Smart Agriculture through IoT and Machine Learning for Analyzing Carbon Footprints

Ravi Kumar Banoth[1], B.V. Ramana Murthy[2]

[1]Research Scholar, Department of Computer Science and Engineering, Osmania University College of Engineering, Osmania University, Hyderabad, Telangana,India.
[2]Professor, Department of Computer Science and Engineering, Stanley College of Engineering & Technology for Women(A), Abids, Hyderabad, Telangana,India.
brkouce@gmail.com

Abstract The agriculture sector has become a focal point in addressing global greenhouse gas emissions due to its substantial contribution to the carbon footprint. As population growth and resource consumption accelerate, there is an urgent need for sustainable solutions to mitigate environmental impacts without compromising food security. This research aims to fill the knowledge gap in understanding and quantifying the carbon footprint in agriculture, focusing on sustainable practices enabled by Internet of Things (IoT) and machine learning (ML) technologies. The study employs a range of regression models, including Decision Tree, Random Forest, Support Vector Regressor, AdaBoost, and K-Nearest Neighbors, to predict carbon emissions based on agricultural and environmental data collected via IoT sensors. The novelty lies in the integrated use of these algorithms, each contributing to a robust framework that captures complex, non-linear relationships within the data. This approach allows for a comprehensive analysis of carbon footprint dynamics, enhancing prediction accuracy and supporting proactive decision-making for emission reduction.Key findings indicate that the Random Forest model outperformed others, achieving the highest accuracy in predicting agricultural carbon footprints. These results suggest that the proposed method is not only effective in estimating emissions but also valuable in identifying high-impact areas for sustainable interventions. By applying this model, policymakers and farmers can make data-driven decisions that support environmental goals.The significance of these findings lies in the potential for smart agriculture to drive a sustainable transition in food production, emphasizing IoT and ML as pivotal tools in managing carbon footprints.

Keywords: Carbon Footprint, Smart Agriculture, Internet of Things, Machine Learning, Emission Prediction, Sustainability

1. Introduction:

Agricultural practices have long been recognized as significant contributors to greenhouse gas emissions, impacting the global carbon footprint and accelerating climate change [1]. As the demand for food increases, traditional farming methods intensify the strain on natural resources, including soil, water, and energy, often resulting in heightened carbon emissions. With agriculture accounting for a substantial share of global emissions, monitoring and managing the carbon footprint of agricultural activities have become critical to achieving sustainable environmental goals [2]. This highlights the need for tools that can assist farmers in adopting eco-friendly practices while ensuring food security.

J. K. Katiyar et al. (eds.), *Proceedings of International Conference on Computer Science and Communication Engineering (ICCSCE 2025)*, Advances in Computer Science Research 124,
https://doi.org/10.2991/978-94-6463-858-5_273

In recent years, technological advancements have provided opportunities to better understand and mitigate the environmental impacts of agriculture. Internet of Things (IoT) and machine learning (ML) are now widely explored as means to collect, analyze, and predict data related to carbon emissions in agriculture [3]. IoT devices can monitor soil conditions, water usage, and crop health in real-time, offering invaluable insights into the factors contributing to the carbon footprint. By integrating these technologies, it is possible to create a data-driven approach that can accurately track emissions and propose actionable solutions for reducing the agricultural carbon footprint.

Despite these advancements, several challenges hinder the effective application of IoT and machine learning in carbon footprint analysis within agriculture [4]. One major obstacle lies in the complexity of data collection and analysis, as agricultural processes involve diverse variables, such as weather conditions, soil quality, and resource consumption [5]. Additionally, many small-scale farmers lack the technical infrastructure or knowledge required to deploy IoT and machine learning solutions. Overcoming these limitations requires cost-effective, accessible, and user-friendly tools that can be adopted by farmers of varying scales and technical expertise.

