Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Explainable Multi-Fidelity Regression for Lithium-Ion Battery Recycling: Maximizing Economic and Environmental Outcomes under U.S. Standards

Shahriar Siddique Ayon
American International University-Bangladesh
Md AKbar Hossain
ahossain@eit.ac.nz
Eastern Institute of Technology
Md. Jisun Abedin Aurnob
American International University-Bangladesh
Majharul Islam Shanto
Troy University
Md. Ebrahim Hossain
American International University-Bangladesh

Research Article

Keywords: Lithium-Ion Battery Recycling, Explainable AI (XAI), Circular Economy, Machine Learning, Sustainability

Posted Date: October 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10872421/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Explainable Multi-Fidelity Regression for Lithium-Ion Battery Recycling: Maximizing Economic and Environmental Outcomes under U.S. Standards

Shahriar Siddique Ayon[1], Md AKbar Hossain[2*], Md. Jisun Abedin Aurnob[1], Majharul Islam Shanto [3], Md. Ebrahim Hossain[1]

[1]Department of Computer Science, American International University-Bangladesh (AIUB), Dhaka, Bangladesh.
[2]School of Computing, Eastern Institute of Technology, Auckland, New Zealand.
[3]Department of Computer Science and Engineering, Troy University, Troy, United States.

*Corresponding author(s). E-mail(s): ahossain@eit.ac.nz;
Contributing authors: shahriarayon63@gmail.com; jisunabedin@gmail.com; mshanto.aiub@gmail.com; mdebrahim4200@gmail.com;

Abstract

The rapid growth of electric vehicles and portable electronics has accelerated end-of-life lithium-ion battery generation, creating pressing environmental and economic challenges. Although recycling is essential for material recovery and circular economy transition, there remains a lack of data-driven tools to jointly predict circularity, processing cost, and profitability under evolving policy frameworks such as U.S. Recycled Content Standards (RCS). Existing studies predominantly rely on static material flow analyses or theoretical projections, failing to capture real-world operational heterogeneity and economic trade-offs. To address this gap, we present an explainable multi-fidelity regression framework that integrates Principal Component Analysis–Information Gain (PCA-IG) feature selection with a novel fidelity-based Bayesian Ridge Regression (BRR) model. The proposed architecture synergizes a computationally efficient low-fidelity predictor with a probabilistic high-fidelity BRR component, enabling rapid yet accurate forecasting across three curated U.S. policy scenario datasets

spanning 2020–2050: material circularity, processing cost, and profitability. Comprehensive benchmarking against Random Forest, Support Vector Regression, and Particle Swarm Optimization–tuned models demonstrates superior generalization, achieving R^2 scores of 0.99 and minimizing mean absolute error to 0.01 (circularity), 0.04 (cost), and 0.11 (profit). Local Interpretable Model-agnostic Explanations (LIME) further disaggregate feature influences, identifying recycling throughput, virgin material value, and direct recycling adoption as primary drivers of economic and environmental performance. These insights are operationalized through a web-based decision-support interface to assist policymakers and industry stakeholders in real-time scenario planning. The proposed framework offers a scalable, interpretable, and deployment-ready solution for optimizing LIB recycling strategies within U.S. regulatory and sustainability contexts.

Keywords: Lithium-Ion Battery Recycling, Explainable AI (XAI), Circular Economy, Machine Learning, Sustainability

1 Introduction

Since the advent of lithium-ion batteries (LIBs) in the 1970s, their demand has surged significantly due to their utilization in electric vehicles (EVs), smartphones, and energy storage systems [1]. As EVs and portable electronics continue to surge in popularity, the demand for LIBs is rising rapidly. Although previously considered an overly ambitious and expensive endeavor, the use of LIBs is steadily increasing the acceptance of EVs [2]. However, this rapid adoption brings a pressing question: what happens when these batteries reach the end of their life? Discarded LIBs pose serious environmental risks, from leaking toxic metals into the soil and water to emitting greenhouse gases when incinerated or dumped in landfills [3, 4].

In response, policymakers and industry leaders are turning to recycling as a sustainable solution. Recycling not only mitigates environmental damage but also recovers critical materials like lithium, cobalt, and nickel—key to building a circular and secure battery supply chain [5, 6]. Encouraged by efforts to combat climate change and global warming, the EVs market in the US achieved a valuation of USD 49.1 billion in 2022. By 2032, it is expected to reach USD 215.7 billion, growing at a compound annual growth rate of 15.5% [7, 8]. Regulatory initiatives such as the U.S. Recycled Content Standards (RCS) and California's battery recycling mandates exemplify the increasing policy emphasis on diverting end-of-life batteries from landfills and promoting systematic resource recovery [9, 10].

However, LIB recycling has significant obstacles in spite of its potential. Conventional methods such as hydrometallurgy and pyrometallurgy frequently have limited scalability, high operating costs, and poor recovery efficiency [11, 12]. However, the industry faces challenges such as low material recovery, high costs, and limited profitability. New, greener recycling methods like chemical leaching and direct recycling are emerging as better alternatives to traditional techniques [13]. At the same time,

machine learning (ML) is transforming the sector by enabling smarter decisions, optimizing recovery processes, and improving real-time performance predictions [14].

While previous studies have projected future material needs and recycling potential for regions like the U.S., European Union, China, and South Korea, these often focus only on the theoretical availability of materials without considering real-world hurdles like costs, collection rates, and processing efficiencies [15]. Advanced ML approaches, including those paired with digital twins and deep learning (DL), help improve battery lifecycle management but still depend heavily on past data and struggle to predict recycling costs or profits in real time [16, 17]. In short, there's still no simple and quick way to reliably predict costs, circularity, and profitability for LIB recycling in the coming years. Since recycling can be expensive or sometimes unprofitable, understanding the key factors that drive profits or reduce costs is crucial. Without this insight, it's challenging for companies and policymakers to make smart decisions in a market and regulatory landscape that's constantly evolving.

This study addresses a critical gap by utilizing three structured datasets to examine how RCS for LIBs in the U.S. may evolve under various future scenarios involving hydrometallurgical, pyrometallurgical, and direct recycling processes. A novel fidelity-based regression model was developed, combining a fast, coarse-grained predictor with a slower, high-precision model. This hybrid approach enables both rapid analysis and highly accurate predictions for key outcomes: material recovery, per-kilogram processing cost, and post-recycling profitability. To uncover the driving factors behind profit and loss, the study applies feature importance analysis and explainable AI (XAI) methods. Additionally, a web-based decision-support tool was built to help stakeholders and policymakers identify which recycling scenarios offer the most cost-effective and sustainable outcomes. The key contributions of this study are summarized as follows:

- This study applies Principal Component Analysis–Information Gain (PCA-IG) based feature selection to filter out less relevant variables, helping focus on the most impactful factors that influence circularity, cost, and profit in LIB recycling.
- The proposed fidelity regression model outperforms traditional methods across all three datasets, delivering the highest accuracy in the LIBs field.
- The proposed model integrates XAI to identify both positively and negatively impactful features, helping to pinpoint which factors drive higher recycling profits or contribute to cost reduction.
- A web-based real-life interface based on our model enables stakeholders to input values, explore future scenarios, and make informed decisions to improve LIB recycling.
- The key feature relationships also reveal which recycling methods and transport options are most efficient in reducing the environmental footprint of LIB recycling.

The remainder of this paper is structured as follows: Section 2 reviews the relevant literature; Section 3 describes the data collection process and methodology; Section 4 presents the experimental results along with a real-world web implementation; Section 5 discusses the findings in detail; and Section 6 concludes the paper and outlines directions for future research.

2 Related Works

Recycling LIBs plays a vital role in boosting profitability, promoting circular resource use, and enhancing cost efficiency within the industry. According to researchers, emerging second-generation recycling technologies that incorporate green chemistry are beginning to replace traditional, high-emission pyrometallurgical methods[18]. The author explored the development of direct recycling techniques aimed at recovering cathode materials of battery-grade quality, which represent a major share of overall battery costs[19]. Existing recycling practices have traditionally focused on extracting high-value metals such as cobalt and copper. However, recent advancements in technology are enabling the recovery of a broader range of materials, thereby promoting the principles of a circular economy[20].

According to experimental characterization, recovery efficiency rates of the cathodic metallic oxide using relatively cheap reagents with low environmental hazardous potential and sub-processes created in accordance with the principles of the Circular Economy are higher than 90% for LCO (with 98% purity in the case of cylindrical LIB cells and 80% for prismatic cells) and 80% for NMC LIB cells (with 85% purity) [21]. Multi-objective goal programming has been applied to optimize recycling strategies, improving operational efficiency while minimizing associated costs [22]. Ongoing technological innovations are also essential for adapting recycling methods to the evolving chemical compositions of LIBs, supporting economically viable processing solutions [23]. Despite these advancements, challenges remain, such as the need for widespread adoption of new technologies and the economic viability of recycling processes.

