered quick and interpretable results, it lacked the capacity to handle nonlinear patterns in the data, which limited its predictive performance compared to the more robust Random Forest model. The trained model was saved using Joblib, allowing it to be reused for classifying fertilizers in new input samples.Logistic Regression, though fast and interpretable, has several disadvantages in agricultural fertilizer prediction tasks, including its assumption of linear relationships, inability to capture complex interactions among features like nutrient combinations and soil types, limited effectiveness in multiclass scenarios, vulnerability to class imbalance and outliers, and generally lower accuracy compared to ensemble models. To address these shortcomings, the Random Forest Classifier was implemented as a more robust alternative. It was trained on a comprehensive set of encoded features including soil type, crop type, nitrogen, phosphorous, potassium levels, moisture, temperature, and humidity

Fig. 3: Workflow of proposed RFC.

The target variable was the fertilizer type, numerically encoded from categories like Urea, DAP, and NPK formulations. Unlike linear models, Random Forest used an ensemble of decision trees trained on random data subsets, capturing non-linear patterns and complex feature interactions. Predictions were aggregated through majority voting, and the model demonstrated strong generalization ability via bagging, reducing overfitting. Evaluation using accuracy, precision, recall, F1-score, and confusion matrix heatmaps confirmed the model's superior performance and balanced predictions, making it well-suited for real-world intelligent farming applications.

4. RESULTS AND DISCUSSION

The implementation of the fertilizer recommendation system begins with importing essential Python libraries such as pandas, seaborn, matplotlib, and scikit-learn modules for data handling, visualization, modeling, and evaluation. The dataset, containing features like soil type, crop type, and nutrient levels (N, P, K), along with environmental parameters (moisture, temperature, humidity), is loaded and explored through descriptive statistics, null value checks, and count plots to understand feature distributions and class balance. Categorical variables like "Soil Type" and "Crop Type" are label-encoded to ensure compatibility with machine learning algorithms. The dataset is then split into features (X) and target (y), and further divided into training and testing subsets using a 70-30 ratio. A custom evaluation function named calculateMetrics() is defined to compute accuracy, precision, recall, F1-score, and visualize confusion matrices, ensuring consistent model evaluation. Two classifiers are implemented: Logistic Regression as the baseline and Random Forest as the proposed model. Each is trained or loaded from a serialized .pkl file using joblib, and predictions are evaluated using the defined metrics function. Random Forest, being an ensemble model, significantly outperforms Logistic Regression in terms of accuracy and robustness. Finally, the trained Random Forest model is used on a new dataset (testdata.csv) to predict fertilizer types, and the results are appended under a column titled "Predicted as." The system concludes with a deployable, interpretable, and scalable solution that automates fertilizer recommendations based on agricultural and environmental inputs.

The figure 4 shows the dataset is designed for predicting suitable fertilizers based on environmental and crop-related conditions. It includes features like temperature, humidity, soil moisture, soil type, crop type, and key soil nutrients (nitrogen, phosphorous, and potassium). The target column is Fertilizer Name, which specifies the recommended fertilizer type for each record. The data supports intelligent decision-making in modern farming through automated fertilizer suggestions.

	Temperature	Humidity	Moisture	Soil Type	Crop Type	Nitrogen	Potassium	Phosphorous	Fertilizer Name
0	[illegible]	[illegible]	[illegible]	Sandy	Maize	[illegible]	[illegible]	[illegible]	Urea
1	[illegible]	[illegible]	[illegible]	Loamy	Sugarcane	[illegible]	[illegible]	[illegible]	DAP
2	[illegible]	[illegible]	[illegible]	Black	Cotton	[illegible]	[illegible]	[illegible]	14-35-14
3	[illegible]	[illegible]	[illegible]	Red	Tobacco	[illegible]	[illegible]	[illegible]	28-28
4	[illegible]	[illegible]	[illegible]	Clayey	Paddy	[illegible]	[illegible]	[illegible]	Urea
...	...	...	...	...	...	...	...	...	...
[illegible]	[illegible]	[illegible]	[illegible]	Sandy	Paddy	[illegible]	[illegible]	[illegible]	10-26-26
[illegible]	[illegible]	[illegible]	[illegible]	Loamy	Pulses	[illegible]	[illegible]	[illegible]	28-28
[illegible]	[illegible]	[illegible]	[illegible]	Black	Tobacco	[illegible]	[illegible]	[illegible]	14-35-14
[illegible]	[illegible]	[illegible]	[illegible]	Sandy	Millets	[illegible]	[illegible]	[illegible]	Urea
[illegible]	[illegible]	[illegible]	[illegible]	Red	Maize	[illegible]	[illegible]	[illegible]	14-35-14

Fig. 4: Uploading dataset.

