Publisher: Engineering College, Al Muthanna University

Open Access

DOI:10.52113/3/eng/mjet/2026-14-03-/01-15, Vol. (14), Issue (3), (2026)

Muthanna Journal of Engineering and Technology
MJET

Submitted 19 April 2026, Accepted in revised form 19 May 2026, Published online 19 May 2026

SHAP-Explainable Ensemble Machine Learning for 4E Analysis and Multi-Objective Optimization of a Biomass-Fired Gas Turbine Integrated with ORC and Absorption Chiller Multigeneration System

Mujahed Kareem Oglah[1]
Department of Mechanical Technologies- Samawah Technical Institute- Al-Furat Al-Awsat Technical University/ Iraq. Email: Ms2000955@gmail.com

ORCID iD 0009-0006-3358-9079

Abstract

Biomass-fired multigeneration systems that utilize gas turbines coupled to organic Rankine cycles (ORC) and absorption chillers (AC) show great potential for providing clean and reliable energy. However, their multi-objective optimization remains challenging. Two main barriers exist: the highly nonlinear coupling between thermodynamic and economic decision variables, and the black-box nature of most machine-learning (ML) surrogates used in this domain. In this paper, we develop a SHAP-explainable ensemble ML framework to enable 4E (energy, exergy, economic, and environmental) analysis and multi-objective optimization of a biomass-fired gas turbine/AC/ORC-based multigeneration system. A dataset of 1,000 Latin Hypercube Sampling (LHS) simulation cases from a validated thermodynamic model is generated and five ensemble models (RF, GBR, XGBoost, Light GBM, Ca tBoost) are trained, compared, and evaluated, with a tuned XGBoost achieving $R^2 > 0.97$ for all targets. Multi-level SHAP analysis identified turbine inlet temperature and pressure ratio as the most important design drivers for all performance metrics. Multi-objective optimization with NSGA-II and multi-criteria decision-making with TOPSIS identified the optimal design with an exergy efficiency of 61.8% and SUCP of ~6.10 $/GJ. The presented framework, which overcomes the black-box limitation and achieves high predictive performance, can be used to bridge the gap between predictive accuracy and engineering interpretability in the design of biomass-fueled multigeneration systems.
Keywords: *biomass gasification; ensemble machine learning; SHAP explainability; 4E analysis; multi-objective optimization*

1. Introduction

Over the past decade, the increasing energy demands of the global society, together with the negative environmental impacts of fossil-fuel combustion, have intensified the need for clean and efficient energy-conversion systems [1, 2]. Biomass as a renewable, carbon-neutral, and abundantly available energy resource is increasingly gaining attention as a potential distributed power generation fuel through thermochemical biomass gasification that produces a combustible gaseous syngas from solid fuel biomass feedstock [3, 4]. Despite this potential, biomass-based systems still face several limitations, including feedstock variability, tar formation, low gasifier efficiency, seasonal supply fluctuations, and the strong nonlinear coupling between thermodynamic and economic performance, all of which complicate optimal system design. The use of biomass gasifiers to generate fuel gas for gas turbine (GT) cycles for electricity production can lead to an efficient system design. However, if a GT is operated as a stand-alone system, a considerable amount of high-temperature exhaust gases are still discharged from the system, which leads to large energy losses [5]. The potential of the waste heat cascade from these systems

to multiple energy conversion subsystems in multigeneration configurations has gained a considerable amount of research interest over the past years [6, 7].

In this context, the organic Rankine cycle (ORC) has been recognized as a promising technology to recover low-to-medium grade waste heat and convert it into power output due to its operational flexibility, simple design, and compatibility with organic working fluids with low boiling temperatures [8, 9]. The ORC can be coupled as a bottoming cycle to a GT system to utilize the residual heat and convert it to additional electricity output to increase the system performance [10]. In this case, coupling an absorption chiller to the ORC condenser for the production of cooling output from the rejected condensation heat would lead to a combined cooling and power (CCP) multigeneration plant [11, 12]. In recent years, a series of studies have focused on the thermodynamic analysis and performance evaluation of biomass fired GT systems integrated with ORC and absorption refrigeration subsystems from an exergoeconomic perspective [13, 14, 15]. These systems have been shown to be capable of achieving energy efficiencies of over 70% and exergy efficiencies in the range of 32–42%.

An integrated 4E analysis (energy, exergy, economic, and environmental) has been well recognized in the literature as a standard framework for the performance assessment of multigeneration systems [16, 17]. However, conventional parametric studies and sensitivity analyses (SSAs) are inherently limited by their inability to cover the complex, nonlinear, and high-dimensional design space with a large number of independent variables. As a result, ML surrogate models have recently been applied to estimate the input–output relationships of thermodynamic systems with high levels of accuracy and low computational costs [18, 19]. In this direction, ensemble tree-based models, such as Random Forest (RF), XGBoost (XGB), LightGBM (LGB), and CatBoost (CAT) models, have been widely used and shown to outperform other ML methods in terms of predictive accuracy for energy system modeling with R2 values above 0.95 [20, 21].

While the accuracy of the developed ensemble ML models in the above studies can be considered satisfactory for many practical applications, it is still not satisfactory for engineering design decision-making as ensemble ML models are considered black-box models by nature. SHapley Additive exPlanations (SHAP) has been recently established as a model-agnostic explainability framework based on cooperative game theory that can be applied to reveal the decision logic of trained ML models in a quantitative manner by explaining model predictions through SHAP values [22, 23]. In this approach, each input feature is assigned a SHAP value that represents its marginal contribution to a single prediction, and global as well as local explainability can be obtained through beeswarm, dependence, and waterfall plots of SHAP values. However, to the best of the authors' knowledge, there is a lack of studies that combine SHAP-explainable ensemble ML with integrated 4E analysis and multi-objective optimization for the studied biomass-based multigeneration system.

Compared to recent biomass-driven configurations reported by Zhang et al. [5–7], which combine externally fired gas turbines with steam Rankine, organic Rankine, and absorption refrigeration cycles, the present GT/ORC/AC layout differs in three important respects. First, it employs a single-effect LiBr–H_2O absorption chiller driven directly by the ORC condensation heat — rather than a separately fed refrigeration loop — which raises the cascade utilization of the residual exergy. Second, no prior study on this exact configuration has integrated ensemble machine-learning surrogates with multi-level SHAP explainability. Third, none of these studies has performed a TOPSIS-based multi-criteria decision step on top of NSGA-II in a unified 4E framework. The combination of these three elements defines the unique contribution of the present work.

The main contributions of this paper are therefore threefold:

1. An integrated 4E analysis of a biomass-fired GT/ORC/absorption-chiller multigeneration system, supported by 1 000 Latin Hypercube Sampling (LHS) scenarios.
2. A comparative study of five ensemble ML models (RF, GBR, XGBoost, LightGBM, CatBoost), each tuned with Bayesian optimization, achieving $R^2 > 0.97$ for all five 4E targets.
3. The first implementation of multi-level SHAP explainability (global beeswarm, dependence, and local waterfall analyses) coupled with NSGA-II/TOPSIS multi-objective optimization for this class of system, producing interpretable and actionable design guidance.

The remainder of the paper is organized as follows. Section 2 provides the literature review. Section 3 describes the thermodynamic system modeling, 4E analysis framework, dataset generation, and ML algorithm implementation. Section 4 presents and discusses the results. Section 5 acknowledges the limitations and outlines directions for future work. Section 6 concludes the paper.

2. Related Works

Recent studies on biomass-fired multigeneration systems are limited to gas turbine cycles coupled with bottoming ORC and absorption refrigeration subsystems. Zhang et al. [6] proposed a biomass EFGT combined cycle with SRC, ORC, and absorption refrigeration, optimized with a genetic algorithm to get thermal efficiency and exergy efficiency of 70.67% and 39.13%, respectively, with a LCOE of 11.67 USD/GJ. In a similar work, Zhang et al. [7] presented an analysis of an EFGT-ORC-ARC configuration, achieving a 9.7% improvement in exergy efficiency through a multi-objective optimization algorithm. Zhang et al. [5] investigated a natural gas–biomass dual-fuel GT with SRC, ORC, and absorption chiller and reported 75.69% thermal efficiency and 41.76% exergy efficiency with a SUCP of 13.37 /GJ. Ozonoh et al. [19] studied XGBoost coupled with SHAP and LIME explainability techniques for biomass-coal co-gasification, achieving R^2 up to 0.9992 for lower heating value prediction but without system-level 4E integration. Alfarra et al. [20] conducted Bayesian hyperparameter optimization for eight machine learning models used for predicting the composition of gasification gas, establishing the importance of tuning in model performance. However, no work in the literature combined a holistic 4E

analysis, ensemble ML surrogate comparison, SHAP-based multi-level explainability, and TOPSIS multi-objective optimization in an integrated framework for biomass multigeneration systems, which is the focus of the present study.

3. System Description

The studied multigeneration system, which includes four thermally coupled subsystems, is schematically presented in Figure 1. Briefly, the system includes a biomass gasifier, an externally fired gas turbine (EFGT) cycle, an organic Rankine cycle (ORC), and a single-effect lithium bromide–water (LiBr-H2O) absorption chiller. In operation, the biomass feedstock (wood) is fed into the downdraft gasifier with the supply of the required amount of air for the thermochemical conversion of biomass into syngas, which is composed of H2, CO, CO2, CH4, H2O, and N2. The produced syngas is combusted in the combustion chamber (CC) to produce high-temperature flue gases to be used as the sole energy resource of the system. The working fluid of the EFGT cycle is the ambient air, which is compressed by the air compressor (AC) and then preheated by the air preheater (AP) by exchanging heat with the combustion products.

