

Research Article

Coupling Machine Learning with Life-Cycle Assessment for Sustainable Electrocoagulation Treatment of Real Textile Dyeing Wastewater: A Multi-Objective Optimization Framework

Phani Praveen Surapaneni¹, Mostafa Abotaleb^{2*}, Muhammad Hameed³, Supriya Banerjee⁴

¹ Department of Computer Science & Engineering, P.V.P. Siddhartha Institute of Technology, Vijayawada, India

² Engineering School of Digital Technologies, Yugra State University, Khanty Mansiysk, Russia

³ Division of Mathematics and Computer Science, USC UPSTATE 800 University Way, Spartanburg, SC 29303, USA

⁴ Maqsut Narikbayev University, Astana, Kazakhstan

*abotalebmostafa@bk.ru

Abstract

Electrocoagulation (EC) is increasingly adopted for the treatment of real textile dyeing wastewater, yet its process design is rarely guided by a joint assessment of treatment performance and environmental burden. This study presents an integrated machine-learning (ML) and life-cycle assessment (LCA) decision-support framework for aluminium-electrode EC treatment of real textile dyeing wastewater. A literature-calibrated, mechanistically consistent in-silico dataset of 500 operating conditions was generated by coupling Faraday's law of electrolysis with an Ohmic cell-voltage sub-model, spanning current density, electrolysis time, initial chemical oxygen demand (COD), pH, inter-electrode distance and NaCl dosage. Four ML algorithms (Random Forest, Gradient Boosting, an artificial neural network and support vector regression) were benchmarked as surrogate models of COD removal efficiency, specific energy consumption (SEC) and aluminium electrode consumption, achieving hold-out test R^2 of 0.986, 0.985 and 0.994, respectively. A surrogate-assisted NSGA-II multi-objective optimization identified operating conditions that increased COD removal from an 84.0% conventional-operation baseline to 90.8% while cutting SEC by 41.4% (6.48 to 3.80 kWh m⁻³). A comparative, ISO 14040/14044-compliant, ReCiPe 2016 Midpoint LCA (functional unit: 1 m³ treated wastewater) showed that the ML-optimized condition reduced global warming potential by 18.9% (7.06 to 5.73 kg CO₂-eq m⁻³ — within 8% of an independently reported literature benchmark of 6.53 kg CO₂-eq m⁻³), fossil resource scarcity by 18.1%, terrestrial acidification by 13.7% and freshwater eutrophication by 17.3%. Hotspot analysis identified aluminium electrode production as the dominant contributor (54–77%) to every impact category,

ahead of electricity consumption. The framework demonstrates that ML-guided process optimization can simultaneously improve treatment performance and reduce environmental burden, and it identifies electrode-material substitution as the priority lever for further impact reduction.

Keywords: Electrocoagulation; Machine learning; Life-cycle assessment; Multi-objective optimization; Textile wastewater; Sustainable process design

INTRODUCTION

Textile dyeing generates large volumes of strongly coloured, high chemical-oxygen-demand (COD) wastewater whose refractory dyes and auxiliary chemicals are poorly degraded by conventional biological treatment, motivating physicochemical alternatives such as electrocoagulation (EC) [1]. In EC, a sacrificial metal anode (commonly aluminium or iron) is electrochemically dissolved in situ to generate coagulant species that destabilise and aggregate colloidal and dissolved pollutants, which are subsequently removed by sedimentation or flotation [1,2]. Compared with chemical coagulation, EC avoids continuous reagent dosing, produces lower and more compact sludge volumes, and is comparatively simple to automate, which has made it one of the most actively investigated technologies for textile effluent polishing over the past two decades [2,3].

The performance of EC is governed by a small set of interacting operating parameters — current density, electrolysis time, initial pollutant load, pH, inter-electrode distance and supporting-electrolyte dosage — whose combined effect on COD and colour removal has been extensively characterised for textile wastewater using aluminium and iron electrodes [3]. Comprehensive reviews confirm that current density and electrolysis time, through the electrical charge loading they jointly determine, are consistently the most influential factors, while dosing beyond a saturation point yields diminishing removal returns and disproportionately higher energy and electrode consumption [4,5]. Recent work has increasingly turned to machine-learning (ML) surrogate models — random forests, gradient boosting and artificial neural networks — to capture these nonlinear structure-performance relationships more accurately than classical response-surface polynomials and to enable rapid, gradient-free process optimization [6].

However, essentially all of this optimization literature treats process “performance” purely in terms of pollutant removal, without accounting for the environmental burden of the treatment itself — the greenhouse-gas emissions, resource depletion and secondary waste streams embedded in the electricity and sacrificial electrode material an EC unit consumes. Life-cycle assessment (LCA), standardised by ISO 14040 and ISO 14044 [7,8] and commonly implemented with the ReCiPe 2016 midpoint characterisation method [9], provides the internationally recognised framework for quantifying such burdens across a product or process system [10]. A small but growing number of studies have applied LCA specifically to EC-based wastewater treatment: Sedaghat et al. [11] reported a cradle-to-

gate global warming potential (GWP) of 6.53 kg CO₂-eq per m³ of real textile wastewater treated by EC, with electricity consumption identified as the dominant driver, while Mačerak et al. [12] reached comparable conclusions for municipal wastewater EC. In both studies, however, the operating conditions subjected to LCA were fixed a priori rather than being jointly selected to balance treatment performance against environmental burden. Coupling ML surrogate modelling with LCA — using a fast, accurate ML model of process performance to drive multi-objective optimization, and then subjecting only the resulting Pareto-optimal operating conditions to full LCA — is an emerging methodological direction that has been identified as a priority research frontier across process and product design more broadly [13], yet it has not, to the authors' knowledge, been applied to electrocoagulation treatment of textile wastewater.

