Journal of Agricultural Chemistry and Environment, 2026, 15(1), 39-62
https://www.scirp.org/journal/jacen
ISSN Online: 2325-744X
ISSN Print: 2325-7458

Hybrid DEA-Machine Learning Methodology for Evaluating and Modeling Technical Efficiency of Rice Farming in Niger

Idrissa Saidou Mahamadou[1,2]

[1]Department of Sociology and Rural Economics, Faculty of Agronomy, Abdou Moumouni University of Niamey, Niamey, Niger
[2]Laboratory of Analysis and Research in Sociology and Rural Economics (LARSER), Niamey, Niger
Email: mi_saidou@yahoo.com

How to cite this paper: Mahamadou, I.S. (2026) Hybrid DEA-Machine Learning Methodology for Evaluating and Modeling Technical Efficiency of Rice Farming in Niger. *Journal of Agricultural Chemistry and Environment*, **15**, 39-62.
https://doi.org/10.4236/jacen.2026.151004

Received: December 24, 2025
Accepted: February 6, 2026
Published: February 9, 2026

Abstract

This study evaluates the technical efficiency of family rice farms in Niger, using a hybrid framework that integrates Data Envelopment Analysis (DEA) with Machine Learning. Primary data were collected from 103 farms in Gaya, Niger, during the 2025 dry season. To account for structural heterogeneity, a typology of production systems was established using Principal Component Analysis (PCA) and Hierarchical Ascendant Classification (HAC). Technical efficiency was assessed through an input oriented Variable Returns to Scale (VRS) DEA model, and the resulting scores were subsequently modeled using a Random Forest algorithm to capture nonlinear relationships and interaction effects. The analysis revealed a mean efficiency score of 0.76, ranging from 0.42 to 1.00, with 18% of farms operating on the efficiency frontier ($\theta = 1$). Commercial farms exhibited the highest performance (median = 0.83), while small family farms recorded lower efficiency (median = 0.71). The Random Forest model demonstrated high predictive accuracy ($R^2 = 0.9168$; MSE = 0.0027). The main determinants of efficiency were: Seed quantity, Farm size and Fertilizer use with respectively a positive effect. Other influential factors included herbicide use and irrigation, while socio demographic variables such as education, gender, and marital status had negligible impact. This hybrid DEA–Machine Learning framework provides both a rigorous benchmark for measuring technical efficiency and a predictive tool for identifying drivers of performance. The findings offer evidence based insights to guide targeted policies aimed at improving resource use, strengthening resilience, and enhancing the sustainability of rice farming systems in Niger.

Keywords

Data Envelopment Analysis, Technical Efficiency, Machine Learning, Rice Farming, Niger

1. Introduction

Agriculture remains a fundamental pillar of economic development, food security, and rural livelihoods in developing countries. It provides the main source of income and employment for the majority of rural households and contributes significantly to national economies [1] [2]. In West Africa, sustained demographic growth and rapid urbanization have sharply increased the demand for staple foods, particularly rice [3]. Today, rice accounts for more than 30% of urban caloric intake in the region (CIRAD, n.d.) and plays a strategic role in poverty reduction, especially when produced under irrigation, where seasonal stability and income generation are enhanced [4]. Rice cultivation in Niger remains an agricultural activity with limited economic weight, although it is strategic for food security. According to the National Rice Development Strategy (SNDR 2021-2030), national rice production is estimated at around 83,000 tons per year, while domestic demand exceeds 500,000 tons. At the macroeconomic level, agriculture contributes about 40% of Niger's GDP, but rice cultivation represents only a marginal share, estimated at less than 1% of GDP. Despite these advantages, productivity remains below potential levels, indicating persistent inefficiencies in the use of land, labor, water, and agro-inputs [5]. Understanding these inefficiencies is crucial for guiding agricultural policies aimed at improving resource allocation, enhancing resilience, and strengthening food security [6]. Technical efficiency analysis offers a robust framework for evaluating how effectively farmers convert inputs into outputs relative to best-practice frontiers. Since the refinement of efficiency concepts in recent literature, frontier methods have remained central to assessing production performance. Data Envelopment Analysis (DEA) continues to be widely applied, as demonstrated by recent studies such as [7] in agricultural contexts and [8] in financial institutions, highlighting its robustness in estimating technical and allocative efficiency. has become one of the most prominent nonparametric methods for examining agricultural efficiency because it accommodates multiple inputs and outputs without requiring a predefined functional form. Previous applications in West Africa and other developing regions consistently highlight significant inefficiencies among smallholder rice producers often linked to poor access to inputs, suboptimal input allocation, and structural constraints [9]-[13]. However, empirical evidence remains limited for irrigated rice systems in Niger, and the determinants of inefficiency remain insufficiently explored. A persistent challenge in frontier analysis lies in explaining variations in efficiency scores. Traditional two-stage approaches typically using Tobit or truncated regression assume linearity and impose restrictive functional forms, limiting their ability to capture the nonlinear interactions and multidimensional constraints that characterize smallholder agriculture. Machine Learning (ML), particularly ensemble methods such as Random Forests, offers powerful tools to overcome these limitations by identifying complex patterns, modeling nonlinear relationships, and ranking the most influential factors driving inefficiency [14]. The issue of technical efficiency in Niger's irrigated rice farming lies in the persistent gap between potential and actual productivity, compounded by structural heterogeneity and nonlinear determinants of performance. Classical DEA approaches provide useful benchmarks but struggle to explain these variations. A hybrid DEA–Machine Learning framework is therefore essential to capture the multidimensional drivers of inefficiency and to generate actionable insights for targeted policy interventions. Moreover, structural heterogeneity among farms stemming from differences in land size, labor organization, and input intensity has received limited empirical attention, although it fundamentally shapes production outcomes and policy responses. This study addresses these gaps by developing a hybrid DEA Machine Learning framework to evaluate and model the technical efficiency of rice farms in Gaya, Niger. Specifically, the study 1) Identifies production system typologies using Principal Component Analysis (PCA) and Hierarchical Ascendant Classification (HAC); 2) Computes farm-level technical efficiency scores using an input-oriented DEA model under Variable Returns to Scale (VRS); and 3) Models the determinants of efficiency using a Random Forest algorithm to capture nonlinear effects and rank variable importance. By integrating frontier efficiency analysis with predictive modeling, this research advances both methodological and empirical knowledge. It demonstrates how modern computational tools can complement classical efficiency analysis in heterogeneous, resource-constrained systems, and provides evidence-based insights to guide interventions aimed at improving the performance and sustainability of irrigated rice farming in Niger. The findings are consistent with the Resource-Based View (RBV), which posits that farm performance depends on the effective mobilization and allocation of scarce resources such as land, labor, and inputs. They also align with the Production Frontier Theory, which emphasizes that efficiency is achieved when producers operate on or close to the best-practice frontier, minimizing input waste while maximizing output. Together, these theories justify the observed efficiency differentials and highlight the importance of targeted resource management and institutional support in enhancing the sustainability of rice farming systems.

2. Methodology

2.1. Study Area

The urban commune of Gaya is located in the extreme south of Niger, within the Dosso region, bordering both Benin and Nigeria. Its strategic transboundary position makes it a key agricultural and commercial hub in the country's southern corridor. Gaya lies within the Sudanian Sahelian agroecological zone, characterized by a prolonged dry season and an average annual rainfall ranging from 700 to 900 mm [15]. The presence of the Niger River provides relatively stable access to water resources, enabling the development of irrigated rice farming, particularly during the dry season when climatic risks are lower and irrigation control is higher [16]. The commune hosts several irrigated perimeters such as Karé Gourou, Tanda, and Bengou where rice cultivation is practiced intensively. These perimeters are equipped with semi-modern hydro-agricultural infrastructure, including pumping stations, drainage systems, and water distribution networks, although maintenance levels vary across farmer groups. Rice farming in Gaya mobilizes significant labor and contributes substantially to rural household incomes, positioning the commune as a focal point for seasonal agricultural activity. Socioeconomically, Gaya is dominated by family-based farming systems, often organized into cooperative groups. The diversity of farming practices, ranging from traditional to semi-intensive systems, reflects varying levels of access to inputs, technical knowledge, and institutional support. This heterogeneity makes Gaya an ideal site for analyzing technical efficiency and modeling agricultural performance.

