Harnessing AI for sustainable agriculture: exergy efficiency optimization in maize *via* neighbourhood component analysis, principal component analysis, extreme gradient boosting

Kutalmis Turhal,[1] Umit Cigdem Turhal,[2] Yasemin Onal[2]

[1]Biosystem Engineering Department, Agriculture and Natural Sciences Faculty;
[2]Electric and Electronic Engineering Department, Engineering Faculty, Bilecik Seyh Edebali University, Bilecik, Turkey

Corresponding author: Umit Cigdem Turhal Electric and Electronic Engineering Department, Engineering Faculty, Bilecik Seyh Edebali University, Bilecik 11210, Turkey. E-mail: ucigdem.turhal@bilecik.edu.tr

Contributions: all authors made a substantive intellectual contribution, read and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Conflict of interest: the authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and materials: All data generated during this study are available on request from the first author.

Highlights

- AI exergy analysis for maize using a variety of inputs: labor, fertilizer, machinery.
- GA-tuned NCA-PCA-XGBoost predicts exergy efficiency and CExE.
- Accuracy: MAE 1.5027, MSE 4.7553, R^2=0.99 on maize data.
- Strong fit in low–moderate efficiency range (0–30%); minimal bias.
- Supports sustainable input planning; reduces waste and fossil use.

Abstract

Conventional analyses of agricultural energy systems often ignore the concepts of exergy degradation and irreversibilities, which results in an incomplete understanding of useful energy loss. Ignoring these factors can lead to poor input and technology choices that ultimately harm sustainability. To fill this gap, we introduce an AI-supported framework designed to predict and explicitly optimize exergy efficiency and cumulative exergy consumption (CExC), specifically in maize cultivation, by using controllable inputs. This research represents a novel endeavor, as it is, to the best of our knowledge, the first study aimed at optimizing exergy efficiency at the crop production level through the application of machine learning techniques to various input variables. This study utilizes a comprehensive input-output dataset comprising 112 observations and encompassing seven input variables -human labor, machinery, diesel fuel, seed, fertilizer,

chemicals, and water- alongside a singular output variable (output energy). The target variables, exergy efficiency and CExC, are calculated by systematically mapping energy flows into the exergy domain. Our methodological approach integrates neighborhood component analysis (NCA) for effective feature selection, principal component analysis (PCA) for dimensionality reduction, and XGBoost for predictive modeling. Furthermore, the optimization of hyperparameters is conducted through a genetic algorithm (GA) to enhance model performance. Implementation of this framework was carried out in Python 3.11, utilizing libraries such as scikit-learn and XGBoost. The evaluation of model performance was conducted using three metrics: mean absolute error (MAE), mean squared error (MSE), and the coefficient of determination (R^2). The study's results indicate a MAE of 1.5027, a MSE of 4.7553, and an R^2 of 0.99. These metrics indicate an exceptional model fit within the low-to-moderate efficiency spectrum (0-30%) while exhibiting minimal bias. Such findings suggest that the proposed framework can effectively inform sustainable input planning strategies aimed at mitigating energy waste and decreasing reliance on fossil fuels, all while preserving agricultural output. Moreover, insights derived from the analysis highlight which adjustments in labor, fertilizer, and machinery contribute most significantly to exergy efficiency gains. Additionally, the integration of economic and CO_2 metrics into the model, as well as the exploration of real-time decision support systems in field trials, would represent valuable advancements in this endeavor. Future work will focus on creating plot-level datasets that include harmonized covariates, such as soil properties, water regime, weather conditions, and pest pressure. These datasets will support causal inference and threshold estimation, fully disaggregated with separate measures for nitrogen (N), phosphorus (P), potassium (K, including K_2O), as well as diesel and electricity usage.

Key words: exergy efficiency; maize production; machine learning; sustainable agriculture; NCA-PCA-XGBoost; data-driven optimization.

Introduction

The increasing energy demand has intensified reliance on fossil fuels, resulting in a range of environmental issues, including acid rain, global warming, and elevated greenhouse gas emissions (Dutta, 2021). Concurrently, inefficient technologies exacerbate energy losses, highlighting an urgent need for enhancements in efficiency to mitigate waste and dependence on fossil fuels (Li *et al*., 2023). Traditional energy-flow analyses in agriculture typically focus on input–output balance metrics (Xu *et al*., 2022), yet they often neglect energy quality and irreversibilities. In contrast, exergy analysis quantifies the potential for useful work and identifies irreversible losses, providing a stronger foundation for optimization (Hercher-Pasteur *et al*., 2020; Chowdhury *et al*., 2020; Wang *et al*., 2020). Complementary indicators such as cumulative exergy consumption (CExC) capture the total exergy extracted from natural resources, serving as a critical signal of environmental depletion (Soleymani *et al*., 2025; Hercher-Pasteur *et al*., 2020). CExC is a key metric in evaluating the sustainability and efficiency of production systems, particularly in agricultural contexts, and is defined as the difference between useful output exergy

(Ex_out) and total input exergy (Ex_in). In other words, it quantifies the amount of resource exergy utilized by a production system that does not translate into useful output. Formally, CExC represents a measure of depletion and irreversibility, capturing the "work potential" consumed and partially lost through inefficiencies during the production process (Liu and Liu, 2020).

A lower CExC indicates a more sustainable use of resources relative to the output produced, signifying reduced resource draw and losses. Conversely, a higher CExC reflects a greater draw on resources coupled with increased losses. The concept of exergy efficiency can be encapsulated in the ratio of Ex_out to CExC (Sejkora *et al*., 2020).

In many empirical datasets, it is observed that Ex_out is significantly smaller than the Ex_in, leading to a strong correlation between CExC and Ex_in, which can result in similar visual representations in graphical analyses. Understanding this relationship is crucial for assessing the overall performance and sustainability of production systems.

Recent exergy-centred applications further map thermodynamic losses at plant, system, and regional scales (Alzaben, 2025; Qi *et al*., 2025; Wang *et al*., 2021) and underscore fuel/fertilizer dominance in crop systems (Hesampour *et al*., 2022). In this framing, exergy efficiency -unlike conventional energy efficiency- explicitly accounts for energy quality and irreversibilities; low exergy efficiency is typically associated with high CExC and unsustainable resource use, whereas higher exergy efficiency aligns with lower CExC and reduced environmental impacts (Amiri *et al*., 2020; Soleymani *et al*., 2025).

Artificial intelligence (AI) is currently being applied in various agricultural tasks, such as recognizing plant diseases and pests, detecting stress and weeds, and providing decision support. Recent notable advancements in this area, based on vision and sensing studies, include the development of a YOLO-based DFN-PSAN classifier (Dai *et al*., 2023), a compact multimodal ITF-WPI framework that integrates images and text (Dai *et al*., 2024), and a lightweight Vision Transformer optimized for edge deployment using Lite-AVPSO proposed by Dai *et al*.(2025). Besides these vision- and sensing-based advances, AI is also increasingly being integrated into energy and exergy assessments within agro-systems, enabling data-driven diagnosis of losses of agricultural applications and optimization of input planning for sustainability. This integration addresses the inherent nonlinearity and heterogeneity of these systems, facilitating data-driven improvements in both efficiency and sustainability (Yang *et al*., 2024; Beni *et al*., 2023). Beyond classical crop-level modelling -e.g., output-energy estimation and energy-use prediction with CD/MLR/MLP/RBF/SVM for potato (Bolandnazar *et al*., 2020), ANN-GA for kiwifruit (Soltanali *et al*., 2017), exergy-SVM-GA for rapeseed (Esmaeilpour-Troujeni *et al*., 2021), exergy-flow analyses in paddy rice (Nikkhah *et al*., 2021), and ELM/SVM for wheat (Mostafaeipour *et al*., 2020)- recent studies also demonstrate AI for resource optimization and emissions/energy modelling across

operations and value chains (Cheema *et al.*, 2025; Balać *et al.*, 2025; Assimakopoulos *et al.*, 2024). At the same time, supply-chain and multi-crop sustainability work continues to expand (Nadi *et al.*, 2022; Yildizhan, 2017; Noorani *et al.*, 2023; Rasoolizadeh *et al.*, 2022), while hybrid AI-thermodynamic frameworks emerge for exergoeconomic/exergoenvironmental optimization (Nabavi-Pelesaraei *et al.*, 2023).

Despite this growing body of work, a notable gap remains: the explicit optimization of exergy efficiency with respect to controllable agricultural inputs, such as labor, fertilizers, and machinery, at the crop-production level has not been addressed. Much of the prior literature emphasizes energy-use prediction, resource scheduling, LCA/carbon footprinting, or descriptive exergy accounting rather than optimizing exergy efficiency itself (Asl *et al.*, 2023).

To address this gap, this study presents an AI-supported framework to optimize exergy efficiency and predict CExC within maize production. Through a structured literature review (Scopus, Web of Science, Google Scholar, TR Dizin; 2000–2025), a maize input–output database was constructed from 112 eligible papers and units were harmonized to MJ ha^{-1}. The dataset comprises seven inputs (human labor, machinery, diesel, seed, fertilizer, chemicals, water) and one output (output energy). A GA-tuned NCA-PCA-XGBoost pipeline was implemented in Python 3.11 (scikit-learn, XGBoost; Jupyter) to optimize exergy efficiency and predict CExC. Performance was evaluated using MAE, MSE, and R^2 metrics. In the database used in the study, electricity use (e.g., electric pumping) could not be distinguished in the source records and is therefore excluded, and the Fertilizer variable combines N-P-K, where K_2O was not reported separately.

Within these constraints, the proposed framework operationalizes exergy-efficiency optimization via controllable inputs, introduces the GA-tuned NCA-PCA-XGBoost methodology to agricultural production, and offers a data-driven instrument to identify high-leverage adjustments that minimize losses, improve resource efficiency, and foster sustainable, low-waste crop cultivation.

The main contributions of the study can be summarized as:

- It is the first study to explicitly optimize exergy efficiency in crop production as a function of controllable inputs (labor, fertilizers, machinery, water, chemicals, seed, diesel), demonstrated on maize.
- It introduces a GA-tuned NCA-PCA-XGBoost pipeline to jointly predict exergy efficiency and CExC, providing interpretable feature ranking (NCA) and compact representations (PCA), with strong performance (MAE 1.5027; MSE 4.7553; R^2 0.99).

- It delivers a deployable, data-driven decision-support tool that identifies high-leverage input adjustments to cut exergy losses, reduce fossil reliance, and improve resource efficiency for sustainable agricultural management.

In the remaining part of the study, the detailed Methodology section including exergy analysis, NCA, PCA, XGBoost and GA algorithms are presented. This is followed by the results and discussion, and concluding with the conclusions.