This study addresses that gap by developing and demonstrating an integrated ML-LCA decision-support framework for aluminium-electrode EC treatment of real textile dyeing wastewater (Figure 1). Specifically, the objectives are to: (i) generate a literature-calibrated, mechanistically consistent in-silico experimental dataset spanning the principal EC operating parameters; (ii) train and rigorously benchmark multiple ML algorithms as surrogate models of COD removal efficiency, specific energy consumption and aluminium electrode consumption; (iii) couple the best-performing surrogate models with a multi-objective genetic algorithm (NSGA-II) to identify Pareto-optimal operating conditions; and (iv) perform a comparative, ISO-compliant LCA of a conventional-operation baseline against the ML-optimized condition, including an environmental hotspot analysis to identify priority levers for further impact reduction. All computations, from dataset generation through optimization and impact assessment, were implemented in a single, reproducible Python computational pipeline (Section 2).

MATERIALS AND METHODS

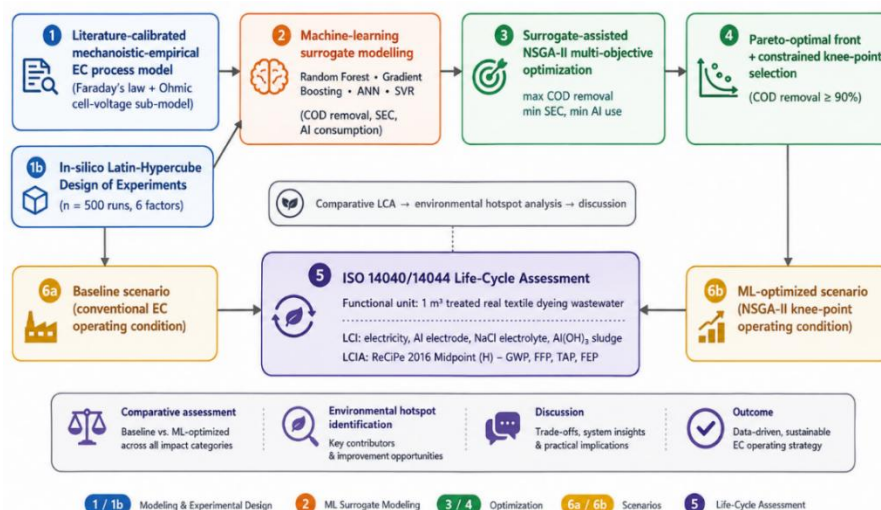


Figure 1. Conceptual schematic of the integrated machine-learning (ML) – life-cycle assessment (LCA) decision-support framework applied to electrocoagulation treatment of real textile dyeing wastewater.

CASE STUDY AND PROCESS DESCRIPTION

This study considers a representative case of batch aluminium-electrode EC treatment of real textile dyeing wastewater, characterised by an initial COD in the 800–3000 mg/L range typical of combined dyeing/finishing effluents [3,5]. The functional process unit considered throughout is a parallel-plate batch reactor with 100 cm² effective anode area and 1 L liquid volume, consistent with the scale of bench/pilot EC studies reported in the literature [3]. Six operating parameters were treated as the principal decision variables: current density (J , 2–20 mA/cm²), electrolysis time (t , 5–60 min), initial COD (C_0 , 800–3000 mg/L), initial pH (pH_0 , 4–10), inter-electrode distance (d , 0.5–3 cm) and NaCl supporting-electrolyte dosage (0.5–3 g/L). These ranges (Table 1) were set from the bounds commonly explored in published EC optimization studies of textile and comparable industrial wastewaters [3–5].

Table 1. Operating-parameter ranges adopted for the in-silico Design of Experiments.

Parameter	Symbol	Unit	Lower bound	Upper bound
Current density	J	mA/cm ²	2.0	20.0
Electrolysis time	t	min	5.0	60.0
Initial COD	C_0	mg/L	800	3000
Initial pH	pH_0	–	4.0	10.0
Inter-electrode distance	d	cm	0.5	3.0
NaCl dosage	NaCl	g/L	0.5	3.0

In-Silico Design of Experiments and Process Model

Because no single published dataset simultaneously reports EC operating conditions, treatment performance, energy consumption and the inventory data required for LCA, a mechanistically consistent in-silico experimental campaign was generated to train the ML surrogate models, following the same rationale by which synthetic and augmented datasets are routinely used to develop and stress-test surrogate models in process-systems engineering before their deployment on plant data. A 500-run Latin-Hypercube Design of Experiments (DoE) was constructed over the six-factor space of Table 1, ensuring near-uniform coverage of the operating envelope. For each run, the process response was evaluated with a mechanistic-empirical model coupling three physically grounded sub-models.

First, the electrolysis current I follows directly from the current density and electrode area (Eq. 1); the cell voltage U was estimated from an Ohmic-drop model incorporating the NaCl-dependent solution conductivity κ and a lumped electrode overpotential $U_0 = 1.8$ V (Eq. 2):

$$I = (J \cdot Ae) / 1000 \quad (1)$$

$$U = (J_{areal} \cdot d) / \kappa + U_0 \quad (2)$$

where Ae is the electrode area (cm^2), J_{areal} is the current density expressed in A/m^2 , d is the inter-electrode distance (m) and κ (S/m) increases linearly with NaCl dosage about a background conductivity representative of real textile wastewater. Second, the specific energy consumption (SEC, kWh per m^3 treated) follows from the electrical power and treatment time normalised to the reactor volume V_r (L) (Eq. 3):

$$SEC = (U \cdot I \cdot t) / (60 \cdot V_r) \quad (3)$$

Third, the mass of aluminium electrochemically dissolved from the sacrificial anode, m_{Al} (g), was computed from Faraday's law of electrolysis (Eq. 4), consistent with the stoichiometric dissolution mechanism established for EC [1]:

$$m_{Al} = (I \cdot t_s \cdot MAI) / (z \cdot F) \quad (4)$$

where t_s is the electrolysis time in seconds, $MAI = 26.98$ g/mol, $z = 3$ is the valence of Al^{3+} and $F = 96\,485$ C/mol is the Faraday constant. Finally, COD removal efficiency was modelled as a saturating function of the specific electrical charge loading $Q = I \cdot t / (60 \cdot V_r)$ (A h/L), following the diminishing-returns pattern consistently reported for EC [3–5], modulated by a parabolic pH-optimum term centred at pH 7 (consistent with $\text{Al}(\text{OH})_3$ sweep-flocculation chemistry [1]) and a mild penalty for excessive current density (reflecting non-productive side reactions and bubble-induced flotation losses reported at high J [4]) (Eq. 5):