2.2. Sources, Tools, and Nature of Data

This study draws on primary data collected from 103 rice farms selected from a population of 731 producers operating within the major irrigated perimeters of the urban commune of Gaya (Karé Gourou, Tanda, and Bengou). The sample size was determined using a probabilistic sampling formula (1) to ensure adequate representativeness of the target population. A stratified random sampling approach was then applied: the total population of 731 producers was first stratified by irrigated perimeter (Karé Gourou, Tanda, and Bengou) to reflect structural heterogeneity. Within each stratum, farms were randomly selected proportionally to the size of the farmer population in that perimeter, until the target sample of 103 farms was reached. This procedure ensured that each farm had an equal probability of selection while maintaining representativeness across the three irrigated Data were collected during the dry season through a structured questionnaire administered digitally using Kobo Collect platform, and subsequently processed and analyzed in R. The survey captured detailed information on the socioeconomic characteristics of farm managers, the mobilization of production factors, and the economic performance of each farming unit. Data collection was conducted in two stages. The first stage involved focus group discussions with representatives of local farmer organizations. The focus group discussions were analyzed using a thematic content analysis approach. Notes and transcripts from the sessions were coded inductively to identify recurring themes related to infrastructure, governance, gender composition, and nursery management practices. These themes were then organized into categories (access to inputs, cooperative governance, irrigation challenges) and triangulated with survey data to provide contextual depth. This qualitative analysis served two purposes: 1) to refine the questionnaire by incorporating locally relevant variables, and 2) to interpret quantitative efficiency results within the broader institutional and organizational context of irrigated rice farming in Gaya. The second stage consisted of individual interviews with rice farm managers, who constitute the primary unit of analysis. The margin of error was set at 9% to balance statistical precision with practical feasibility. While a 5% margin would have required a substantially larger sample size, the 9% threshold allowed us to obtain a representative sample of 103 farms across the three irrigated perimeters, within the logistical and resource constraints of fieldwork. This level of precision is consistent with accepted practice in exploratory agricultural studies where margins between 5%-10% are considered adequate.

$$n = \frac{t_p^2 * P(1-P) * N}{t_p^2 * P(1-P) + (N-1) * y^2} \tag{1}$$

- ❖ Target population (N): 731 (total rice farmers,provided by the farmers organization)
- ❖ Confidence level (tp): 1.96 (corresponding to 95%)
- ❖ Estimated proportion (P): 0.5 (to maximize variance)
- ❖ Margin of error (y): 9% (0.09)

2.3. Analytical Methods

2.3.1. Typology of Rice Farming Systems

A farm typology was developed to reduce structural heterogeneity and ensure more meaningful comparisons among production units. The typology was constructed through a Principal Component Analysis (PCA) applied to key structural, technical, and socioeconomic variables (Table 1), followed by a Hierarchical Ascendant Classification (HAC) using Ward's method to derive homogeneous clusters of farms. The selected variables capture essential features of production systems, including input intensity, land and labor resources, and household characteristics. Socioeconomic attributes such as education level, gender of the farm manager, and income diversification were incorporated because they reflect managerial capacity and household resilience factors known to indirectly influence production performance. For example, education may enhance input allocation decisions, while income diversification can reduce exposure to production risks and stabilize resource use. The combined PCA–HAC approach provides a robust multivariate framework for identifying structurally consistent groups of farms.

Table 1. Variables Used in the combined PCA-HAC approach

Variable	Description	Typology	Interpretation
Area	Size of the farm in hectares.	Structural variable (farm size)	Precision farming techniques for land management
Seed_kg	Quantity of seeds used (kg).	Input variable	Use of high-yield or drought resistant seeds
Fertilizer_kg	Quantity of fertilizer used (kg).	Input variable	Organic fertilizers and or smart fertilization methods
Urea_kg	Quantity of urea used (kg).	Input variable	Urea efficient practices or alternatives

Continued

Herbicide_kg	Quantity of herbicide used (kg).	Input variable	Integrated pest management (IPM)
Miscellaneous_Cost	Other production-related costs (currency).	Economic variable	Financial management and cost cutting strategies
Labor_Pre	Labor used before rice planting (in hours).	Labor variable	Labor saving technologies or mechanization
Labor_Rice	Labor used during rice cultivation (in hours).	Labor variable	Labor efficiency through automation
Labor_Post	Labor used after rice harvest (in hours).	Labor variable	Post harvest technology to reduce labor
Age_of_Farm_ Manager	Age of the farm manager.	Socio-demographic variable	Training and capacity building for younger managers
Education_Level	Education level of the farm manager.	Socio-demographic variable	Vocational training for better management
Gender_of_Farm_ Manager	Gender of the farm manager.	Socio-demographic variable	Gender inclusive policies and support
Household_Income_ Sources	Number of income sources within the household.	Socio-economic variable	Diversification strategies to reduce risk
Number of active household members	Count of individuals in the household who actively contribute to farm work (field labor, herding, processing, marketing).	Socio-economic variable	Indicates household labor capacity.

2.3.2. Hybrid DEA-Machine Learning Approach

The hybrid DEA-Machine Learning (DEA-ML) approach combines the complementary strengths of production econometrics. While Data Envelopment Analysis (DEA) provides a robust and economically interpretable measure of technical efficiency (2), Machine Learning algorithms enable the identification, ranking, and prediction of complex determinants of efficiency in nonlinear and multidimensional contexts.

The input variables in this DEA model include:

- Cultivated land area (hectares),
- Quantity of seeds (kg),
- Quantity of fertilizer (kg),
- Quantity of urea and herbicides (kg),
- Irrigation fees,
- Miscellaneous expenses,
- Labor required during the different phases of rice cultivation (pre-rice, rice, and post-rice).

The final output is measured by the total rice production (kg). DEA relies on two main models: the CCR model and the BCC model (Charnes *et al.* 1978). The CCR model assumes constant returns to scale and the BCC model allows for variable returns to scale, which is often more suitable in agriculture where productivity may vary with farm size [17].

2.3.3. Integration of Socio-Economic Factors

Socio-economic variables such as the farm manager's education level, marital status, and the number of active household members can also be incorporated into the DEA analysis to better understand their influence on farm efficiency. Previous studies have shown that higher levels of education and more structured family labor organization are associated with better resource utilization and more efficient farm management [18].

$$Eff\left(y_i\right)=\frac{\sum_{r=1}^{R}\lambda_{ryir}}{\sum_{j=1}^{M}\lambda_{jxij}};\quad \sum_{R=1}^{R}\lambda_r=1 \quad \text{with} \quad \lambda_r \geq 0, \forall r \tag{2}$$

Machine Learning into the DEA framework enhances both the explanatory and predictive power of efficiency models. According to the support of [19], the hybrid approach can be implemented through two complementary strategies: 1) ML-Enhanced DEA: Machine Learning algorithms are applied upstream to select relevant explanatory variables, reduce dimensionality, or correct noisy data prior to the implementation of DEA. This preprocessing step improves the robustness and interpretability of efficiency estimates by refining the input space and minimizing distortions caused by outliers or irrelevant features. 2) DEA-Informed ML: DEA efficiency scores are used as target variables in supervised learning models to identify nonlinear and interaction-based determinants of performance. This integrated approach enables the capture of the multidimensional complexity of agricultural systems, which are characterized by strong interdependencies among biophysical, socioeconomic, and environmental inputs. Furthermore, ML models offer the possibility of conducting sensitivity analyses and feature importance assessments, thereby highlighting the most influential factors driving farm-level efficiency. The analytical process is structured into three successive stages:

Step 1: Farm Segmentation

In the first stage, farm heterogeneity is explicitly addressed through an unsupervised classification approach. Hierarchical Ascendant Classification (HAC), following dimensionality reduction via Principal Component Analysis (PCA), is applied to a set of structural, agronomic, and socio-economic variables in order to identify homogeneous groups of rice farms. This segmentation reduces intra-group variability and improves the comparability of decision-making units within each group, thereby enhancing the precision and robustness of subsequent efficiency estimates. Such a typology-based approach is particularly recommended in heterogeneous production environments, where pooling structurally different farms may bias efficiency measurement [20].