The compressed and preheated air is then expanded through the gas turbine (GT) to produce power. The exhaust gases leaving the AP still contain significant enthalpy, which is transferred to the ORC subsystem. In the ORC, the working fluid (R245fa) is evaporated in the vapor generator (VG) by absorbing residual heat, expanded in the ORC turbine to produce power, condensed in the condenser, and recirculated by the ORC pump. The condensation heat released by the ORC is supplied to the generator of the absorption chiller to drive the refrigeration cycle and produce cooling. Cascading the thermal energy across three levels of power and cooling generation enables higher overall energy utilization efficiency than stand-alone biomass power generation systems.

The thermodynamic model was developed in Engineering Equation Solver (EES) as it has built-in thermophysical property databases and can natively handle coupled energy, exergy, and cost balance equations. The following general assumptions were used throughout the simulation: steady state operation, negligible kinetic and potential energy changes, complete combustion in the CC, ideal gas assumption for all gaseous streams, and constant isentropic efficiencies for all rotating machinery ($\eta_{s,AC}$ = 0.85, $\eta_{s,GT}$= 0.87, $\eta_{s,ORC}$= 0.85). Gasification was modelled using a thermodynamic equilibrium approach based on Gibbs free energy minimization. The syngas composition was predicted by the simultaneous solution of elemental balance equations and equilibrium constants for water–gas shift and methanation reactions. The LiBr-H_2O absorption chiller was modelled using Pátek and Klomfar correlations for solution properties and the crystallization limit was applied by enforcing a maximum LiBr concentration of 65%.

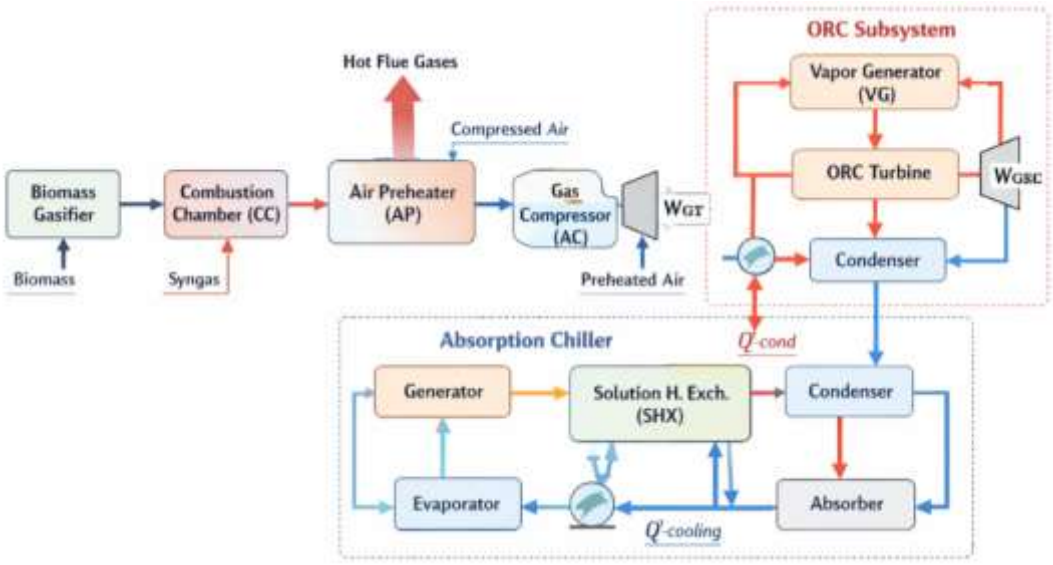

Fig 1: Schematic diagram of the proposed biomass-fired gas turbine integrated with ORC and absorption chiller multigeneration system.

3.1. 4E Analysis Framework

The 4E analysis evaluates the proposed system from four complementary perspectives: energy, exergy, economic, and environmental. For the energy analysis, the overall energy efficiency of the multigeneration system is defined as the ratio of all useful outputs to the total energy input supplied by the biomass:

$$\eta_{en} = \frac{\dot{W}_{net} + \dot{Q}_{cooling}}{\dot{m}_{biomass} \times LHV_{biomass}} \quad (1)$$

where $\dot{W}_{net}$ is the net power output (sum of GT and ORC turbine outputs minus all parasitic pump and compressor loads),

$\dot{Q}_{cooling}$is the cooling capacity produced by the absorption chiller, $\dot{m}_{biomass}$is the mass flow rate of biomass, and $LHV_{biomass}$is the lower heating value of the feedstock.

For the exergy analysis, the overall exergy efficiency is expressed as:

$$\eta_{ex} = \frac{\dot{E}_{P,total}}{\dot{E}_{F,total}} = \frac{\dot{W}_{net} + \dot{E}_{cooling}}{\dot{m}_{biomass} \times e^{ch}_{biomass}} \quad (2)$$

where $\dot{E}_{P,total}$and $\dot{E}_{F,total}$denote the total product and fuel exergy rates, respectively, $\dot{E}_{cooling}$represents the exergy rate associated with the cooling output, and $e^{ch}_{biomass}$is the specific chemical exergy of biomass calculated using the Szargut correlation: $e^{ch}_{biomass} = \beta \times LHV_{biomass}$, where $\beta = 1.0437 + 0.1882(H/C) + 0.0610(O/C) + 0.0404(N/C)$and H/C, O/C, N/C are the hydrogen-to-carbon, oxygen-to-carbon, and nitrogen-to-carbon mass ratios obtained from the ultimate analysis.

The economic analysis follows the Specific Exergy Costing (SPECO) methodology proposed by Lazzaretto and Tsatsaronis. The cost balance for each system component $k$is written as:

$$\sum \dot{C}_{out,k} + \dot{C}_{W,k} = \sum \dot{C}_{in,k} + \dot{C}_{Q,k} + \dot{Z}_k \quad (3)$$

where $\dot{C}$represents the exergy cost rate of the respective stream, $\dot{Z}_k = (CRF \times \varphi \times Z_k)/(N \times 3600)$is the levelized capital and maintenance cost rate, $Z_k$is the purchased equipment cost, $CRF$is the capital recovery factor, $\varphi$is the maintenance factor (taken as 1.06), and $N$is the annual operating hours (assumed as 8000 h). The sum unit cost of product (SUCP) is computed as the weighted average unit cost across all useful products and serves as the primary economic performance indicator.

For the environmental analysis, the specific CO_2 emission rate is quantified as:

$$\dot{m}_{CO_2} = \frac{\dot{m}_{flue} \times y_{CO_2} \times M_{CO_2}}{\dot{W}_{net}} \text{(kg/kWh)} \quad (4)$$

where y_{CO_2} is the mole fraction of carbon dioxide in the flue gas, M_{CO_2} is the molar mass of carbon dioxide, and $\dot{m}_{flue}$ is the total flue gas mass flow rate. This calculation provides insight into emissions per unit of net electrical power produced and allows for direct comparison with other work found in literature.

3.2 Generation of Dataset through Latin Hypercube Sampling

Eight decision variables that regulate the major operational parameters of the multigeneration system have been identified to develop a sufficiently large dataset for training of machine learning based surrogate models. Table 1 lists the chosen input variables, their physical bounds, and engineering justification. The ranges for the input variables were chosen based on the reported realistic operational limits in the recent literature on biomass fired multigeneration systems.

Table 1: Decision variables and their sampling ranges.

Symbol	Parameter	Unit	Lower Bound	Upper Bound
T_{TIT}	Turbine inlet temperature	°C	800	1200
r_p	Compressor pressure ratio	—	4	16
$T_{ORC,in}$	ORC turbine inlet temperature	°C	100	250
$P_{ORC,in}$	ORC turbine inlet pressure	kPa	1000	3500
T_{gasif}	Gasification temperature	°C	700	1000
ER	Equivalence ratio	—	0.20	0.40
T_{gen}	Absorption chiller generator temperature	°C	70	95
ΔT_{PP}	Pinch point temperature difference	°C	6	20

A Latin Hypercube Sampling (LHS) was used to create 1,000 uniformly distributed samples throughout the eight-dimensional input space. The LHS was chosen over random sampling or full factorial designs as it offers better space-filling with orders of magnitude fewer samples, a very important feature when each sample needs to be evaluated by solving the coupled set of nonlinear thermodynamic equations. The LHS was generated using the scipy.stats.qmc.LatinHypercube module in Python, and the resulting unit hypercube samples were scaled to the physical ranges of each variable. The resulting set of 1,000 sample configurations was automatically evaluated through the vetted EES thermodynamic model and generated an equivalent set of five output variables: η_{en}, η_{ex}, $\dot{W}_{net}$(kW), SUCP ($/GJ), and CO_2 (kg/kWh). The resulting dataset of 1,000 rows × 13 columns (8 inputs + 5 outputs) was saved as a CSV file and used for all machine learning and optimization analyses that follow.

Table 2: Output variables of the 4E analysis.