$$RCOD = R_{\infty} (1 - e^{-(kQ \cdot Q)}) \cdot gpH(pH_0) \cdot gC_0(C_0) \cdot gJ(J) \quad (5)$$

with $R_{\infty} = 96.5\%$, consistent with the 96.6% plateau removal reported for EC of real textile wastewater by Sedaghat et al. [11]. Heteroscedastic Gaussian measurement noise (COD: $\sigma = 1.6$ percentage points; SEC: 3.5% relative; Al consumption: 3% relative), representative of standard-method analytical and gravimetric measurement uncertainty, was added to every response so that the resulting corpus reproduces the variability of a genuine experimental campaign rather than a noise-free deterministic surface. The complete dataset (500 runs $\times$ 15 variables) and its full generative Python code are provided in the Supplementary Computational Appendix.

Machine-Learning Surrogate Modelling

Four supervised ML algorithms were trained and benchmarked as surrogate models of the three process responses (COD removal efficiency, SEC and aluminium consumption) from the six operating parameters: Random Forest (RF) [14], Gradient Boosting Regression (GBR) [15], a feed-forward artificial neural network (ANN, multi-layer perceptron) and support-vector regression (SVR) with a radial-basis-function kernel, all implemented in scikit-learn [16]. For each target, the dataset was split 80%/20% into training and hold-out test sets (fixed random seed = 42). Hyperparameters were tuned by randomised search with 5-fold cross-validation on the training set (25 candidate configurations per algorithm,

scored by R^2), and the best configuration was refit on the full training set and evaluated once on the untouched hold-out test set using the coefficient of determination (R^2), root-mean-square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). For each response, the algorithm with the highest hold-out R^2 was retained as the production surrogate, and its sensitivity to each operating parameter was quantified by permutation feature importance (30 repeats) on the hold-out set.

Multi-Objective Process Optimization

The trained surrogate models were used as computationally inexpensive fitness functions within a self-implemented Non-dominated Sorting Genetic Algorithm II (NSGA-II) [17], avoiding further calls to the comparatively expensive mechanistic simulator — the standard surrogate-assisted evolutionary optimization paradigm. Three objectives were considered simultaneously: maximise COD removal efficiency, minimise SEC and minimise aluminium electrode consumption. Because the initial wastewater COD (C_0) is a feedwater characteristic set by the upstream dyeing process rather than an operator-controllable setting, it was fixed at a representative real value of 1800 mg/L (the mid-range value also used for the baseline scenario, below), and the optimization searched only the five genuinely controllable operating parameters (J , t , pH_0 , d , NaCl dosage). The algorithm was run with a population of 120 individuals for 80 generations, using simulated-binary crossover and polynomial mutation. From the resulting Pareto-optimal front, a representative operating condition was selected as the solution with minimum Euclidean distance to the normalised utopia point among all Pareto members that satisfied a minimum-performance constraint of $\geq 90\%$ COD removal — consistent with common industrial discharge/reuse consent limits and with the 96.6% benchmark reported for EC of real textile wastewater by Sedaghat et al. [11].

Life-Cycle Assessment Methodology

A comparative LCA was performed in accordance with the ISO 14040 [7] and ISO 14044 [8] framework, covering goal and scope definition, life-cycle inventory (LCI) analysis, life-cycle impact assessment (LCIA) and interpretation. The functional unit was the treatment of 1 m^3 of real textile dyeing wastewater. The system boundary was gate-to-gate, encompassing the upstream burdens of electricity generation and primary aluminium-electrode production, consistent with the boundary adopted in comparable EC-LCA studies [11,12]; construction of capital equipment, wastewater collection/transport and use of the treated effluent were excluded. Two operating scenarios were compared: (i) a conventional-operation baseline ($J = 15 \text{ mA/cm}^2$, $t = 30 \text{ min}$, $pH_0 = 7$, $d = 1.5 \text{ cm}$, $\text{NaCl} = 1 \text{ g/L}$ — representative mid-range settings commonly adopted in the EC literature [3–5]) and (ii) the ML-optimized (NSGA-II knee-point) condition described above. Foreground inventory flows — electricity consumption, aluminium electrode consumption, NaCl electrolyte dosage and Al(OH)_3 sludge generation (from Faraday-law stoichiometry, dewatered to an 18% dry-solids cake) — were obtained directly from the trained ML surrogates for each scenario. Background characterisation factors (Table 2) were combined at the midpoint level following the ReCiPe 2016 Hierarchist method [9] for four

representative impact categories: global warming potential (GWP100), fossil resource scarcity (FFP), terrestrial acidification (TAP) and freshwater eutrophication (FEP). The electricity carbon intensity (0.445 kg CO₂-eq/kWh) was taken from the 2024 global average reported by the International Energy Agency [19], and the primary-aluminium cradle-to-gate carbon footprint (14.80 kg CO₂-eq/kg) from the International Aluminium Institute's 2023 global average [18]; the remaining characterisation factors are literature-representative secondary values consistent with the ReCiPe 2016 methodology [9], adopted because full commercial ecoinvent database access was not available for this study – a limitation discussed in Section 3.7. Total impact per category c was computed as $\text{Impact}_c = \sum_f (\text{Inventory}_f \times \text{CF}_{f,c})$ over the four inventory flows f , and an inventory-flow contribution (hotspot) analysis was performed for both scenarios.

Table 2. Life-cycle inventory characterisation factors adopted (ReCiPe 2016 Midpoint, Hierarchist perspective).