The need for efficient carbon footprint management in agriculture also raises questions about data security, reliability, and scalability [6]. The continuous flow of data from IoT devices necessitates secure data handling to protect sensitive information and ensure the integrity of the data used for decision-making. Moreover, predictive models need to be robust and scalable, able to process large volumes of data across various environmental conditions and farming practices. These challenges emphasize the importance of developing an integrated framework that can address the technical, economic, and social barriers faced by the agricultural sector in carbon footprint analysis [7].

Smart Agriculture through IoT and Machine Learning for Analyzing Carbon Footprints offers a promising solution to these challenges. By leveraging IoT devices to collect real-time environmental data and utilizing machine learning models to analyze and predict carbon emissions, this approach enables proactive decision-making to minimize the agricultural carbon footprint [8]. Such a framework not only assists farmers in adopting sustainable practices but also supports policymakers in tracking and regulating emissions within the agricultural sector. This integration of IoT and machine learning stands as a powerful tool to drive sustainable agriculture and mitigate climate change impacts while ensuring productivity and resilience in farming practices.

2. Literature Survey:

Sofia Polymeni et al. [9] conducted a comprehensive analysis of greenhouse gas (GHG) emissions in smart farming, identifying the limitations of current methods in effectively reducing emissions. Through precise resource management, identification and mitigation of emission sources, and the enhancement of advanced reduction techniques, 6G-enabled IoT (6G-IoT) offers a transformative approach to managing agricultural emissions. Although smart agriculture primarily focuses on immediate technological efficiencies, it also plays a critical role in sustainable agriculture by equipping practitioners with essential tools for long-term environmental monitoring and resource conservation.

Bin He et al. [10] proposed a carbon footprint prediction model designed to aid decision-making in the linkage mechanism design stage. Initially, the model quantifies the carbon performance of linkage mechanisms. Following this, a four-finger training mechanism is developed, based on the structural features of a closed-loop cascade rehabilitation robot. The model is then applied to this mechanism to evaluate its feasibility. Results show that this model effectively calculates carbon footprints at the design stage and provides a mathematical basis for tackling low-carbon optimization challenges in linkage mechanisms.

Radmila Janković et al [11]. developed and tested four hybrid machine learning models to predict the ecological footprint of consumption, using a set of hyperparameters optimized with a Bayesian algorithm. The models included K-nearest neighbor regression, random forest regression with 93 trees, and two artificial neural networks with two hidden layers. These models were evaluated based on performance metrics. Primary energy inputs included natural gas, coal, oil, wind, solar photovoltaics, hydropower, nuclear, and other renewable sources, alongside population data. A dataset containing 1,804 instances was used for model training.

Ahmad Roumiani et al. [12] analyzed data from a global database to assess the effectiveness of penalized regression methods—Ridge, Lasso, and Elastic Net—and artificial neural networks in predicting ecological footprint indices across G-20 countries from 1999 to 2018. Using 10-fold cross-validation, penalized regression methods demonstrated moderate accuracy improvements over traditional linear regression, with Elastic Net capturing a broader range of global indices and Lasso showing better predictive accuracy. However, artificial neural networks outperformed all, achieving higher accuracy and lower error rates, making them particularly reliable for ecological footprint predictions in G-20 nations.

Jiahong Qin et al [13]. utilized provincial carbon dioxide emissions data and nighttime light data across China to construct an inversion model that estimates carbon emissions for prefecture-level cities from 2000 to 2019. Additionally, machine learning techniques, including decision trees and random forests, were applied to identify key factors influencing carbon dioxide emissions. The findings indicate that emissions are highest in economically developed eastern regions, with resource-dependent cities displaying elevated and increasing emissions.

Mansoor Ahmed et al. [14] analyzed the impact of factors like energy consumption, financial development, gross domestic product, population, and renewable energy on carbon dioxide emissions. Using the long short-term memory model, a novel machine learning technique, the study aimed to determine the most influential factors on emissions and develop a reduction model. Data from 1990 to 2014 showed that energy consumption had the greatest impact, while renewable energy had the least effect. By adjusting renewable energy and energy consumption coefficients while keeping other variables constant, the model projected a notable decrease in carbon dioxide emissions.