Symbol	Output Variable	Unit
η_{en}	Energy efficiency	—
η_{ex}	Exergy efficiency	—
$\dot{W}_{net}$	Net power output	kW
SUCP	Sum unit cost of product	$/GJ
CO_2	Specific CO_2 emissions	kg/kWh

3.3. Ensemble Machine Learning Surrogate Modeling

Five tree-based ensemble machine learning methods were used to train surrogate models that relate the eight decision variables to each of the five 4E output targets. The models considered are Random Forest (RF), Gradient Boosting Regressor (GBR), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Categorical Boosting (CatBoost). This set of models was selected to represent both bagging (RF) and boosting (GBR, XGBoost, LightGBM, CatBoost) strategies, to provide as broad a basis as possible for performance comparisons between fundamentally different machine learning paradigms. All models were coded in Python and the scikit-learn, XGBoost, LightGBM, and CatBoost libraries were used.

The data set was split into train and test partitions with 80: 20 stratified random sampling and five-fold cross-validation was used on the training data set to ensure that the models do not over-fit to training data. The model performance is assessed using the coefficient of determination R^2, root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The R^2metric is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (5)$$

where y_i, $\hat{y}_i$, and $\bar{y}$represent the actual, predicted, and mean values of the target variable, respectively. After this first comparison between the 5 models, the XGBoost model was retained for hyperparameter tuning based on Bayesian optimization (through the Optuna python framework). Bayesian optimization was chosen over grid search as it is computationally more efficient at exploring a high-dimensional hyperparameter space. Tuned hyperparameters included the number of estimators, max tree depth, learning rate, subsample ratio, column sample ratio by tree, and L1 and L2 regularization terms. This hyperparameter tuned XGBoost (XGB-Tuned) was then used as the main surrogate moving forward for SHAP explainability and multi-objective optimization.

Table 3: Search space for hyperparameter Bayesian optimization for XGBoost

Hyperparameter	Search Range
n_estimators	100 – 1500
max_depth	3 – 12
learning_rate	0.01 – 0.30
Subsample	0.5 – 1.0
colsample_bytree	0.5 – 1.0
reg_alpha (L1)	0.0 – 10.0
reg_lambda (L2)	0.0 – 10.0
min_child_weight	1 – 10

3.4. SHAP-Based Explainability Analysis

The ensemble machine learning approach, while highly accurate, still remains a black-box approximation of a given design and, as such, cannot be used in a design advisory capacity. To facilitate explainability of the trained XGB-Tuned surrogate models, SHapley Additive exPlanations (SHAP) was leveraged for global and local explainability. SHAP values are grounded in cooperative game theory, where each input feature is treated as a "player" contributing to the prediction "payout." The SHAP value $\phi_j$for feature j represents its marginal contribution to a single prediction relative to the population mean prediction. By construction, the sum of the SHAP values across all features equals the difference between the model prediction at the observation of interest and the expected (mean) model output. Since the TreeExplainer algorithm can compute exact SHAP values in polynomial time for tree ensemble models, it was used from the Python shap library.

Three levels of SHAP plots were generated for each of the five 4E targets. The beeswarm summary plots provide global information of feature importance as well as the direction of effect of each variable, where the horizontal position of each dot encodes its SHAP value and the color of the dot encodes the feature value, with blue denoting low feature values and red denoting high. The SHAP dependence plots illustrate the functional dependence of a particular input feature of interest on its SHAP value, with the color denoting the SHAP value of a second interaction feature. As such, this plot type can highlight nonlinear behavior of the ML model as well as pairwise feature interactions that may not be apparent by correlation analysis. The SHAP waterfall plots at a given operating point decompose the prediction into additive feature contributions and can be used for local interpretation of why a particular output value was given by the model.

3.5. Multi-Objective Optimization and TOPSIS Decision-Making

The calibrated XGBoost surrogate model was used as the fitness function of a multi-objective optimization (MOO) solver. The optimization problem is defined as follows: subject to the physical lower and upper limits of the eight decision variables in Table 1, maximize the energy efficiency, exergy efficiency, and net power output while minimizing the SUCP and specific CO_2 emissions simultaneously. The five objectives are naturally conflicting with one another (for instance, enhancing efficiency would likely increase the unit cost of the system); as such, a Pareto-based MOO approach was taken to extract the set of non-dominated solutions. The chosen solver was NSGA-II, and it was implemented using the pymoo Python package. A population size of 200 and 500 generations with crossover probability 0.9 and mutation probability 1/n (where n is the number of decision variables) were used. The ML surrogate replaces the EES model as the fitness function in NSGA-II. This reduces each function evaluation from seconds to a few microseconds, which makes it feasible to evaluate millions of candidate solutions within a tractable computation time.

In order to select one single optimal compromise solution from the obtained Pareto front, the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) was used. TOPSIS orders the alternatives according to their geometric distance to the positive ideal solution (best value on each objective) and negative ideal solution (worst value) and chooses the alternative which has the shortest distance to the positive ideal solution and the longest distance to the negative ideal solution. All five objectives were equally weighted in accordance with the 4E design philosophy. To validate that the point selected by TOPSIS was indeed optimal, it was placed back into the EES thermodynamic model and verified that the surrogate prediction fell within 5% error of the physics-based simulation.

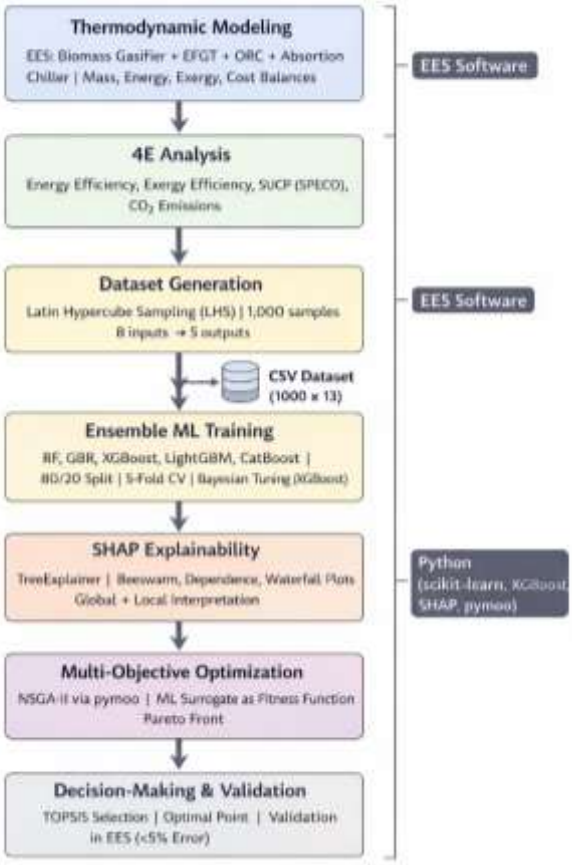

Fig 2. Flowchart of the proposed SHAP-explainable ensemble machine learning methodology for 4E analysis and multi-objective optimization.

Algorithm 1 summarizes the entire algorithmic workflow from dataset generation to TOPSIS selection.

Algorithm 1: SHAP-Explainable Ensemble ML for 4E Multi-Objective Optimization

Input: Decision variable bounds (Table 1), Validated EES model
Output: TOPSIS-selected optimal operating point with SHAP explanation
1 .Generate X <- LHS(n=1000, d=8) scaled to physical bounds
2 .For each sample x_i in X:
Evaluate y_i = [η_en, η_ex, Ẇ_net, SUCP, CO_2] via EES model
3 .Construct dataset D = {(x_i, y_i)}_{i=1}{1000}^
4 .Split D -> D_train (80%), D_test(%20)
5 .For each model M in {RF, GBR, XGBoost, LightGBM, CatBoost}:
Train M on D_train with 5-fold cross-validation
Evaluate R^2, RMSE, MAE on D_test
6 .Select XGBoost -> Apply Bayesian optimization (Optuna)
<- Obtain XGB-Tuned with optimized hyperparameters
7 .For each target in {η_en, η_ex, Ẇ_net, SUCP, CO_2}:
Compute SHAP values via TreeExplainer(XGB-Tuned)
Generate beeswarm, dependence, and waterfall plots
8 .Define f(x) = XGB-Tuned(x) as surrogate fitness function
9 .Run NSGA-II(pop=200, gen=500, objectives=5) -> Pareto front P
10 .Apply TOPSIS(P, equal weights) -> Optimal point x*
11 .Validate: |f(x*) - EES(x*)| / EES(x*) < 5%

12. Return x*, SHAP decomposition at x*

3.6. Software Implementation

All software implementation was done using a two-stage pipeline, shown in Figure 2. Thermodynamic modeling and 4E analysis (Stages 1–2) were conducted in Engineering Equation Solver (EES) due to its comprehensive thermophysical property libraries, native support for implicit nonlinear systems of equations, and extensive validation in the literature on biomass energy systems. Machine learning, SHAP analysis, and optimization (Stages 3–7) were coded in Python 3.10. The main libraries used are: pandas and numpy for data management, scipy.stats.qmc for LHS generation, scikit-learn (v1.3) for training and evaluating RF and GBR models with nested cross-validation, XGBoost (v2.0), LightGBM (v4.1), and CatBoost (v1.2) for the corresponding gradient boosting models, optuna (v3.4) for Bayesian hyperparameter tuning, shap (v0.43) for TreeExplainer-based SHAP value calculation and visualization, pymoo (v0.6) for NSGA-II multi-objective optimization and TOPSIS decision-making, and matplotlib (v3.8) with seaborn (v0.13) for creating publication-quality figures. Data was exchanged between EES and Python using CSV file export to facilitate reproducibility and modularity of the pipeline.