ML is transforming lithium battery recycling by enhancing predictive capabilities and streamlining recycling operations. Recent studies indicate that incorporating ML methodologies into recycling systems promotes sustainability,increases process efficiency, and improves material recovery outcomes [24]. Researchers have employed algorithms such as AdaBoost, Gaussian Processes, and convolutional neural networks (CNNs) to assess battery health, with CNNs demonstrating superior predictive accuracy [25]. These predictive models, further refined through particle swarm optimization (PSO), have significantly enhanced the estimation of LIB state of health (SOH) [26]. Other studies underscore the role of artificial intelligence (AI)—especially ML and DL—in automating complex recycling tasks like sorting, classification, and disassembly, thereby minimizing human error and increasing operational efficiency. Moreover, AI-enabled systems are equipped with advanced sensing technologies to ensure precise characterization of materials [27]. Although notable progress has been made, persistent challenges—including data integrity, scalability, and adherence to regulatory standards—continue to hinder development.

Numerous regression models have been studied for their capacity to predict battery performance and recycling viability, including Random Forest, Linear Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and PSO regression. While serving as a fundamental benchmark, LR is often employed for comparison; however, its performance declines when applied to the inherently non-linear characteristics of battery behavior [28]. According to other studies, SVM has been applied to

estimate the state of charge (SOC), exhibiting strong accuracy in real-time implementations, though it may be outperformed by ensemble methods like Gaussian Processes [29]. The combination of traditional ML and DL techniques offers a workable solution for real-time LIB health monitoring, illustrating the harmony between interpretability and complexity [30]. Despite their promising ability to forecast battery performance and assess recycling feasibility, regression models continue to face limitations stemming from suboptimal data quality and elevated model complexity.

Accurately predicting the Remaining Useful Life (RUL) of LIBs is essential for ensuring safety, enhancing reliability, and enabling cost-effective maintenance in electric power systems. The authors proposed a hybrid model that integrates the Whale Optimization Algorithm (WOA), Variational Mode Decomposition (VMD), and Long Short-Term Memory (LSTM) networks to tackle the challenge of capacity recovery, which can distort RUL predictions [31]. In a similar vein, introduced an RUL prediction framework utilizing WOA to optimize the parameters of LSTM [32]. This model, validated using the NASA battery dataset, exhibited better convergence performance than alternative metaheuristic algorithms such as PSO and Artificial Bee Colony (ABC). To further mitigate issues related to generalization and sensitivity to noise, developed a model incorporating the Improved Sparrow Search Algorithm (ISSA), VMD, and LSTNet, enhanced with attention mechanisms and skip connections [33].

The integration of ML regression techniques with XAI has become increasingly critical in advancing the efficiency and sustainability of LIB recycling processes. Contemporary research underscores the expanding role of AI in battery waste management, highlighting its capacity to optimize recycling operations, minimize environmental footprint, and promote sustainable practices through enhanced data interpretation and automated decision-making [34]. Recent developments in RUL prediction for EV LIBs have introduced ensemble-based ML models that leverage the complementary strengths of multiple algorithms to improve predictive reliability and robustness [35]. Moreover, AI and ML methodologies are being widely deployed in renewable energy applications, where they bolster forecasting precision, refine energy allocation, and enhance storage system performance through adaptive control and predictive analytics [36].

XAI in ML is being used more and more in LIB recycling to improve operational efficiency and predictive maintenance. Methods such as Shapley values enable the identification of influential features that determine a battery's RUL, thereby supporting more informed recycling and maintenance decisions [37]. Research further indicates that AI-enabled recycling frameworks contribute to the circular economy by facilitating the efficient recovery of valuable materials, reducing environmental harm and resource depletion [38]. Additionally, SHAP-based XAI to dissect the influence of operational features—temperature, cycle index, voltage, and current—on model predictions, highlighting temperature as the dominant factor shaping discharge capacity outcomes [39]. Addressing these technological challenges, therefore, is critical for unlocking the full potential of AI in achieving sustainable recycling systems. Although XAI presents valuable benefits for LIB recycling, it also introduces critical challenges—including the imperative for strong cybersecurity protocols and the delicate balance between predictive accuracy and model transparency.

The lack of data-driven predictive modeling limits previous research, such as that done by Dunn j. et al., [40], from shedding light on the viability of recycled content standards for cobalt, lithium, and nickel in the United States through material flow analysis, techno-economic assessment, and life cycle assessment. Interestingly, they don't use XAI or ML to predict operational efficiency, performance, or economic feasibility. The estimation of battery health, regression modeling for recycling optimization, and the impact of operational variables (such as cycle index, current, and temperature) on results are also ignored in these studies. On the other hand, our study fills these gaps by using a fidelity-based ensemble regression model enhanced with interpretability-enhancing XAI. Using the same dataset as Dunn j., et al.[40], our approach offers a more granular and predictive model that strengthens profitability forecasting, promotes transparency, and informs circular economy strategies through robust analytics.

3 Research Methodology

This section outlines the research approach for LIB, including key study components. The main steps involve data collection, preprocessing, feature engineering, model selection, and hyperparameter tuning. Additionally, XAI helps interpret the model's predictions, and a React-based web application is used to test its performance in real-world scenarios. Figure 1 provides a clear summary of the methodology, visually representing the proposed approach.

3.1 Data Collection and Description

Our research utilizes a well-curated dataset from a publicly available study [40] hosted on Dryad [41]. We use three well-organized datasets, which explored how RCS for LIBs in the U.S. might evolve from 2020 to 2050. The first dataset, OUTPUT_MFA, tracks how materials flow over time—like how much is recycled, how much energy is retired, and how circular the system becomes under different policy and tech scenarios. The second, OUTPUT_Everbatt, breaks down the recycling costs ($/kg) by process type, disassembly levels, and transport methods. The third, OUTPUT_sensitivity, shows how profitability shifts depending on economic and transportation assumptions. These datasets span a 30-year period and include over 800 scenario variations, specifically designed to support policy analysis and advanced research. For easier reference and understanding, we renamed the 'OUTPUT_MFA', 'OUTPUT_Everbatt', and 'OUTPUT_sensitivity' datasets to Circularity dataset, Cost dataset, and Profit dataset, respectively. The Circularity dataset contains 14 features, while the Cost dataset and Profit dataset each have 10 features. Tables 1, 2, and 3 present detailed overviews of the features in the Circularity, Cost, and Profit datasets, respectively.

3.2 Data Preprocessing

For the preprocessing of the three LIB recycling datasets, we ensured that there were no missing values and removed any unnecessary columns. Some categorical variables in each dataset were transformed into numerical values using one-hot encoding

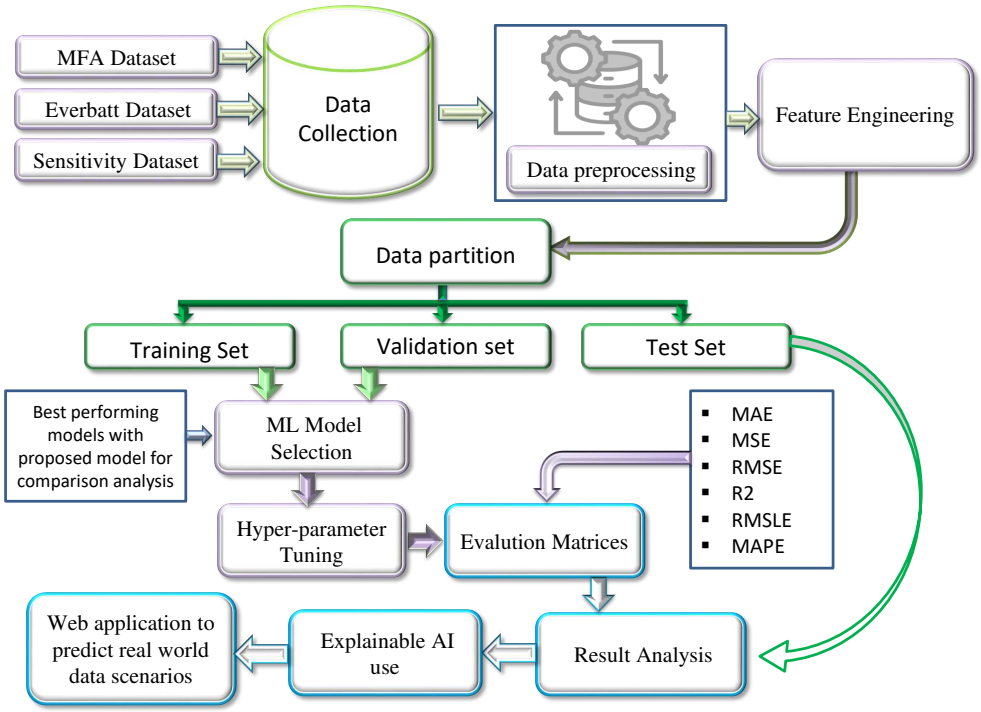


Fig. 1: Proposed Analytical Pipeline for Predicting Circularity, Cost, and Profit in LIB Recycling.