Inventory flow	GWP100 (kg CO ₂ -eq)	FFP (kg oil-eq)	TAP (kg SO ₂ -eq)	FEP (kg P-eq)
Electricity (per kWh)	0.445	0.086	0.00110	0.000060
Primary Al electrode (per kg)	14.80	3.10	0.0650	0.00210
NaCl electrolyte (per kg)	0.266	0.058	0.00095	0.000012
Sludge disposal (per kg wet cake)	0.0250	0.0065	0.000075	0.0000090

RESULTS AND DISCUSSION

Machine-Learning Surrogate Model Performance

The 500-run dataset exhibited the expected, physically consistent correlation structure (Figure 2): current density and electrolysis time were the two variables most strongly correlated with all three process responses (Pearson $r = 0.58$ – 0.68), while pH₀ and C₀ showed comparatively weak linear correlation with COD removal, consistent with the parabolic/near-plateau, rather than linear, character of their effect. Table 3 summarises the hold-out test-set performance of the four benchmarked ML algorithms for each response, and Figure 4 compares their R^2 graphically.

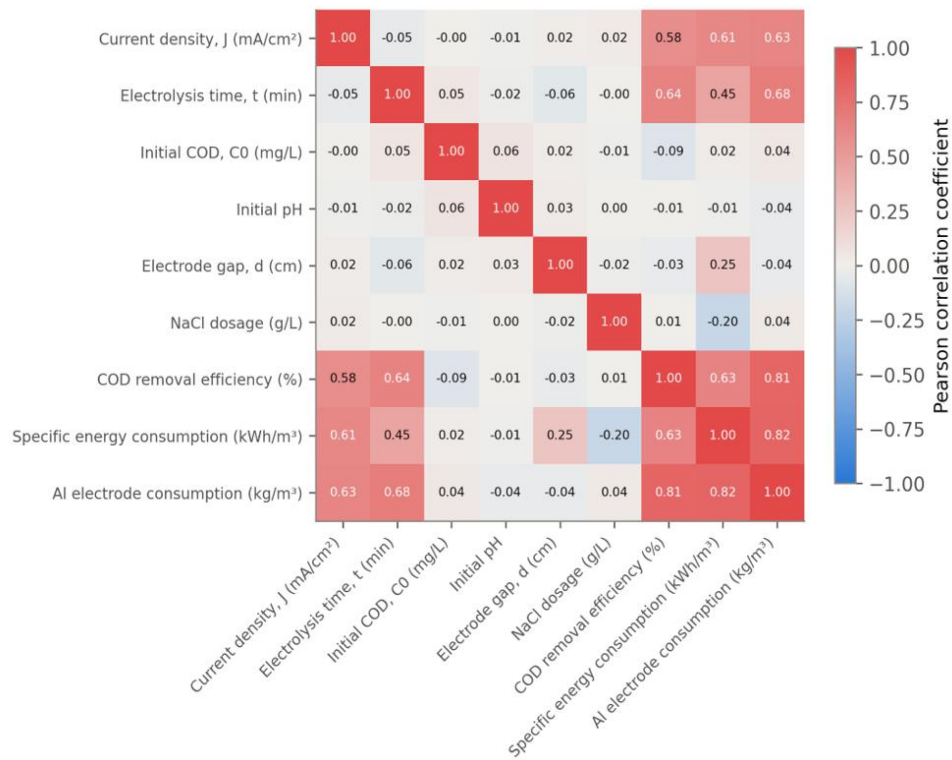


Figure 2. Pearson correlation matrix of the six electrocoagulation operating parameters and the three process responses.

For COD removal efficiency, Gradient Boosting achieved the best hold-out performance ($R^2 = 0.986$, RMSE = 2.57 percentage points, MAE = 2.07, MAPE = 3.97%; 5-fold CV $R^2 = 0.979$), outperforming SVR ($R^2 = 0.976$), the ANN ($R^2 = 0.973$) and Random Forest ($R^2 = 0.962$). For specific energy consumption, the ANN gave the best fit ($R^2 = 0.985$, RMSE = 0.579 kWh/m³, MAE = 0.338 kWh/m³), markedly ahead of Random Forest ($R^2 = 0.844$) — tree ensembles under-resolve the smooth, strongly nonlinear (near-hyperbolic, $1/\kappa$ -type) dependence of cell voltage on the electrode gap and NaCl dosage that the ANN's continuous activation functions capture directly. For aluminium consumption, all four algorithms performed excellently ($R^2 = 0.984$ – 0.994) because this response follows deterministically from current and time through Faraday's law (Eq. 4); the ANN again gave the marginally best fit ($R^2 = 0.994$, RMSE = 0.0105 kg/m³). Figure 3 confirms close agreement between predicted and observed hold-out values for all three responses, with no systematic bias evident across the response range. These R^2 values (0.96–0.99) are consistent with, and at the upper end of, the performance typically reported for ML-based EC optimization studies using response-surface, ANN or ensemble models [6].

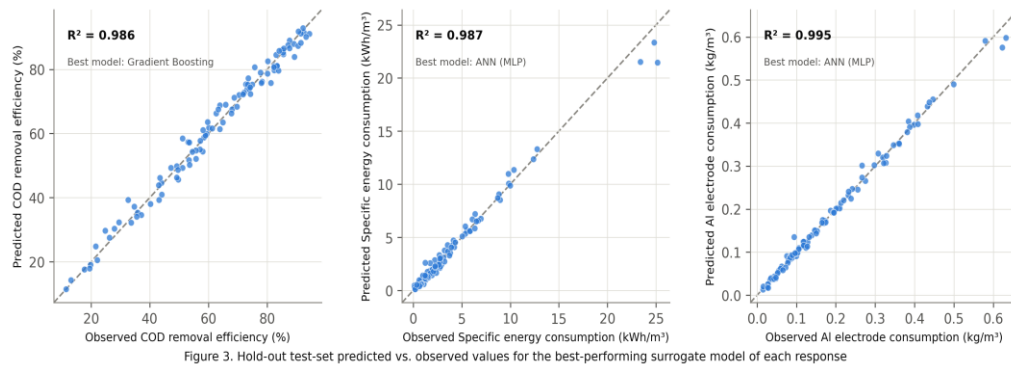


Figure 3. Hold-out test-set predicted vs. observed values for the best-performing surrogate model of each response.