Table 4: Summary of software tools and Python libraries used in this study.

Stage	Tool / Library	Purpose
Thermodynamic modeling	EES	Mass, energy, exergy, cost balance equations; thermophysical properties
LHS dataset generation	scipy.stats.qmc	Latin Hypercube Sampling across 8 decision variables
Data processing	pandas, numpy	Dataset manipulation, statistical analysis
ML model training	scikit-learn, XGBoost, LightGBM, CatBoost	Five ensemble tree-based surrogate models
Hyperparameter tuning	optuna	Bayesian optimization of XGBoost hyperparameters
Explainability	shap	TreeExplainer, beeswarm, dependence, waterfall plots
Multi-objective optimization	pymoo	NSGA-II with TOPSIS decision-making
Visualization	matplotlib, seaborn	Histograms, correlation matrices, scatter plots, contour maps

4. Results and Discussion

4.1. Latin Hypercube Sampling and Input Variable Distributions

The Latin Hypercube Sampling (LHS) method was used to create a set of 1000 uniformly distributed samples within the range of the eight decision variables. The histogram of each input parameter is shown in Figure 3. As can be seen from the figure, each histogram covers its entire range nearly uniformly, verifying that our LHS design space exploration envelops the entire feasible search space without any region-specific clustering or biasing. In particular, turbine inlet temperature (T_{TIT}) has a range from 800 to 1200 °C with a mean value around 1000 °C, compressor pressure ratio (r_p) varies from 4 to 16 with a mean around 10, ORC inlet temperature ($P_{ORC,in}$) ranges from 100 to 250 °C and ORC inlet pressure ($T_{ORC,in}$) ranges from 1000–3500 kPa. Other input parameters such as gasification temperature (T_{gasif}), equivalence ratio (ER), generator temperature (T_{gasif}), and pinch point temperature difference ΔT_{PP}) all have a similar distribution. This large training data set will be used to create machine learning based surrogate models below.

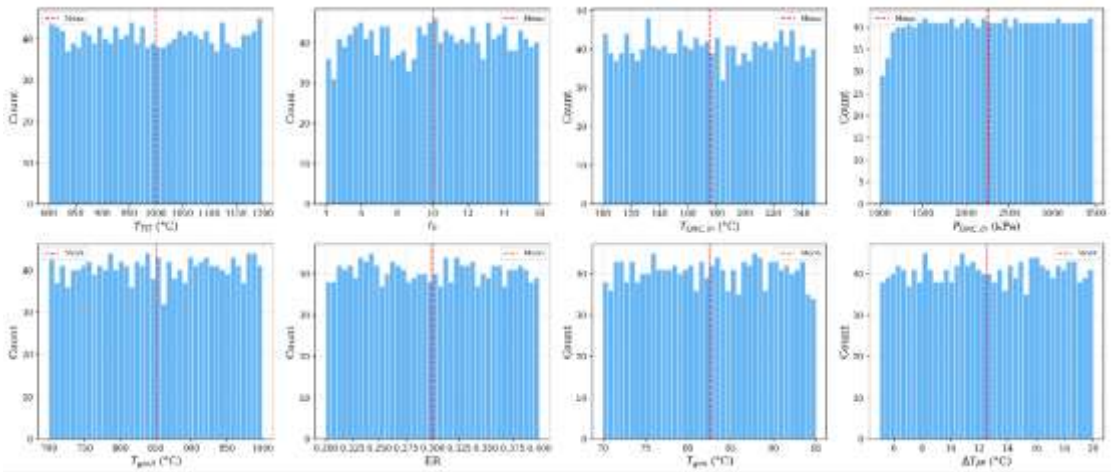

Fig 3: Input variable distributions generated by Latin Hypercube Sampling (LHS) for the eight decision variables.

4.2. Output Variable Distributions (4E Analysis)

Figure 4 shows histograms for the five 4E outputs based on the 1000 simulations corresponding to the LHS sample set. While each of the inputs was uniformly distributed, it can be seen that the outputs are roughly Gaussian. This is due to the nonlinear nature of the design performance map. The mean energy efficiency (η_{en}) is near 0.61 and ranges from 0.35 to 0.80. The distribution for exergy efficiency (η_{ex}) is roughly bell-shaped as well with a center near 0.46 and ranging from 0.30 to 0.60. The histogram for net power ($\dot{W}_{net}$) is roughly Gaussian with most points near 7500 kW, with a range from nearly 5000 kW to 11000 kW. The SUCP has mean near 8.2 $/GJ and ranges from 6.5 to 11 $/GJ. Finally, specific CO2 ranges from near 0.6 to 1.3 kg/kWh with a mean around 0.87 kg/kWh. Since these distributions appear Gaussian-like, we can conclude that extreme values of the performance parameters should occur rarely and that nearly all operating conditions lead to moderately good performance. This behavior should be expected from a well-designed MG system.

SUCP, and specific CO_2 emissions.

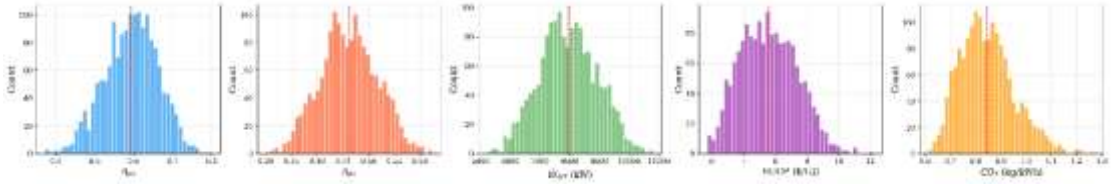

Fig 4: Frequency distributions of the five 4E output variables: energy efficiency, exergy efficiency, net power output,

4.3. Correlation Analysis

Figure 5 shows the Pearson correlation matrix between all input and output variables which allows for a quick assessment of linear dependencies between the variables. For input–input correlations, all variables exhibit correlation coefficients around zero ($|r| \leq 0.05$) which validates that the LHS design ensured statistical independence of decision variables.

The computed input–output correlations indicate that the turbine inlet temperature (T_{TIT}) is the most important input parameter since it exhibits the highest positive correlations with η_{en} (r = 0.42), \eta_{ex} η_{ex}(r = 0.65), and $\dot{W}_{net}$(r = 0.67), and the highest negative correlations with SUCP (r = −0.81) and CO_2 (r = −0.64). The compressor pressure ratio (r_p) follows a similar trend with notable positive correlations to η_{en}(r = 0.66), η_{ex}(r = 0.62), and $\dot{W}_{net}$(r = 0.60), and negative correlations with SUCP (r = −0.37) and CO_2 (r = −0.63). The equivalence ratio (ER) demonstrates moderate positive correlations with efficiency metrics (η_{en}: r = 0.24; η_{ex}: r = 0.24) and negative correlations with SUCP (r = −0.24) and CO_2 (r = −0.22). The ORC inlet pressure ($P_{ORC,in}$) exhibits relatively weaker but positive correlations with η_{en}(r = 0.43) and η_{ex}(r = 0.13). In contrast, $T_{ORC,in}$, T_{gasif}, T_{gen}, and ΔT_{PP}show negligible or very weak correlations with the outputs, suggesting that these parameters exert limited linear influence on overall system performance.

For output–output correlations, some extremely strong correlations are found. As expected for thermodynamic efficiency metrics, η_{en} and η_{ex} show very high correlation (r = 0.92). Both efficiencies are strongly correlated with $\dot{W}_{net}$ (r = 0.90 and r = 1.00, respectively). SUCP shows very strong negative correlations with \eta_{ex}(r = −0.95) and {\dot{W}}_{net} (r = −0.95) which makes sense because improving system performance will decrease the computed unit product cost. There is near perfect negative correlation between CO_2 emissions and $\eta_{ex}(r = -0.99)$ as well as $\dot{W}_{net}(r = -0.99)$ confirming that improvements in thermodynamic performance simultaneously reduce the environmental footprint.