and label encoding techniques. Specifically, in the Circularity dataset, one-hot encoding was applied to the 'recycling.process', 'Sales.Scenario', and 'Cathode.Scenario' columns, while label encoding was used for the 'scenario.percent.repurposed', 'scenario.reuse.lifespan', and 'material' columns. In the Cost dataset, one-hot encoding was performed on the 'Type' and 'Location of recycling' columns, and label encoding was applied to the 'Category' and 'material' columns. For the Profit dataset, one-hot encoding was used for the 'Type', 'Cathode Scenario', 'Location scenario', and 'location of recycling' columns, while the 'scenarios' column was label encoded. After applying the encoding techniques, the Circularity dataset increased from 14 to 18 features, the Cost dataset expanded from 10 to 13 features, and the Profit dataset grew from 10 to 15 features. Next, we applied outlier detection using the Interquartile Range (IQR) technique to enhance the analysis. The IQR is a robust method for detecting outliers in a dataset [42]. It is calculated as the difference between the 75th percentile (Q3) and the 25th percentile (Q1):

$$\text{IQR} = Q_3 - Q_1 \tag{1}$$

Outliers are identified as values outside the bounds:

Table 1: Circularity Dataset Features and Descriptions with Statistics.

Feature Name	**Data Type**	**Description**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Range**
Year	integer	Year of analysis	2035	8.94	2020	2050	30
Scenario percent repurposed	string	Repurposing percentage scenario	-	-	-	-	-
Scenario reuse lifespan	string	Reuse lifespan scenario	-	-	-	-	-
Material	string	Type of material analyzed	-	-	-	-	-
Recycling-process	string	Type of recycling process used (direct, hydro, pyro)	-	-	-	-	-
Sales. Scenario	string	Sales scenario type (SDS, STEPS)	-	-	-	-	-
Cathode. Scenario	string	Cathode chemistry type (LFP, NCX)	-	-	-	-	-
kWh. retired	float	Total retired kWh for the year and scenarios	167,834,669	199,768,033	2,317,375	1,067,348,183	1,065,030,808
kg. retired	float	Total retired battery weight (kg)	131,402,459	303,249,778	192,985	3,419,526,039	3,419,333,054
kg. recycled.-weight	float	Total recovered material (kg)	87,737,116	230,604,040	0	3,077,573,435	3,077,573,435
kWh	float	Total kWh demand for the year and scenarios	609,583,478	487,128,769	19,444,307	1,548,418,148	1,528,973,841
kg.demand	float	Total material demand (kg)	525,365,514	969,200,062	1,194,515	5,632,861,430	5,631,666,915
Circularity (%)	float	Circularity percentage	0.15	0.13	0.00	1.04	1.04
Collection_rate	float	Battery collection rate	0.775	0.11	0.60	0.95	0.35

$$\text{Lower Bound} = Q_1 - 1.5 \times \text{IQR} \tag{2}$$

$$\text{Upper Bound} = Q_3 + 1.5 \times \text{IQR} \tag{3}$$

Data points below the lower bound or above the upper bound are considered outliers. After outlier removal, the Circularity dataset decreased from 187,488 to 146,803

Table 2: Cost Dataset Features and Descriptions with Statistics.

Feature Name	Data Type	Description	Mean	Std. Dev.	Min	Max	Range
Category	string	Defines the category of output number	-	-	-	-	-
Output	float	Cost-related values (e.g., recycling, collection, and disassembly costs)	136.08	441.70	0.00	2078 .06	2078 .06
Type	string	Type of recycling used (direct, hydro, pyro)	-	-	-	-	-
Material	string	Material being recycled	-	-	-	-	-
Year	integer	Year of analysis	2035	8.94	2020	2050	30
Location of retired supply	string	Location where supply is retired	-	-	-	-	-
Recycling Throughput	integer	Recycling plant throughput per year (metric tons)	13,928 .57	14,805 .74	1,000	50,000	49,000
Cathode Scenario	integer	Cathode chemistry scenario (1 = NCX, 2 = LFP)	1.50	0.50	1	2	1
Location scenario	integer	Recycling and transportation mode scenario	2.00	0.82	1	3	2
Location of recycling	string	Recycling location	-	-	-	-	-

datapoints, the Cost dataset from 294,252 to 166,284, and the Profit dataset from 8,586 to 6,222 datapoints. Also, 80% of the data was allocated for training, 20% for testing, and 20% of the training data was further reserved for validation. Data partition details are shown in Table 4.

3.3 Feature Engineering

After completion of data preprocessing and splitting, we proceeded with feature engineering. We employed a Principal Component Analysis combined with Information Gain (PCA-IG) for feature importance analysis. This approach accurately identifies the most influential features by integrating PCA with Information Gain [43]. PCA is applied to reduce dimensionality, and Information Gain is then used to evaluate the relevance of the transformed features to the target variable. PCA-IG rankings were derived exclusively from the training subset to ensure that feature importance estimates remained blind to the validation and test distributions.

Given a dataset X with N samples and d features, PCA tries to develop a transformation matrix W so that,

$$X_{new} = X \cdot W \tag{4}$$

PCA reduces the original data matrix X (size $n \times d$) by transforming it with W (size $d \times k$), where k is the number of principal components. The resulting X_{new} (size $n \times k$) captures the dataset's variance using the most significant components. Equation 4 describes PCA, whereas Information Gain for a feature A is defined as,

Table 3: Profit Dataset Features and Descriptions with Statistics.

Feature Name	**Data Type**	**Description**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Range (Max - Min)**
Type	string	Type of recycling used (Pyro, Hydro, Direct)	-	-	-	-	-
Year	integer	Year of analysis	2035	9	2020	2050	30
Location of retired supply	string	Location where supply is retired	-	-	-	-	-
Recycling Throughput	integer	Recycling throughput per year (metric tons)	-	-	-	-	-
Cathode Scenario	string	Cathode chemistry scenario (LFP, NCX)	-	-	-	-	-
Location scenario	string	Recycling and transportation mode scenario	-	-	-	-	-
Location of recycling	string	Location where recycling occurs	-	-	-	-	-
Scenarios	string	Cost and profit scenarios	-	-	-	-	-
Virgin material value ($)	float	Value of virgin materials recovered from recycling ($/kg)	6	2	3	13	10
Cost to recycle ($)	float	Cost of recycling materials ($/kg)	5	1	4	8	4
Profit ($)	float	Profit or loss from recycling ($/kg)	1	2	-2	8	10

Table 4: Dataset Split Sizes and Target Variables.

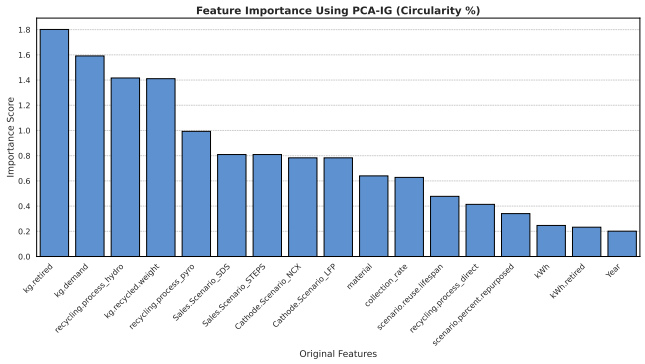
Dataset	**Target Column**	**Training Set Size**	**Validation Set Size**	**Test Set Size**
Circularity Dataset	circularity (%)	93,953	23,489	29,361
Cost Dataset	Output	106,421	26,606	33,257
Profit Dataset	Profit ($)	3,981	996	1,245

$$IG\left(A\right) = H\left(Y\right) - H\left(Y|A\right) \tag{5}$$

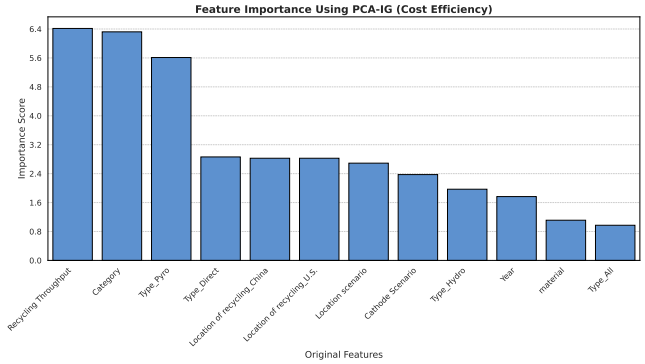
Where $H\left(Y\right)$ is the target variable's entropy and $H\left(Y|A\right)$ is the target variable's conditional entropy given the feature A.