Methodology

Dataset description and exergy analysis

For this study, a structured search was conducted in Scopus, Web of Science, Google Scholar, and TR Index for the period from 2000 to 2025. The search strings were '(maize OR corn) AND (energy OR exergy) AND (tillage OR seeding) AND (Turkey OR global)'. A full-text eligibility assessment was performed after title/abstract screening. Inclusion criteria: i) maize studies conducted under field conditions, ii) input–output energy calculations and data convertible to MJ ha^{-1}, iii) statement of unit and moisture basis. Exclusion criteria: greenhouse/controlled environment, economic models only, unit inconsistency, or missing main variable. A total of 112 studies were included in the data set. Data extraction was performed by a two-stage check in Excel; all inputs were converted to MJ ha^{-1}, yields were adapted to a 14% moisture basis, fertilizers were collected with N/P_2O_5/K_2O mappings and labeled under 'Fertilizer'. The Fertilizer variable combines nitrogen (N), phosphorus (P), and potassium (K), but it did not separately report the specific contribution of potassium oxide (K_2O). Additionally, the available records did not allow for disaggregating electricity use, such as electric pumping; therefore, this aspect has been excluded from the analysis.

The inputs -human labor, machinery, diesel fuel, seed, fertilizer (N/P_2O_5/K_2O), chemicals, water, and electricity- along with the output (energy), were converted to MJ ha^{-1} using a unique coefficient table from 2021 to 2025. Maize yields were standardized to 14% moisture content. We applied unit and outlier validations, such as winsorization and log-scale adjustments.

The collected data was systematically structured in a tabular format, where columns represent different attributes of input variables, and rows correspond to individual data points. This structured approach ensures efficient organization and easy accessibility for analysis. The input space for the feature selection process consists of row vectors, expressed as $Xm = (x_{m1},.....x_{mn})$, n number of features) facilitating accurate data processing and model optimization. In the agricultural production process, it is important to calculate energy efficiency coefficients to determine the input and output energy equivalents. However, the energy equivalents can vary depending on the specific processes involved. Agricultural systems typically involve human

labor, machinery, fuel, etc. The energy efficiency coefficients used in this study is given in Table 1.

Exergy analysis is a valuable thermodynamic tool used to assess the quality and efficiency of energy use. Unlike conventional energy analysis, which focuses solely on the total energy consumed, exergy analysis differentiates between the energy that can be transformed into useful work and the energy that is lost due to inefficiencies. Exergy analysis assesses how effectively energy is utilized, distinguishing between useful exergy and irreversibilities, which are energy losses that reduce system efficiency. From the useful exergy, important metrics such as exergy efficiency and CExC are derived to evaluate the sustainability of the production process. Below are the key exergy formulations utilized in this optimization framework.

Total exergy Input $Ex_{in} = \sum_{i=1}^{n}(m_i.\psi_i)$refers to the overall energy available from various inputs necessary for the process that is analyzed (Eq. 1). These inputs contribute to the overall energy consumption of the agricultural system.

$$Ex_{in} = \sum_{i=1}^{n}(m_i.\psi_i) \quad \text{(Eq. 1)}$$

where, m_i :quantity of input $i, (kg, L, h, m^3, \text{kWh})$, ψ_i:specific exergy value of input$i, (J/unit)$, n: number of inputs. For the agricultural production process of maize the exergy contributions of the specific inputs can be expressed (Eq. 2).

Human labor: $Ex_{labor} = Hours \times \psi_{labor}$
Diesel fuel: $Ex_{fuel} = Liters \times \psi_{fuel}$
Machinery: $Ex_{machinery} = kg \times \psi_{machinery}$ (Eq. 2)
Nitrogen fertilizers: $Ex_N = kg \times \psi_N$
Phosphate fertilizers: $Ex_P = kg \times \psi_P$
Potassium fertilizers: $Ex_K = kg \times \psi_K$
Water: $Ex_{water} = m^3 \times \psi_{water}$
Electricity: $Ex_{electricity} = kWh \times \psi_{electricity}$ $(\psi_{electricity} = 3.6MJ/kWh)$

$$Ex_{in} = Ex_{labor} + Ex_{fuel} + Ex_{machinery} + Ex_N + Ex_P + Ex_K + Ex_{water} + Ex_{electricity}$$

Exergy output (Ex_{out}) refers to the useful exergy content of maize yield (Eq. 3):

$$Ex_{out} = M_{maize}.\psi_{maize}$$

(Eq. 3)

where M_{maize} =maize yield (kg), ψ_{maize} =specific exergy of maize (J/kg).

CExC represents the total exergy depletion in the production process (Eq. 4):

$$CE\ x\ C = Ex_{in} - Ex_{out} \quad \text{(Eq. 4)}$$

Exergy efficiency η_{ex} in maize production measures the effectiveness of converting input exergy into valuable output (Eq. 5). Higher values indicate better conversion efficiency and lower resource waste.

$$\eta_{ex} = \left(\frac{Ex_{out}}{Ex_{in}}\right) \times 100 \quad \text{(Eq. 5)}$$

where η_{ex} = exergy efficiency (%).

Exergy loss (Ex_{loss}) represents the wasted energy due to inefficiencies. It must be minimized for sustainability.

$$(Ex_{loss}) = Ex_{in} - Ex_{out} \quad \text{(Eq. 6)}$$

Neighbourhood component analysis, principal component analysis, extreme gradient boosting based prediction model

To enhance exergy efficiency, a machine learning optimization approach using NCA-PCA-XGBoost is introduced. The exergy efficiency optimization diagram for maize production is illustrated in Figure 1. This method improves predictive modeling and optimization processes, helping to reduce energy waste by identifying inefficiencies and dynamically adjusting system parameters. Through optimization, energy waste is minimized, which directly contributes to sustainability by promoting better resource management and reducing the environmental impact of maize production. This structured framework emphasizes the critical role of exergy analysis and machine learning in optimizing agricultural energy efficiency, ensuring sustainable resource utilization and minimal environmental impact.

NCA is a supervised machine learning method used for selecting features and reducing dimensionality. Unlike conventional techniques such as PCA, which aim to maximize data variance, NCA prioritizes feature selection based on model performance in classification or regression. Its goal is to uncover the most relevant features and enhance predictive accuracy. In this way, it improves the ability of models to classify or forecast outcomes effectively. NCA

functions by transforming the feature space to improve class separation or enhance the accuracy of regression tasks. It does this by minimizing the leave-one-out classification error, which helps ensure that instances with similar characteristics are positioned closer together. A key benefit of NCA is its capability to automatically evaluate the importance of features, allowing for the elimination of irrelevant or redundant ones. This process not only simplifies the model but also increases computational efficiency.

In the fields of agricultural production and energy modeling, NCA is key to optimizing feature selection for predicting exergy efficiency. By eliminating less important variables, it boosts the model's accuracy and robustness. This process is particularly effective when paired with advanced machine learning models such as XGBoost. Together, these techniques facilitate improved decision-making, encourage sustainable resource use, and enhance energy efficiency assessments in intricate agricultural systems. The pseudocode code of NCA is given below:

"Algorithm 1 — NCA-based feature selection

Input: X (m×n), labels y, target dim d ≤ n, lr η, temp τ, reg λ, max_iters, tol
Output: A (d×n)

```
Standardize X (fit on train only)
Initialize A ← I_n (first d rows)
For t = 1..max_iters:
    Z ← A X
    Compute pairwise distances on rows of Z; set self-distance to +∞
    Row-wise softmax P over −distance/τ; set p_ii = 0
    L ← sum_i sum_{j in same-class} P_ij − λ‖A‖²_F
    A ← A + η · ∂L/∂A
    If improvement < tol: break
Return A"
```

Feature ranking using principal component analysis

Feature selection algorithms are commonly used in machine learning to select a subset of relevant features from a larger set of available features. This helps to improve the performance of the machine learning model by reducing the number of features to be processed and eliminating irrelevant or redundant features that can affect the accuracy of the model. The algorithm selects the features that are most relevant to the target concept, based on their relationships with each other. By doing so, the model can achieve the highest possible accuracy, which is crucial in applications where precision is of utmost importance (Kohaviand John, 1997).

The main motivation behind using PCA in this study is to obtain a better representation of the input features in a space different from the original feature space. The aim is to identify features that are more informative, independent, and orthogonal. This approach leads to a more efficient and effective analysis, enabling valuable insights and better decision-making. PCA is a common technique used in contemporary data analysis and is employed by nearly all scientific disciplines. The primary objective of PCA is to identify the most significant basis to re-express a particular data set. The new basis is expected to reveal hidden structures in the data set while eliminating noise. It is a mathematical technique that is commonly used in data analysis and machine learning. It works by transforming data into a new space that is characterized by eigenvectors of X. By doing so, PCA identifies features that explain the most variance in the new space. If the first principal component (PC) covers a large percentage of the variance, the loads associated with that component can indicate the importance of features in the original X space. This technique is particularly useful in dimensionality reduction, where it can help to simplify complex datasets by identifying the most important features. It also has several applications, including data compression, feature extraction, and data visualization.

Let $X(X \in R^{m \times n})$ be the training data, the mean $(\bar{X})$ and the $n \times n$ dimensional covariance matrix $(\sum)$ can be found in Eq.7 and Eq.8 respectively. Eigenvalue decomposition is then applied to the covariance matrix as given in Eq.9 and eigenvalues and their corresponding eigenvectors are calculated. The first eigenvector corresponding to the maximum eigenvalue gives the first PC with the maximum variance.

$$\overline{\mathrm{X}} = E[X] \quad \text{(Eq. 7)}$$

$$Cov\sum = E[(X - \overline{\mathrm{X}})(X - \overline{\mathrm{X}})^T] \quad \text{(Eq. 8)}$$

$$[V, \Lambda]eig(\sum) \quad \text{(Eq. 9)}$$

Here V is the eigenvector matrix and Λ is the corresponding eigenvalues matrix.If an A matrix is constructed that columns are the eigenvectors, whose eigenvalues are ordered in the descending order of the data covariance matrix given in Eq. 8, then this matrix can be used to transform the original data into a new feature space as given in Eq. 10. This process is commonly known as the eigendecomposition or PCA of the data.

$$Z = A'X \quad \text{(Eq. 10)}$$

One of the key properties of the covariance matrix is that eigenvectors corresponding to different eigenvalues are orthogonal. If matrix$A$ is orthogonal, then the $k.th.$column corresponds to the $k.th$eigenvector of the covariance matrix. To transform the original features into a new orthogonal space Z, an orthonormal linear transformation is applied. This ensures that the new feature space Z remains orthogonal, which can be advantageous for certain analyses. The original input space (S) is transformed into a lower dimensional feature space (F) using PCA as indicated

in Figure 2. Then Eq. 10 is used to construct the transformation matrix *A* and the feature space *Z*. The first maximum of three PCs was selected to obtain the *A* matrix. This space was used for the regression analysis to estimate the output energy using the XGBoost method.

Extreme gradient boosting

Extreme gradient boosting (XGBoost), introduced by Chen and Guestrin (2016), is a highly efficient ensemble learning algorithm that enhances gradient boosting machines (GBM) by optimizing both computational speed and predictive accuracy. It follows a boosting framework, where multiple decision trees are trained sequentially, each refining the errors of its predecessors (Figure 3). XGBoost employs classification and regression trees (CART) as its base learner.