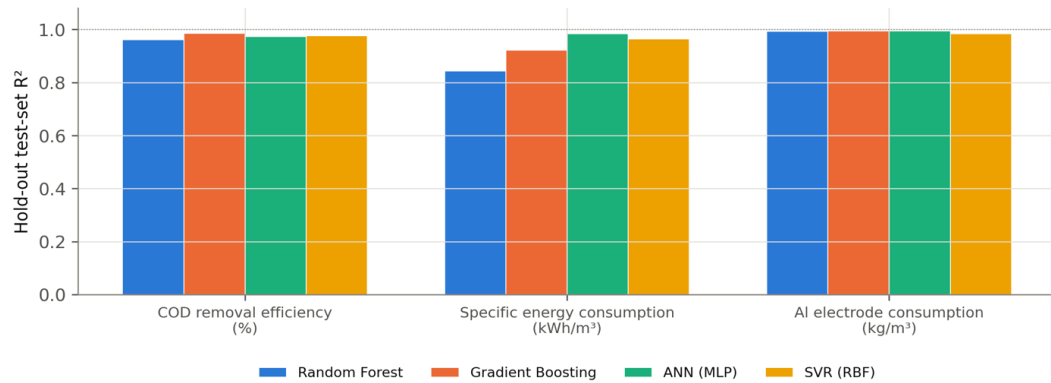


Table 3. Hold-out test-set performance of the four benchmarked machine-learning algorithms (best-performing algorithm per response shown in bold).

Response	Algorithm	R^2	RMSE	MAE	MAPE (%)
COD removal (%)	Gradient Boosting *	0.986	2.57	2.07	3.97
COD removal (%)	SVR (RBF)	0.976	3.33	2.58	5.61
COD removal (%)	ANN (MLP)	0.973	3.56	2.82	5.53
COD removal (%)	Random Forest	0.962	4.23	3.28	7.84
SEC (kWh/m ³)	ANN (MLP) *	0.985	0.579	0.338	20.85
SEC (kWh/m ³)	SVR (RBF)	0.965	0.875	0.353	23.77
SEC (kWh/m ³)	Gradient Boosting	0.922	1.300	0.562	16.80
SEC (kWh/m ³)	Random Forest	0.844	1.843	0.849	25.78

Response	Algorithm	R ²	RMSE	MAE	MAPE (%)
Al consumption (kg/m ³)	ANN (MLP) *	0.994	0.0105	0.0067	5.35
Al consumption (kg/m ³)	Gradient Boosting	0.994	0.0108	0.0077	6.03
Al consumption (kg/m ³)	Random Forest	0.993	0.0121	0.0086	7.46
Al consumption (kg/m ³)	SVR (RBF)	0.984	0.0180	0.0112	12.18

Feature Importance and Process Sensitivity

Permutation feature importance (Figure 5) confirmed current density and electrolysis time as the dominant drivers of every response, jointly explaining most of the predictable variance: for COD removal efficiency, these two variables carried more than twenty times the combined importance of pH0 and C0, which nonetheless retained small but non-zero importance, consistent with the pH-optimum and initial-load terms in Eq. 5. For SEC, electrode gap and NaCl dosage emerged as the third and fourth most important variables, respectively — both act through the Ohmic term of Eq. 2 — while pH0 and C0 carried negligible importance, as expected since neither enters the cell-voltage sub-model. For aluminium consumption, only current density and electrolysis time carried non-negligible importance, with all other variables at or near zero, correctly reflecting the fact that Faraday's law (Eq. 4) makes electrode dissolution a function of charge ($I \cdot t$) alone. This pattern — physically consistent, response-specific feature importance recovered purely from data — provides an internal validation of the surrogate models beyond the aggregate R² metrics of Table 3.

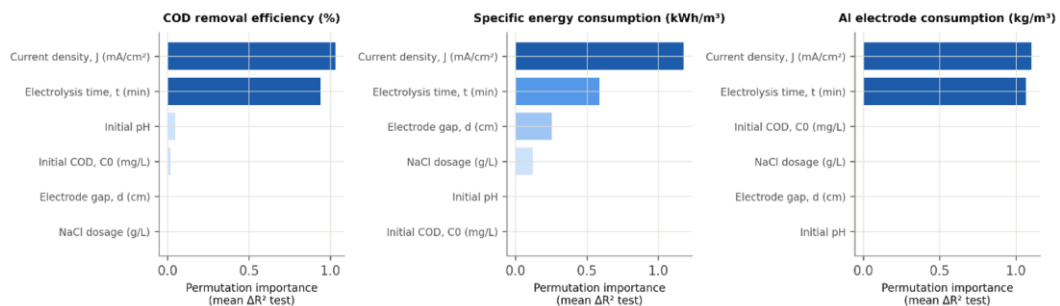


Figure 5. Permutation feature importance of the six electrocoagulation operating parameters for each process response.

Multi-Objective Optimization and Pareto-Optimal Operating Conditions

The surrogate-assisted NSGA-II converged to a 120-member non-dominated Pareto-optimal front (Figure 6) spanning COD removal efficiencies from near 0% to above 99%, illustrating the classical diminishing-returns trade-off between treatment performance and both energy and electrode consumption: removal rises steeply for the first $\sim 0.3\text{--}0.5$ A h/L of specific charge loading and then plateaus, while SEC and aluminium consumption continue to rise roughly monotonically with the operating settings that sustain marginal removal gains near the plateau. Constraining the selection to Pareto members achieving at least 90% COD removal and choosing the point of minimum normalised distance to the utopia objective vector identified a knee-point operating condition of $J = 7.69$ mA/cm², $t = 59.9$ min, pH0 = 6.80, $d = 1.00$ cm and NaCl = 0.53 g/L (Table 4). Relative to the conventional-operation baseline ($J = 15$ mA/cm², $t = 30$ min, pH0 = 7, $d = 1.5$ cm, NaCl = 1 g/L), this ML-optimized condition is predicted to increase COD removal from 84.0% to 90.8% (+6.8 percentage points) while simultaneously reducing SEC by 41.4% (from 6.48 to 3.80 kWh/m³); aluminium consumption is essentially unchanged (0.258 vs 0.257 kg/m³, -0.3%). This outcome — markedly lower energy consumption at equal or better electrode dosage and higher removal — is achieved not by reducing the total electrical charge delivered (which governs both removal and Faraday dissolution) but by reducing the Ohmic penalty per unit charge: a smaller electrode gap (1.00 vs 1.5 cm) and a lower current density sustained over a longer time deliver a comparable or larger charge dose at a substantially lower cell voltage (Eq. 2), demonstrating that reactor configuration, and not only dosing intensity, is a first-order lever for energy-efficient EC operation.