Step 2: Intra-Group Efficiency Estimation

Given the context of Gaya, which is marked by limited access to productive resources characterizing smallholder farming systems, an input-oriented DEA model was adopted. This approach is consistent with empirical studies conducted in West African agricultural contexts, where producers primarily seek to optimize the use of scarce inputs in order to sustain or increase output levels [21]. For each homogeneous group, DEA scores (θ_i) are estimated separately using the input-oriented Variable Returns to Scale (VRS) model. This multi-level approach ensures that performance measurement is tailored to the internal structure of each farm type [18]. In addition to the input-oriented DEA specification, which is appropriate in resource-constrained contexts where farmers seek to minimize (3) input use, an output-oriented DEA model can also be applied for robustness. The output-oriented approach evaluates the potential proportional increase in rice production that each farm could achieve with its current level of inputs. Comparing results across orientations allows us to verify whether efficiency scores remain consistent and to identify whether constraints are primarily input-related or output-related. This dual specification strengthens the robustness of the analysis and provides a more comprehensive understanding of farm performance.

$$\underset{\theta,\lambda}{\min.} \quad \theta \tag{3}$$

subject to:

$$\sum_{j=1}^{n} \lambda_j Y_{j \geq Y_i},$$

$$\sum_{j=1}^{n} \lambda_j X_{j \leq \theta X_i},$$

$$\sum_{j=1}^{n} \lambda_j = 1,$$

$$\lambda_j \geq 0, \forall_j = 1, \cdots, n,$$

où:

- Y_i represents the output vector (rice production),
- X_i the input vector (land area, seeds, fertilizer, labor, herbicide, and urea),
- λ_j the weights assigned to the reference farms,
- and $\theta \in [0,1]$, the technical efficiency score of farm *i*.

A value of $\theta = 1$ indicates a technically efficient farm (located on the production frontier), whereas a value of $\theta < 1$ reflects a potential for improvement equivalent to $1 - \theta$ expressed as the proportion of input reduction required to reach the frontier.

Step 3: Predictive Modeling of Efficiency Determinants

The technical efficiency scores obtained (θ_i) (4)are subsequently used as the target variable in a supervised Machine Learning model, specifically a Random Forest algorithm. This model aims to identify and rank the key determinants of efficiency across heterogeneous farming systems. The explanatory variables include technical inputs, as well as the socioeconomic and demographic characteristics of farm managers. The general form of the predictive model is expressed as:

$$\theta_i = f\left(X_i, Z_i\right) + \int_i \tag{4}$$

where:

- θ_i: Technical efficiency score of farm *i*,

- X_i: Technical inputs for farm i,
- Z_i: Socio-demographic and economic characteristics of farm manager,
- ϵ_i: Error term, representing unmodeled factors that influence efficiency,
- $f(X_i, Z_i)$: The predictive function is a complex mapping of the explanatory variables X_i and Z_i for each farm ii, where $f(\cdot)$ denotes the function learned by the algorithm. In this context, X_i represents the technical inputs, and Z_i captures the socio-demographic and economic characteristics of the farm manager. The Random Forest algorithm models this relationship to estimate the technical efficiency score θ_i.

This modeling approach allows for the exploration of nonlinear relationships and interaction effects that may not be captured by traditional econometric methods. Moreover, the Random Forest algorithm provides variable importance metrics, enabling the identification of the most influential factors driving efficiency at the farm level. Variable importance metrics derived from the model such as the increase in mean squared error (MSE) or node purity enable the identification of key drivers of efficiency. These indicators highlight the most influential factors, providing actionable insights into the levers that can improve farm performance.

2.3.4. Random Forest Model

The Random Forest algorithm operates by constructing an ensemble of independent decision trees to predict the technical efficiency score θ_i (5). For each tree, a random subset of explanatory variables is selected, and predictions are made through successive decision splits. The final output of the Random Forest model is the average of the predictions generated by all individual trees, thereby reducing variance and improving generalization performance.

$$\hat{\theta}_i = \frac{1}{N}\sum_{j=1}^{N} f_j\left(X_i, Z_i\right) \tag{5}$$

where:

- $\hat{\theta}_i$: Prediction of the technical efficiency score for farm i,
- N: Number of trees in the forest,
- $f_j\left(X_i, Z_i\right)$: Prediction of tree j for farm i.

The Random Forest model typically minimizes the Mean Squared Error (MSE) to adjust the weights of the trees within the forest:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} f_j\left(\theta_i, \hat{\theta}_i\right)^2 \tag{6}$$

where:

- n is the total number of farms in the sample,
- θ_i is the actual efficiency score for the farm i,
- $\hat{\theta}_i$ is the predicted efficiency score for farm i.

The model predicts each farm's capacity to reduce input use while maintaining its current level of rice production. This reflects the potential for technical efficiency gains without compromising output. The explanatory variables enabled the identification of key factors influencing the technical efficiency of rice farms, offering insights into the structural and managerial levers that can enhance performance.

3. Results

3.1. Typology of Rice Farms

Contributions of the Variables

To account for structural heterogeneity among rice farms in Gaya, a Principal Component Analysis (PCA) was conducted using 17 agronomic, economic, and socio-demographic variables. The first three principal components explained 54.16% of the total variance, with eigenvalues of 5.83, 2.59, and 1.33 respectively (Table 2). These components revealed distinct dimensions of farm organization and resource use. The first dimension, which accounted for 32.41% of the total variance, was primarily shaped by agronomic and economic variables. High contributions were observed for fertilizer quantity (14.40%), production output (13.60%), land area (13.58%), total input cost (12.29%), seed quantity (11.45%), irrigation fees (10.19%), and urea application (9.92%). This dimension reflects a gradient of input intensity and productive capacity, distinguishing farms with higher resource allocation and output levels. The second dimension, explaining 14.38% of the variance, was dominated by labor-related variables. The most influential were labor during the rice phase (28.19%), post-rice phase (23.73%), pre-rice phase (21.11%), and the number of active household members (13.30%). This axis captures the structure and timing of labor deployment across production stages. While labor intensity varied significantly across farms, its contribution to technical efficiency was modest, suggesting that labor quantity alone is insufficient without strategic allocation and timing. The last Dimension, accounting for 7.37% of the variance, was defined by socio-demographic and residual economic traits. Education level (25.37%), age of the farm manager (14.34%), number of income sources (13.99%), gender of the farm manager (10.19%), miscellaneous expenses (11.37%), and marital status (7.59%) were the leading contributors. This component reflects household diversity and financial behavior.

Table 2. Variable contributions to the first three principal components.

Variables	**Dim.1**	**Dim.2**	**Dim.3**
Land area (hectares)	13.576228978	0.146471329	1.146624098
Seed quantity (kg)	11.446160431	0.973173316	2.360895549
Fertilizer quantity (kg)	14.404852613	2.210933229	0.415439750
Urea quantity (kg)	9.922612510	1.534849799	0.106320267
Herbicide quantity (kg)	4.376654770	0.019333184	2.761079819
Irrigation fee (SAI)	10.191356301	0.328838526	0.084539013
Total input cost	12.294892900	0.483282614	0.685989992
Miscellaneous expenses	1.472306268	3.231476963	11.370144563
Production output (kg)	13.595289195	1.137239730	0.008346001

Continued

Labor—Pre-rice phase	1.841270344	21.110041257	4.814192118
Labor—Rice phase	1.642533294	28.188296521	0.191600531
Labor—Post-rice phase	0.882736540	23.723921420	0.111940002
Marital status	0.003326181	0.764690994	7.589805718
Education level	0.387249016	0.198381044	25.372544626
Gender of farm manager	0.443629787	0.001328294	10.193688877
Number of active household members	2.193024920	13.300806388	4.456433538
Age of farm manager	1.193400511	1.574026123	14.335181579
Number of household income sources	0.132475442	1.072909269	13.995233959
Eigenvalue	5.83368606	2.58894631	1.32613874
Percentage of variance	32.4093670	14.3830351	7.3674374
cumulative percentage of variance	32.40937	46.79240	54.15984

Hierarchical Ascendant Classification

The Hierarchical Ascendant Classification (HAC), implemented using Ward's method, identified three homogeneous clusters of rice farms based on their structural and socioeconomic characteristics:

- Cluster 1—Commercial or Semi-Intensive Farms (F1): This group comprises relatively large farms managed by experienced producers. These farms exhibit high levels of input use, more intensive production practices, and stronger integration into commercial markets.
- Cluster 2—Small Family Farms (F2): These farms are small in size, rely predominantly on family labor, and use inputs more sparingly. Production practices are generally traditional, and resource constraints are more pronounced than in other groups.
- Cluster 3—Medium-Sized Farms (F3): Representing the largest share of the sample, this cluster includes farms of intermediate size, with an average cultivated area of 0.31 hectares. Their input use and production intensity fall between the two previous groups.