Similar to gradient boosting (GB), it iteratively improves weak learners, but unlike traditional boosting methods, XGBoost utilizes a second-order Taylor expansion on the loss function. This approach leverages both the first and second derivatives, enhancing prediction accuracy and model optimization. Additionally, regularization terms (L_1 and L_2) are included in the loss function to control model complexity and prevent overfitting effectively. A key advantage of XGBoost lies in its parallel processing capability, allowing for fast training on large-scale datasets while maintaining high efficiency. Additionally, it features automated missing value handling, eliminating the need for extensive preprocessing.

By assigning higher weights to misclassified instances, XGBoost captures complex patterns and thus proves highly effective for both classification and regression problems. These optimizations make XGBoost one of the most powerful and widely used machine learning algorithms today. The objective function departs from the plain squared error approach by incorporating both a training loss, which measures the fit of the model, and a regularization term, which discourages overly complex structures.The objective function for the XGBoost algorithm can be expressed as follows in Eq. 11 (Mitchell and Frank, 2017):

$$Obj = \sum_i L\left(y_i, \hat{y}_i\right) + \sum_k \Omega(f_k) \quad \text{(Eq. 11)}$$

In the XGBoost algorithm, L denotes a convex, differentiable loss function that evaluates the error between predictions and true labels for each training sample, while $\Omega(f_k)$captures the structural complexity of the decision tree f_k (Chen and Guestrin, 2016). $\Omega(f_k)$is obtained using Eq. 12.

$$\Omega(f_k) = \gamma.T + \frac{1}{2}\lambda\omega^2 \quad \text{(Eq. 12)}$$

Here Tdenotes the number of leaves in the tree f_k, ω isthe leaf weights, which correspond to the predicted values assigned to the leaf nodes. Incorporating $\Omega(f_k)$into the objective function enforces the optimization of simpler trees while ensuring minimization of $L\left(y_i, \hat{y}_i\right)$, thereby

mitigating overfitting. γ the regularization parameter penalizing the number of leaves used to avoid the overfitting, and λ is the L_2 regularization term to control leaf weights. Specifically, $\gamma.T$ imposes a constant penalty for each additional leaf, whereas $\lambda\omega^2$constrains extreme weight values.

Since boosting proceeds iteratively, the objective function at iteration m can be expressed in terms of the prediction from the previous iteration $\hat{y}_i^{(m-1)}$, adjusted by the contribution of the newly added tree f_k and is obtained using Eq. 13.

$$Obj^m = \sum_i L(y_{i,}\hat{y}_i^{(m-1)} + f_k(x_i)) + \sum_k \Omega(f_k) \quad \text{(Eq. 13)}$$

Expanding the function using a second-order Taylor series enables straightforward handling of different types of loss functions as shown in Eq. 14.

$$Obj^m \cong \sum_i \left[L\left(y_{i,}\hat{y}_i^{(m-1)}\right) + g_i f_k(x) + \frac{1}{2} h_i f_k(x)^2\right] + \sum_k \Omega(f_k) + constant \quad \text{(Eq. 14)}$$

Here, g_i and h_i denote the first and second derivatives of the loss function with respect to the prediction $\hat{y}_i^{(t-1)}$ from the previous iteration as shown in Eq. 15:

$$g_i = \frac{\partial L(y_i,\hat{y}_i^{(m-1)})}{\partial \hat{y}_i^{(m-1)}},\ h_i = \frac{\partial^2 L(y_i,\hat{y}_i^{(m-1)})}{\partial (\hat{y}_i^{(m-1)})^2} \quad \text{(Eq. 15)}$$

It should be noted that the previous prediction $\hat{y}_i^{(m-1)}$ remains fixed throughout this optimization process. The objective function can then be simplified by removing constant terms:

$$Obj^m = \sum_i \left[g_i f_k(x) + \frac{1}{2} h_i f_k(x)^2\right] + \sum_k \Omega(f_k) \quad \text{(Eq. 16)}$$

The objective function can subsequently be reformulated as a sum of over the tree's leaves, including the regularization term from Eq. (12).If the sums of the derivatives at each leaf are taken and Eq. 12 is rearranged:

$$Obj^m = \gamma.T + \sum_{i=1}^{T} \left[\left(\sum_{i\in I_j} g_i\right).w_j + \frac{1}{2}\left(\left(\sum_{i\in I_j} h_i\right) + \lambda\right).{w_j}^2\right] \quad \text{(Eq. 17)}$$

Consequently, XGBoost achieves improved optimization, prevents overfitting through regularization, and offers efficient scalability for large datasets.

Genetic algorithm

Genetic algorithm is a robust optimization method inspired by the principles of natural selection and genetics. It is widely applied in machine learning, research, and problem-solving, particularly for complex optimization tasks. GA enhances a system, method, or approach by searching for optimal solutions through key genetic operations such as selection, crossover, and mutation. These mechanisms enable it to iteratively refine a population of potential solutions, improving performance over multiple generations. Optimization typically involves maximizing or minimizing objective functions by adjusting input parameters within a defined search space.

GA effectively explores this search space, identifying efficient solutions to improve system performance and resource allocation.

The concept of GA originated from Holland (1973), drawing inspiration from Darwin's theory of natural selection. This approach relies on random variations and the principle of survival of the fittest to drive the optimization process. Similar to biological evolution, GA starts with the random generation of solutions, and the 'fitness' of each solution determines its likelihood of reproduction. Over generations, the fittest solutions evolve, leading to progressively optimized results until an optimal solution is discovered. This capability makes GA particularly effective for addressing computationally intensive problems, where traditional optimization methods may face challenges.

Performance evaluation metrics

Many indices are used in the literature to evaluate the performance of machine learning models. In this study, three indices were used: coefficient of determination (R^2), mean squared error (MSE), and mean absolute error (MAE). The MSE shows the difference between the actual and the predicted values. The MAE shows the mean error between the actual and the forecasted values. Larger R^2 value, lower MSE and MAE values indicate higher accuracy and higher performance of the ML model used. R^2, MSE, and MAE values are obtained using Eqs. 18 to Eq. 20.

$$R^2 = \left(\frac{\sum_{j=1}^{n}(y_{j,Exp} - \overline{y_{Exp}}) - (y_{j,Pre} - \overline{y_{Pre}})}{\sqrt{\sum_{j=1}^{n}(y_{j,Exp} - \overline{y_{Exp}})^2} \sqrt{\sum_{j=1}^{n}(y_{j,Pre} - \overline{y_{Pre}})^2}} \right)^2 \quad \text{(Eq. 18)}$$

$$MSE = \sum_{j=1}^{n} (y_{j,Exp} - y_{j,Pre})^2 \quad \text{(Eq. 19)}$$

$$MAE = \frac{1}{n} \sum_{j=1}^{n} \left| y_{j,Exp} - y_{j,Pre} \right| \quad \text{(Eq. 20)}$$

where $y_{j,Exp}$ and $\overline{y_{Exp}}$ are the experimental and mean experimental values, and $y_{j,Pre}$ and $\overline{y_{Pre}}$ are the predicted values and the mean predicted values.

Results and Discussion

Data preprocessing and analysis

The optimization process depends on the quality and comprehensiveness of the input data. The data collection process was structured, covering various data sources, key variables, and preprocessing techniques to ensure accuracy. The dataset includes both input and output variables that capture the energy flow in maize production. The input variables consist of human labor, machinery, fertilizers (such as nitrogen and phosphate), chemicals, seeds, diesel fuel, oil,

and water. Initially, the data were standardized to ensure accuracy and consistency. Z-score normalization was applied to variables with different scales. This prevents features with large values from overly affecting the model. This preprocessing step enhances the robustness of the dataset, making it more suitable for predictive modeling.

The distribution of exergy efficiency bar graph in Figure 2 highlights the variability and trends in exergy efficiency across different maize production scenarios.It provides insights into efficiency consistency and highlights potential areas for optimization. Meanwhile, the average exergy contribution of each input factor bar graph in Figure 3 complements this analysis by depicting the relative impact of various energy inputs on overall exergy efficiency, helping to identify key influencing factors in the production process.

The bar graph in Figure 2 illustrates the distribution of exergy efficiency (%) in maize production. It reveals that many cases have very low efficiency (0-10%). This indicates substantial energy losses. The distribution is right-skewed, with multiple peaks, suggesting distinct efficiency groups. While most cases cluster around 10-20% efficiency, there are also higher-efficiency outliers (40-70%), highlighting variability in energy utilization.

Concurrently, it seems that most farms have less than 1% exergy efficiency. This is primarily due to the way this metric is defined. It compares the useful exergy produced during a harvest to the total exergy consumed. When the Ex_in significantly exceeds the exergy present in the output, the amount consumed overshadows the useful portion, causing the efficiency value to diminish and approach zero.

This phenomenon, characterized by a left-tail concentration of efficiencies below 1%, can be attributed to scenarios involving substantial embodied exergy inputs -such as those from fertilizers, diesel, mechanization, and in some cases, the production of machinery- paired with relatively modest output yields. Factors like drought, pest pressures, or timing issues may contribute to these modest yields. Furthermore, the inclusion of protective inputs, such as various chemical treatments, increases the input exergy without a proportional increase in output exergy.

The second bar chart in Figure 3 presents the average exergy contribution of each input factor. It shows that Diesel Fuel (Ex_Dizel_Fuel) is the highest energy consumer, exceeding 6 million joules, followed by machinery (Ex_Machinery) and fertilizer (Ex_Fertilizer) as the next most significant contributors. Seed energy (Ex_Seed) also represents a notable share, whereas water (Ex_Water) and chemicals (Ex_Chemicals) have relatively lower exergy consumption. These findings indicate that fuel efficiency, optimized fertilizer application, and improved mechanization strategies could play a crucial role in enhancing overall energy efficiency in maize production. By addressing low-efficiency cases and targeting high-energy-consuming inputs, agricultural practices can be made more sustainable. This also improves cost-effectiveness.

The correlation matrix of exergy factors and exergy efficiency in Figure 4 quantifies the strength and direction of correlations, helping to identify key influencing factors. A positive correlation suggests that as one variable increases, exergy efficiency also improves, while a negative correlation indicates that an increase in one factor leads to a decline in efficiency. Variables with a correlation close to zero have little to no direct impact on exergy efficiency.

Figure 4 shows that CExC tracks Ex_in very closely. Across the studies we reviewed, Ex_out is typically much smaller and far less variable than Ex_in. Because CExC = Ex_in − Ex_out, when Ex_out is small and low-variance, CExC is almost the same as Ex_in; hence CExC and Ex_in are highly collinear and appear nearly identical in correlation matrices and scatter plots. This resemblance stems from the variables' definitions and relative scales, not from any analytic error.