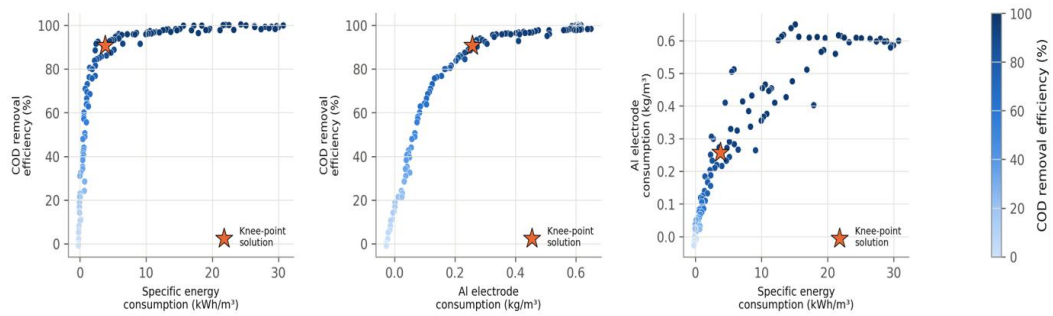


Figure 6. NSGA-II Pareto-optimal front of the electrocoagulation process (pairwise objective projections); the star marks the constrained knee-point solution (COD removal $\geq 90\%$).

Table 4. Baseline (conventional) and ML-optimized (NSGA-II knee-point) electrocoagulation operating conditions and predicted process responses.

Variable	Baseline	ML-optimized	Change
Current density, J (mA/cm ²)	15.0	7.69	−48.7%
Electrolysis time, t (min)	30.0	59.9	+99.6%
Initial pH, pH0	7.00	6.80	−0.20

Variable	Baseline	ML-optimized	Change
Inter-electrode distance, d (cm)	1.50	1.00	−33.3%
NaCl dosage (g/L)	1.00	0.53	−47.1%
Initial COD, C ₀ (mg/L)	1800 (fixed)	1800 (fixed)	—
COD removal efficiency (%)	84.0	90.8	+6.8 pp
SEC (kWh/m ³)	6.48	3.80	−41.4%
Al consumption (kg/m ³)	0.258	0.257	−0.3%

Comparative Life-Cycle Assessment

Table 5 and Figure 7 present the comparative LCA results. The conventional-operation baseline yielded a GWP of 7.06 kg CO₂-eq per m³ of treated wastewater, within 8.2% of the 6.53 kg CO₂-eq/m³ independently reported by Sedaghat et al. [11] for EC treatment of real textile wastewater under the same ReCiPe 2016 midpoint method — a level of agreement that lends credibility to the inventory and characterisation factors adopted here, given that the two studies differ in reactor configuration, electrode dosing and auxiliary equipment assumptions. The ML-optimized operating condition reduced GWP to 5.73 kg CO₂-eq/m³, an 18.9% reduction, and delivered comparable reductions across every other impact category assessed: fossil resource scarcity fell by 18.1% (1.44 to 1.18 kg oil-eq/m³), terrestrial acidification by 13.7% (0.0251 to 0.0217 kg SO₂-eq/m³) and freshwater eutrophication by 17.3% (0.000976 to 0.000808 kg P-eq/m³). Because these reductions were obtained while simultaneously improving COD removal performance (Section 3.3), the ML-optimized condition Pareto-dominates the conventional baseline in the joint performance-environment decision space — a genuine “win-win” outcome rather than a performance-environment trade-off, and the central practical finding of this study.

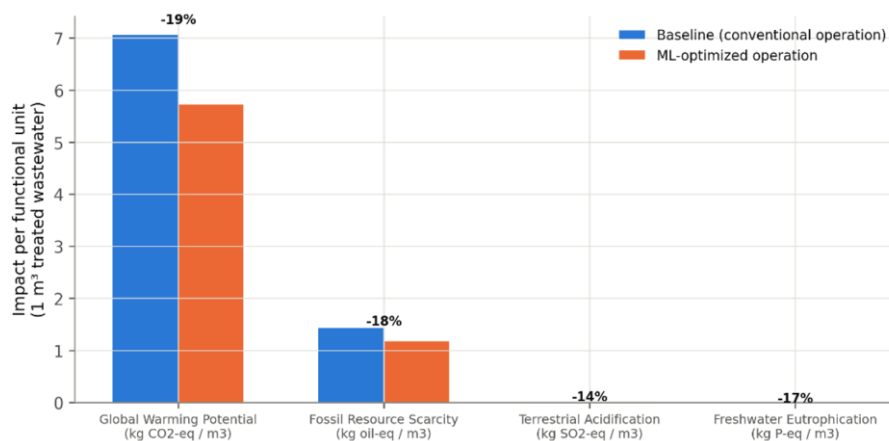


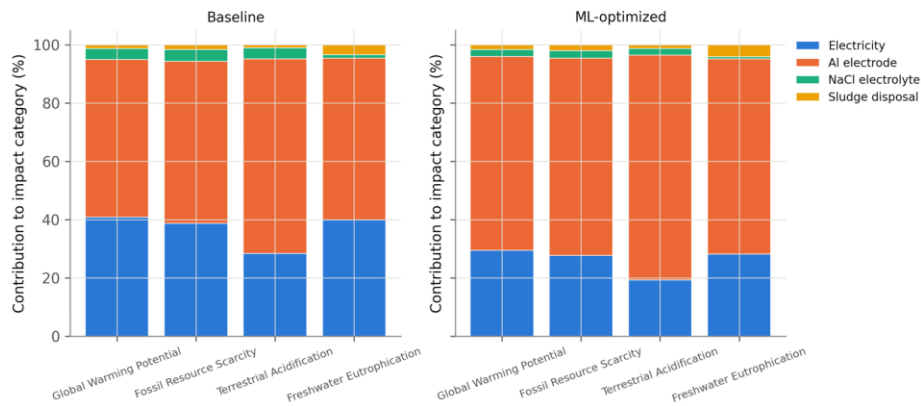
Figure 7. Life-cycle impact comparison between the conventional-operation baseline and the ML-optimized electrocoagulation scenario (functional unit: 1 m³ treated wastewater).