These clusters reflect distinct production logics and resource endowments within the irrigated perimeters of Gaya. They provide a meaningful basis for analyzing efficiency differentials and for designing targeted interventions aimed at improving farm performance (see Figure 1).

3.2. Socioeconomic Characteristics of Farm Managers by Rice Farm Type

Table 3 indicates that the majority of farm managers lack formal education across all farm categories (67.2% in medium-sized farms, 64.3% in commercial farms, and 68.4% in small family farms), with no statistically significant differences between groups ($\chi^2 = 4.540$; p = 0.338). This pervasive absence of formal schooling reflects structural constraints in rural Niger, where agricultural knowledge is predominantly transmitted informally through family and community networks rather than vocational or technical training. Similarly, marital status and gender distributions do not vary significantly across farm types ($p > 0.05$), with most managers being married men, underscoring the persistence of male dominance in land access and decision-making within irrigated perimeters. In contrast, household income diversification differs significantly across farm categories ($\chi^2 = 16.42$; $p = 0.012$). Medium-sized farms demonstrate greater diversification, combining rice cultivation with complementary activities such as petty trade, livestock rearing, and seasonal migration. These findings suggest that income diversification may represent a critical pathway for enhancing economic sustainability in rice-based systems.

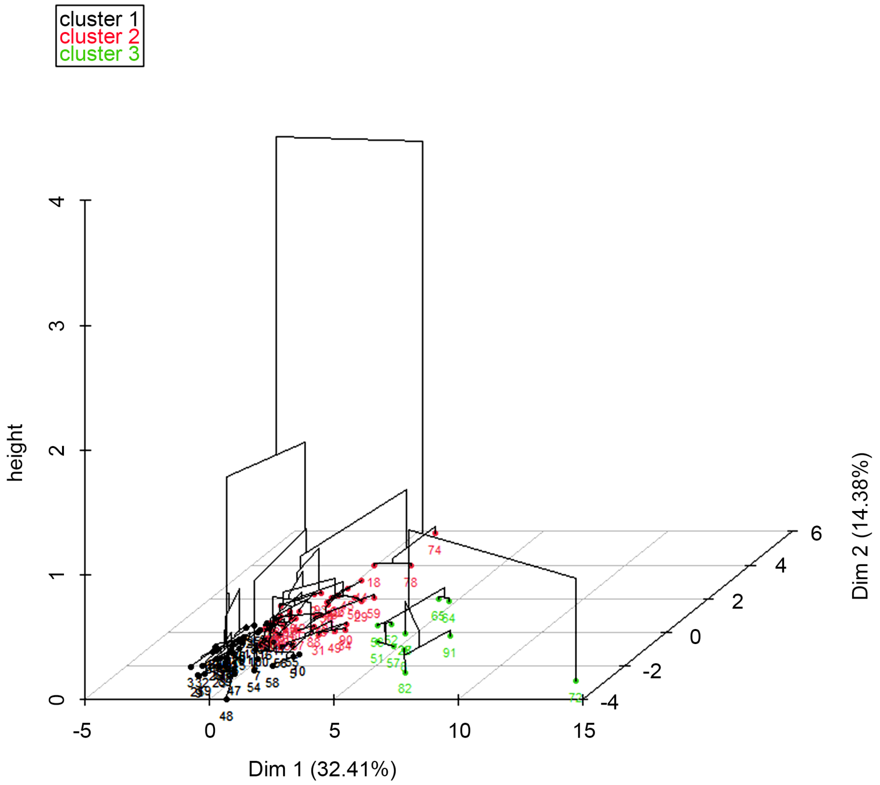

Figure 1. Rice Farms groups.

Table 3. Socioeconomic characteristics of farm managers by rice farm type

Variables	Modalities	Farm type			(Chi-squared (χ^2)). Sig
		F3	F1	F2	
Education level	No formal education	67.2%	64.3%	68.4%	(4.540) 0.338
	Primary	22.4%	35.7%	31.6%	
	Secondary	10.4%	0.0%	0.0%	
Marital status	Unmarried	1.5%	0.0%	0.0%	(0.498) 0.780
	Married	98.5%	100.0%	100.0%	
Gender of farm manager	Male	94.0%	100.0%	100.0%	(2.05); 0.358
	Female	6.0%	0.0%	0.0%	

Continued

Number of household income sources	One (1)	7.5%	0.0%	10.5%	(16.42); 0.012
	Two (2)	49.3%	64.3%	21.1%	
	Three (3)	37.3%	35.7%	36.8%	
	Four (4)	6.0%	0.0%	31.6%	

3.3. Analysis of Technical Indicators by Rice Farm Typology

Figure 2 illustrates that commercial or semi-intensive farms (Cluster 1), characterized by larger landholdings, make substantial use of inputs such as seeds, fertilizers, and urea. This intensive input use enables them to achieve high rice yields ranging from 2000 to 6000 kg, with some farms reaching up to 9000 kg. These farms are typically managed by older farm heads, reflecting greater agricultural experience. In contrast, small family farms (Cluster 2), with smaller plots, rely on moderate input levels and produce lower rice yields, generally between 1500 and 4000 kg. These farms are often managed by younger producers and follow less intensive, more traditional farming practices. Medium-sized farms (Cluster 3) represent a middle ground, with moderate land area and input use, yielding between 3000 and 6000 kg of rice. The farm managers in this group tend to be of intermediate age, suggesting a balance between traditional knowledge and innovative management approaches.

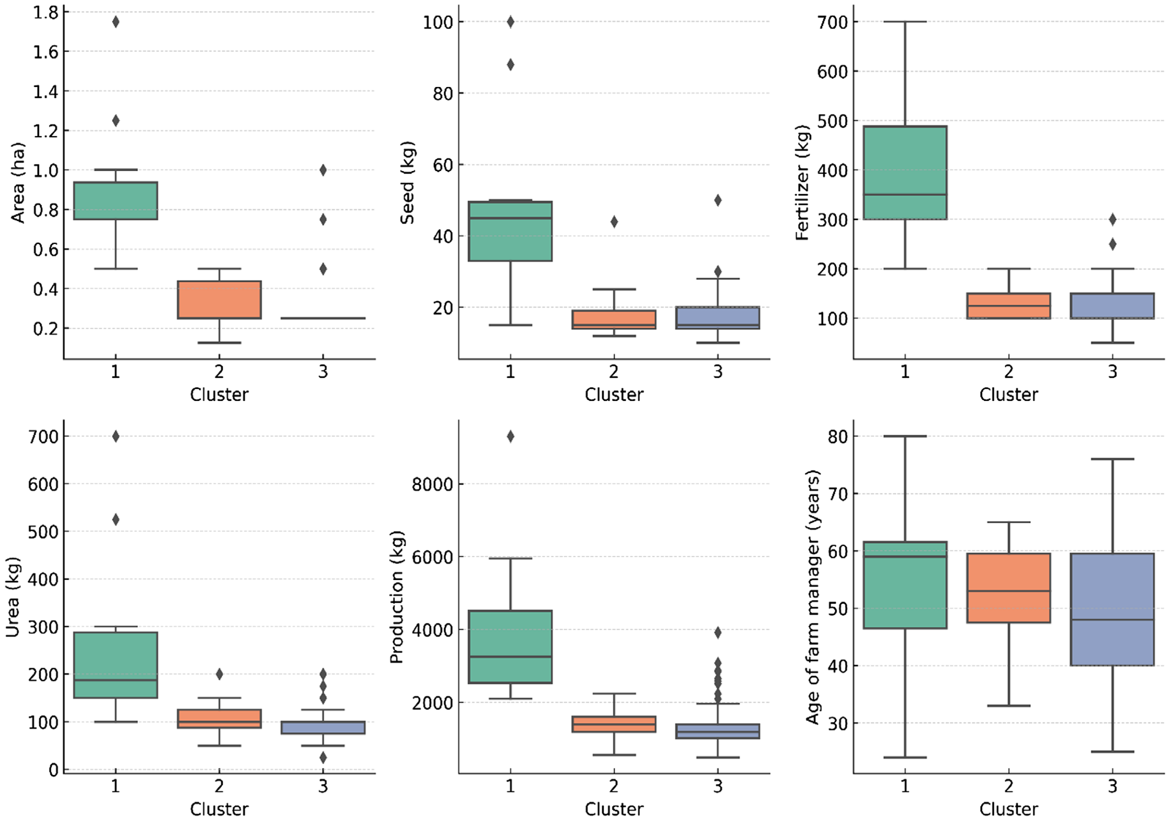

Figure 2. Technical indicators by rice farm typology.