It shows that Ex_out has the highest positive correlation (0.81) with exergy efficiency, indicating that improving energy recovery significantly enhances overall efficiency. Water consumption (Ex_Water) also has a moderate positive correlation (0.49), suggesting that effective irrigation management may contribute to better energy utilization. Conversely, Ex_HumanLabor (-0.09), Ex_Seed (-0.12), and Ex_Chemical (-0.25) exhibit weak negative correlations, implying that increasing these inputs does not necessarily improve efficiency and may even reduce it. Additionally, CExC shows a moderate negative correlation (-0.51), indicating that higher energy consumption does not always translate into better efficiency. These findings suggest that optimizing energy output while minimizing unnecessary inputs such as excessive labor, chemicals, and seed usage can significantly enhance energy efficiency.

Under the observed conditions, fertilizer and mechanization exhibit a strong negative relationship with exergy efficiency -consistent with diminishing returns when high-embodied-exergy inputs increase faster than useful output. However, when total input energy is kept constant, the fertilizer-efficiency link largely vanishes, suggesting that the original correlation is mainly due to scale effects rather than fertilizer-specific inefficiency. Conversely, the relationship between machinery and exergy efficiency remains minimal, exhibiting a weak overall association (with the full-sample partial correlation near zero). Notably, in the efficiency range of 0-30%, the partial Spearman correlation coefficient is approximately -0.09. This modest correlation reflects the potential benefits associated with timeliness or loss prevention in specific contexts, yet it remains insignificant when considered across the broader dataset.

Overall, these findings suggest that while both fertilizer and mechanization exhibit negative associations with exergy efficiency, the underlying dynamics are more complex and influenced by factors such as input scale and operational context.

To reduce scale bias and explore underlying mechanisms, we present partial correlations that account for external inputs such as Ex_in. This is accompanied by a series of robustness checks,

shown in Figure 5 for fertilizer and Figure 6 for machinery. The analyses include baseline partials, 5% winsorization, log-scale partials, and evaluations limited to the 0-30% efficiency range. Findings indicate that excessive application of fertilizer or mechanization concerning output is associated with a significant decline in efficiency. Conversely, appropriately implemented mechanization can yield modest efficiency improvements.

The relationship between chemical inputs and agricultural exergy can be characterized by a high correlation with the Ex_in and a low correlation with Ex_out. This phenomenon arises from the fact that chemical inputs possess a high exergy value per unit, which significantly contributes to an increase in the Ex_in. However, the primary agronomic function of these inputs tends to be protective, focusing on loss avoidance rather than enhancing yield directly. The effectiveness of chemical applications is often contingent on factors such as pest and disease pressure, as well as the timing of the applications.

Moreover, the variability observed across different studies tends to weaken the direct correlation between the Ex_in and the Ex_out. Additionally, chemical usage may serve as a proxy for unobserved stress conditions, which clarifies the strong association with the Ex_in while simultaneously explaining the weaker correlation with the Ex_out.

Optimized neighbourhood component analysis, principal component analysis, extreme gradient boosting model performance

Using NCA followed by PCA is deliberate and addresses two complementary needs. NCA is supervised: it uses the target to learn directions that best separate efficiency levels, up-weighting informative inputs and down-weighting noisy or irrelevant ones. This preserves interpretability at the raw-feature level (critical for agronomic decisions) by yielding clear variable weights that indicate what matters. On its own, however, NCA can leave residual multicollinearity and cross-study noise in the transformed space.

PCA then operates on this outcome-aligned subspace to orthogonalize and compress it, removing remaining correlation and measurement noise. This improves numerical conditioning and the bias–variance trade-off in our highly correlated inputs (fuel, machinery, labor, chemicals) and makes the downstream XGBoost stage more stable across seeds and folds. By contrast, PCA alone is blind to the target and may keep variance that does not help prediction; NCA alone can still leave correlated, high-variance directions.

The order also matters: doing PCA first could discard task-relevant structure before the supervised step sees it, while doing NCA after PCA would optimize on components that are harder to map back to actionable inputs. Our NCA-PCA sequence keeps actionable rankings at the input level (via NCA) and supplies a compact, low-variance basis for modeling (*via* PCA). In

practice, we tune the number of NCA and PCA components by cross-validation to control overfitting (Tuncer *et al*., 2020).

Recent exergy-centered studies largely determines loss locations or optimizes single processes -e.g., crop-plant exergy flow and regional indicators (Alzaben, 2025; Qi *et al*., 2025), process-level improvements such as rice drying (Wang *et al*., 2021), and case studies highlighting the thermodynamic role of fuel and fertilizer (Hesampour *et al*., 2022). In parallel, machine-learning studies have targeted adjacent sustainability goals- resource optimization in potato cultivation, tractor CO_2/energy modeling, and value-chain efficiency (Cheema *et al*., 2025; Balać *et al*., 2025; Assimakopoulos *et al*., 2024). Hybrid thermodynamic frameworks have meanwhile emphasized exergoeconomic / exergoenvironmental assessment rather than field-level control of exergy efficiency (Nabavi-Pelesaraei *et al*., 2023; Yang *et al*., 2024; Beni *et al*., 2023).

Against this backdrop, our study contributes by: i) explicitly optimizing exergy efficiency with respect to controllable inputs (labor, fertilizer, machinery/diesel, water, chemicals, seed) while predicting plot-scale CExC; ii) assembling a harmonized multi-study maize dataset, moving beyond single-site/process analyses; and iii) deploying a GA-tuned NCA–PCA–XGBoost pipeline that attains MAE = 1.50, MSE = 4.76, and R^2 = 0.99, with robust parity and residual diagnostics.

Importantly, the NCA stage yields interpretable input-importance rankings that surface high-leverage factors. Our partial-residual analysis indicates a near-zero marginal association for fertilizer once total input is controlled, whereas machinery/fuel shows a small positive association -consistent with reports of fuel's dominance and with diminishing fertilizer returns under specific baselines (Hesampour *et al*., 2022; Yang *et al*., 2024; Qi *et al*., 2025). Thus, beyond describing patterns, we quantify how input re-allocation can raise exergy efficiency, offering a data-driven basis for exergy-aware input planning and low-waste cultivation.

To mitigate leakage, all preprocessing (scaling) and modeling steps (NCA, PCA, GA hyperparameter search) were encapsulated in a scikit-learn Pipeline fitted only on training folds within a nested cross-validation scheme; an independent test set was held out for final evaluation. To manage complexity, XGBoost was regularized via max_depth, min_child_weight, subsample, colsample_bytree, and L1/L2 penalties; early stopping was triggered by validation loss. Diagnostics -including parity plots and residual checks (Figures 5 and 6)- showed homoscedastic, near-normal residuals, and test-set performance closely matched validation results, indicating no evidence of data leakage or overfitting (Figure 7).

The graph shown Figure 8 compares the predicted and actual exergy efficiency values. It demonstrates the accuracy of the NCA-PCA-XGBoost model. The NCA-PCA-XGBoost model demonstrated a strong correlation in estimating exergy efficiency in maize production. This is evident in the scatter plot aligned along the diagonal that compares predicted and actual values.

The model effectively captured the underlying patterns, resulting in predictions that closely align with actual exergy efficiency values. The linear trend observed in the plot indicates that the model performs well. However, minor deviations at higher efficiency levels (specifically above 40%) suggest some variance in predictions. This may warrant further refinement.This could be due to data imbalance, as the model was trained with fewer high-efficiency cases. Another possibility is the existence of nonlinear relationships that the model struggles to capture effectively. Further improvements in the model's ability to handle high-exergy scenarios could be achieved by incorporating additional predictive features or augmenting the dataset to ensure a balanced representation of all efficiency levels.

The graph in Figure 9 illustrates the residual distribution, highlighting the prediction errors and evaluating the model's performance regarding bias and variance. The histogram of residual distributions supports this assessment, showing that the errors are centered around zero and exhibit a nearly normal distribution. The normal-like shape of the residuals suggests that the model does not suffer from extreme bias. The symmetric shape of the residuals implies balanced predictions. However, a few outliers suggest occasional challenges in capturing more complex variations. Overall, the NCA-PCA-XGBoost framework serves as a reliable tool for optimizing exergy efficiency, reducing prediction errors, and contributing to improved sustainability and resource management in agricultural production.

The confidence interval plot given in Figure 10 further reinforces the findings from the prediction accuracy analysis. The narrower error bars for low-efficiency predictions indicate that the model is highly confident in its estimates in this range. However, as efficiency values increase, the confidence intervals widen, signifying greater uncertainty. This suggests that the model has learned to generalize well for lower exergy efficiencies but struggles with high-efficiency predictions, potentially due to insufficient training samples in this range. The wider intervals could also result from higher variance in exergy efficiency factors at larger efficiency values. This suggests that additional influencing variables might need to be considered. To address this, data augmentation, refined feature selection, and additional hyperparameter tuning could help improve confidence in the high-exergy range.

The performance of the proposed hybrid regression model for maize exergy analysis (optimized NCA-PCA-XGBoost) was compared with several commonly used regression methods, including linear regression (LR), ridge regression (RR), lasso regression (LaR), random forest (RF), ELM, and LightGBM. The proposed algorithm demonstrated superior performance compared to all other algorithms, as measured by metrics such as MSE, MAE, and R^2according to Table 2.

The performances of different methods and the presented method previously used in the literature for exergy efficiency estimation are given in Table 2. Accordingly; the optimized NCA-

PCA-XGBoost achieves the lowest MAE of 1.5027, demonstrating minimal prediction errors, and the lowest MSE of 4.7553, which signifies reduced large deviations. Furthermore, the highest R^2 score of 0.99 illustrates its strong ability to explain variance in the data, making it the most effective model. Among the other approaches, LightGBM performs competitively, with a MAE of 2.3801, an MSE of 6.3241, and an R^2 score of 0.97, showcasing its capacity to capture complex patterns. The regression models, including LR, RR, and LaR show similar performance, with R^2 values around 0.94. However, their higher MAE (approximately 3.0) and MSE (approximately 18.6-18.8) indicate comparatively lower accuracy.

RF and ELM demonstrate the weakest performance. RF has the highest MSE of 31.5415, while ELM exhibits the highest MAE of 3.5409 and the lowest R^2 of 0.89. These results suggest inconsistency in predictions.Overall, the findings validate the effectiveness of the proposed hybrid approach in optimizing exergy efficiency prediction, making it a valuable tool for improving sustainability and resource management in maize production.

Figure 11 shows a comparative evaluation of various machine learning algorithms used for estimation of exergy efficiency in corn production. Figure 11 consists of three subplots representing (a) MA, (b) MSE and (c) R^2. Lower MAE and MSE values indicate improved prediction accuracy, while higher R^2 values indicate stronger explanatory ability of the models. As can be seen from the results, the proposed method achieves superior accuracy with minimum prediction error, significantly outperforming the traditional models in all evaluation metrics. These results highlight the robustness and effectiveness of the proposed approach in terms of data-driven exergy efficiency optimization.