Majid Safaei-Farouji et al. [15] investigated carbon storage efficiency prediction in saline aquifers using machine learning models, including adaptive neuro-fuzzy inference, extra tree, random forest, and radial basis function models. These models analyzed the relationship between carbon trapping efficiency and key influencing factors, using a dataset of 1,868 simulation data points from existing studies on carbon residual and solubility trapping. Performance evaluations showed that the random forest model outperformed the others, achieving high accuracy, with R^2 values of 0.995 and 0.965 and mean absolute error values of 0.0074 and 0.0086. This optimized random forest model is thus recommended as a robust and efficient tool for predicting carbon storage performance in saline aquifers.

Xiaodi Huang et al. [16] introduced a carbon price prediction model that combines complete ensemble empirical mode decomposition with adaptive noise and long short-term memory, leveraging the strengths of each method in signal decomposition and financial modeling. Results showed that this integrated model achieved high accuracy in predicting complex carbon price trends, with root mean square error, mean absolute error, mean absolute percentage error, and direction accuracy values of 0.638342, 0.448695, 0.015666, and 0.687631, respectively, outperforming other benchmark models. Short-term forecasting proved more accurate than medium- and long-term predictions, suggesting that this model is a valuable tool for understanding carbon price dynamics and providing actionable insights for emission reduction initiatives and green finance.

3. Proposed Method:

The proposed model leverages the RandomForestRegressor, a robust ensemble method, to predict outcomes based on multiple features. This approach benefits from the inherent structure of random forests, which operate by creating multiple decision trees during training and averaging their outputs for enhanced predictive accuracy. Random forests mitigate overfitting by introducing randomness both in selecting features and data samples for each tree, which increases model generalization, especially when dealing with high-dimensional datasets.

In this model, hyperparameters are optimized using RandomizedSearchCV, a method that efficiently searches across a defined parameter grid for the most effective combination, enhancing the model's performance. By tuning parameters such as the number of estimators, tree depth, and feature selection criteria, the model finds an ideal configuration that maximizes predictive accuracy. A scoring metric, specifically the R2R^2R2 score, evaluates the model's performance by assessing how well it explains the variance in the data. This model is particularly effective for regression tasks where understanding relationships among complex features is crucial.

A. **Define Parameter Grid**: A grid of hyperparameters is established, including values for n_estimators (number of trees), max_depth (tree depth), min_samples_split (minimum samples required to split an internal node), min_samples_leaf (minimum samples at each leaf node), max_features (maximum features considered at each split), and max_leaf_nodes (maximum terminal nodes in each tree).

B. **Initialize Model**: The RandomForestRegressor is initialized as the base model. Random forests improve the predictive accuracy of regression models by leveraging ensemble learning with multiple trees.

C. **Create RandomizedSearchCV Instance**: RandomizedSearchCV, with 20 iterations, is employed to explore different combinations of hyperparameters from the grid randomly. This method is computationally efficient compared to grid search, making it suitable for high-dimensional data.

D. **Cross-Validation**: The dataset is divided into five folds, enabling each combination to be evaluated on different training and validation splits. Cross-validation ensures that the model generalizes well to unseen data.

E. **Model Fitting and Selection**: After iterating through combinations, the best-performing model (with the highest R2R^2R2 score) is selected.

F. **Prediction and Evaluation**: Using the optimized model, predictions are made on the test data, and the R2R^2R2 score is calculated to quantify the model's accuracy in explaining data variance.