3.4. Distribution of DEA Scores by Farm Type and Share of Efficient Units

Figure 3 shows that commercial or intensive farms (Cluster 1) exhibit the highest technical efficiency scores (average DEA score of 0.83) and the largest share of efficient units (27%). Their median score surpasses that of medium-sized farms (0.76) and small family farms, reflecting the advantage of larger landholdings, higher input intensity, and greater managerial experience. In the context of Gaya, these farms benefit from better access to irrigation infrastructure and markets, which allows them to optimize resource use more effectively. In contrast, small family farms (Cluster 2) display greater variability and relatively lower DEA scores (average of 0.71), with only 12% of units classified as efficient. This lower performance may be linked to constraints in input access, reliance on family labor, and limited financial capacity. Such conditions often lead to suboptimal resource allocation and reduced resilience to shocks, highlighting the structural vulnerability of smallholders in Gaya's irrigated perimeters. Medium-sized farms (Cluster 3) occupy an intermediate position, with balanced input use and production levels, and a moderate efficiency profile. Their average score and share of efficient units suggest a more stable and adaptive management approach. In Gaya's transboundary economy, these farms often combine traditional practices with incremental innovations, enabling them to maintain efficiency despite resource constraints. This intermediate performance underscores their potential as a target group for policy interventions, as they are well positioned to scale up efficiency gains if provided with improved access to inputs, training, and institutional support.

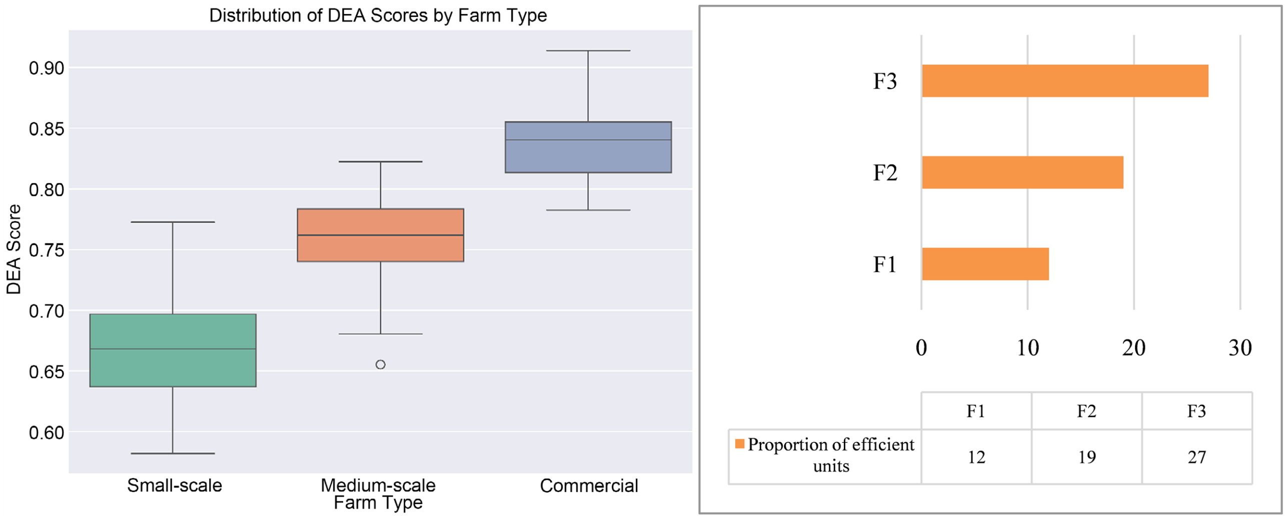

Figure 3. Distribution of DEA scores by farm type and share of efficient units.

3.5. Analysis of Observed and Predicted DEA Scores Using the Random Forest Model

The results reveal a strong concordance between observed and predicted DEA scores, as evidenced by a high coefficient of determination ($R^2 = 0.9168$) and a very low Mean Squared Error (MSE = 0.0027). To validate model robustness, we report the Out-of-Bag (OOB) error estimate, which provides an internal cross-validation during Random Forest training. The OOB error rate was 0.087, corresponding to a mean squared error (MSE) of 0.0031, confirming that predictive accuracy remained stable when tested on unseen data. As an additional check, a 5-fold cross-validation was performed, where the dataset was partitioned into five subsets and iteratively trained and tested. The average cross-validated R^2 was 0.89, with a mean absolute error (MAE) of 0.041, closely aligned with the original model performance. These results demonstrate that the Random Forest model generalizes well and is not overfitting the sample, thereby strengthening confidence in the identified determinants of technical efficiency. The correlation between observed and predicted scores reaches 0.9575, indicating the model's exceptional ability to capture efficiency variations based on the selected explanatory variables. The dispersion of points around the ideal diagonal line (y = x) remains minimal, further confirming the model's precision and stability in predicting technical efficiency across rice farms (**Figure 4**). From an economic perspective, this high level of predictive accuracy suggests that the Random Forest model successfully identifies the structural and managerial factors that drive efficiency in Gaya's irrigated rice systems.

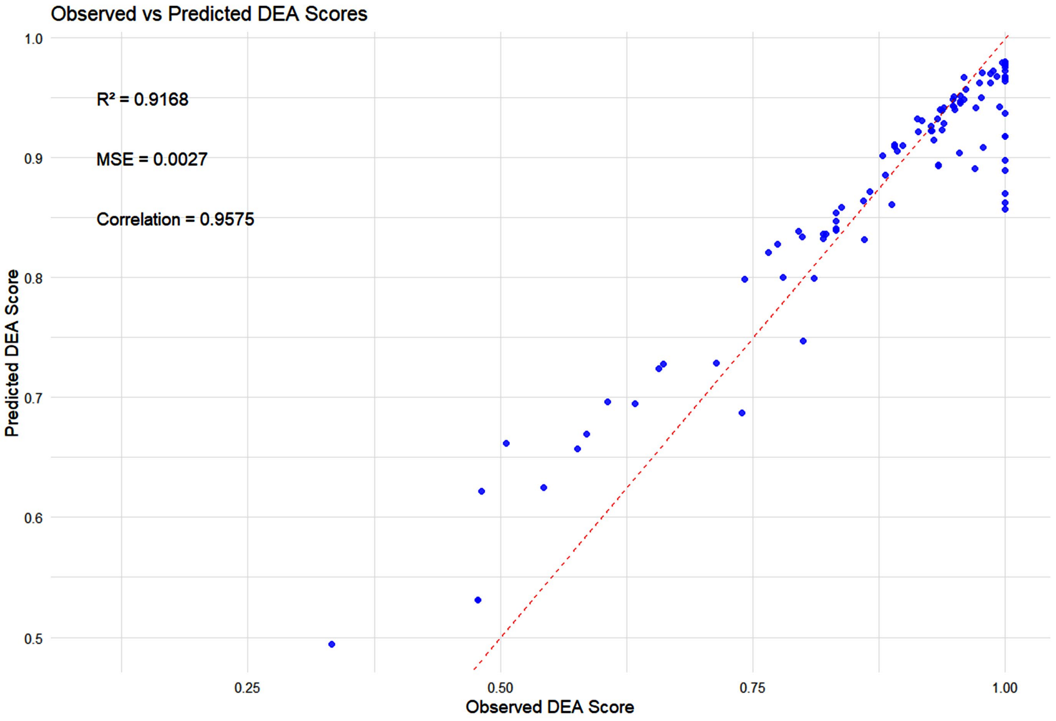

Figure 4. Observed vs. predicted DEA scores using the random forest model.