This hybrid approach is effective for large, complex datasets with high dimensionality. Figure 2 presents the feature importance analysis based on PCA-IG for the three datasets.

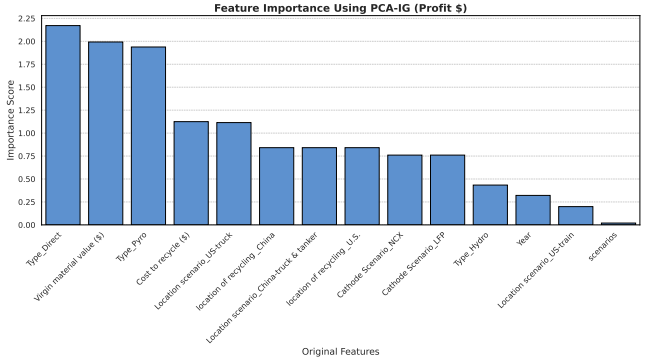
In figure 2a, which shows the Circularity dataset, the importance scores highlight the most influential features for predicting the target variable. The top features include kg.retired (1.801464), kg.demand (1.591443), and recycling.process_hydro (1.416497).



(a) Circularity dataset feature importance



(b) Cost dataset feature importance



(c) Profit dataset feature importance

Fig. 2: Feature importance analysis for different datasets: (a) Circularity, (b) Cost, and (c) Profit.

Other important features are kg.recycled.weight (1.410843), recycling.process_pyro (0.993109), and Sales.Scenario_STEPS (0.809103). On the other hand, features like

Year (0.201031) and kWh.retired (0.232636) had lower importance, showing they were less relevant for prediction in this dataset.

In figure 2b, the feature importance scores for the Cost dataset highlight key predictors, with Recycling Throughput (6.417708), Category (6.322921), and Type_Pyro (5.612451) being the most influential. Other significant features are Type_Direct (2.863190), Location of recycling (China and U.S., both 2.826638), and Location scenario (2.690714), while Year (1.766049), material (1.111236), and Type_All (0.970906) have lower importance.

In figure 2c for the Profit dataset, the most important features are Type_Direct (2.174265), Virgin material value ($1.995465), and Type_Pyro (1.937538). Other significant features include Cost to recycle ($1.127116) and Location scenario_US-truck (1.117374). Features such as Location of recycling_China (0.841526), Location of recycling_U.S. (0.841526), Cathode Scenario_NCX (0.758868), and Type_Hydro (0.431217) show lower importance, while Location scenario_US-train (0.201644) and scenarios (0.020522) have the least impact on the prediction task.

3.4 ML Model Selection

After data preparation, the most effective ML and DL models were chosen for analysis. These models were implemented using popular Python libraries to assess different aspects of the prediction task.

Random Forest Regression (RFR) is an ensemble learning method that builds multiple decision trees and averages their predictions to enhance accuracy and reduce overfitting. Linear Regression (LR) models the relationship between a dependent variable and one or more independent variables by fitting a linear equation to minimize the error. Bayesian Ridge Regression (BRR) applies a probabilistic approach to linear regression, incorporating a prior distribution on the coefficients to prevent overfitting and improve generalization. Support Vector Regression (SVR) uses the principles of Support Vector Machines to find a hyperplane that best fits the data while maintaining a margin of tolerance for errors, making it robust to outliers. K-Nearest Neighbors (KNN) regression predicts a value by averaging the target values of the k closest data points, making it simple yet effective for non-linear relationships. The Stepwise Regression Backpropagation Neural Network (SR-BPNN) model combines LassoCV for feature selection with a backpropagation neural network to improve predictive accuracy in regression tasks by fine-tuning both the features and the neural network parameters. For each dataset, the best-performing machine learning model was further optimized using swarm intelligence algorithms, including PSO, Artificial Bee Colony (ABC), Grey Wolf Optimizer (GWO), and Genetic Algorithm (GA), to enhance predictive performance. Among all optimization techniques, PSO consistently achieved the most significant performance improvement. PSO is a swarm intelligence method where particles collaborate and learn from each other to find the best solution [44]. We then applied our proposed fidelity-based model, customized with each dataset's best-performing ML model, for analysis and comparison. The following section provides a detailed overview of our proposed model.

3.5 Proposed Fidelity Regression Model

The Fidelity-Based Regression Model integrates high-fidelity (HF) and low-fidelity (LF) models to optimize predictive performance. HF models, while more accurate, incur higher computational costs, whereas LF models are computationally efficient but less precise. By combining both models, this approach strikes a balance between accuracy and computational efficiency, resulting in a robust and effective predictive framework.

In this method, the LF model offers a quick initial approximation, which is then refined by the more accurate HF model, ensuring a hierarchical structure that optimizes both speed and precision for superior predictive performance.

3.5.1 General Regression Formulation

In a typical regression problem, the target variable y is predicted as a function of input features X and model parameters θ:

$$y = f(X, \theta) + \epsilon \tag{6}$$

Where y is the target variable, X is the matrix of input features, and $f(X, \theta)$ is the model function. The term ϵ represents the error (residuals).

3.5.2 Low-Fidelity and High-Fidelity Models

The LF Model $f_{LF}(X, \theta_{LF})$, which is computationally efficient but may sacrifice accuracy for speed, and the HF Model $f_{HF}(X, \theta_{HF})$, which provides higher accuracy but requires greater computational resources for predictions. The overall prediction $\hat{y}$ is obtained by combining the outputs of both models:

$$\hat{y} = \alpha f_{LF}(X, \theta_{LF}) + (1 - \alpha) f_{HF}(X, \theta_{HF}) + \epsilon \tag{7}$$

Where α is the weighting factor, controlling the contribution of the LF and HF models to the final prediction, with $0 \leq \alpha \leq 1$. $f_{LF}(X, \theta_{LF})$ is the output of the LF model, and $f_{HF}(X, \theta_{HF})$ is the output of the HF model.

3.5.3 Loss Function and Model Optimization

To combine the model effectively, we minimize a loss function that accounts for both the prediction error and the complexity of the models. The loss function can be defined as:

$$L(\theta_{LF}, \theta_{HF}) = \sum_{i=1}^{n} \Big(y_i - \big(\alpha f_{LF}(X_i, \theta_{LF}) \quad + (1 - \alpha) f_{HF}(X_i, \theta_{HF}) \big) \Big)^2 \tag{8}$$
$$+ \lambda \left(\|\theta_{LF}\|_2^2 + \|\theta_{HF}\|_2^2 \right)$$

In this equation, y_i represents the observed target values for the i-th data point, and X_i denotes the input feature vectors for the same data point. The term λ is a regularization parameter that prevents overfitting by penalizing large model parameters through L2 regularization. Additionally, $\|\theta_{LF}\|_2^2$ and $\|\theta_{HF}\|_2^2$ are the L2 regularization terms applied to the LF and HF model parameters, respectively.

3.5.4 Optimization Method

The Fidelity-Based Regression Model was optimized using standard ML algorithms, with BRR demonstrating the best overall performance. The optimization process employed a probabilistic regularization framework combined with GridSearchCV-based hyperparameter tuning to identify the optimal parameter configuration, ensuring an effective balance between bias and variance for improved generalization capability. Algorithm 1 presents the architecture of the proposed customized fidelity-based regression framework applied to the BRR model.

Algorithm 1 Fidelity Regression with Bayesian Ridge and Grid Search Optimization

1: Generate feature data X over the range $[0, 10]$
2: Compute high-fidelity target: $y_{\text{high}} = \sin(X) + \cos(X)$
3: Compute low-fidelity target: $y_{\text{low}} = \sin(X)$
4: Split the data into training, validation, and test sets
5: Standardize the feature sets using `StandardScaler`
6: Initialize a `BayesianRidge` regression model
7: Define the hyperparameter grid:
8: $\alpha_1, \alpha_2 \in \{10^{-6}, 10^{-3}, 10^{-1}\}$
9: $\lambda_1, \lambda_2 \in \{10^{-6}, 10^{-3}, 10^{-1}\}$
10: Perform `GridSearchCV` with 5-fold cross-validation to identify the best hyperparameters
11: Fit the model to the training data
12: Retrieve the best estimator from `grid_search.best_estimator_`
13: Predict the target values for the test set using the optimized model
14: Compute and output evaluation metrics (e.g., MAE, RMSE, R^2)

3.6 Hyperparameter Tuning

Once the models were selected, the next crucial step was fine-tuning the hyperparameters to boost performance. Table 5 provides an overview of the hyperparameters used across all models. The proposed Fidelity-Based BRR utilizes $alpha_1$, $alpha_2$, $lambda_1$, and $lambda_2$ values of $[1e-6, 1e-3, 1e-1]$, with hyperparameters tuned via GridSearchCV. Across the three datasets, PSO-optimized and Fidelity-based BRR models were applied based on their superior performance in each dataset scenario.