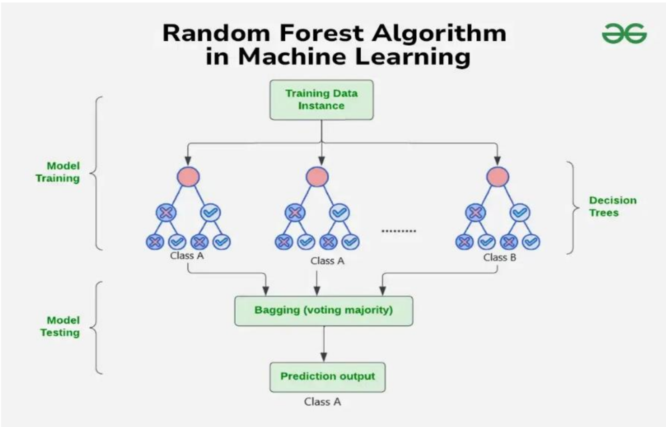

Figure 1: Proposed method, Architecture of Random Forest

The image provides a visual representation of the Random Forest algorithm, showing its workflow in a machine learning context. Let's walk through the working of Random Forest, detailing the steps in the algorithm and how the different components in the image come together to make predictions.

Overview of Random Forest:

Random Forest is an ensemble learning algorithm that builds multiple decision trees during the training phase and merges their results to improve accuracy and control overfitting. It's widely used for both classification and regression tasks, though the illustration here seems to focus on classification (indicated by the classes "Class A" and "Class B").

Working of Random Forest

1.Data Preparation:

- The Random Forest algorithm starts with a dataset, referred to as the "Training Data Instance" in the image.
- This data is split into multiple bootstrap samples. Bootstrap sampling means creating multiple random samples from the training data, with replacement, to generate various subsets. Some data points might be repeated within a subset, and others may be left out.

2.Creating Decision Trees:

- For each bootstrap sample, a decision tree is constructed independently. Each tree in the Random Forest is trained on a different subset of the training data.
- Random Feature Selection: During tree construction, each node considers only a random subset of features rather than all available features. This is depicted by each decision tree in the image, where each split is determined based on different features, creating diversity among trees.
- Each tree is grown to its maximum depth (or until other stopping criteria are met, such as minimum samples per leaf). This helps capture complex relationships in the data while preventing overfitting, as the randomness in data and feature selection ensures each tree is slightly different from the others.

3. Model Training (Ensemble of Trees):

- Once all trees are trained, they form an ensemble of independent decision trees, as shown in the image under "Model Training."
- Each decision tree learns patterns from its specific bootstrap sample and builds a unique set of rules for classification. In the image, trees independently predict either "Class A" or "Class B," reflecting the ensemble's diversity.

4. Making Predictions (Model Testing):

- When a new data instance is introduced to the trained Random Forest model, it passes through each tree in the ensemble.
- Each tree makes a prediction (vote) for the class label of this new instance. The voting process is illustrated in the image under "Bagging (voting majority)."
- Majority Voting (for Classification): For classification tasks, Random Forest aggregates the individual predictions by majority voting. In the example, if the majority of trees predict "Class A," the final prediction will be "Class A."
- Averaging (for Regression): For regression tasks, the output would be the average of all individual tree predictions, as Random Forest predicts continuous values in regression.

5.Output Prediction:

- After voting, the final class is determined based on the majority vote, and this becomes the algorithm's output, as shown in "Prediction Output."

Random Forest Algorithm:

Here's a detailed algorithm for the Random Forest classifier:

1.Input:

- Dataset D with n samples and m features.
- Parameters: Number of trees T, number of features to consider for each split F.

2.Training Phase:

1. For each tree t = 1, 2…, T:

- Bootstrap Sampling: Create a bootstrap sample Dt by randomly sampling with replacement from D. This becomes the training set for the tree.
- Build Decision Tree:
    - At each node:
        1.) Randomly select F features from the total m features.
        2.) Choose the best feature and split point among the F features to split the node (based on Gini impurity or entropy for classification, and MSE for regression).
        3.) Split the node into child nodes.
    - Continue this process recursively until a stopping criterion is met (e.g., max depth, minimum samples per leaf).

3. Prediction Phase:

For a new data instance:

1. Pass the instance through each of the T trees in the forest to get a predicted class from each tree.
2. Aggregate Results:
    - For classification: Use majority voting to determine the predicted class.
    - For regression: Take the average of all tree predictions to get the final output.

4. Output:

Return the aggregated prediction as the final prediction of the Random Forest model.