3.6. Analysis of Variables Importance in Predicting Technical Efficiency

The Random Forest regression model was employed to complement the DEA results by identifying the determinants of technical efficiency among family rice farms. Variable importance was assessed using the increase in mean squared error (IncMSE) and node purity indices, which quantify the predictive contribution of each factor to efficiency scores. The analysis revealed that seed quantity (17.44) is the most influential variable, confirming that the choice and dosage of seeds are central to production efficiency. This finding aligns with agronomic literature emphasizing the role of seed quality and planting density in yield optimization. Farm size (7.63) ranked second, suggesting that scale effects and land allocation strategies significantly influence efficiency, consistent with the hypothesis of economies of scale in smallholder systems. Other agronomic inputs also demonstrated positive contributions. Herbicide (6.25) use and fertilizer (6.08) application were strong predictors, highlighting the importance of weed control and nutrient management. The ability to optimize direct costs and production-related fees (5.57) contributed positively, indicating that farms which strategically allocate financial resources achieve higher efficiency. Urea application (3.14) further reinforced the role of targeted fertilizer use. Additionally, household income diversification (1.18) and family labor during the rice production phase (0.11) showed modest but positive effects, suggesting that economic resilience and timely labor allocation can support technical performance. In contrast, socio-demographic variables such as marital status, education level, and gender of the household head exhibited negligible or negative importance. Similarly, labor-related variables outside the production phase contributed little to efficiency, implying that efficiency is more dependent on the quality and timing of input use than on household composition.

Table 4. Variables Importance and theirs ranks.

Variable	IncMSE I	IncNodePurity	Rank
Seed quantity (kg)	17.4468387	0.560904153	1
Farm size (ha)	7.6345068	0.200337689	2
Herbicide quantity (liters or kg)	6.2564666	0.127726848	3
Fertilizer quantity (kg)	6.0814333	0.175834486	4
Ability to optimize direct costs (FCFA/ha)	5.5720796	0.262230645	5
Irrigation fee (FCFA)	4.6538309	0.192231486	6
Urea quantity (kg)	3.1469459	0.073189218	7
Number of household income sources (count)	1.1893500	0.026701936	8
Labor during rice production and maintenance phase (man-days/ha)	0.1152123	0.034089415	9
Marital status (categorical)	0.0000000	0.014420372	10
Education level (categorical)	−0.0883418	0.015512092	11
Labor during post-rice phase (man-days/ha)	−0.4249930	0.027074891	12
Gender of household head (categorical)	−0.6850047	0.001270608	13
Age of household head (years)	−0.7763113	0.102975605	14
Number of active household members (count)	−1.5618457	0.043026255	15
Labor during pre-rice phase (man-days/ha)	−1.6353117	0.048736864	16
Miscellaneous expenses (FCFA)	−2.9545572	0.101345495	17

4. Discussion

This study employed Principal Component Analysis (PCA) and Hierarchical Ascendant Classification (HAC) to identify the main determinants of technical efficiency among rice farms in Gaya, Niger. The PCA results showed that the first three principal components accounted for 54.16% of the total variance in farm characteristics. These components were mainly driven by farm size, input use, and socio-economic attributes, consistent with [22], who identified farm size and input intensity as major determinants of efficiency in rice production. The first component, which explained the highest share of variance (32.41%), was dominated by agronomic and economic variables such as seed quantity, fertilizer use, land area, and input costs, indicating that these factors are central to productivity in Gaya's rice production systems. Following dimensionality reduction, HAC classified the farms into three distinct groups: commercial or semi-intensive farms, small family farms, and medium-sized farms. Commercial or semi-intensive farms were characterized by larger cultivated areas, high input intensity, and experienced management, which translated into higher yields. These findings corroborate previous evidence showing that input intensification enhances agricultural performance [23] and [24]. Small family farms used moderate input levels and relied on traditional practices, which limits their capacity to achieve optimal production outcomes, as suggested by [25]. Medium-sized farms represented an intermediary production model, combining moderate land size with balanced input use, thereby benefiting from flexible management while maintaining relatively high efficiency [26]. Across all farm categories, a low level of formal education was observed among farm managers, with more than two-thirds having no formal schooling. This finding aligns with studies conducted in rural Sub-Saharan Africa [27], which identify low educational attainment as a major constraint to the adoption of innovations and efficient resource management. While education is not a direct driver of productivity, it plays an important role in facilitating the uptake of modern agricultural technologies and in improving farm recordkeeping and interactions with support institutions [28]-[30]. Similarly, the predominance of married male managers reflects the patriarchal structure of the local agricultural system, in which decision-making and land management remain male-dominated [31]. This social homogeneity may limit managerial diversity and reduce women's involvement in rice production, despite their significant contributions to labor and post-harvest processes. Income diversification differed significantly across farm categories ($\chi^2 = 16.42$; p = 0.012). Medium-sized farms showed higher diversification, combining rice cultivation with activities such as livestock rearing, market gardening, and petty trade, thereby strengthening their economic resilience. This observation is consistent with [32], who demonstrated that diversification enhances rural household resilience in contexts of structural vulnerability. Pluriactivity thus constitutes a key mechanism for risk mitigation, enabling households to diversify income sources and reduce dependence on rice cultivation alone. Medium-sized farms were more resilient to climatic shocks and market fluctuations, whereas commercial farms that specialize exclusively in rice production were more vulnerable to hydrological variability, yield losses, and price instability. This vulnerability is reinforced by a social system where decision-making remains centered on male heads of households, limiting women's participation in strategic management [31] [33]. Commercial farms' reliance on high input levels such as improved seeds, fertilizers, urea, and herbicides resulted in higher yields (2000 - 8000 kg/ha), in line with findings by [32] and [34]. The positive relationship between input intensity and productivity is well established, provided that inputs are efficiently managed and supported by appropriate technologies and institutional frameworks. Empirical evidence from Senegal demonstrates that agricultural intensification enhances productivity outcomes in smallholder systems [35]. In contrast, small family farms applied lower quantities of inputs (15 - 20 kg of seeds and <100 kg of fertilizer), resulting in lower yields (1500 - 4000 kg/ha). Their reliance on household labor and conservative management practices limits productivity [25]. The DEA results revealed substantial heterogeneity in technical efficiency across farm types. Commercial or semi-intensive farms achieved the highest efficiency scores (mean = 0.83), reflecting their superior capacity to transform inputs into outputs. These trends align with earlier studies showing that larger farms benefit from better access to productive resources, inputs, credit, and innovations [36] and [37]. Conversely, small family farms showed lower efficiency and greater score variability (approximately 12% classified as efficient), a result attributed to structural constraints such as limited land size, insufficient capital, and restricted access to mechanization and financial services [38] [39]. The Random Forest model exhibited excellent predictive performance, explaining 91.68% of the variance in technical efficiency scores. The close alignment of predicted values with the 1:1 line demonstrates the model's robustness and accuracy. These results highlight the potential of machine learning methods as reliable tools for estimating technical efficiency without extensive field measurements [40]. The Random Forest variable importance analysis showed that technical efficiency is primarily driven by input-related factors, cost optimization capacity, land structure, and income diversification. Seed quantity was the most influential predictor, consistent with research emphasizing the importance of planting density and seed quality in maximizing yields [41] [42]. Farm size was the second most important variable, supporting evidence of scale economies in smallholder agriculture [43] [44]. Herbicide use and fertilizer application also contributed significantly, echoing findings from Senegal [45] and China [46]. Economic variables related to the management of direct production costs were positively associated with efficiency, in line with [47]. Income diversification also had a positive effect, consistent with [48] and [49], who argue that diversified households exhibit greater resilience and production efficiency. Labor dedicated to rice production also contributed positively, reflecting the importance of timely operations during critical phenological stages. In contrast, socio-demographic variables (marital status, education, gender) exhibited negative or negligible importance, supporting evidence that technical efficiency is primarily shaped by agronomic and economic decisions rather than household composition [50].

5. Contribution and Novelty

This study introduces a novel and integrated methodological framework for analyzing technical efficiency in heterogeneous agricultural systems by combining Principal Component Analysis (PCA), Hierarchical Ascendant Classification (HAC), Data Envelopment Analysis (DEA), and Machine Learning techniques. The primary methodological novelty lies in the structured sequencing of these tools to address key limitations of conventional efficiency studies. First, the study innovates by incorporating multivariate classification prior to efficiency estimation. Through PCA and HAC, farms are grouped into homogeneous typologies before applying DEA, thereby reducing bias associated with structural heterogeneity among decision-making units. This step enhances the robustness and interpretability of efficiency scores, an issue often overlooked in standard DEA applications. Second, the study employs an input-oriented DEA model under Variable Returns to Scale (VRS), which is particularly suited to resource-constrained smallholder farming systems. This specification allows efficiency to be assessed independently of scale effects while reflecting farmers' realistic objective of minimizing input use for a given level of output. Third, a major methodological innovation is the integration of a Random Forest algorithm as a second-stage analytical tool to explain and predict DEA efficiency scores. Unlike traditional econometric approaches commonly used in DEA second-stage analysis, the Machine Learning framework captures nonlinear relationships and complex interactions among agronomic, economic, and socio-demographic variables. The high predictive accuracy achieved demonstrates the effectiveness of this approach in modeling agricultural efficiency outcomes. Finally, the use of variable importance measures derived from Random Forest provides a transparent and data-driven ranking of efficiency determinants, offering deeper insights than conventional regression coefficients. By transforming DEA from a purely diagnostic tool into a predictive and scalable framework, this study extends the methodological frontier of efficiency analysis and offers a replicable approach for future research in data-scarce and heterogeneous agricultural contexts.