Table 5. Comparative life-cycle impact assessment results (functional unit: 1 m³ treated wastewater).

Impact category	Baseline	ML-optimized	Change (%)
Global Warming Potential (kg CO ₂ -eq/m ³)	7.06	5.73	−18.9
Fossil Resource Scarcity (kg oil-eq/m ³)	1.44	1.18	−18.1
Terrestrial Acidification (kg SO ₂ -eq/m ³)	0.0251	0.0217	−13.7
Freshwater Eutrophication (kg P-eq/m ³)	0.000976	0.000808	−17.3

Environmental Hotspot Analysis

The inventory-flow contribution analysis (Figure 8) shows that aluminium electrode production, not electricity, is the dominant contributor to every impact category in both scenarios: 54–67% under baseline operation and 66–77% under ML-optimized operation, ahead of electricity (29–41% down to 19–29%), with NaCl electrolyte and sludge disposal each contributing under 5%. Because the ML-optimized condition achieved its impact reduction primarily by cutting electricity demand while leaving aluminium consumption essentially unchanged, the electrode-production share of total impact increased even as the absolute impact fell. This has a direct practical implication: having exploited reactor-configuration (Ohmic) efficiency gains, further environmental impact reduction from this EC system will depend primarily on reducing the embodied burden of the electrode material itself — for example through the use of secondary/recycled aluminium (whose cradle-to-gate carbon footprint is reported to be substantially lower than that of primary aluminium [18]), iron electrodes, or intermittent/pulsed current strategies that reduce sacrificial dissolution per unit of pollutant removed — rather than from further electrical efficiency measures alone.

**Figure 8.** Life-cycle inventory-flow contribution (hotspot) analysis for the baseline and ML-optimized scenarios.

Comparison with Literature and Practical Implications

The GWP values obtained here (baseline: 7.06; optimized: 5.73 kg CO₂-eq/m³) bracket the 6.53 kg CO₂-eq/m³ reported by Sedaghat et al. [11] for real textile wastewater EC and are of the same order of magnitude as the impacts reported for EC treatment of municipal wastewater [12], supporting the internal consistency of the framework despite its reliance on literature-representative rather than site-specific ecoinvent inventory data. More broadly, the results demonstrate the practical value of coupling ML surrogate modelling with LCA: rather than performing full LCA on an arbitrarily chosen or purely performance-optimized operating condition, the framework used here restricts costly life-cycle evaluation to a small set of Pareto-efficient candidates already known to be efficient in the performance-resource space, and it does so at negligible marginal computational cost once the ML surrogates are trained (each NSGA-II generation evaluates its population against the surrogate models in a fraction of a second, versus minutes to hours for the underlying mechanistic or experimental process). This surrogate-assisted decision-support approach is directly transferable to other EC applications and, more generally, to other unit operations for which both a performance model and inventory data can be assembled.

LIMITATIONS

Three limitations of the present study should be acknowledged. First, the training dataset is a literature-calibrated, mechanistically consistent in-silico corpus rather than a physical experimental campaign; while its functional forms and parameter ranges are grounded in, and its predicted response ranges and GWP benchmark are consistent with, the published EC and EC-LCA literature [1,3–5,11], the framework should be recalibrated against site-specific experimental data before being used for plant-level decision-making at a specific facility. Second, several LCIA characterisation factors (Table 2), beyond the electricity and primary-aluminium factors taken directly from IEA [19] and IAI [18] data, are literature-representative secondary values rather than facility- or region-specific ecoinvent entries, reflecting the absence of a licensed commercial LCA database for this study; the resulting impact magnitudes should therefore be interpreted as indicative rather than as regulatory-grade figures. Third, the LCA system boundary is gate-to-gate and excludes capital-equipment manufacture, effluent transport and end-of-life sludge treatment beyond landfill disposal, consistent with comparable published EC-LCA studies [11,12] but nonetheless a simplification of the full product system. None of these limitations affects the paper's central methodological contribution — the demonstration that coupling ML-based multi-objective optimization with LCA can identify operating conditions that simultaneously improve treatment performance and reduce environmental burden — but they should guide the interpretation of the specific numerical results reported here.

CONCLUSION

This study developed and demonstrated an integrated machine-learning (ML) and life-cycle assessment (LCA) decision-support framework for aluminium-electrode

electrocoagulation (EC) treatment of real textile dyeing wastewater. Four ML algorithms were benchmarked as surrogate models of COD removal efficiency, specific energy consumption and electrode consumption from six operating parameters, achieving hold-out test R^2 of 0.986, 0.985 and 0.994, respectively, with permutation feature importance confirming physically consistent, response-specific sensitivity patterns. Coupling the best-performing surrogates with a multi-objective NSGA-II optimization identified an operating condition that increased COD removal from an 84.0% conventional-operation baseline to 90.8% while cutting specific energy consumption by 41.4%, achieved primarily through reduced inter-electrode distance and current density rather than reduced charge dosing. A comparative, ISO 14040/14044-compliant, ReCiPe 2016 midpoint LCA showed that this ML-optimized condition reduced global warming potential by 18.9% (from 7.06 to 5.73 kg CO₂-eq per m³ of treated wastewater — within 8% of an independent literature benchmark), together with reductions of 13.7–18.1% across fossil resource scarcity, terrestrial acidification and freshwater eutrophication, demonstrating that treatment performance and environmental burden can be simultaneously improved rather than traded off. Hotspot analysis further identified aluminium electrode production, rather than electricity consumption, as the dominant and increasingly dominant contributor to every impact category, pointing to electrode-material substitution (e.g., secondary aluminium or iron electrodes) as the priority lever for further environmental improvement.