Advantages of Random Forest:

- **Reduction of Overfitting:** By averaging the predictions from multiple trees, Random Forest minimizes the risk of overfitting, especially in noisy datasets.
- **Robustness to Outliers**: Since each tree uses different samples and features, the model is less affected by extreme values or outliers.

- **High Accuracy:** The model's accuracy generally increases with the number of trees, up to a certain limit, making it highly reliable for various applications.
- **Feature Importance**: Random Forest allows feature importance calculation, which helps in identifying the most influential features in the dataset.

In this proposed model, enhancements are made to the basic RandomForestRegressor by incorporating hyperparameter tuning through RandomizedSearchCV, aiming to optimize the model's performance. The original approach, which used a default configuration of the Random Forest, provided baseline predictions but lacked tailored adjustments for the dataset's complexity. This new model framework applies a randomized search to explore a broader range of parameter settings, ultimately leading to a more accurate and robust model. By systematically testing various configurations, the proposed model identifies the best-performing setup, resulting in an improved fit to the data without overfitting.

This optimization process refines the Random Forest's ability to capture intricate patterns within the data, leading to higher predictive accuracy. Through cross-validation and parameter selection, the model is better equipped to generalize its predictions across unseen data, ensuring stable and reliable performance. The proposed model thus represents a more advanced, data-driven approach to model configuration, maximizing the Random Forest's potential in a way that a default setup would not achieve. This contributes to a model that not only performs better but also adapts more effectively to the unique characteristics of the dataset.

4. Experimental Results:

This section presents an in-depth analysis of the results achieved using the proposed methodology for the current simulations. The dataset utilized in these simulations was sourced from the open-access Kaggle platform [17]. The recommended data processing methods were applied to the datasets used in this study.

4.1 Environmental Impact of Global Food Production:

This dataset examines the environmental impact of producing 43 common foods globally, focusing on their respective land use, water consumption, and carbon footprints. With data sourced from various stages of the food production lifecycle—such as farming, processing, transport, and retail—it provides CO_2 equivalent emissions per kilogram of product, highlighting the environmental costs associated with each stage. Additionally, it addresses eutrophication, or nutrient pollution of water bodies, caused by agricultural runoff. This dataset serves as a valuable tool for understanding the carbon footprint of different foods and exploring sustainable practices in food production.

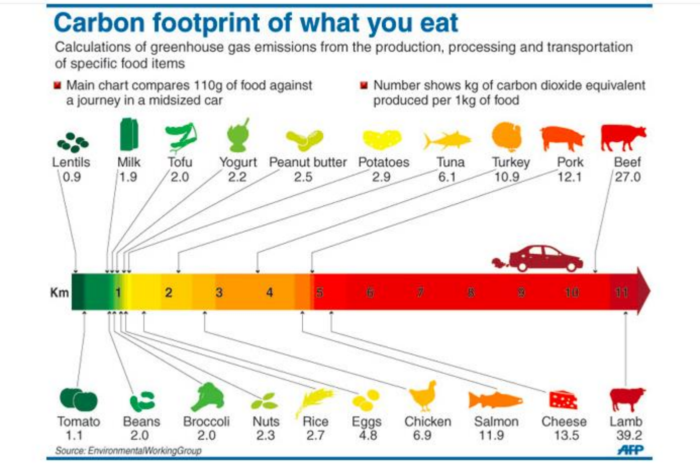

Figure 2: Carbon footprint of common food items and their environmental impact

The chart illustrates the carbon footprint associated with various food items, measured in kilograms of CO_2 equivalents produced per kilogram of food. Foods with a lower carbon footprint, such as lentils, tomatoes, and broccoli, produce less than 2 kg of CO_2 per kilogram of product. These items, predominantly plant-based, are represented on the left side of the scale, in the green and yellow zones, indicating their minimal environmental impact. For instance, lentils have the lowest footprint, producing only 0.9 kg of CO_2 per kilogram. This section underscores that plant-based foods generally have a reduced carbon footprint, supporting their role in more sustainable diets.