Conclusions

This study evaluated where exergy is used in maize production and how it can be reduced without affecting output. Diesel fuel and machinery are the main sources of total exergy use, with fertilizers also contributing. However, when total input stays the same, the impact of fertilizer on efficiency is small and depends on the context, while machinery can offer modest improvements by improving timing and reducing losses. Seed planting and fuel-heavy operations are consistently the most effective points for intervention across our analysis. At the same time, excessive human labor, chemical use, and high seeding rates tend to lower exergy efficiency, whereas water use has a positive effect (Juárez-Hernández *et al.*, 2019) - highlighting that improving energy recovery (getting more useful output from inputs) is key for overall efficiency improvements.

In optimization experiments, the proposed model consistently outperformed common regression methods achieving a MAE of 1.5027, a MSE of 4.7553, and an R^2 value of 0.99.

Additionally, residual diagnostics demonstrated well-behaved and approximately balanced errors (Figure 9 and Figure 1).

This study demonstrates that maize exergy efficiency can be predicted from controllable inputs -labor, machinery, diesel, seed, fertilizer, chemicals, and water- and, overall, this approach provides a reliable predictive tool for exergy-aware energy management in agriculture, supporting more sustainable and resource-efficient farming practices. This provides a proof-of-concept for data-driven optimization and highlights the factors most closely associated with efficiency. The findings are meant to support decision-making, not to prescribe actions. Correlation does not imply causation, and local (plot-level) trials remain crucial. Accordingly, this approach should be viewed as a screening and guidance tool that helps prioritize low-waste, high-impact adjustments tailored to local conditions.

The compiled dataset based on the literature has certain limitations that restrict its prescriptive use. These limitations include cross-study variability and missing covariates such as soil properties, weather conditions, water management practices, and pest pressure. Additionally, the dataset aggregates fertilizer reporting (combining nitrogen, phosphorus, and potassium) without distinguishing potassium oxide separately. It also does not cleanly differentiate electricity use from fuel consumption, lacks specific records for input dosage and timing at the plot level, and contains inconsistencies related to unit and moisture harmonization. These gaps hinder the ability to make robust causal claims and establish farm-specific thresholds.

Several practical signals emerge from the data. Establishing a good stand -through seed quality, planting depth, and planting rate- consistently proves to be a high-return strategy (Tian *et al.*,2022). Practicing fuel and machinery discipline by minimizing the number of passes, using the right equipment and settings, and ensuring timely operations can significantly reduce diesel consumption and losses. Additionally, the dataset indicates a neutral average marginal effect of fertilizer, highlighting the necessity to identify local thresholds for dosage and timing. This interpretation aligns with agronomic literature on diminishing marginal returns and on the interactions among rate, timing, and water management, which jointly shape the efficiency response to fertilization (Tian *et al.*, 2022). It is also important to keep fertilizer components separated by nitrogen, phosphorus, and potassium, and to consider split applications when appropriate for the context.

To turn these diagnostics into actionable recommendations, future efforts should focus on creating specialized datasets at the plot level, covering multiple regions. These datasets should include comprehensive variables such as soil characteristics, water regimes, pest and climate factors, and detailed nutrient information (including specific measurements of K_2O). Additionally,

it is important to differentiate between electric and fuel energy consumption and to incorporate precise data on timing, dosage, and mechanization (including passes, overlaps, and idling).

Furthermore, integrating economic and CO_2 metrics, along with testing real-time decision support in field trials, will facilitate farm-specific optimization that is causally grounded. This approach will establish clear thresholds for adjusting inputs effectively.

References

Alzaben, H., Fraser, R. 2025. Energy and Exergy analyses applied to a crop plant system. Thermo 5:3.

Amiri, Z., Asgharipour, M.R., Campbell, D.E., Armin, M. 2020. Extended exergy analysis (EAA) of two canola farming systems in Khorramabad, Iran. Agric. Syst. 180:102789.

Asl, J.H., Asakereh, A., Nejad, N.B. 2023. Evaluation of sugarcane production based on the analysis of cumulative exergy and environmental effects (case study of Agri-industry of Mirza Kochakh Khan). J. Agric. Sci. Sustain. Prod. 33:293-309.

Assimakopoulos, F., Vassilakis, C., Margaris, D., Kotis, K., Spiliotopoulos, D. 2024. Artificial intelligence tools for the agriculture value chain: Status and prospects. Electronics 13:4362.

Azizpanah, A., Taki, M. 2025. Evaluating the sustainability of sugar beet production using life cycle assessment approach. Sugar Tech. 27:78-93.

Balać, N., Mileusnić, Z., Dragičević, A., Milanović, M., Rajković, A., Miodragović, R., Ećim-Đurić, O. 2025. Implementation of XGBoost models for predicting CO2 emission and specific tractor fuel consumption. Agriculture 15:1209.

Beni, M.S., Parashkoohi, M.G., Beheshti, B., Ghahderijani, M., Bakhoda, H. 2023. Application of machine learning to predict of energy use efficiency and damage assessment of almond and walnut production. Environ. Sustain. Ind. 20:100298.

Bolandnazar, E., Rohani, A., Taki, M. 2020. Energy consumption forecasting in agriculture by artificial intelligence and mathematical models. Energ. Sources A 42:1618-1632.

Cheema, S.J., Karbasi, M., Randhawa, G. S., Liu, S., Esau, T. J., Grewa, K.S., et al. 2025. A state-of-the-art novel approach to predict potato crop coefficient (Kc) by integrating advanced machine learning tools. Smart Agr. Technol. 11:100896.

Chen, T., Guestrin, C. 2016. Xgboost: A scalable tree boosting system. Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, New York. pp. 785-794.

Chowdhury, T., Chowdhury, H., Ahmed, A., Park, Y.K., Chowdhury, P., Hossain, N., Sait, S.M. 2020. Energy, exergy, and sustainability analyses of the agricultural sector in Bangladesh. Sustainability 12:4447.

Dai, G., Fan, J., Dewi, C. 2023. ITF-WPI: Image and text based cross-modal feature fusion model for wolfberry pest recognition. Comput. Electron. Agric. 212:108129.

Dai, G., Tian, Z., Wang, C., Tang, Q., Chen, H., Zhang, Y. 2025. Lightweight vision transformer with lite-avpso hyperparameter optimization for agricultural disease recognition. IEEE Internet Things. Online Ahead of Print.

Dai, G., Tian, Z., Fan, J., Sunil, C.K., Dewi, C. 2024. DFN-PSAN: Multi-level deep information feature fusion extraction network for interpretable plant disease classification. Comput. Electron. Agric. 216:108481.

Dutta, A. 2021. Energy conservation and its impact on climate change. In: P.K. Sikdar, editors Environmental management: issues and concerns in developing countries. Springer. pp. 139-150.

Esmaeilpour-Troujeni, M., Rohani, A., Khojastehpour, M. 2021. Optimization of rapeseed production using exergy analysis methodology. Sustain. Energy Technol. 43:100959.

Hercher-Pasteur, J., Loiseau, E., Sinfort, C., Hélias, A. 2020. Energetic assessment of the agricultural production system: A review. Agron. Sustain. Dev. 40:1-23.

Hesampour, R., Hassani, M., Yildizhan, H., Failla, S., Gorjian, S. 2022. Exergoenvironmental damages assessment in a desert-based agricultural system: A case study of date production. Agron. J. 114:3155-3172.

Holland, J.H. 1973. Genetic algorithms and the optimal allocation of trials. SIAM J. Comput. 2:88-105.

Hulmani, S., Salakinkop, S.R., Somangouda, G. 2022. Productivity, nutrient use efficiency, energetic, and economics of winter maize in south India. PLoS One 17:e0266886.

Jamali, M., Soufizadeh, S., Yeganeh, B., Emam, Y. 2021. A comparative study of irrigation techniques for energy flow and greenhouse gas (GHG) emissions in wheat agroecosystems under contrasting environments in south of Iran. Renew. Sustain. Energ. Rev. 139:110704.

Juárez-Hernández, S., Usón, S., Pardo, C. S. 2019. Assessing maize production systems in Mexico from an energy, exergy, and greenhouse-gas emissions perspective. Energy 170:199-211.

Khanali, M., Akram, A., Behzadi, J., Mostashari-Rad, F., Saber, Z., Chau, K.W., Nabavi-Pelesaraei, A. 2021. Multi-objective optimization of energy use and environmental emissions for walnut production using imperialist competitive algorithm. Appl. Energ. 284:116342.

Kohavi, R., John, G.H. 1997. Wrappers for feature subset selection. Artif. Intell. 97:273–324.

Li, H., Chen, C., Umair, M. 2023. Green finance, enterprise energy efficiency and green total factor productivity: evidence from China. Sustainability 15:11065.

Liu, H., Liu, S. 2020. Exergy analysis in the assessment of hydrogen production from UCG. Int. J. Hydrogen Energy 45:26890-26904.

Marina, I., Grujić Vučkovski, B. 2022. Energetic efficiency of raspberry production in protected Arewa facility type tunnel. West. Balk. J. Agric. Econom. Rural Dev. 4:119-133.

Mitchell, R., Frank, E. 2017. Accelerating the XGBoost algorithm using GPU computing. PeerJ Comput. Sci. 3:e127.

Mostafaeipour, A., Fakhrzad, M.B., Gharaat, S., Jahangiri, M., Dhanraj, J.A., Band, S.S., Mosavi, A. 2020. Machine learning for prediction of energy in wheat production. Agriculture 10:517.

Nabavi-Pelesaraei, A., Ghasemi-Mobtaker, H., Salehi, M., Rafiee, S., Chau, K.W., Ebrahimi, R. 2023. Machine learning models of exergoenvironmental damages and emissions social cost for mushroom production. Agronomy 13:737.

Nadi, F., Górnicki, K. 2022. Evaluation of sustainability of wheat-bread chain based on the second law of thermodynamics: a case study. Sustainability 14:14229.

Nikkhah, A., Kosari-Moghaddam, A., Troujeni, M.E., Bacenetti, J., Van Haute, S. 2021. Exergy flow of rice production system in Italy: Comparison among nine different varieties. Sci. Total Environm. 781:146718.

Noorani, M.H., Asakereh, A., Siahpoosh, M.R. 2023. Investigating cumulative energy and exergy consumption and environmental impact of sesame production systems, a case study. Int. J. Exergy 42:96-114.

Qi, H., Dong, Z., You, X., Zhou, S., Zhu, Y. 2025. Extended exergy accounting of agricultural resources in China's four provinces of mountains and rivers. Sci. Rep. 15:22213.

Rashidi, K., Azizpanah, A., Fathi, R., Taki, M. 2024. Efficiency and sustainability: Evaluating and optimizing energy use and environmental impact in cucumber production. Environ. Sustain. Ind. 22:100407.

Rasoolizadeh, M., Salarpour, M., Borazjani, M.A., Nikkhah, A., Mohamadi, H., Sarani, V. 2022. Modeling and optimizing the exergy flow of tropical crop production in Iran. Sustain. Energ. Technol. Assess. 49:101683.

Sejkora, C., Kühberger, L., Radner, F., Trattner, A., Kienberger, T. 2020. Exergy as criteria for efficient energy systems - a spatially resolved comparison of the current exergy consumption, the current useful exergy demand and renewable exergy potential. Energies 13:843.