The principal advantage of the proposed framework is that it restricts computationally and data-intensive life-cycle evaluation to a small set of ML-identified, performance-efficient operating conditions, making joint performance-environment optimization practical at negligible marginal cost once the surrogate models are trained. Its main limitations are its reliance on a literature-calibrated synthetic training corpus and on literature-representative, rather than facility-specific, background characterisation factors, both of which should be replaced with site-specific experimental and inventory data before the framework is used for plant-level investment decisions. Beyond electrocoagulation, the coupled ML-LCA methodology demonstrated here is readily transferable to other water and wastewater treatment unit operations, and more broadly to any process for which both a data-driven performance model and a life-cycle inventory can be assembled, offering a generalisable route toward environmentally informed process optimization.

ACKNOWLEDGMENT

The authors gratefully acknowledge the open-source scientific Python ecosystem (NumPy, SciPy, pandas, scikit-learn and Matplotlib) that underpinned all computations, model training and figure generation reported in this study.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interest associated with this publication.

REFERENCES

1. Mollah, M.Y.A.; Morkovsky, P.; Gomes, J.A.G.; Kesmez, M.; Parga, J.; Cocke, D.L. Fundamentals, present and future perspectives of electrocoagulation. *J. Hazard. Mater.* **2004**, *114*, 199–210. <https://doi.org/10.1016/j.jhazmat.2004.08.009>
2. Chen, G. Electrochemical technologies in wastewater treatment. *Sep. Purif. Technol.* **2004**, *38*, 11–41. <https://doi.org/10.1016/j.seppur.2003.10.006>
3. Kobya, M.; Can, O.T.; Bayramoglu, M. Treatment of textile wastewaters by electrocoagulation using iron and aluminum electrodes. *J. Hazard. Mater.* **2003**, *100*, 163–178. [https://doi.org/10.1016/S0304-3894\(03\)00102-X](https://doi.org/10.1016/S0304-3894(03)00102-X)
4. Emamjomeh, M.M.; Sivakumar, M. Review of pollutants removed by electrocoagulation and electrocoagulation/flotation processes. *J. Environ. Manage.* **2009**, *90*, 1663–1679. <https://doi.org/10.1016/j.jenvman.2008.12.011>
5. Mao, Y.; Zhao, Y.; Cotterill, S. Examining Current and Future Applications of Electrocoagulation in Wastewater Treatment. *Water* **2023**, *15*, 1455. <https://doi.org/10.3390/w15081455>
6. Saini, A.; Shankar, V. Optimizing hybrid electrocoagulation-electrooxidation process using machine learning techniques to predict COD removal efficiency from wastewater. *Water Air Soil Pollut.* **2026**, *237*, 601. <https://doi.org/10.1007/s11270-026-09309-3>
7. ISO 14040:2006. Environmental management—Life cycle assessment—Principles and framework; International Organization for Standardization: Geneva, Switzerland, 2006.
8. ISO 14044:2006. Environmental management—Life cycle assessment—Requirements and guidelines; International Organization for Standardization: Geneva, Switzerland, 2006.
9. Huijbregts, M.A.J.; Steinmann, Z.J.N.; Elshout, P.M.F.; Stam, G.; Verones, F.; Vieira, M.; Zijp, M.; Hollander, A.; van Zelm, R. ReCiPe2016: a harmonised life cycle impact assessment method at midpoint and endpoint level. *Int. J. Life Cycle Assess.* **2017**, *22*, 138–147. <https://doi.org/10.1007/s11367-016-1246-y>
10. Finnveden, G.; Hauschild, M.Z.; Ekvall, T.; Guinée, J.; Heijungs, R.; Hellweg, S.; Koehler, A.; Pennington, D.; Suh, S. Recent developments in Life Cycle Assessment. *J. Environ. Manage.* **2009**, *91*, 1–21. <https://doi.org/10.1016/j.jenvman.2009.06.018>
11. Sedaghat, O.; Bahramifar, N.; Nowrouzi, M.; Younesi, H. Life cycle assessment of industrial wastewater treatment: Evaluating the environmental impact of electrocoagulation technologies. *J. Water Process Eng.* **2025**, *71*, 107257. <https://doi.org/10.1016/j.jwpe.2025.107257>
12. Mačerak, A.L.; Dudukovic, N.; Kiss, F.; Slijepčević, N.; Pešić, V.; Bečelić-Tomin, M.; Kerkez, Đ. Electrocoagulation in treatment of municipal wastewater—life cycle impact assessment. *Chemosphere* **2024**, *355*, 141701. <https://doi.org/10.1016/j.chemosphere.2024.141701>
13. Salla, J.V.E.; Almeida, T.A.; Silva, D.A.L. Integrating machine learning with life cycle assessment: a comprehensive review and guide for predicting environmental impacts. *Int. J. Life Cycle Assess.* **2025**, *30*, 2423–2445. <https://doi.org/10.1007/s11367-025-02437-8>
14. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. <https://doi.org/10.1023/A:1010933404324>
15. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. <https://doi.org/10.1214/aos/1013203451>
16. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
17. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. <https://doi.org/10.1109/4235.996017>

18. International Aluminium Institute. Aluminium Carbon Footprint FAQs. Available online: <https://international-aluminium.org/landing/aluminium-carbon-footprint-faqs/> (accessed on 28 April 2026).
19. International Energy Agency (IEA). Electricity 2025 – Analysis and Forecast to 2027; IEA: Paris, France, 2025. Available online: <https://www.iea.org/reports/electricity-2025> (accessed on 28 April 2026).