3.7 Evaluation Matrices

Each model's performance was assessed using six key regression metrics. These include the Mean Absolute Error (MAE), which measures the average absolute difference between predicted and actual values; Mean Squared Error (MSE), which computes the average squared difference; Root Mean Squared Error (RMSE), providing the square root of MSE for interpretability; R-squared (R^2), indicating the proportion of variance explained by the model; Root Mean Squared Logarithmic Error (RMSLE), assessing the ratio of predicted to actual values; and Mean Absolute Percentage Error (MAPE),

Table 5: Model Hyperparameter Configuration.

Model Name	**Hyper-parameters Setting for All Models**
RFR	$n_{\text{estimators}} = 100$, $max_{\text{depth}} = \text{None}$, $random_{\text{state}} = 42$
LR	$fit_{\text{intercept}} = \text{True}$, $normalize = \text{False}$, $copy_{\text{X}} = \text{True}$
BRR	$\alpha_1 = 1e-6$, $\alpha_2 = 1e-6$, $\lambda_1 = 1e-6$, $\lambda_2 = 1e-6$
SVR	$kernel = \text{rbf}$, $C = 1.0$, $epsilon = 0.1$
KNN	$n_{\text{neighbors}} = 5$, $weights = \text{uniform}$
SR-BPNN Regressor	cv = 5, max_iter = 10000, $random_{\text{state}}$ = 42, hidden_layer_sizes = (100, 50, 25), activation = relu, solver = adam, learning_rate_init = 0.001
PSO-Based BRR	$\alpha_1, \alpha_2 \in \{1e-6, 1e-3, 1e-1\}$ (search space for PSO)
Proposed Fidelity-Based BRR	$\alpha_1, \alpha_2, \lambda_1, \lambda_2 \in \{1e-6, 1e-3, 1e-1\}$ (for GridSearchCV)

representing the average percentage error between predictions and actual outcomes. The optimal performance is characterized by lower values for MAE, MSE, RMSE, and RMSLE, while higher values for (R^2) and lower percentages for MAPE indicate better model performance. The corresponding equations for these metrics are provided below.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{9}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{10}$$

$$RMSE = \sqrt{MSE} \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{12}$$

$$RMSLE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\log(1+y_i) - \log(1+\hat{y}_i))^2} \tag{13}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{14}$$

Where, N represents the total number of samples, y_i denotes the actual values, $\hat{y}_i$ signifies the predicted values, and $\bar{y}_i$ indicates the mean of the actual values.

3.8 Explainable Artificial Intelligence (XAI)

XAI helps users understand what influences a ML model's results. While there are several tools like SHAP, Shapash, and LIME, we chose LIME because it gives clear and simple explanations for each prediction [45, 46]. LIME makes complex models easier to understand by using simpler models to explain their behavior, helping to clarify why a particular prediction was made. The explanation model for LIME is as follows:

$$\hat{g}(x) = argmin_{g \epsilon G} L(f, g, \pi_{x'}) + \Omega(g) \tag{15}$$

In equation 15, the explanation model $\hat{g}(x)$ is designed to locally approximate the behavior of complex models by using a set G of simpler models, typically linear ones. It employs a loss function $L(f, g, \pi_{x'})$ to balance the model's accuracy with its simplicity and interpretability.

3.9 Web-based Evaluation Framework

To facilitate user interaction with the developed prediction models and evaluate their usability in practical settings, a web-based evaluation framework was implemented. The system enables real-time prediction of circularity, cost, and profitability of LIB recycling processes under different scenarios and parameters, aligning with the research focus on U.S. RCS. Figure 3 depicts the structured workflow stages in the web application.

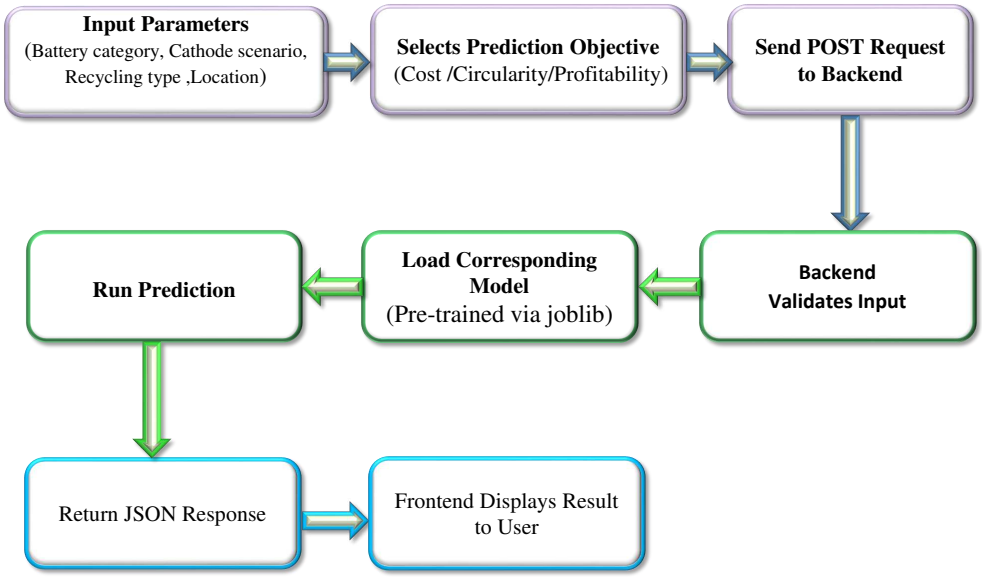


Fig. 3: Sequential Stages in the Web Application Workflow.

The backend of the application was developed using Flask, a lightweight Python web framework, which serves three pre-trained fidelity regression models—one each for circularity, cost, and profitability. These models were serialized using the joblib module and are loaded into memory upon server initialization for efficient response handling. The system exposes a unified RESTful API route (/predict/¡model_name¿)

to receive POST requests containing user-defined feature values. The route performs validation and invokes the appropriate model to return prediction results in JSON format. Cross-Origin Resource Sharing (CORS) is enabled to allow communication between the frontend and backend components.

On the frontend, a responsive user interface was developed using React.js, allowing users to input relevant parameters such as battery category, cathode scenario, recycling type (Hydro, Pyro, or Direct), and location configurations for the U.S. and China. The interface provides toggleable options to select the desired prediction objective—Circularity, Cost, or Profitability—before invoking the backend to retrieve and display the prediction result.

4 Experimental Results and Analysis

This section compares the prediction performance of various ML models across the Circularity, Cost, and Profit datasets. Training, validation, and test performance graphs are analyzed to evaluate the proposed model, along with correlation strategies to identify key feature relationships. Additionally, LIME explanations provide an in-depth analysis of feature importance. The novelty of the proposed fidelity regression model is assessed across different domain datasets, and a web application is developed for real-world analysis of LIB recycling, including circularity, cost, and profit evaluation. The results highlight the superior performance of the proposed model, as detailed below.

4.1 Experimental Setup

In this study, we used Google Colab as our main platform for building, training, and testing ML models. The free version of Colab gave us access to a high-performance environment, including an NVIDIA Tesla T4 GPU, 12 GB of RAM, and a virtual CPU—all without the need for expensive local hardware. We wrote all our code in Python 3.8, and relied on Scikit-learn (v1.0.2) for key ML tasks like feature selection, regression, and performance evaluation[47, 48].

Model selection and hyperparameter tuning were conducted using 5-fold cross-validation on the training set only. The held-out test set was used solely for final evaluation and was never accessed during model development or optimizer selection. We've documented our package versions and hardware details to make our experiments easy to reproduce.

4.2 Performance of ML Models without Feature Selection

The prediction results for the regression analysis of LIB recycling, focusing on circularity, cost, and profit predictions, as shown in Tables 6, 7, and 8. These results are based on the test data.

Table 6 presents the performance of various ML models tested on the circularity dataset without feature selection. The BRR performed well with MAE 0.0123 and R^2 0.9853. LR and RFR showed weaker performance, both with MAE 0.0540 and R^2 around 0.65. SVR improved slightly with RMSE 0.0681 and R^2 0.7131. KNN performed better with MAE 0.0128 and R^2 0.9672 but had higher MAPE. SR-BPNN

gave strong results with low error and R^2 0.98. PSO-based BRR reduced MAPE to 5.22. The best outcome came from the proposed Fidelity-based BRR with the highest R^2 of 0.99 and low overall errors.