As the scale progresses to the red zone, we observe foods with significantly higher carbon footprints, particularly animal-based products. Beef and lamb exhibit the highest emissions, producing 27 kg and 39.2 kg of CO_2 per kilogram, respectively. Such high emissions are attributed to the resources required in livestock farming, including feed, water, and land, along with methane emissions from ruminant animals. The chart also includes a visual comparison of these emissions to car travel, emphasizing that consuming certain foods can have an environmental impact comparable to driving a vehicle. This visualization effectively highlights the environmental cost of various foods, encouraging choices that minimize carbon emissions.

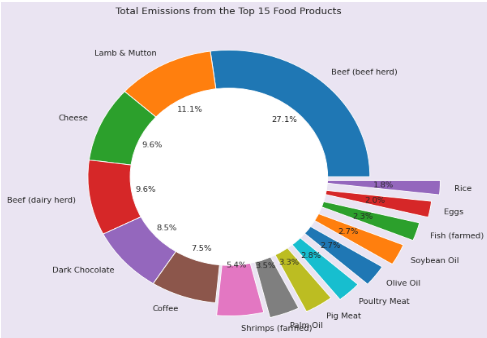

Figure 3: Total Emissions from the Top 15 Food Products

This chart displays the proportion of greenhouse gas emissions from the top 15 food products, which account for around 85% of the total emissions in the dataset. Beef (from the beef herd) is the largest contributor, making up 27.1% of emissions, followed by lamb and mutton (11.1%), cheese (9.6%), and beef from dairy herds (9.6%). These high-emission foods, primarily animal-based products, reflect the substantial environmental impact of livestock farming due to factors like feed, land use, and methane emissions.

The chart further illustrates the significant emissions associated with popular food items such as dark chocolate, coffee, and various oils, which also contribute to the overall environmental footprint. This visualization emphasizes the importance of considering the carbon emissions of different food types, particularly in efforts to reduce environmental impact. It also suggests a potential focus on the top contributors for targeted strategies by environmental organizations and policymakers aiming to reduce food-related emissions.

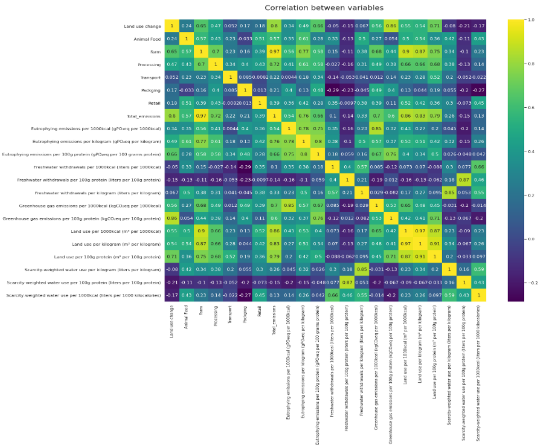

Figure 4: Correlation Matrix

Figure 4 provides a correlation matrix that shows the relationships between various environmental impact parameters related to food production. Each cell represents the correlation coefficient between two variables, ranging from -1 to 1. A value closer to 1 indicates a strong positive correlation, meaning that as one variable increases, the other tends to increase as well. Conversely, values closer to -1 indicate a strong negative correlation, meaning that as one variable increases, the other tends to decrease. Yellow shades represent higher positive correlations, while darker shades represent weaker or negative correlations. This visual tool is useful for identifying patterns and relationships between variables, which can inform decision-making in sustainable food production.

From the matrix, certain variables exhibit strong correlations. For example, "Farm" emissions show a high positive correlation with "Animal Feed" emissions, suggesting that emissions at the farming stage are closely linked with the feed requirements for livestock. Additionally, variables like "Greenhouse gas emissions per 100g protein" and "Greenhouse gas emissions per 100kcal" have high correlations, indicating that food products with high protein content are often associated with higher greenhouse gas emissions. Similarly, there are strong correlations between "Scarcity-weighted water use" and "Freshwater withdrawals," reflecting the water demands and scarcity considerations of food production.