Soleymani, M., Asakereh, A., Safaieenejad, M. 2025. Optimization of cumulative energy, exergy consumption and environmental life cycle assessment modification of corn production in Lorestan Province Iran. Optimization 15:23-46.

Soltanali, H., Nikkhah, A., Rohani, A. 2017. Energy audit of Iranian kiwifruit production using intelligent systems. Energy 139:646-654.

Tian, P., Liu, J., Zhao, Y., Huang, Y., Lian, Y., Wang, Y., Ye, Y. 2022. Nitrogen rates and plant density interactions enhance radiation interception, yield, and nitrogen use efficiencies of maize. Front. Plant Sci. 13:974714.

Tuncer, T., Ertam, F., Dogan, S. 2020. Automated malware recognition method based on local neighborhood binary pattern. Multimed. Tools Appl. 79:27815-27832.

Tutar, H., Eren, O., Er, H., Gonulal, E., Gokdogan, O. 2025. Field-based experimental greenhouse gas emissions and energy use efficiency study of sorghum x sudan grass hybrid growth in a semi-arid region. Energy 315:134450.

Wang, G., Wu, W., Fu, D., Xu, W., Xu, Y., Zhang, Y. 2021. Energy and exergy analyses of rice drying in a novel electric stationary bed grain-drying system with internal circulation of the drying medium. Foods 11:101.

Wang, W., Deng, S., Zhao, D., Zhao, L., Lin, S., Chen, M. 2020. Application of machine learning into organic Rankine cycle for prediction and optimization of thermal and exergy efficiency. Energy Convers. Manag. 210:112700.

Wang, X., Zhang, W., Lakshmanan, P., Qian, C., Ge, X., Hao, Y., et al. 2021. Public–private partnership model for intensive maize production in China: A synergistic strategy for food security and ecosystem economic budget. Food Energy Secur. 10:e317.

Xu, X., Ma, F., Zhou, J., Du, C. 2022. Control-released urea improved agricultural production efficiency and reduced the ecological and environmental impact in rice-wheat rotation system: A life-cycle perspective. Field Crops Res. 278:108445.

Yang, C., Nutakki, T.U.K., Alghassab, M.A., Alkhalaf, S., Alturise, F., Alharbi, F.S., Abdullaev, S. 2024. Optimized integration of solar energy and liquefied natural gas regasification for sustainable urban development: Dynamic modeling, data-driven optimization, and case study. J. Clean. Prod. 447:141405.

Yildizhan, H. 2017. Thermodynamics analysis for a new approach to agricultural practices: case of potato production. J. Clean. Prod. 166:660-667.

Table 1. Energy coefficients of inputs-outputs in maize production.

Items	Unit	Energy coefficients (MJ/unit)	Reference
Inputs			
Human labour	H	1.96	Rashidi *et al*., 2024
Machinery	H	62.7	Jamali *et al*., 2021
Nitrogen	kg	66.14	Azizpanah and Taki, 2025
Phosphate	kg	12.44	Khanali *et al*., 2021
Chemicals	kg	238	Marina and Grujić Vučkovski, 2022
Seed	kg	14.7	Wang *et al*., 2021
Diesel fuel	L	56.31	Hulmani *et al*., 2022
Oil	L	38.41	Pourmehdi and Kheiralipour, 2024
Irrigation	m^3	0.63	Tutar *et al*., 2025
Output			
Maize	kg	14.7	Hulmani *et al*., 2022

Table 2. Comparison of exergy efficiency prediction performance in maize production.

Method	MAE	MSE	R^2
LR	2.9904	18.6880	0.94
RR	3.0473	18.8596	0.94
LaR	3.0129	18.6390	0.94
RF	3.1588	31.5415	0.90
ELM	3.5409	34.0257	0.89
LightGBM	2.3801	6.3241	0.97
NCA-PCA-XGBoost (proposed method)	1.5027	4.7553	0.99

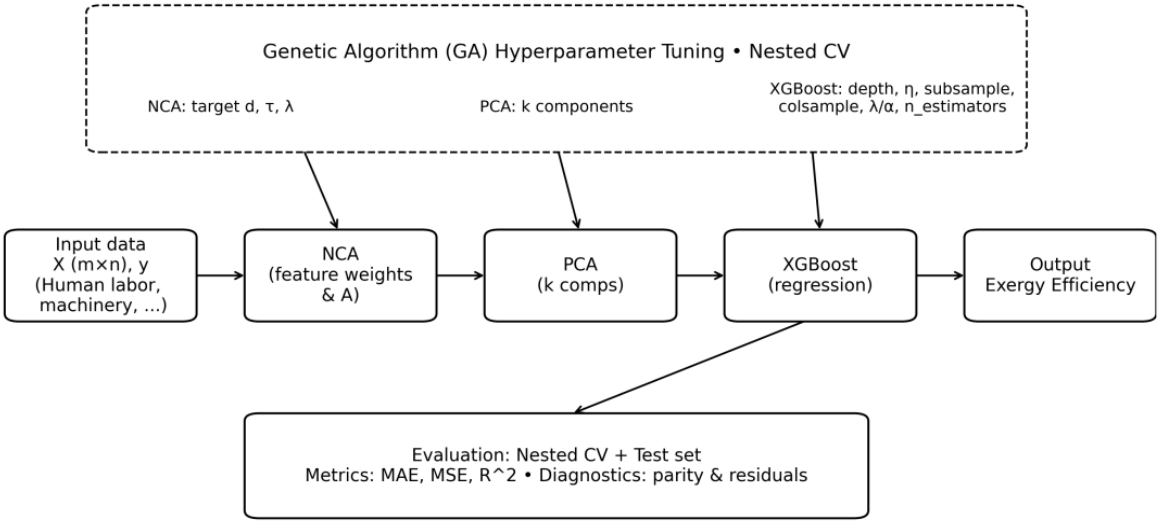

Figure 1. GA-optimized NCA-PCA-XGBoost based exergy efficiency optimization framework.

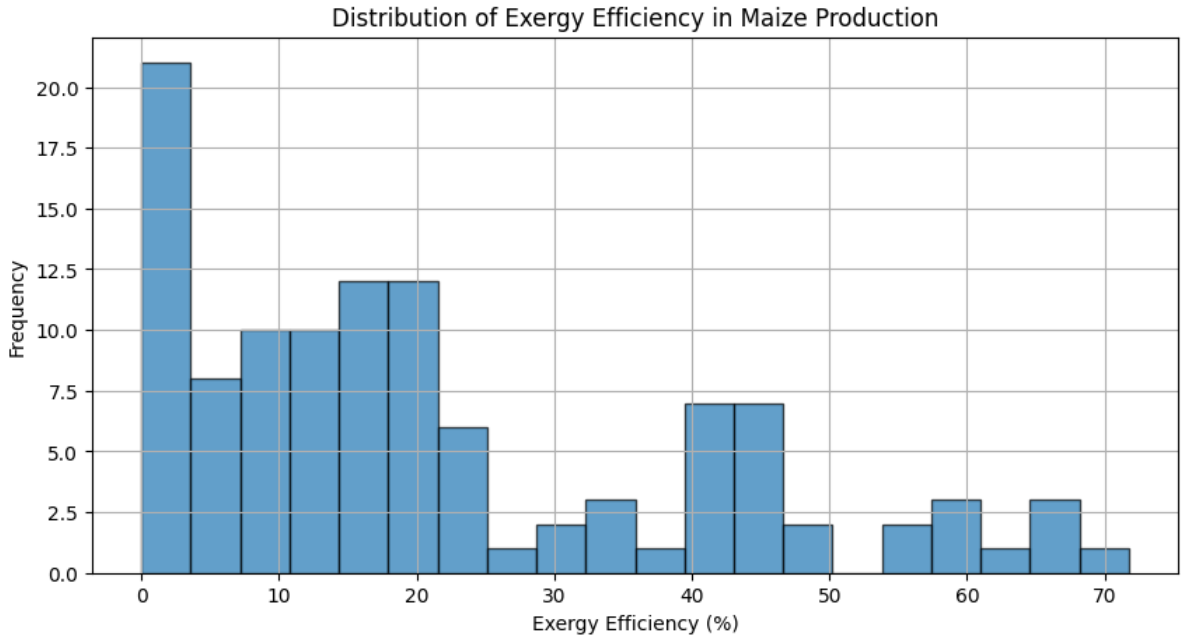

Figure 2. The exergy efficiency distribution in maize production.

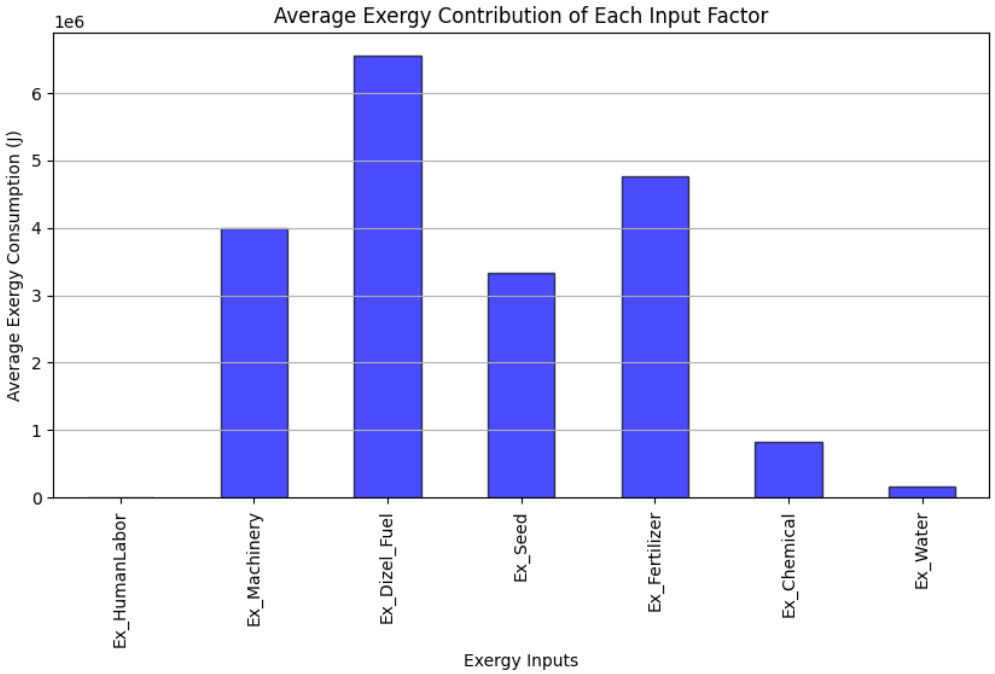

Figure 3. The average exergy contribution of each input factor in maize production.