Table 6: Circularity Dataset Test Results Without Feature Selection.

Model	MAE	MSE	RMSE	R^2	RMSLE	MAPE (%)
BRR	0.0123	0.0002	0.0154	0.9853	0.0127	31.19
LR	0.0540	0.0056	0.0751	0.6508	0.0629	179.79
RFR	0.0540	0.0057	0.0752	0.6498	0.0629	180.25
SVR	0.0551	0.0046	0.0681	0.7131	0.0560	166.63
KNN	0.0128	0.0005	0.0230	0.9672	0.0194	39.26
SR-BPNN Regression	0.0100	0.0000	0.0100	0.9800	0.0100	17.39
PSO-based BRR	0.0100	0.0000	0.0200	0.9800	0.0100	5.22
Proposed Fidelity - based BRR	0.0100	0.0000	0.0100	0.9900	0.0100	25.90

Table 7 highlights how different ML models performed on the cost dataset without feature selection. The BRR delivered strong results with MAE 0.11 and R^2 0.98. In contrast, LR and RFR performed poorly, with high MAEs of 0.67 and low R^2 scores of 0.10. SVR struggled even more, with negative R^2 and a MAPE over 1000%. KNN and SR-BPNN showed moderate results, with better error rates but still high MAPE. PSO-based BRR improved accuracy further, but the best results came from the proposed Fidelity-based BRR, which had the lowest errors and the best overall consistency, including a MAPE of 135.22 and R^2 of 0.98.

Table 7: Cost Dataset Test Results without Feature Selection.

Model	MAE	MSE	RMSE	R^2	RMSLE	MAPE (%)
BRR	0.11	0.02	0.14	0.98	0.08	497.82
LR	0.67	0.85	0.92	0.10	0.45	4876.80
RFR	0.67	0.85	0.92	0.10	0.45	4876.63
SVR	0.55	1.13	1.06	-0.20	0.49	1150.50
KNN	0.26	0.24	0.49	0.74	0.22	746.35
SR-BPNN Regression	0.25	0.27	0.52	0.71	0.20	178.69
PSO-based BRR	0.12	0.07	0.11	0.98	0.09	152.45
Proposed Fidelity-based BRR	0.08	0.05	0.05	0.98	0.06	135.22

Table 8 shows that most models performed reasonably well on the profit dataset. BRR stood out with solid accuracy (MAE 0.1631, R^2 0.9677), while SVR and KNN showed higher error rates and less consistency. SR-BPNN and RFR gave balanced results, but PSO-based BRR brought notable improvements (MAE 0.15, MAPE 56.08%). The best overall performance came from the Proposed Fidelity-based BRR, with the lowest error (MAE 0.13) and highest R^2 of 0.98.

Table 8: Profit Dataset Test Results without Feature Selection.

Model	MAE	MSE	RMSE	R^2	RMSLE	MAPE (%)
RFR	0.26	0.10	0.32	0.92	0.13	68.37
LR	0.2095	0.0780	0.2792	0.9376	0.1151	65.57
BRR	0.1631	0.0403	0.2009	0.9677	0.0826	59.13
SVR	0.2407	0.0945	0.3074	0.9244	0.1283	85.28
KNN	0.2999	0.1460	0.3820	0.8832	0.1594	102.29
SR-BPNN Regression	0.2314	0.0988	0.3143	0.9210	0.1237	74.86
PSO-based BRR	0.15	0.04	0.20	0.97	0.08	56.08
Proposed Fidelity-based BRR	0.13	0.03	0.17	0.98	0.08	54.79

4.3 Performance of ML Models with Feature Selection

After applying feature selection, the regression analysis results for LIB recycling show improved performance in circularity, cost, and profit predictions across all datasets, as reflected in Tables 9,10, and11.

Table 9 presents the circularity dataset results after feature selection. The BRR improved with lower MAE 0.0079 and high R^2 0.9895. LR and RFR remained limited, both showing MAE 0.0439 and R^2 0.6516. SVR performed slightly better than LR and RFR, with MAE 0.0422 and R^2 0.7159. KNN showed a strong improvement, reaching MAE 0.0049 and R^2 0.9662. SR-BPNN continued to perform well with low MAE 0.01 and R^2 0.98. PSO-based BRR further enhanced accuracy with MAE 0.01 and MAPE 5.90%. The best overall result came from the Proposed Fidelity-based BRR, which achieved the lowest MAPE of 4.15% and the highest R^2 of 0.99.

Table 9: Circularity Dataset Test Results with Feature Selection.

Model	MAE	MSE	RMSE	R^2	RMSLE	MAPE (%)
BRR	0.0079	0.0001	0.0100	0.9895	0.0084	20.39
LR	0.0439	0.0033	0.0577	0.6516	0.0494	135.17
RFR	0.0439	0.0033	0.0577	0.6516	0.0494	135.18
SVR	0.0422	0.0027	0.0521	0.7159	0.0438	109.43
KNN	0.0049	0.0003	0.0180	0.9662	0.0160	20.22
SR-BPNN Regression	0.01	0.00	0.01	0.98	0.01	15.02
PSO-based BRR	0.01	0.00	0.01	0.99	0.01	5.90
Proposed Fidelity-based BRR	0.01	0.00	0.01	0.99	0.01	4.15

Table 10 shows that after feature selection, the proposed Fidelity-based BRR gave the best results with the lowest MAE of 0.04, RMSE of 0.05, and MAPE of 15.85%, along with a strong R^2 of 0.99. BRR and PSO-based BRR also performed very well. In contrast, LR, RFR, and SVR showed poor accuracy with high error and extremely large MAPE values. KNN and SR-BPNN showed moderate improvements but still trailed behind the top-performing models.

Table 10: Cost Dataset Test Results with Feature Selection.

Model	MAE	MSE	RMSE	R^2	RMSLE	MAPE (%)
BRR	0.05	0.01	0.07	0.99	0.06	15.85
LR	0.6705	0.8511	0.9225	0.10	0.4469	1846.39
RFR	0.6704	0.8511	0.9226	0.10	0.4469	1846.34
SVR	0.53	1.09	1.12	-0.18	0.41	1048.80
KNN	0.2565	0.2441	0.4940	0.7411	0.2243	444.04
SR-BPNN Regression	0.22	0.24	0.49	0.69	0.21	157.26
PSO-based BRR	0.05	0.00	0.07	0.99	0.06	15.85
Proposed Fidelity-based BRR	0.04	0.00	0.05	0.99	0.03	15.85

Table 11 summarizes the profit prediction results after feature selection. The proposed Fidelity-based BRR outperformed all models with the lowest MAE of 0.11, RMSE of 0.12, RMSLE of 0.05, and the highest R^2 of 0.99, along with the lowest MAPE of 48.53%. PSO-based BRR also performed well with MAE 0.14 and R^2 0.98. BRR and SVR showed solid performance (MAE 0.1472 and 0.1540, R^2 0.9732 and 0.9697, respectively), while LR, RFR, KNN, and SR-BPNN lagged slightly behind, with higher MAEs and MAPEs. Overall, preprocessing significantly boosted the accuracy of the top-performing models, especially the proposed one.

Table 11: Profit Dataset Test Results with Feature Selection.

Model	MAE	MSE	RMSE	R^2	RMSLE	MAPE (%)
RFR	0.22	0.08	0.29	0.93	0.12	65.71
LR	0.1871	0.0537	0.2318	0.9570	0.1073	81.92
BRR	0.1472	0.0335	0.1831	0.9732	0.0766	62.06
SVR	0.1540	0.0379	0.1946	0.9697	0.0811	72.08
KNN	0.2121	0.0716	0.2676	0.9427	0.1190	79.49
SR-BPNN Regression	0.1941	0.0562	0.2371	0.9550	0.1112	98.43
PSO-based BRR	0.14	0.03	0.17	0.98	0.08	56.90
Proposed Fidelity-based BRR	0.11	0.01	0.12	0.99	0.05	48.53

Figure 4 illustrates how accurately our proposed model predicts outcomes by comparing actual vs. predicted values across the training, testing, and validation sets for all three preprocessed datasets. Figure 4a (Circularity dataset) compares actual and predicted circularity (%) values, with color-coded lines for different data splits. The strong alignment across subsets demonstrates the model's high predictive accuracy and generalization ability. Figure 4b (Cost dataset) illustrates the predictive performance of the Fidelity based BRR model for cost efficiency across train, validation, and test sets. The close alignment between actual and predicted values, especially in validation and test sets, demonstrates the model's ability to capture cost efficiency trends while maintaining generalization. Notable variance in peaks reflects complex cost fluctuations, which the model tracks effectively, showcasing its robustness and adaptability. Figure 4c (Profit dataset) illustrates the Fidelity based BRR model's

accurate profit predictions across train, validation, and test sets, showcasing strong generalization and robustness in capturing financial variations.