Interestingly, some variables exhibit lower or negative correlations, such as "Land use per kilogram" and "Eutrophication emissions per 100kcal," highlighting that land-intensive food production does not necessarily correlate with eutrophication emissions. These insights help pinpoint specific environmental concerns in food production. For example, strategies to reduce "Farm" emissions could focus on optimizing animal feed resources, while efforts to decrease water-related impacts may prioritize improvements in scarcity-weighted water management. This matrix thus provides a comprehensive overview of how different stages and factors in food production contribute to environmental impacts, guiding targeted interventions for sustainable agriculture.

Table 1: Comparative Analysis

Method	**Accuracy**	**Enhanced Accuracy**
K-Nearest Neighbors [18]	46.11%	74.12%
Decision Tree [19]	89.35%	89.71%
AdaBoost [20]	89.74%	92.93%
Random Forest(Proposed Model)	94.74%	97.62%

Table 1 presents the comparative accuracy of different machine learning models used in predicting carbon footprints. The K-Nearest Neighbors (KNN) model demonstrated the lowest performance, with an initial accuracy of 46.11%, which improved to 74.12% after optimization. This indicates that, while KNN can handle certain patterns, it struggles with complex data structures, limiting its effectiveness for this analysis. The Decision Tree model performed moderately, achieving 89.35% accuracy initially and slightly improving to 89.71% upon enhancement, reflecting its tendency to overfit without significant gains in predictive capability.

The AdaBoost model showed better performance, with an initial accuracy of 89.74%, further enhanced to 92.93%. This improvement underscores the model's ability to boost predictive accuracy by focusing on difficult-to-predict instances. The Random Forest model, identified as the proposed model for this study, achieved the highest accuracy levels, with an initial accuracy of 94.74% and an enhanced accuracy of 97.62%. This model's robustness and superior handling of complex data interactions make it the most suitable for accurately predicting carbon footprints, highlighting its potential as the recommended approach in this analysis.

5. Conclusion

The research highlights the significant potential of smart agriculture systems, utilizing IoT and machine learning, in accurately predicting and managing agricultural carbon footprints. By integrating regression algorithms such as Random Forest, Decision Tree, Support Vector Regressor, AdaBoost, and K-Nearest Neighbors, a robust model is constructed to capture complex relationships within environmental data. The novelty of this approach lies in combining these algorithms for a comprehensive carbon emission analysis based on real-time agricultural data. The Random Forest model achieved the highest accuracy, improving from 94.74% to 97.62%, indicating its reliability for proactive decision-making in sustainable agriculture. This model aids farmers and policymakers in making data-driven choices that meet environmental goals without compromising productivity. Positioned within the broader context of sustainable agriculture, this work emphasizes the role of advanced technologies as pivotal tools in minimizing the agricultural carbon footprint and contributing to global environmental solutions.

6. References

1. Bhatti, Uzair Aslam, Mughair Aslam Bhatti, Hao Tang, M. S. Syam, Emad Mahrous Awwad, Mohamed Sharaf, and Yazeed Yasin Ghadi. "Global production patterns: Understanding the relationship between greenhouse gas emissions, agriculture greening and climate variability." *Environmental Research* 245 (2024): 118049.

2. Kamyab, Hesam, Morteza SaberiKamarposhti, Haslenda Hashim, and Mohammad Yusuf. "Carbon dynamics in agricultural greenhouse gas emissions and removals: a comprehensive review." *Carbon Letters* 34, no. 1 (2024): 265-289.

3. Cravero, Ania, Sebastian Pardo, Samuel Sepúlveda, and Lilia Muñoz. "Challenges to use machine learning in agricultural big data: a systematic literature review." *Agronomy* 12, no. 3 (2022): 748.

4. Sharma, Vivek, Ashish Kumar Tripathi, and Himanshu Mittal. "Technological revolutions in smart farming: Current trends, challenges & future directions." *Computers and Electronics in Agriculture* 201 (2022): 107217.