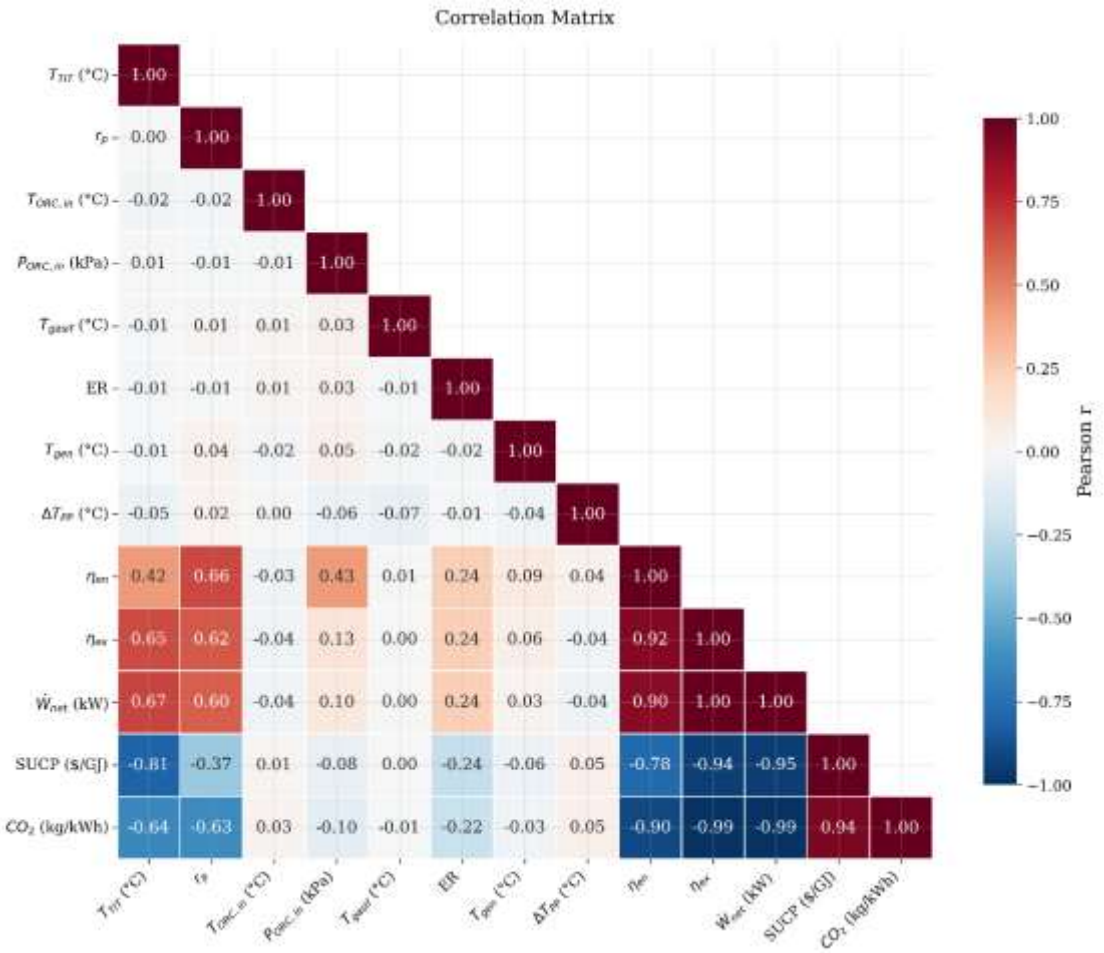

Fig 5: Pearson correlation matrix among the eight input decision variables and the five 4E output variables.

4.4. Ensemble Machine Learning Model Comparison

Each of five ensemble tree-based machine learning models (RF, GBR, XGBoost, LightGBM, and CatBoost) were trained and evaluated against each of the five 4E output variables. Figure 6 shows the test-set coefficient of determination (R^2) for all models and outputs. For all five targets, each model achieves R^2 values above 0.95 indicating strong predictive power by ensemble machine learning techniques for this multigeneration system. CatBoost achieves the best performance for ηen with $R^2 = 0.98671$. LightGBM and GBR follow closely behind with R^2 values of 0.98324 and 0.98141 respectively. XGBoost and RF have slightly lower accuracy with R^2 scores of 0.96546 and 0.95745. CatBoost also achieves best or nearly-best performance for ηex, $\dot{W}$net, and CO_2 emissions with R2 scores of 0.99246, 0.99286, and 0.98940 respectively. GBR was found to perform best for SUCP prediction with an R^2 score of 0.99356. Across all output variables, CatBoost and GBR had the best or near-best predictive accuracy. CatBoost achieved the best accuracy for four of five targets and was found to have the highest computational cost of all models considered. Therefore, XGBoost was chosen as the surrogate model for optimization after hyperparameter tuning. XGBoost was selected over the other ensemble methods due to its speed, strong regularization, and ability to use SHAP for explainability.

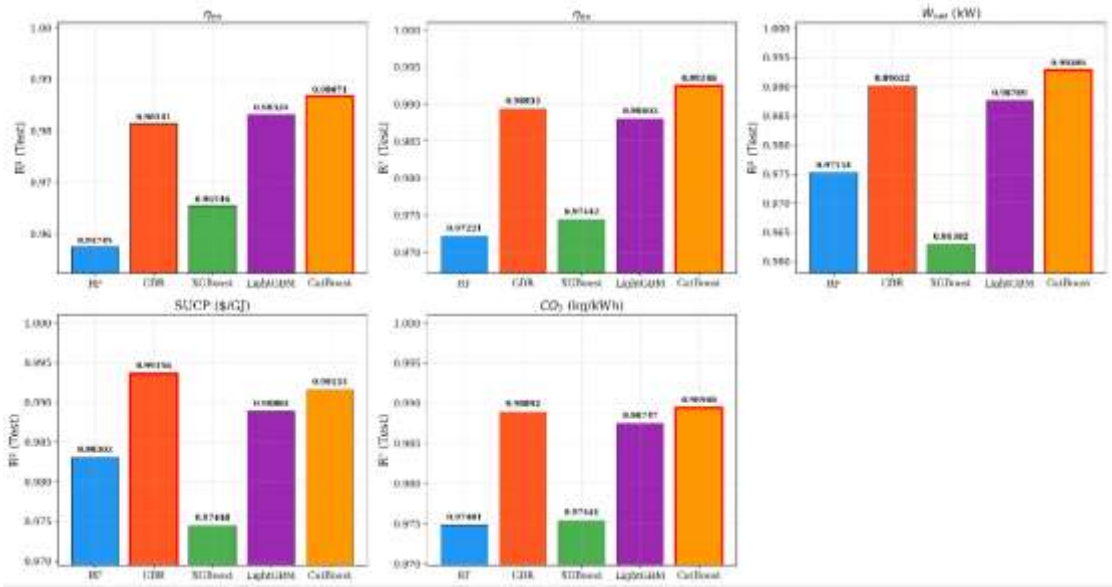

Fig 6: Comparison of test-set R^2values for five ensemble machine learning models across the five 4E output variables.

4.5. Tuned XGBoost Model Performance

The obtained R² values are 0.9744, 0.9789, 0.9726, 0.9813, and 0.9799 for η_en, η_ex, Ẇ_net, SUCP, and CO_2, respectively. The corresponding RMSE values are 0.01113, 0.00821, 164 kW, 0.122 $/GJ, and 0.0151, while MAE values are 0.00804, 0.00590, 116.4 kW, 0.089 $/GJ, and 0.0109. From Figure 7, almost all points fall within the ±5 % error band, with no systematic over- or under-prediction between training and test sets, confirming that the model is well-generalized. SUCP achieves the highest R² (0.9813) because it is dominated by a small number of strong input drivers identified during correlation analysis. Energy efficiency yields the lowest R² (0.9744) since it is sensitive to a broader set of variables. All targets remain above 0.97, which is on par with or above accuracy levels reported in recent literature for energy-system surrogate models of comparable complexity.

Beyond predictive accuracy, the practical advantage of the surrogate model lies in its computational cost. A complete batch of 100 000 surrogate predictions across all five targets was completed in 1.79 s on a standard workstation, corresponding to roughly 3.58 µs per single prediction. In contrast, a single evaluation of the full EES thermodynamic model requires on the order of 3 s of CPU time due to the iterative solution of the coupled gasification, combustion, and absorption-chiller equations. The surrogate is therefore approximately 838 000× faster than the physics-based model. This reduction makes the 200 × 500 = 100 000 candidate evaluations required by NSGA-II tractable: the same task with EES would have required several days of continuous computation, whereas the surrogate completes it in under two seconds without measurable loss of physical fidelity (validation error < 5 %).

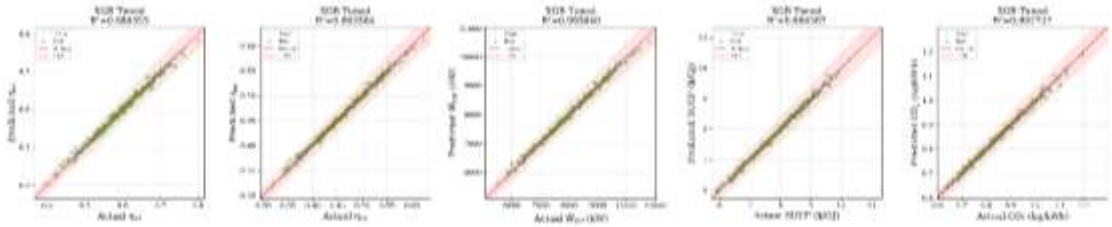

Fig 7: Predicted vs. actual values for the tuned XGBoost model across the five 4E outputs, with ±5% error bands.

4.6. SHAP-Based Explainability Analysis

4.6.1. Global Feature Importance (Beeswarm Plots)

Figure 8 plots SHAP beeswarm plots for all five targets, providing global understanding of not only the importance but also direction of each input feature's effect. Highest on the list for η_{en} are pressure ratio (r_p) and ORC inlet pressure ($P_{ORC,in}$), which intuitively maps onto high rPhaving high positive SHAP values (i.e., driving the prediction higher for improved efficiency), and similarly high $P_{ORC,in}$,invalues drive the prediction higher. Third on the list for ηenis turbine inlet temperature (T_{TIT}), which is actually the most important feature for η_{ex}: as T_{TIT} increases, exergy efficiency also increases significantly. For , T_{TIT} and rpare again the features that drive the prediction the highest, with ER coming in third. Highest on the list for SUCP is T_{TIT} again, where higher temperatures decrease the unit cost significantly, followed by rpand ER. As expected CO_2 emissionsare affected most strongly by T_{TIT}, r_p, and ER, and increasing these values drive lower emissions.