6. Limitations and Future Research

While the Random Forest model provided strong predictive accuracy and highlighted key determinants of efficiency, the relatively small sample size (103 farms) compared to the number of predictors (17 variables) raises methodological considerations. A low observations-to-variables ratio can affect the stability of variable importance rankings, particularly for weaker socio-demographic predictors. Strong technical factors such as seed quantity, farm size, and fertilizer use consistently emerge as robust determinants, but the relative importance of less influential variables may fluctuate across validation runs. This limitation suggests that the rankings in Table 4 should be interpreted as indicative rather than definitive. Future studies with larger samples are needed to confirm the robustness of these findings and to strengthen the reliability of policy recommendations derived from the model. While the sample remains statistically representative of the target population, the relatively high margin of error may reduce the precision of subgroup comparisons and limit the generalizability of findings beyond the study area. Future research could reduce this margin by increasing sample size or by employing stratified sampling to ensure more balanced representation across farm types. Despite this limitation, the results provide valuable insights into efficiency patterns and determinants in irrigated rice farming systems, but they should be considered as indicative rather than exact measures.

7. Conclusions and Recommendations

This study assessed the technical efficiency of family rice farms in the urban commune of Gaya using a hybrid framework that combines Data Envelopment Analysis (DEA) and Machine Learning. The results reveal marked heterogeneity across farm profiles: commercial and semi-intensive farms achieved higher DEA scores and a greater proportion of efficient units, while small family farms lagged behind. Principal Component Analysis (PCA) identified farm size, input intensity, socio-economic characteristics, and production volume as key dimensions of variation. The Random Forest model demonstrated strong predictive accuracy, confirming the robustness of these explanatory variables. The predominance of technical factors such as input costs, fertilizer use, and farm size relative to the modest role of socio-demographic variables underscores that rational intensification and structured production practices are central to efficiency gains. These findings carry important policy implications. First, efficiency improvements from higher input use must be weighed against marginal costs, suggesting that subsidies or credit schemes should be carefully designed to ensure fiscal sustainability. Second, the distinction between technical efficiency (minimizing inputs for a given output) and allocative efficiency (choosing the optimal input mix given prices) is critical: while DEA captures technical efficiency, the Random Forest results highlight the decisive role of input costs, pointing to the need for interventions that also address allocative efficiency to prevent resource misallocation. Finally, the heterogeneity among farm types calls for differentiated strategies: commercial farms may benefit from cost-optimization measures, whereas small family farms require support in both input access and managerial capacity. Linking efficiency improvements to broader objectives such as income diversification and resilience is essential to ensure that productivity gains contribute to equitable and sustainable rural development. In this way, the DEA-ML framework provides a rigorous foundation for designing policies that balance efficiency enhancement with long-term economic sustainability. Based on these findings, the following recommendations are proposed:

- Targeted capacity building: Medium-sized farms and small family units should benefit from extension programs focused on input optimization and integrated resource management.

- Support for rational intensification: Commercial or intensive farms, although more efficient, should be guided in managing scale effects and diversifying practices to avoid the risks of overcapitalization.
- Extension of the approach to other agroecological zones: The DEA-ML methodology can be replicated in other regions of Niger or West Africa to support territorial planning and data-driven agricultural policies.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Pawlak, K. and Kołodziejczak, M. (2020) The Role of Agriculture in Ensuring Food Security in Developing Countries: Considerations in the Context of the Problem of Sustainable Food Production. *Sustainability*, **12**, Article 5488. https://doi.org/10.3390/su12135488

[2] Thornton, P.K., Loboguerrero, A.M., Campbell, B.M., Kavikumar, K.S., Mercado, L. and Shackleton, S. (2019) Rural Livelihoods, Food Security and Rural Transformation Under Climate Change. https://gca.org/wp-content/uploads/2020/12/RuralLivelihoodsFoodSecurityRural-Transformation_V2.pdf

[3] Mendez del Villar, P. and Bauer, J. (2013) Rice in West Africa: Dynamics, Policies and Trends. *Cahiers Agricultures*, **22**, 336-344. https://doi.org/10.1684/agr.2013.0657

[4] International Water Management Institute (IWMI) (2007) Water for Food, Water for Life: A Comprehensive Assessment of Water Management in Agriculture. https://archive.iwmi.org/assessment/files_new/newsroom/waterforfoodwaterfor-life_2007_SIWI.pdf

[5] Moussa, S.T., Zerbo, P. and Sanou, Y. (2022) Evaluation de la Performance de L'irrigation en Utilisant L'indice du Taux de Satisfaction en eau du Périmètre Irrigué de Saga dans la Vallée du Fleuve Niger. *International Journal of Progressive Sciences and Technologies*, **32**, 453-465.

[6] Latruffe, L. (2010) Efficacité technique des exploitants agricoles, décisions d'utilisation des facteurs de production, et influence des politiques publiques. Thèse de doctorat, Université d'Angers—SMART LERECO. https://hal.inrae.fr/tel-02819656v1/document

[7] Boakye, K., Lee, Y., Annor, F.F., Dadzie, S.K.N. and Salifu, I. (2024) Data Envelopment Analysis (DEA) to Estimate Technical and Scale Efficiencies of Smallholder Pineapple Farmers in Ghana. *Agriculture*, **14**, Article 1032. https://doi.org/10.3390/agriculture14071032

[8] Darko, E., Saghi-Zedek, N. and Thenet, G. (2025) Technical and Economic Efficiency Measurement of African Commercial Banks Using Data Envelopment Analysis (DEA). *Management Science Letters*, **15**, 143-154. https://doi.org/10.5267/j.msl.2024.5.007

[9] Habibullah, M.S. and Ismail, M.M. (1994) Production Frontier and Technical Efficiency: The Case for Beekeeping Farms in Malaysia. B*angladesh Journal of Agricultural Economics*, **XVII**, 31-43.

https://ageconsearch.umn.edu/record/202688?v=pdf

[10] Son, T.V.H., Coelli, T. and Fleming, E. (1993) Analysis of the Technical Efficiency of State Rubber Farms in Vietnam. *Agricultural Economics*, **9**, 183-201. https://doi.org/10.1111/j.1574-0862.1993.tb00267.x

[11] Obwona, M. (2000) Determinants of Technical Efficiency Differentials amongst Small and Medium Scale Farmers in Uganda: A Case of Tobacco Growers. Final Report Presented at the AERC Bi-Annual Workshop, Grand Regency Hotel, Nairobi, Kenya. African Economic Research Consortium. https://aercafrica.org/

[12] Boubacar, O., Hui-Qiu, Z., Rana, M.A. and Ghazanfar, S. (2016) Analysis on Technical Efficiency of Rice Farms and Its Influencing Factors in South-Western of Niger. *Journal of Northeast Agricultural University* (*English Edition*), **23**, 67-77. https://doi.org/10.1016/s1006-8104(17)30009-0

[13] Idrissa, S.M., Soumana, B., Nouhou, M.O. and Sanouna, A. (2022) Economic Performance Determinants of Farms in the Context of Health and Climate Change in Rural District of Kourthèye in Niger. *International Journal of Innovation and Applied Studies*, **37**, 23-35.

[14] Vlaicu, P.A. and Matei, B. (2025) Applications of Machine Learning Technology in Agricultural Data Mining. *Applied Sciences*, **15**, Article 5286. https://doi.org/10.3390/app15105286

[15] INS (2022) Annuaire statistique régional. Institut National de la Statistique du Niger. https://www.stat-niger.org/?page_id=1555

[16] Mossi, M.I., Haougui, A., Sido, YA., Basso, A., Bibata, A., Bizo, N.M., Salami, I. and Salmou, A. (2015) On-Farm Testing of the System of Rice Intensification (SRI) in Lowlands Ecology in Niger. *Journal of Research in Biology*, **9**, 2693-2700. http://jresearchbiology.com/documents/RA0698.pdf

[17] Banker, R.D., Charnes, A. and Cooper, W.W. (1984) Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, **30**, 1078-1092. https://doi.org/10.1287/mnsc.30.9.1078

[18] Coelli, T., Rao, D.S.P., O'Donnell, C.J. and Battese, G.E. (2005) An Introduction to Efficiency and Productivity Analysis. Springer.