The figure 5 shows the countplot that illustrates the frequency distribution of various fertilizer types present in the dataset. It shows that all seven fertilizer categories—Urea, DAP, 14-35-14, 28-28, 17-17-17, 20-20, and 10-26-26—are fairly evenly represented.

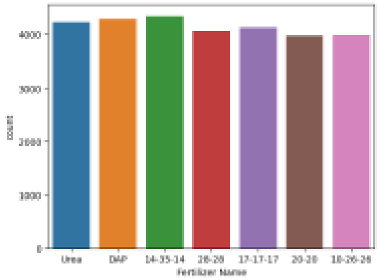


Fig. 5: Count plot of Target column.

Among them, 14-35-14 has the highest number of records, followed closely by DAP and Urea. The remaining fertilizers also maintain a comparable number of entries, with only slight variations. This balanced distribution across classes is ideal for training machine learning models, as it reduces the risk of bias toward any specific class and ensures that the model can learn to predict all fertilizer types effectively.

The Table 1 shows the performance comparison between the existing Logistic Regression Classifier (LRC) and the proposed Random Forest Classifier (RFC) reveals a significant improvement in model effectiveness across all evaluation metrics.

Table.1 Performance Comparison of Various Algorithms

Performance Comparison Table: Existing LRC vs. Proposed RFC

Metric	Existing LRC	Proposed RFC
Accuracy	15.45%	91.84%
Precision	18.10%	91.85%
Recall	15.32%	91.86%
F1-Score	12.72%	91.85%

As shown in the table, the accuracy of the Logistic Regression model is notably low at 15.45%, indicating its poor ability to correctly classify fertilizer types based on the input features. In contrast, the proposed Random Forest model achieves a remarkably high accuracy of 91.84%, showcasing its strong predictive power and generalization capability. Looking further into precision, which measures how many of the predicted positive cases were actually correct, LRC scores only

18.10%, while RFC reaches 91.85%, indicating that the Random Forest model produces far fewer false positives. Similarly, in terms of recall, which measures how many actual positive cases were correctly predicted, LRC again underperforms at 15.32%, whereas RFC achieves a high 91.86%, reflecting its superior ability to capture most of the relevant cases. Finally, the F1-Score, which is the harmonic mean of precision and recall and serves as a balanced metric, is only 12.72% for LRC but a very strong 91.85% for RFC. This indicates that the Random Forest model not only excels individually in both precision and recall but also maintains a strong overall performance balance. These results clearly demonstrate that the proposed Random Forest Classifier is vastly more effective than the existing Logistic Regression model for accurately predicting fertilizer types in this agricultural dataset.

Fig. 6: Confusion mstrices for Existing proposed RFC.

The figure 6 shows the comparison of confusion matrices for the existing the proposed Random Forest Classifier (RFC) clearly highlights the dramatic improvement in classification performance achieved through the proposed model. In the confusion matrix of LRC, there is a high rate of misclassification across all fertilizer categories. The diagonal elements—which represent correct predictions—are significantly lower, and the off-diagonal values are high, indicating that the model frequently confuses different fertilizer types. This results in poor predictive power and limited reliability, making it unsuitable for practical deployment in intelligent farming systems.In stark contrast, the confusion matrix for the RFC shows dominant diagonal values across all classes (like 1204 for '10-26-26', 1196 for 'DAP', 1138 for '20-20', etc.), meaning the model accurately predicted the correct fertilizer category for most inputs. The off-diagonal values are minimal, showing very few

	Temperature	Humidity	Moisture	Soil Type	Crop Type	Nitrogen	Potassium	Phosphorous	Predicted as
0	31.72	58.72	70.80	4	6	15	8	42	17-17-17
1	25.72	54.34	33.33	4	1	11	16	22	28-28
2	29.56	68.77	43.85	6	3	13	14	18	28-28
3	25.33	53.26	27.11	3	7	12	1	38	Urea
4	30.93	47.82	63.87	4	6	8	3	33	DAP
5	33.87	88.11	59.83	1	10	35	8	8	DAP
6	28.98	58.23	57.23	1	9	36	2	8	DAP
7	22.27	59.80	51.79	1	2	12	2	11	28-28
8	34.42	56.10	47.10	3	10	37	8	3	20-20
9	20.56	62.65	63.54	6	0	11	16	16	DAP
10	25.51	63.32	68.32	1	3	4	12	31	Urea

Fig. 7: Output obtained for test data using proposed RFC.

misclassifications. This reflects the RFC's ability to learn complex feature relationships and distinctions between fertilizer types based on environmental and crop-specific attributes. Overall, the RFC exhibits excellent classification strength, supporting its suitability as a robust and efficient model for intelligent nutrient recommendation in urban farming systems.