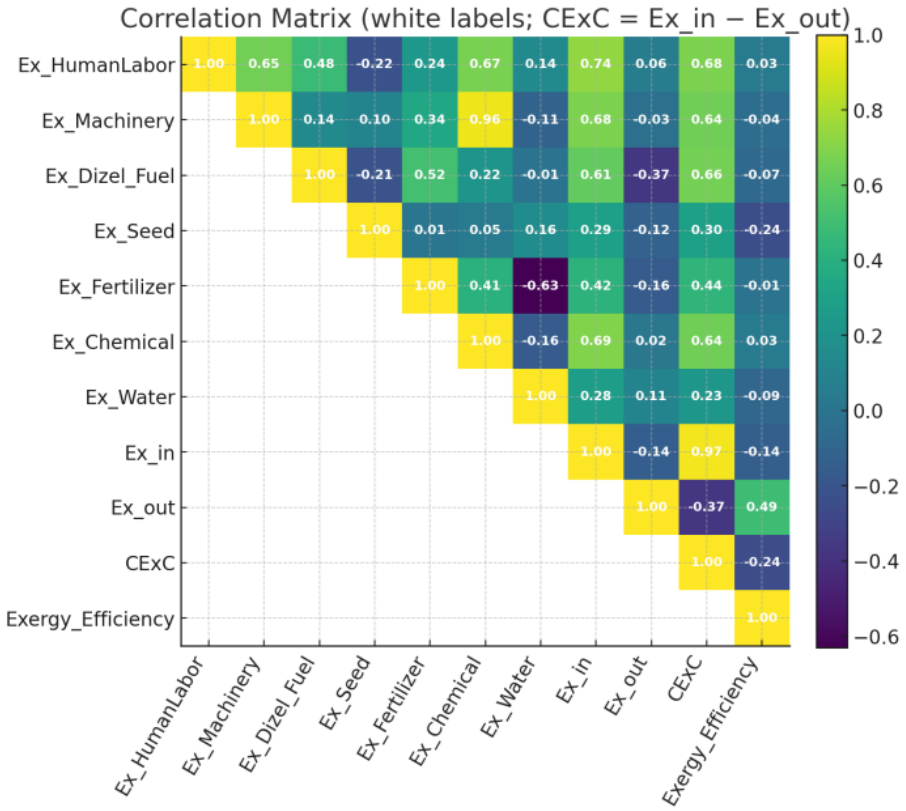

Figure 4. The relationship between exergy inputs and exergy efficiency in maize production process.

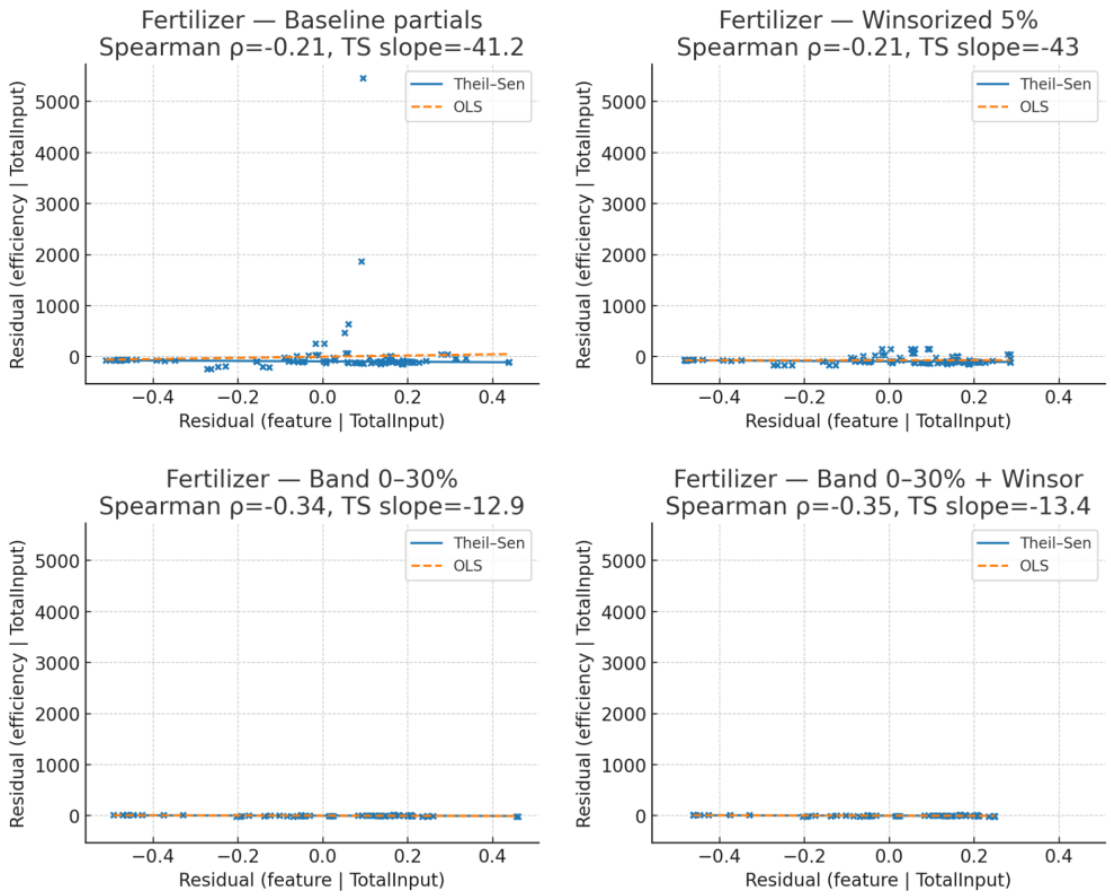

Figure 5. Fertilizer-partial residuals (control: total ınput).

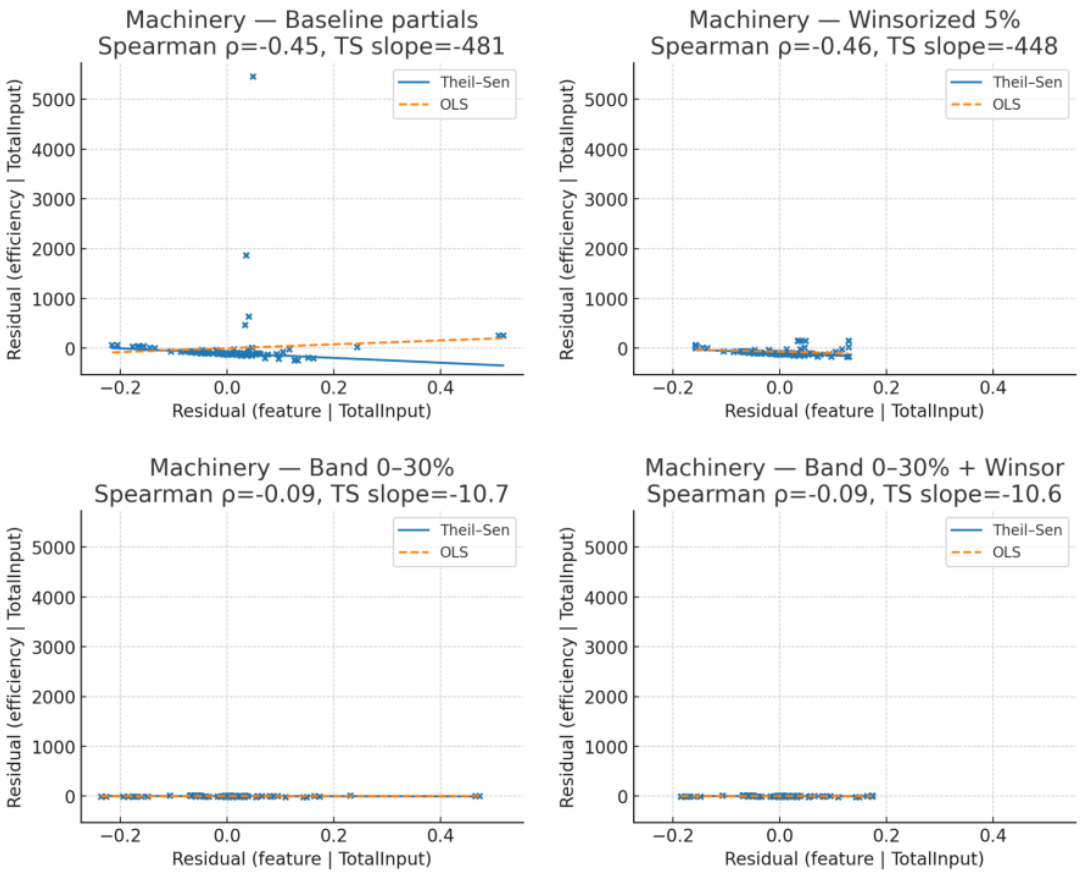

Figure 6. Machinery-partial residuals (control: total ınput).

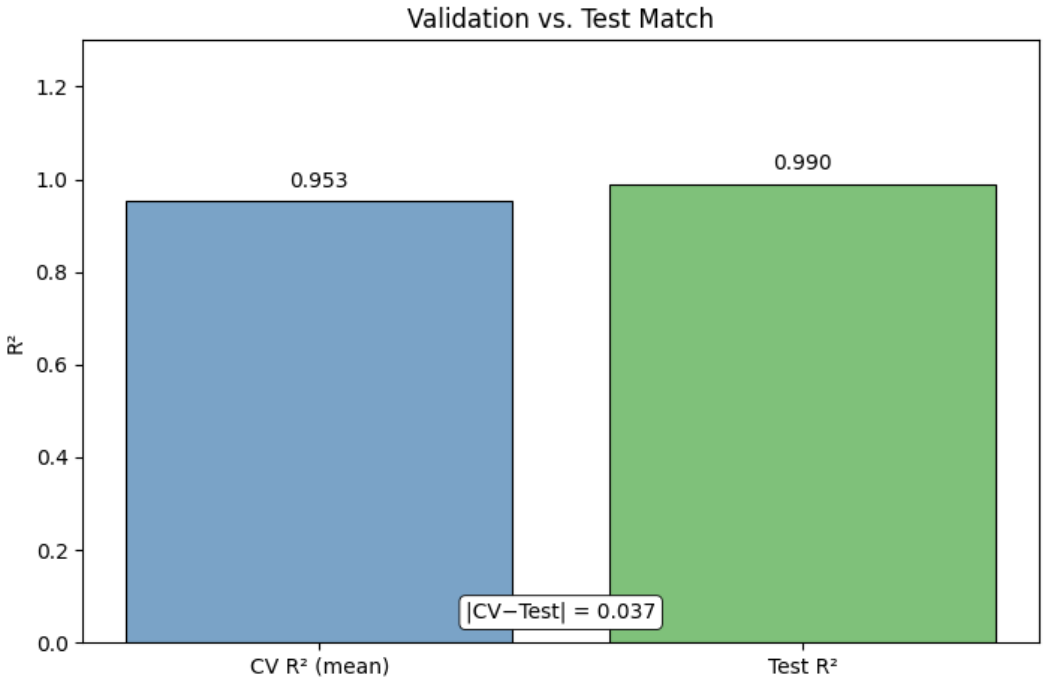

Figure 7. The bar graph comparing the mean R^2 of 5-fold cross-validation (CV) with the independent test R^2: demonstration of validation-test agreement.