[19] Assaf, A.G. and Josiassen, A. (2015) Frontier Analysis: A State-Of-The-Art Review and Meta-Analysis. *Journal of Travel Research*, **55**, 612-627. https://doi.org/10.1177/0047287515569776

[20] Tone, K. and Tsutsui, M. (2010) Dynamic DEA: A Slacks-Based Measure Approach. *Omega*, **38**, 145-156. https://doi.org/10.1016/j.omega.2009.07.003

[21] Farrell, M.J. (1957) The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A* (*General*), **120**, 253-287. https://doi.org/10.2307/2343100

[22] Zhang, X., Li, Y. and Wang, Z. (2015) Analysis of the Efficiency of Rice Production in China: A DEA Approach. *Agricultural Systems*, **137**, 1-9.

[23] Ebeke, C. and Etoundi, M.N. (2025) Crop Productivity in Sub-Saharan Africa: The Role of Research and Development. IMF Working Paper WP/25/249. https://www.imf.org/-/media/files/publications/wp/2025/english/wpiea2025249-source-pdf.pdf

[24] Manyanga, M., Pedzisa, T. and Hanyani-Mlambo, B. (2023) Adoption of Agroecological Intensification Practices in Southern Africa: A Scientific Review. *Cogent Food & Agriculture*, **9**, Article ID: 2261838. https://doi.org/10.1080/23311932.2023.2261838

[25] Coelli, T.J., Rao, D.S., O'Donnell, C.J. and Battese, G.E. (2002) An Introduction to Efficiency and Productivity Analysis. 2nd Edition, Springer.

[26] Esquivel, K.E., Carlisle, L., Ke, A., Olimpi, E.M., Baur, P., Ory, J., *et al.* (2021) The "Sweet Spot" in the Middle: Why Do Mid-Scale Farms Adopt Diversification Practices at Higher Rates? *Frontiers in Sustainable Food Systems*, **5**, Article 734088. https://doi.org/10.3389/fsufs.2021.734088

[27] Bationo, A., Waswa, B., Okeyo, J.M., Maina, F. and Kihara, J. (2011) Innovations as Key to the Green Revolution in Africa. Springer. https://link.springer.com/book/10.1007/978-90-481-2543-2

[28] Mu, H. and Jones, K. (2020) The Role of Education in Agricultural Technology Adoption: A Systematic Review. Harper Adams University, School of Sustainable Food and Farming. https://cdn.harper-adams.ac.uk/document/page/705_The-role-of-education-in-agricultural-technology-a.pdf

[29] Fadeyi, O.A., Ariyawardana, A. and Aziz, A.A. (2022) Factors Influencing Technology Adoption among Smallholder Farmers: A Systematic Review in Africa. *Journal of Agriculture and Rural Development in the Tropics and Subtropics*, **123**, 13-30. https://doi.org/10.17170/kobra-202201195569

[30] Basir, M.S., Buckmaster, D., Raturi, A. and Zhang, Y. (2024) From Pen and Paper to Digital Precision: A Comprehensive Review of On-Farm Recordkeeping. *Precision Agriculture*, **25**, 2643-2682. https://doi.org/10.1007/s11119-024-10172-7

[31] FAO (2011) Save and Grow: A Policymaker's Guide to the Sustainable Intensification of Smallholder Crop Production. FAO. https://openknowledge.fao.org/server/api/core/bitstreams/77bd1c67-4e4c-4b1d-b836-bede1d3bf9dc/content

[32] Bélières, J.F. and Andrianantoandro, R. (2015) Typologie des exploitations agricoles et performances de la riziculture irriguée. CIRAD Working Paper. CI-RAD. https://agritrop.cirad.fr/579746/

[33] Alobo, S. (2017) Options de survie, processus de changement et transformation structurelle dans les exploitations agricoles familiales en Afrique subsaharienne. Thèse de doctorat, INRA. https://univ-tlse2.hal.science/INRA/tel-04543616v1

[34] FAO (2011) The State of Food and Agriculture 2010-2011: Women in Agriculture—Closing the Gender Gap for Development. FAO. https://www.fao.org/3/i2050e/i2050e.pdf

[35] Ndip, F.E. and Sakurai, T. (2025) Enhancing Agricultural Intensification through Contract Farming: Evidence from Rice Production in Senegal. *Agriculture & Food Security*, **14**, Article No. 6. https://doi.org/10.1186/s40066-025-00525-4

[36] Goudjo, G.G. (2023) Accès au financement et productivité de l'agriculture familiale au Bénin: cas des petites exploitations. Ph.D. Thesis, Université d'Abomey-Calavi. https://hal.science/tel-04496575

[37] Amoussouga, G.A. (2021) Innovation technologique, efficacité technique et convergence de la productivité agricole dans l'UEMOA. *Annales de l'Université de Lomé, Série Sciences Économiques et de Gestion*, **16**, 55-67.

[38] Saito, K. (2013) Smallholder Farmers and Sustainable Agriculture. FAO. https://openknowledge.fao.org/server/api/core/bitstreams/5ed2c62d-f6aa-4165-b05d-d1fddb1ab352/content

[39] Ngom, C.A.B., Sarr, F. and Fall, A.A. (2016) Mesure de l'efficacité technique des riziculteurs du bassin du fleuve Sénégal. *Économie rurale*, No. 355, 91-105. https://doi.org/10.4000/economierurale.5021

[40] Li, L. (2021) Accurate and Timely Crop Yield Forecasts Using Machine Learning

Techniques. *Agricultural and Forest Meteorology*, **307**, Article ID: 108482.

[41] Ling, Y., Zhang, H. and Chen, L. (2024) Optimum Seeding Density and Seedling Age for Rice Yield. *Frontiers in Plant Science*, **15**, Article 1123.

[42] Zhu, X. (2025) Optimizing Planting Density to Enhance Rice Productivity. *Rice Genomics and Genetics*, **16**, 71-85. https://doi.org/10.5376/rgg.2025.16.0007

[43] Muteti, F.N., Akite, I., Mpofu, T.P. and Mugonola, B. (2023) Determinants of Technical Efficiency among Smallholder Upland Rice Farmers in Northern Uganda—A Cobb–douglas Stochastic Frontier Approach. *SN Business & Economics*, **4**, Article No. 4. https://doi.org/10.1007/s43546-023-00597-z

[44] Anang, B.T., Bäckman, S. and Sipiläinen, T. (2016) Technical Efficiency and Its Determinants in Smallholder Rice Production in Northern Ghana. *Journal of Developing Areas*, **50**, 311-328.

[45] Diallo, M.F. and Garba, A.A. (2024) Effect of Seed and Fertilizer Subsidies on the Technical Efficiency of Rice Farmers in Senegal. *Journal of Agricultural Science*, **16**, 81-93. https://doi.org/10.5539/jas.v16n10p81

[46] Sun, Z. and Li, X. (2021) Technical Efficiency of Chemical Fertilizer Use and Its Influencing Factors in China's Rice Production. *Sustainability*, **13**, Article 1155. https://doi.org/10.3390/su13031155

[47] Mukwalikuli, M. (2018) Determinants of Technical Efficiency of Smallholder Rice Farming in Zambia. Neliti Research Repository.
FAO (2023) Crop Diversification Increases Productivity and Stabilizes Income of Smallholders. Food and Agriculture Organization of the United Nations.
https://openknowledge.fao.org/server/api/core/bitstreams/0b710e4a-e688-4171-a30a-0dcfe2b73477/content

[48] Hien, T. (2025) Relationship between Crop Diversification and Farm Efficiency: Does Farm Size Matter? *European Review of Agricultural Economics*, **52**, 334-377. https://doi.org/10.1093/erae/jbaf016

[49] Yang, F., Wu, G.Y., Riaz, N. and Radlińska, K. (2024) Technical Efficiency and Farm Size in the Context of Sustainable Agriculture. *Agricultural Economics*, **70**, 446-456. https://doi.org/10.17221/158/2024-AGRICECON