4.4 Using LIME to Interpret Proposed Model's Predictions

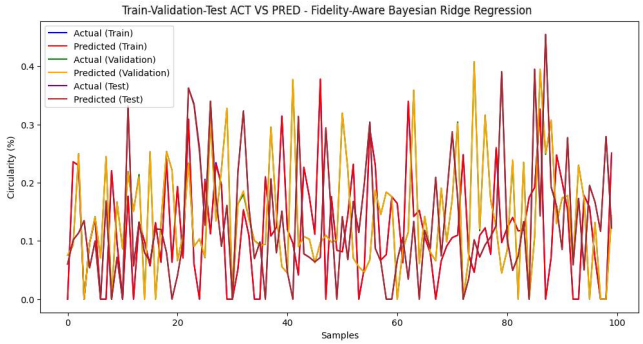
In this section, we are diving into feature analysis with XAI techniques. LIME does a great job of highlighting both positive and negative feature impacts. Figures 5a, 5b & 5c shows how LIME's tabular explainer highlights key features for predictions across three datasets, with prediction probabilities on the left, feature weights in the middle, and feature values table on the right. In Figure 5a, the LIME explainer for the Circcularity dataset shows that the Model predicts a value of 0.06 on a scale from 0.00 to 0.45. The plot highlights the key features influencing this prediction, with both positive (orange) and negative (blue) contributions. The top positive features are "recycling.process_pyro" with 1.39, "material" with 1.43, and "scenario.reuse.lifespan" with 1.22, which drive the prediction upward. On the negative side, "scenario.percent.repurposed" with -1.18, "recycling.process_direct" with -0.70, and "recycling.process_hydro" with -0.70 significantly reduce the predicted value.

In Figure 5b, the LIME plot for the Cost dataset highlights the ML model's prediction of 0.00 on a scale from 0.00 to 3.33. Positive factors like "Cathode Scenario" (1.00) and "Type_Hydro" (1.62) push the prediction higher. Meanwhile, negative factors like "Location scenario" (-1.22), "Type_Direct" (-0.57), and "Category" (-0.48) pull the prediction down. Smaller influences come from "Recycling Throughput" and "Location of recycling_China," which slightly lower the result.

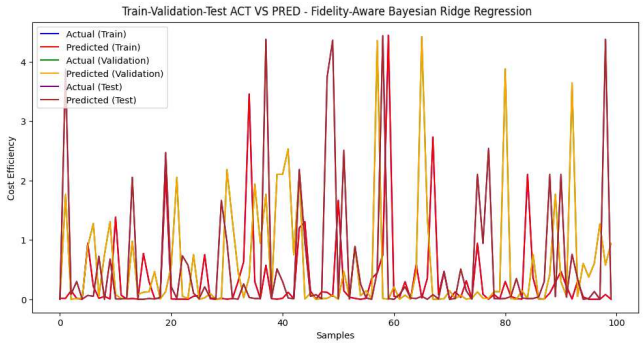
In Figure 5c, for the Profit Dataset, the ML model predicts a target value of 1.90, within a range of -2.55 to 4.84. Key features influencing the prediction include "Virgin material value ($)" (1.80), "Cost to recycle ($)" (1.34), and "Location scenario_China-truck & tanker" (1.17), with higher values boosting the prediction. In contrast, "Type_Hydro" (-0.85), "Location scenario_US-train" (-0.64), and "Location of recycling_U.S." (-1.17) negatively affect the prediction.

Figure 6 shows the LIME feature importance plot for predictions across three datasets using the proposed Model. Green indicates positive, while red shows negative impactful features. In Figure 6a, the LIME plot for the Circularity dataset shows that 'kg.recycled.weight', 'material', and 'recycling.process_direct' are the key positive features that boost the prediction. Specifically, when 'kg.recycled.weight' ranges from -0.46 to 0.15, 'material' is above 0.90, and 'recycling.process_direct' is -0.70 or lower, the prediction increases significantly. On the other hand, 'Year', 'kWh.retired', and 'kg.demand' are the top negative features. When 'Year' falls between -0.80 and -0.07, and 'kWh.retired' ranges from -0.73 to -0.42, the prediction slightly decreases as these values increase. Additionally, other features have a minor influence on the prediction but with very limited impact.

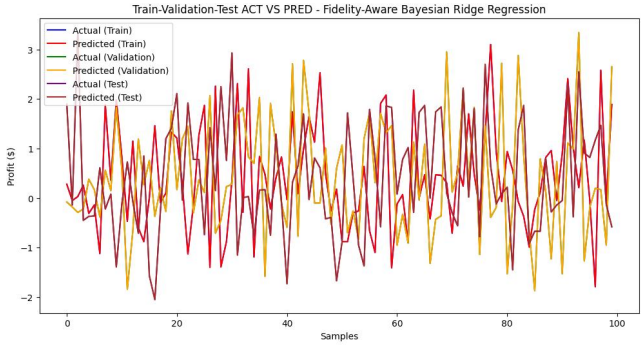
In Figure 6b, the Cost dataset shows the feature importance with the x-axis ranging from -0.4 to 0.2. The most negative features are 'material' (-0.4) and 'Type_Direct' (-0.2), which significantly lower the prediction when 'material' is between 0.37 and 0.85, and 'Type_Direct' is -0.57 or lower. On the positive side, 'Type_Pyro' (0.14), 'Category' (0.11), and 'Type_All' (0.5) are the top contributors. The prediction increases



(a) Circularity Dataset: Actual Vs. Predicted Values.



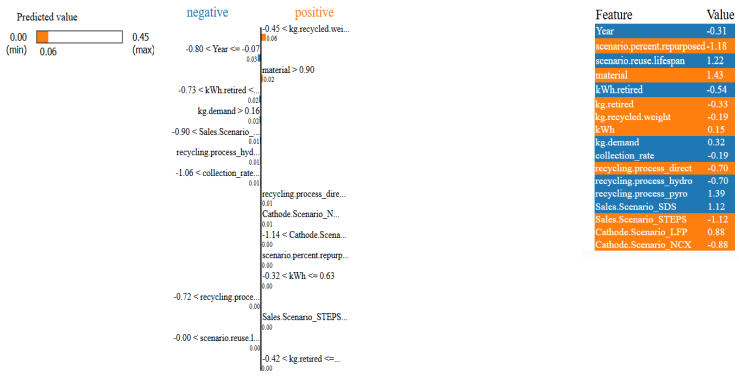
(b) Cost Dataset: Actual Vs. Predicted Values.



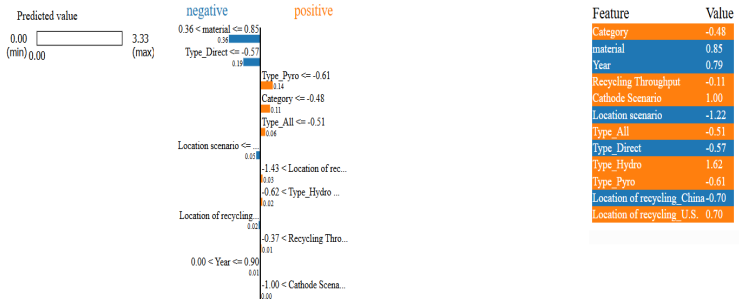
(c) Profit Dataset: Actual Vs. Predicted Values.

Fig. 4: Actual vs. Predicted Performance of the Proposed Fidelity-Based BRR Model.

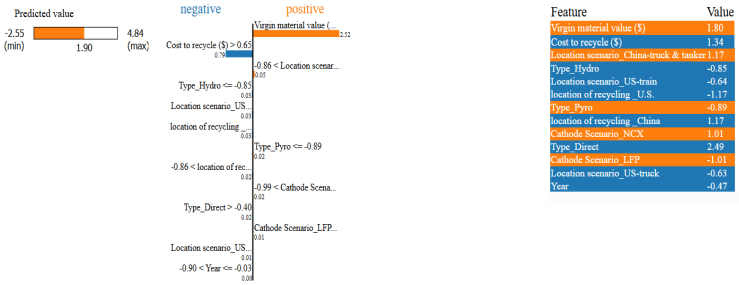
notably when 'Type_Pyro' is -0.61 or lower and 'Category' is -0.48 or lower, while other features have minimal impact.



(a) Circularity Dataset LIME Tabular Plot.



(b) Cost Dataset LIME Tabular Plot.