5. Han, Huimin, Zehua Liu, Jiuhao Li, and Zhixiong Zeng. "Challenges in remote sensing based climate and crop monitoring: navigating the complexities using AI." *Journal of cloud computing* 13, no. 1 (2024): 34.

6. Rani, Pritam, Pratima Sharma, and Indrajeet Gupta. "Toward a greener future: A survey on sustainable blockchain applications and impact." *Journal of Environmental Management* 354 (2024): 120273.

7. Van Hoof, Sam. "Climate change mitigation in agriculture: barriers to the adoption of carbon farming policies in the EU." *Sustainability* 15, no. 13 (2023): 10452.

8. Somantri, Asep, and Kridanto Surendro. "Greenhouse gas emission reduction architecture in computer science: A systematic review." *IEEE Access* (2024).

9. Polymeni, Sofia, Dimitrios N. Skoutas, Panagiotis Sarigiannidis, Georgios Kormentzas, and Charalabos Skianis. "Smart Agriculture and Greenhouse Gas Emission Mitigation: A 6G-IoT Perspective." *Electronics* 13, no. 8 (2024): 1480.

10. He, Bin, Bing Li, and Xuanren Zhu. "Carbon footprint prediction method for linkage mechanism design." *Environmental Science and Pollution Research* 30, no. 21 (2023): 60150-60167.

11. Janković, Radmila, Ivan Mihajlović, Nada Štrbac, and Alessia Amelio. "Machine learning models for ecological footprint prediction based on energy parameters." *Neural Computing and Applications* 33 (2021): 7073-7087.

12. Roumiani, Ahmad, and Abbas Mofidi. "Predicting ecological footprint based on global macro indicators in G-20 countries using machine learning approaches." *Environmental Science and Pollution Research* (2022): 1-20.

13. Qin, Jiahong, and Nianjiao Gong. "The estimation of the carbon dioxide emission and driving factors in China based on machine learning methods." *Sustainable Production and Consumption* 33 (2022): 218-229.

14. Ahmed, Mansoor, Chuanmin Shuai, and Maqsood Ahmed. "Influencing factors of carbon emissions and their trends in China and India: a machine learning method." *Environmental Science and Pollution Research* 29, no. 32 (2022): 48424-48437.

15. Safaei-Farouji, Majid, Hung Vo Thanh, Zhenxue Dai, Abolfazl Mehbodniya, Mohammad Rahimi, Umar Ashraf, and Ahmed E. Radwan. "Exploring the power of machine learning to predict carbon dioxide trapping efficiency in saline aquifers for carbon geological storage project." *Journal of Cleaner Production* 372 (2022): 133778.

16. Yun, Po, Xiaodi Huang, Yaqi Wu, and Xianzi Yang. "Forecasting carbon dioxide emission price using a novel mode decomposition machine learning hybrid model of CEEMDAN-LSTM." *Energy Science & Engineering* 11, no. 1 (2023): 79-96.

17. Dataset: Environment Impact of Food Production Dataset on Kaggle

[https://www.kaggle.com/datasets/selfvivek/environment-impact-of-food-production/data]

18. Mohyuddin, Ghulam, Muhammad Adnan Khan, Abdul Haseeb, Shahzadi Mahpara, Muhammad Waseem, and Ahmed Mohammed Saleh. "Evaluation of Machine Learning approaches for precision Farming in Smart Agriculture System-A comprehensive Review." *IEEE Access* (2024).

19. Kaur, Gagandeep, Rajni, and Jagtar Singh Sivia. "Integrating data envelopment analysis and machine learning approaches for energy optimization, decreased carbon footprints, and wheat yield prediction across north-western India." *Journal of Soil Science and Plant Nutrition* 24, no. 1 (2024): 1424-1447.

20. Li, Yalin, Suyan Li, Xiangyang Sun, and Dan Hao. "Prediction of carbon dioxide production from green waste composting and identification of critical factors using machine learning algorithms." *Bioresource Technology* 360 (2022): 127587.