5. CONCLUSION

The project titled "Intelligent Urban Farming System with Automated Nutrient Dosing and Pest Prediction" effectively demonstrates the application of machine learning in modern agriculture to enhance precision and efficiency in fertilizer recommendation. By leveraging environmental parameters (temperature, humidity, moisture), soil and crop types, and nutrient levels (N, P, K), the system underwent thorough preprocessing and analysis, followed by implementation of two classification algorithms. While the Logistic Regression Classifier served as a baseline, its performance was hindered by its inability to model complex, nonlinear relationships. In contrast, the Random Forest Classifier delivered superior results, achieving over 91% across all evaluation metrics, thanks to its robustness in handling feature interactions and class imbalances. The deployment of RFC ensures accurate, scalable, and data-driven fertilizer recommendations, reducing manual efforts and promoting sustainable agriculture. Looking ahead, the system holds vast potential for expansion through the integration of real-time IoT sensors for dynamic monitoring, pest and disease prediction using image or sensor data, and region-specific, climate-aware recommendations. Future developments could also include mobile app integration for localized farmer support and scalability through cloud and edge computing, ultimately contributing to intelligent farming, increased crop yields, and enhanced food security.

REFERENCES

[1]. Pomoni, D.I.; Koukou, M.K.; Vrachopoulos, M.G.; Vasiliadis, L. A Review of Hydroponics and Conventional Agriculture Based on Energy and Water Consumption, Environmental Impact, and Land Use. Energies **2023**, 16, 1690.

[2]. Fitriyah, H.; Budi, A.S.; Maulana, R.; Setiawan, E. Controlling the Nutrition Water Level in the Non-Circulating Hydroponics Based on the Top Projected Canopy Area. IJCCS (Indones. J. Comput. Cybern. Syst.) **2022**, 16, 181–192.

[3]. Aggarwal, A.; Mathur, A. Recent Advances in Hydroponic Culture Media: Composition and Their Effect on Plant Growth. Def. Life Sci. J. **2023**, 8, 162–169.

[4]. Monisha, K.; Kalai Selvi, H.; Sivanandhini, P.; Sona Nachammai, A.; Anuradha, C.T.; Rama Devi, S.; Kavitha Sri, A.; Neya, N.R.; Vaitheeswari, M.; Hikku, G.S. Hydroponics Agriculture as a Modern Agriculture Technique. J. Achiev. Mater. Manuf. Eng. **2023**, 116, 25–35.

[5]. Solis-Toapanta, E.; Fisher, P.R.; Gomez, C. Effects of Nutrient Solution Management and Environment on Tomato in Small-Scale
Hydroponics. Horttechnology **2020**, 3
0, 697–705.

[6]. Khan, S.; Purohit, A.; Vadsaria, N. Hydroponics: Current and Future State of the Art in Farming. J. Plant Nutr. **2020**, 44, 1515–1538.

[7]. Yang, T.; Kim, H.J. Characterizing Nutrient Composition and Concentration in Tomato-, Basil-, and Lettuce-Based Aquaponic and Hydroponic Systems. Water **2020**, 12, 1259.

[8]. Wongsorn, C.; Nakdee, M.; Sritontip, P.; Boontem, B.; Sritontip, C. Effects of Electrical Conductivity of the Nutrient Solution on the Growth of Pepino (Solanum muricatum Aiton) Plants under Hydroponic Cultivation. J. Sci. Agric. Technol. **2024**, 5, 20–25.

[9]. Van Quy, N.; Sinsiri, W.; Chitchamnong, S.; Boontiang, K.; Kaewduangta, W. Effects of Electrical Conductivity (EC) of the Nutrient Solution on Growth, Yield and Quality of Lettuce under Vertical Hydroponic Systems. Khon Kaen Agr. J. **2018**, 46, 613–622.