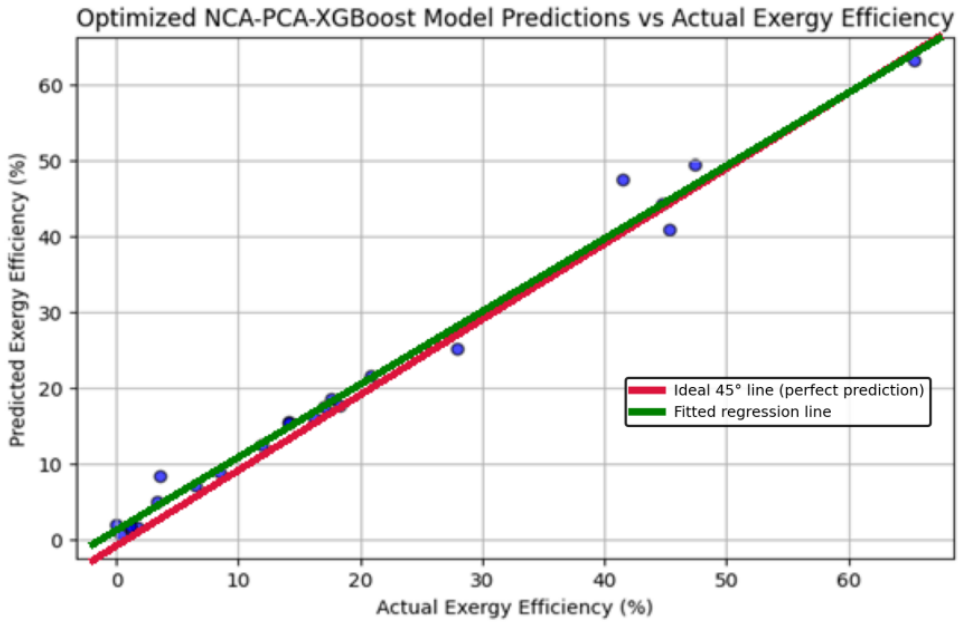

Figure 8. Predicted *vs* actual exergy efficiency using optimized NCA-PCA-XGBoost model.

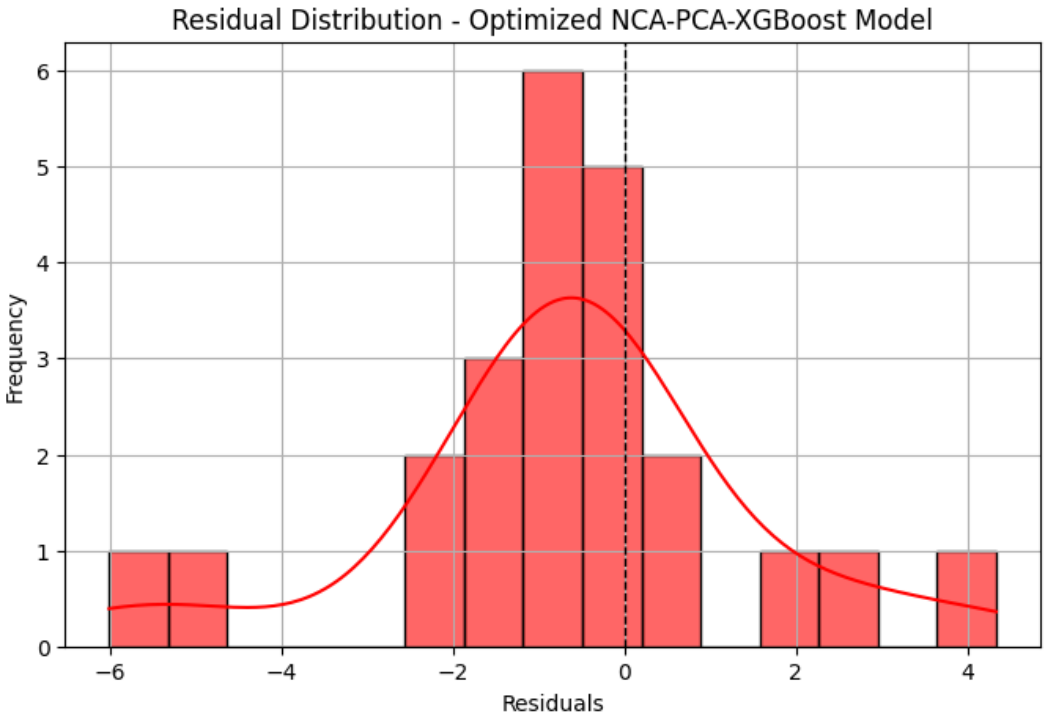

Figure 9. Residual analysis of the optimized NCA-PCA-XGBoost model.

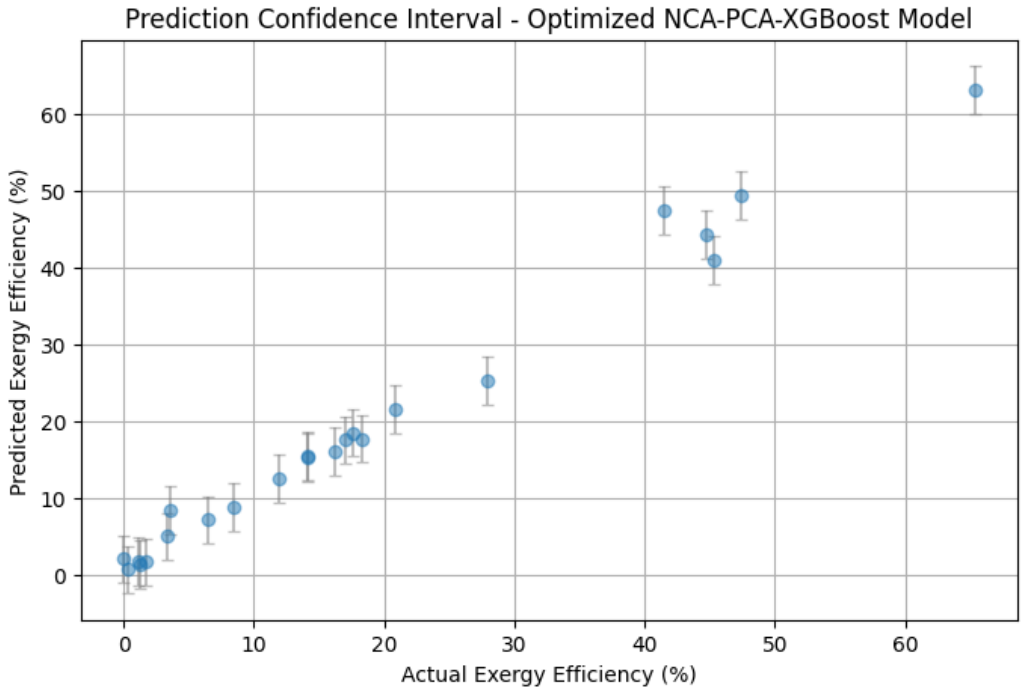

Figure 10. The confidence interval plot of the optimized NCA-PCA-XGBoost hybrid method.

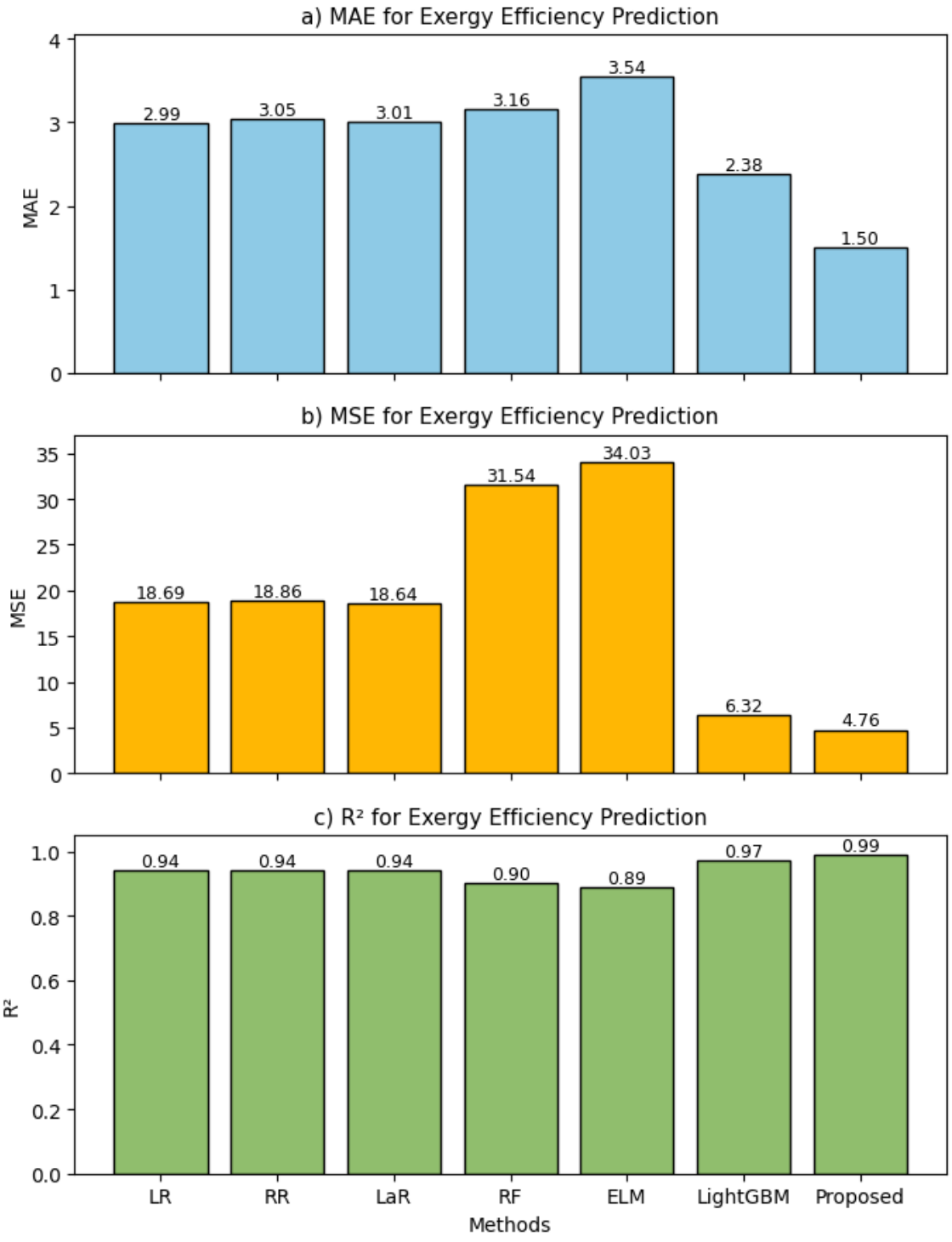

Figure 11. Graph of exergy efficiency estimation performances of different methods in terms of MAE, MSE and R^2 metrics.