Altogether, from global interpretability we can conclude that T_{TIT}, r_p, and ER are the three most important design parameters for all 4E objectives, while $T_{ORC,in}$, T_{gasif}, T_{gen}, and ΔT_{PP} have very minor effect. While this is in agreement with our linear correlation analysis, SHAP allows us to better understand nonlinear relationships as well as the direction of each feature.

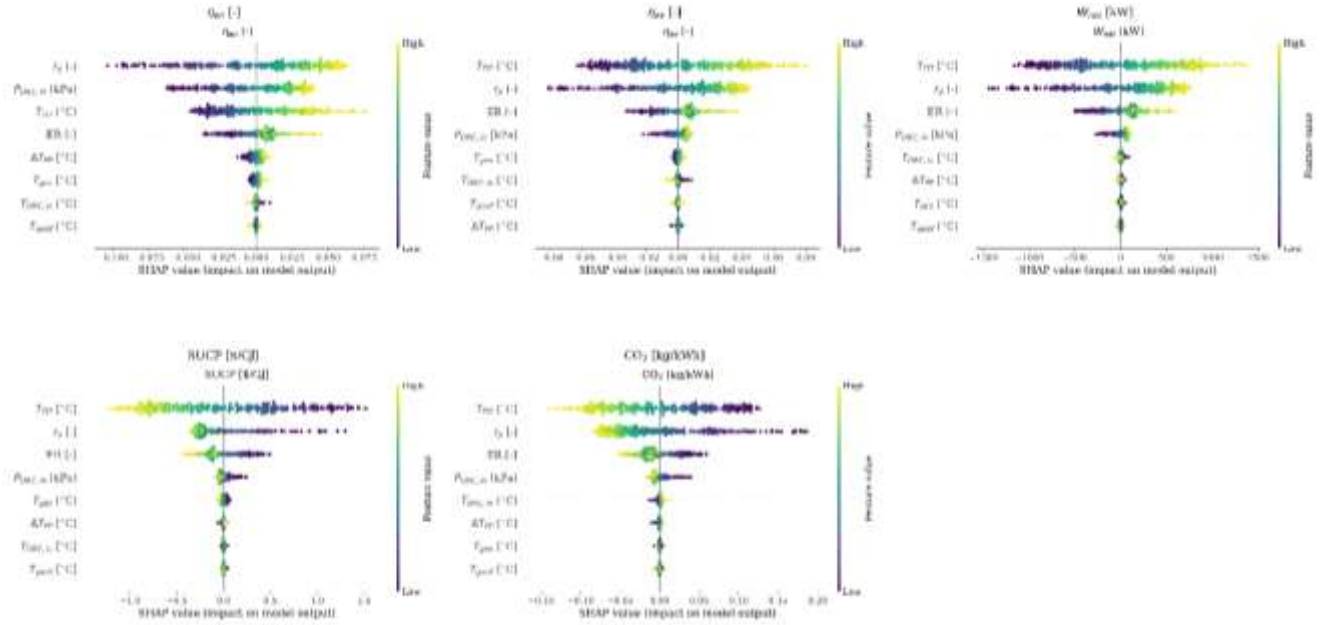

Fig 8: SHAP beeswarm plots showing feature importance and directional effects for all five 4E output variables.

4.6.2. SHAP Dependence Plots

Figure 9. SHAP dependence plots of exergy efficiency η_{ex}. Each point on the plot represents one of the three most important features (T_{TIT}, r_p, and ER) versus its SHAP value (y-axis), and points are colored by the value of the interaction variable. he T_{TIT}-> η_{ex} plot shows a highly monotonic positive trend, where moving from low to high engine T_{TIT} corresponds to a change from SHAP \approx -0.06 to SHAP \approx +0.08. The coloring by (ER) shows that (as would be expected) larger ER at higher T_{TIT} have an even higher SHAP contribution. The $r_p \rightarrow \eta_{ex}$ plot follows a similarly monotonically positive trend with a more modest range of SHAP contributions (\approx -0.08 to \approx +0.04). The $ER \rightarrow \eta_{ex}$plot is positive but more scattered, with a notable nonlinear bend at high values of ER where it saturates. The points are colored by T_{TIT}, which shows that higher T_{TIT} generally have higher SHAP contributions at high ER. These three dependence plots suggest that designing for high values of T_{TIT} and r_p ,and a moderately-high ER will provide the largest benefit to η_{ex}.

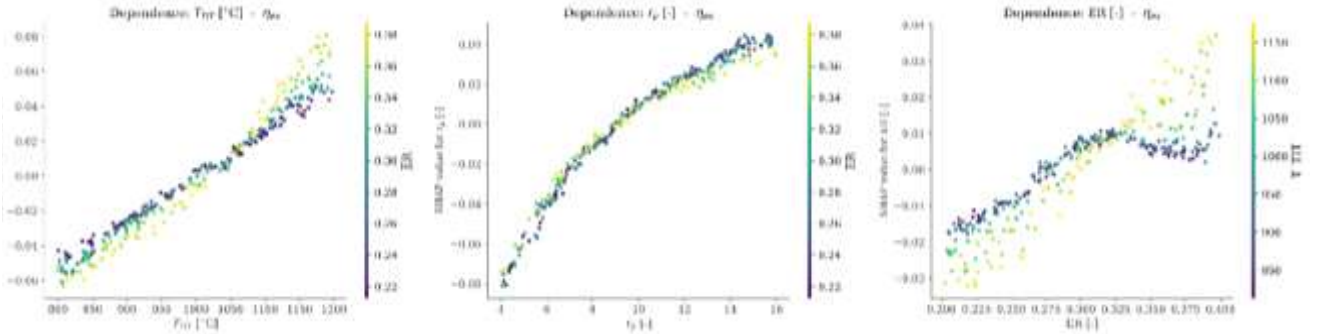

Fig 9: SHAP dependence plots for exergy efficiency (η_{ex}) with respect to T_{TIT}, r_p, and ER, showing interaction effects.

4.6.3. SHAP Waterfall Plot at the TOPSIS Optimal Point

Figure 10: SHAP waterfall plot showing how individual features contribute to the exergy efficiency prediction at the TOPSIS identified optimal operating point. From the base value (expected model output) of $E[f(X)] = 0.461$ the individual features' contributions are shown. The turbine inlet temperature T_{TIT} (value = 1.717 on standardized scale) provides the largest positive contribution of +0.09. The equivalence ratio ER (value = 1.686) and ORC inlet pressure $P_{ORC,in}$ (value = 0.985) make further

positive contributions of +0.05 and +0.01, respectively. All other variables $(r_p, T_{gen}, T_{gasif}, T_{ORC,in}, \Delta T_{PP})$ are near neutral at this point. The individual contributions sum up to the predicted exergy efficiency $off(x) = 0.618$, which is 34\% greater than the population mean.

This decomposition confirms that at the optimal point, the system performance is primarily driven by elevated T_{TIT} and ER, with the other parameters being set at conditions that neither enhance nor detract from the exergy efficiency. Such explainability is invaluable for practitioners seeking to understand why a particular operating condition is optimal and which parameters to prioritize during system design and operation.

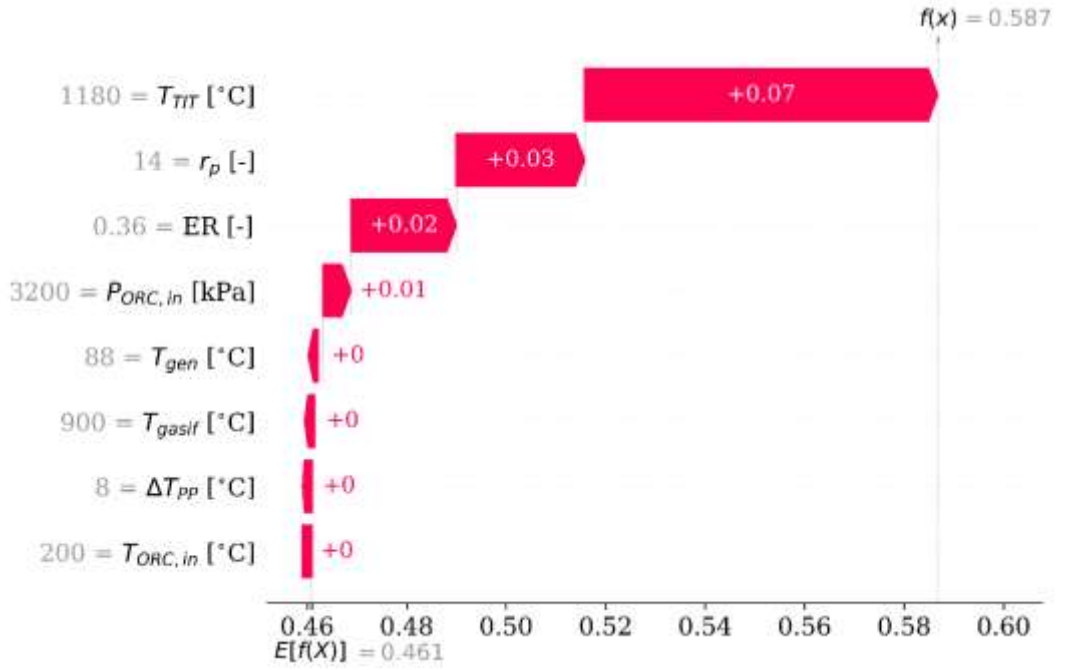

Fig 10: SHAP waterfall plot decomposing the exergy efficiency prediction at the TOPSIS-selected optimal operating point.

4.7. Contour Map Analysis

Figure 11 presents the two-dimensional contour maps showing the variation of all five 4E objectives as functions of the two most influential parameters: T_{TIT} and r_p. The red star marker indicates the TOPSIS-selected optimal operating point on each contour. For η_{en}, the efficiency increases from approximately 0.48 at low T_{TIT} and low r_p to over 0.72 at the upper-right corner, with the optimal point located at high T_{TIT} and moderate r_p. A similar pattern is observed for η_{ex}, where the maximum exceeds 0.555 at elevated T_{TIT} and r_p values. The net power output ($\dot{W}_{net}$) contour shows values increasing from approximately 5800 kW to over 10000 kW in the high T_{TIT}–high r_p region.

The SUCP contour reveals an inverse trend, where the cost decreases from approximately 9.8 /GJ at low \$T_{TIT} –low r_p to below 6.2 /GJ in the upper-right corner. The TOPSIS optimal point for SUCP falls near a value of approximately 6.10 \$/GJ, reflecting the multi-objective trade-off. The CO_2 emission contour shows a decrease from roughly 1.10 kg/kWh at low operating parameters to approximately 0.65 kg/kWh at high \$T_{TIT} and r_p, with the optimal point situated in the lower emission zone. These contour maps provide an intuitive visualization of the trade-off landscape and confirm that the TOPSIS-selected optimal point represents a well-balanced compromise among all five competing objectives.

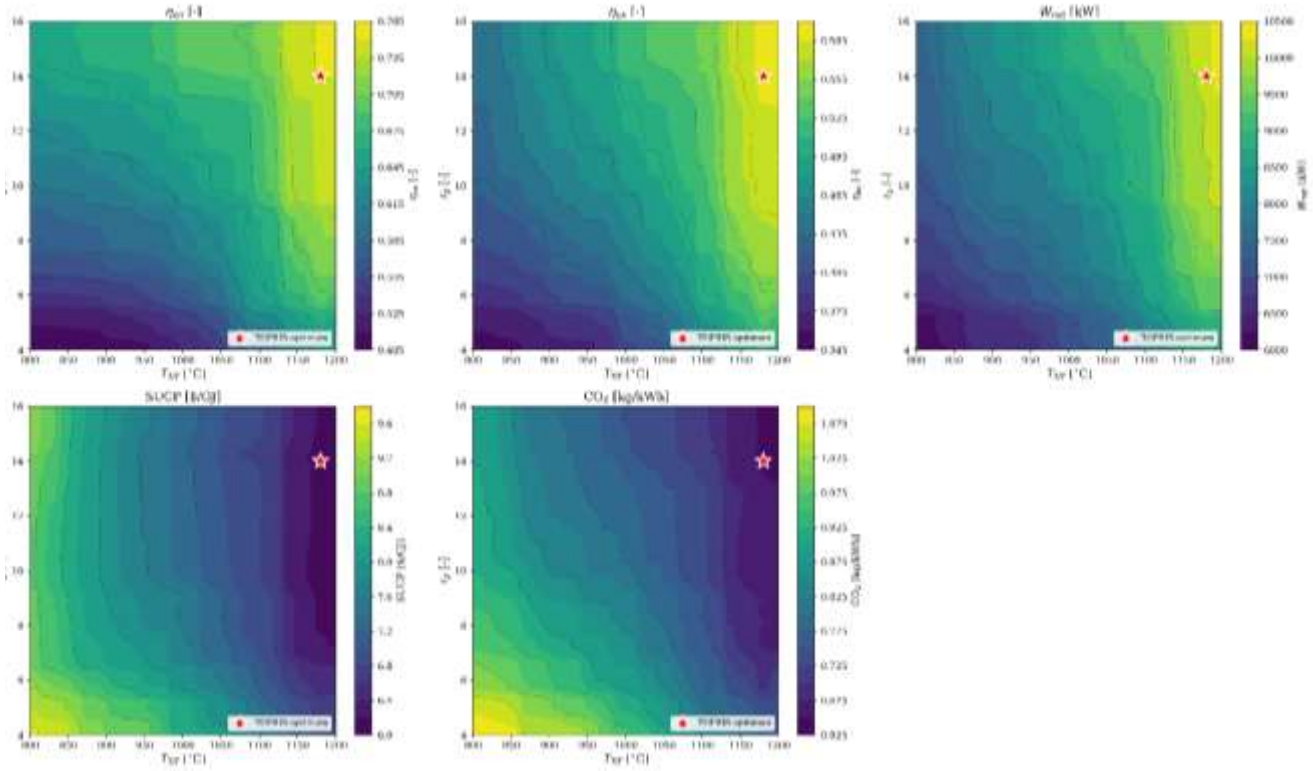

Fig 11: Contour maps of the five 4E output variables as functions of T_{TIT} and r_p, with the TOPSIS-selected optimal point marked.

4.8. Economic Sensitivity Analysis

To assess the robustness of the TOPSIS-selected optimal point against market fluctuations, a one-at-a-time sensitivity study was performed on the two most influential economic inputs: the biomass fuel price and purchased equipment cost (PEC). Each parameter was varied by ±10 %, ±20 %, and ±30 % around its baseline value, while all other inputs were held at their TOPSIS-optimal values, and the corresponding change in the SUCP was recorded.

The baseline SUCP at the TOPSIS-optimal operating point is 6.10 $/GJ. The results indicate that a ±30 % variation in biomass price causes the SUCP to shift by approximately ±6.30 % (range: 5.72–6.48 $/GJ), while a ±30 % variation in the total PEC produces a change of about ±9.00 % (range: 5.55–6.65 $/GJ). The exergy efficiency and CO_2 emission rate remain unchanged, since they depend on thermodynamic rather than economic parameters. These results demonstrate that the optimum is economically stable: even under aggressive cost fluctuations, the SUCP stays within a competitive 5.55–6.65 $/GJ range, well below the 11.67–13.37 $/GJ values reported by Zhang et al. [5,6] for comparable systems. The full sensitivity sweep is presented as a tornado chart in Figure 12.

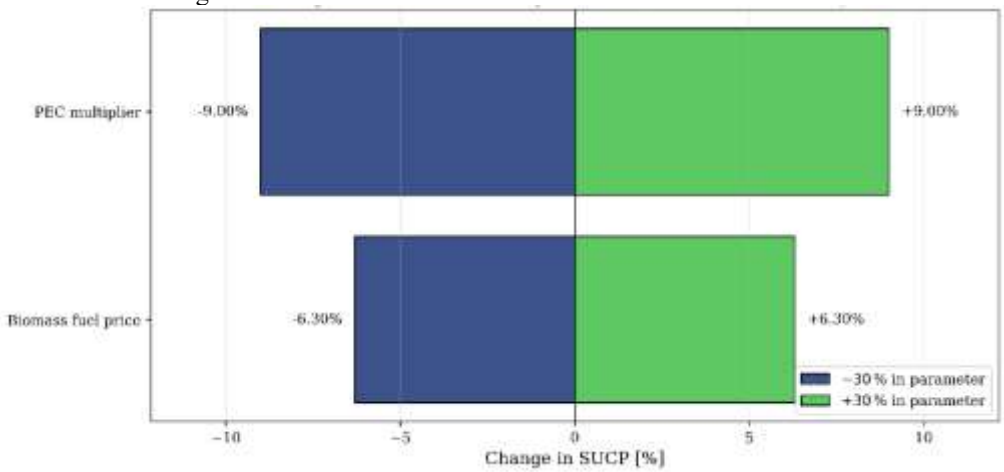

Fig. 12: Tornado chart of one-at-a-time economic sensitivity of SUCP to biomass fuel price and purchased equipment cost (PEC) at the TOPSIS-optimal operating point.

4.9. Comparison with Related Works

To contextualize the contributions and performance of the present study, a comparison with five recent related works is presented in Table 5. The comparison covers system configuration, analysis methodology, ML approach, best ML accuracy (R^2), and key performance indicators.

Table 5: Comparison of the present study with recent related works.

Aspect	Present Study	Zhang et al. (2024a)	Zhang et al. (2024b)	Zhang et al. (2024c)	Ozonoh et al. (2024)	Alfarra et al. (2025)
System	Biomass GT + ORC + Absorption Chiller	Biomass EFGT + SRC + ORC + ARC	Biomass EFGT + ORC + ARC	NG+Biomass DFGT + SRC + ORC + SEAC	Biomass-coal co-gasification	Biomass gasification
Analysis	4E (Energy, Exergy, Economic, Environmental)	Thermodynamic + Exergoeconomic	Thermodynamic + Exergoeconomic	Thermodynamic + Exergoeconomic	ML prediction + SHAP	ML prediction
ML Models	RF, GBR, XGBoost, LightGBM, CatBoost	None	None	None	LR, SVR, GPR, XGBoost, CatBoost	RF, XGBoost, LightGBM, GBR, KNN, DT and others
Explainability	SHAP (beeswarm, dependence, waterfall)	None	None	None	SHAP + LIME	None
Optimization	TOPSIS multi-objective	GA multi-objective	GA multi-objective	GA multi-objective	Evolutionary algorithm	Bayesian hyperparameter optimization
Best ML R^2	0.9813 (SUCP, XGB-Tuned)	N/A	N/A	N/A	0.9992 (LHV)	Not reported uniformly
η_{en}(%)	~61 (mean); up to ~72	70.67	Not explicitly reported	75.69	N/A	N/A
η_{ex}(%)	~46 (mean); 61.8 (TOPSIS optimal)	39.13	Up to 9.7% improvement from baseline	41.76	N/A	N/A
SUCP / LCOE	~6.10 $/GJ (TOPSIS optimum)	11.67 USD/GJ	SUCP as objective	13.37 $/GJ	N/A	

A major novelty of the current study is that it bridges three important aspects at the same time, for the first time, to the best of the authors' knowledge: (1) full 4E assessment of a biomass-fired GT–ORC–absorption chiller multigeneration plant, (2) comparative and tuning analysis of five ensemble state-of-the-art ML models with $R^2 > 0.97$ for all considered targets, and (3) the use of multi-level SHAP explainability (beeswarm, dependence, and waterfall plots) with TOPSIS-based multi-objective optimization. No other study reported in the literature has, to date, focused on all three.

5. Limitations and Future Work

Despite the demonstrated accuracy and interpretability of the proposed framework, several limitations should be acknowledged, each pointing to a clear direction for future research.

(i) Steady-state assumption. The current 4E analysis is restricted to steady-state operation. Real biomass multigeneration plants experience transient phenomena during start-up, shut-down, and partial-load operation, where gasifier dynamics, turbine inertia, and absorption-chiller crystallization risks become important. Future work will extend the framework by coupling a dynamic model with a recurrent or LSTM-based surrogate to capture time-dependent behavior.

(ii) Single feedstock. Only wood biomass was considered. The performance map could differ noticeably for agricultural residues, energy crops, or municipal solid waste due to changes in moisture content, ash composition, and lower heating value. A multi-feedstock dataset will be generated in future work to train a more generalizable surrogate.

(iii) Equilibrium gasification model. The Gibbs-free-energy minimization approach used here neglects kinetic and tar-related effects. Coupling the framework with a CFD or kinetic gasifier model would yield more realistic syngas predictions, particularly at low gasification temperatures.

(iv) Static economic parameters. Capital, operation, and biomass prices were assumed to be constant in the baseline analysis. Although a deterministic sensitivity study (Section 4.8) confirmed economic stability, a future stochastic 4E framework based on Monte-Carlo simulation, combined with the present SHAP–NSGA-II–TOPSIS pipeline, would allow full propagation of market and policy uncertainties into the design recommendations.

(v) Single decision-maker weighting. Equal weights were applied in TOPSIS. Future work will incorporate AHP- or entropy-based weighting and multi-stakeholder preference scenarios to test the sensitivity of the optimal point to subjective judgment.

6. Conclusions

A new framework of SHAP-explainable ensemble machine learning with 4E analysis and multi-objective optimization was presented for a biomass-fired gas turbine coupled with ORC and absorption chiller multigeneration system. The five ensemble models were compared in a systematic manner and the tuned XGBoost is observed to yield test-set $R^{2} > 0.97$ for all the five targets. The SHAP-based contribution plots indicate the turbine inlet temperature, compressor pressure ratio, and equivalence ratio to be the key driving factors for the system performance. The TOPSIS-based optimal point shows an overall well-balanced compromise among the three objectives of thermodynamic efficiency, economic cost, and environmental impact. The proposed framework can be used to convert the black-box ML surrogates to interpretable design advisors that can provide actionable engineering insight for sustainable design and operation of biomass multigeneration systems.

References

[1] Kumar, J., & Vyas, S. (2025). Comprehensive review of biomass utilization and gasification for sustainable energy production. *Environment, Development and Sustainability*, *27*, 1–40. https://doi.org/10.1007/s10668-023-04127-7

[2] Abouemara, K., Shahbaz, M., McKay, G., & Al-Ansari, T. (2024). The review of power generation from integrated biomass gasification and solid oxide fuel cells: Current status and future directions. *Fuel*, *360*, 130511. https://doi.org/10.1016/j.fuel.2023.130511

[3] Osman, A. I., et al. (2024). Cutting-edge biomass gasification technologies for renewable energy generation and achieving net zero emissions. *Energy Conversion and Management*, *321*, 119005. https://doi.org/10.1016/j.enconman.2024.119005

[4] Taqvi, S. A. A., et al. (2024). State-of-the-art review of biomass gasification: Raw to energy generation. *ChemBioEng Reviews*, *11*(4), e202400003. https://doi.org/10.1002/cben.202400003

[5] Zhang, X., Guo, H., He, Z., & Wang, J. (2024). Exergoeconomic evaluation of a cogeneration system driven by a natural gas and biomass co-firing gas turbine combined with a steam Rankine cycle, organic Rankine cycle, and absorption chiller. *Processes*, *12*(1), 82. https://doi.org/10.3390/pr12010082

[6] Zhang, X., Guo, H., He, Z., & Wan, J. (2024). Exergoeconomic analysis and optimization of a biomass integrated gasification combined cycle based on externally fired gas turbine, steam Rankine cycle, organic Rankine cycle, and absorption refrigeration cycle. *Entropy*, *26*(6), 511. https://doi.org/10.3390/e26060511

[7] Zhang, X., Guo, H., & Wan, J. (2024). Investigation of a biomass-driven cogeneration system integrated with an externally fired gas turbine, organic Rankine cycle, and absorption refrigeration cycle: Thermodynamic and exergoeconomic analyses and optimization. *Sustainability*, *16*(11), 4495. https://doi.org/10.3390/su16114495

[8] Sprouse, C. (2024). Review of organic Rankine cycles for internal combustion engine waste heat recovery: Latest decade in review. *Sustainability*, *16*(5), 1924. https://doi.org/10.3390/su16051924

[9] Peris, B., Navarro-Esbrí, J., Molés, F., & Mota-Babiloni, A. (2015). Organic Rankine cycle for waste heat recovery in a refinery. *Industrial & Engineering Chemistry Research*, *54*(14), 3600–3611. https://doi.org/10.1021/acs.iecr.5b03381

[10] Tocci, L., Pal, T., Pesmazoglou, I., & Franchetti, B. (2017). Small scale organic Rankine cycle (ORC): A techno-economic review. *Energies*, *10*(4), 413. https://doi.org/10.3390/en10040413

[11] Ren, J., Qian, Z., Fei, C., Lu, D., Zou, Y., Xu, C., & Liu, L. (2023). Thermodynamic, exergoeconomic, and exergoenvironmental analysis of a combined cooling and power system for natural gas–biomass dual fuel gas turbine waste heat recovery. *Energy*, *269*, 126823. https://doi.org/10.1016/j.energy.2023.126823

[12] Ahmadi, P., Dincer, I., & Rosen, M. A. (2012). Exergo-environmental analysis of an integrated organic Rankine cycle for trigeneration. *Energy Conversion and Management*, *64*, 447–453. https://doi.org/10.1016/j.enconman.2012.06.001

[13] Khanmohammadi, S., Atashkari, K., & Kouhikamali, R. (2015). Exergoeconomic multi-objective optimization of an externally fired gas turbine integrated with a biomass gasifier. *Applied Thermal Engineering*, *91*, 848–859. https://doi.org/10.1016/j.applthermaleng.2015.08.080

[14] El-Sattar, H. A., Kamel, S., Vera, D., & Jurado, F. (2020). Tri-generation biomass system based on externally fired gas turbine, organic Rankine cycle and absorption chiller. *Journal of Cleaner Production*, *260*, 121068. https://doi.org/10.1016/j.jclepro.2020.121068

[15] Roy, D., Samanta, S., & Ghosh, S. (2019). Techno-economic and environmental analyses of a biomass based system employing solid oxide fuel cell, externally fired gas turbine and organic Rankine cycle. *Journal of Cleaner Production*, *225*, 36–57. https://doi.org/10.1016/j.jclepro.2019.03.261

[16] Lazzaretto, A., & Tsatsaronis, G. (2006). SPECO: A systematic and general methodology for calculating efficiencies and costs in thermal systems. *Energy*, *31*(8–9), 1257–1289. https://doi.org/10.1016/j.energy.2005.03.011

[17] Szargut, J., Morris, D. R., & Steward, F. R. (1988). *Exergy analysis of thermal, chemical, and metallurgical processes*. Hemisphere Publishing.

[18] Bongomin, O., et al. (2024). Exploring insights in biomass and waste gasification via ensemble machine learning models and interpretability techniques. *International Journal of Energy Research*, *2024*, 6087208. https://doi.org/10.1155/2024/6087208

[19] Ozonoh, M., Oboirien, B. O., & Daramola, M. O. (2024). Machine learning optimization for enhanced biomass-coal co-gasification. *Renewable Energy*, *228*, 120682. https://doi.org/10.1016/j.renene.2024.120682

[20] Alfarra, F., Ozcan, H. K., Cihan, P., Ongen, A., Guvenc, S. Y., & Ciner, M. N. (2025). Bayesian hyperparameter optimization of machine learning models for predicting biomass gasification gases. *Applied Sciences*, *15*(3), 1018. https://doi.org/10.3390/app15031018

[21] Hasanzadeh, A., et al. (2022). Comparative analysis between 10 MW gas turbine and hybrid SOFC-GT systems using 4E criteria and ANN-Grey Wolf optimizer. *International Journal of Hydrogen Energy*, *47*(39), 17121–17149. https://doi.org/10.1016/j.ijhydene.2022.03.190

[22] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774.

[23] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785