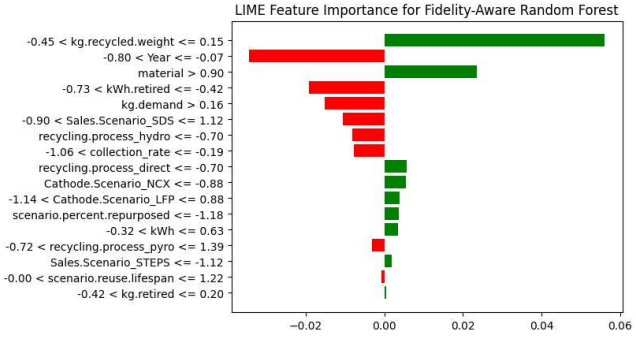


(c) Profit Dataset LIME Tabular Plot.

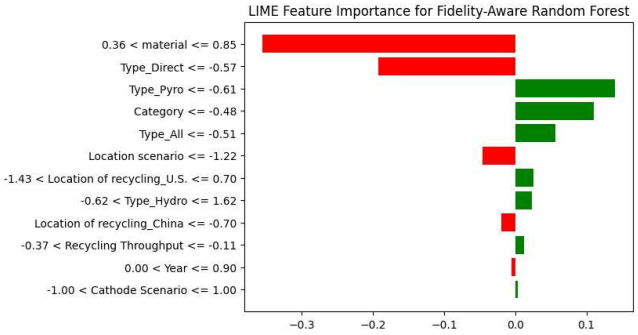
Fig. 5: LIME Tabular Explanations for Circularity, Cost, and Profit Datasets.

In Figure 6c, the Profit dataset highlights that 'Virgin material value ($)' has the strongest positive impact (2.5), while 'Cost to recycle ($)' has the highest negative impact (-0.6). The prediction increases significantly when 'Virgin material value ($)'

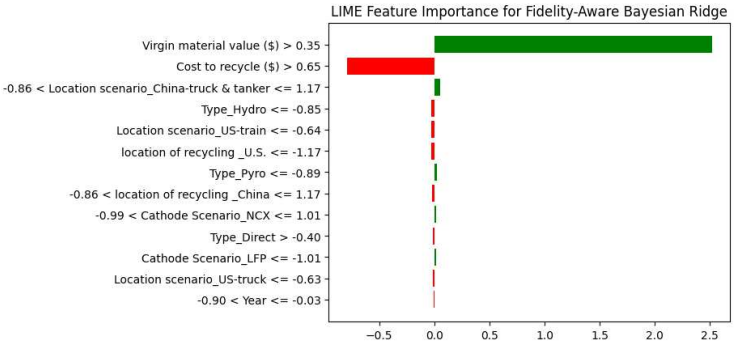
is below 0.35, while it drops when 'Cost to recycle ($)' is below 0.65. Other features also play a role, but their impact is relatively minor.



(a) Circularity Dataset Feature Importance Plot.



(b) Cost Dataset Feature Importance Plot.



(c) Profit Dataset Feature Importance Plot.

Fig. 6: LIME Feature Importance Analysis for Circularity, Cost, and Profit Datasets.

4.5 Key Feature Relationships and Insights

In the Key Feature Relations and Insights Outcomes section, we explored how top features from the three LIB recycling datasets relate to our target variables. Figures 7, 8, and 9 reveal new insights into the recycling process. Figure 7 shows the relationship between retired energy (kWh), recycled material (kg), and circularity (%) across SDS and STEPS sales scenarios in the Circularity Battery Recycling Dataset. The interactive bubble scatter plot, created with Plotly Express, uses log scales to reveal patterns, with clusters around values like 10^6 kWh and 10^7 kg. Bubble size represents circularity, and color indicates the recycling process (blue for direct, red for hydro, green for pyro). The SDS scenario shows better overall material recovery, higher retired energy levels, and improved circularity, as reflected by larger bubble sizes, compared to the STEPS scenario, which has lower recovery and smaller circularity percentages. In both scenarios, hydro and pyro recycling are the leading methods, with pyro recycling taking a slightly bigger role in SDS, suggesting it scales better with sustainability-focused policies. Direct recycling remains limited in both cases, indicating it's not yet widely adopted or scalable. Overall, SDS stands out as the more effective approach for enhancing battery recycling efficiency.

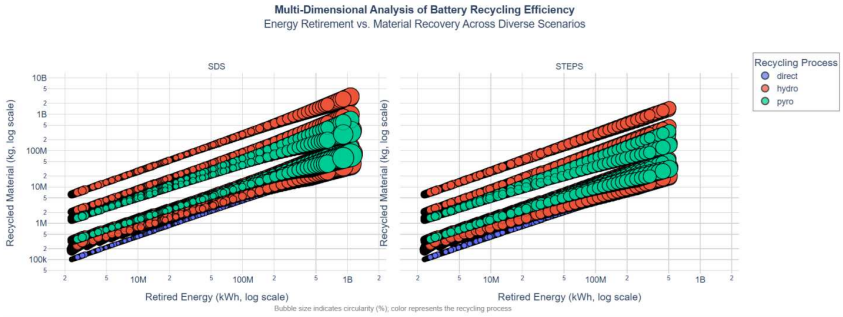


Fig. 7: Retired Energy vs. Recycled Material.

Figure 8 highlights key insights from the Cost Dataset with two graphs. In Figure 8a, a bar chart shows how much material is recovered through Direct, Hydro, and Pyro recycling methods. The y-axis represents either the recovered value ($/kg) or the material quantity, while the x-axis categorizes the recycling methods. Different colors indicate recovered materials like Graphite, Copper, Aluminum, NMC types, and metal ions (Co^2+, Mn^2+, Ni^2+, etc.). The graph shows that Graphite ($0.196/kg) and Copper ($0.162/kg) are the most valuable recovered materials, best preserved by Direct and Hydro recycling. Pyro recycling is less effective, recovering Copper, Co^2+, and Ni^2+ but losing Graphite and other materials. Hydro recycling stands out as the most versatile, recovering metal ions (Co^2+, Mn^2+, Ni^2+), while Direct recycling excels at preserving high-value materials like Graphite, Copper, and NCA ($0.159/kg) but is less effective for metal ions. Due to its high-temperature processing, Pyro recycling has lower material recovery efficiency and should only be considered

when other methods are not feasible. For the best balance of value and sustainability, combining Direct and Hydro recycling is recommended.

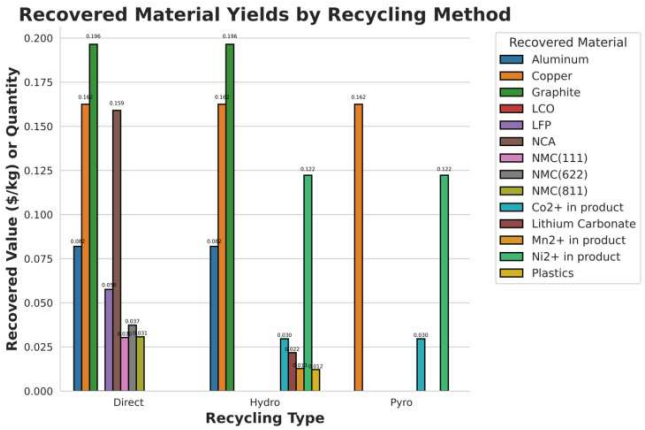
Figure 8b presents a bar chart comparing the cost breakdown of different battery recycling methods (All, Direct, Hydro, and Pyro) in $/kg, with each cost component represented by a different color in the legend. The cost analysis shows that Direct recycling is the most expensive ($12.41/kg), mainly due to high material processing costs ($7.95/kg). Hydro recycling is the most cost-effective ($8.38/kg) with lower process costs ($2.77/kg), making it the cheapest option. Pyro recycling falls in between ($9.38/kg), with higher costs than Hydro but lower than Direct. The "All" method, which averages costs across all methods, is the least expensive ($5.00/kg) since it only includes collection and disassembly. Overall, Hydro is the most cost-effective, while Direct offers higher material recovery at a higher cost, and Pyro strikes a balance between cost and efficiency.

Figure 9a & 9b shows insights from the Profit Dataset. In Figure 9a, it illustrates the profit trends for three transportation scenarios: US-truck, US-train, and China-truck & tanker, from 2020 to 2050. The China-truck & tanker scenario starts with the highest profit at $2.0, gradually declining but staying above the average profit line of $0.69. Meanwhile, US-truck and US-train profits begin at $0.5 and $0.9, respectively, but steadily decrease, falling below the average profit line after 2030 and turning negative by 2050. The widening confidence intervals show increasing uncertainty, likely due to factors like rising costs and market shifts. This analysis highlights China's advantage in transportation efficiency, while the US faces growing challenges in maintaining profitability, underlining the need for strategic innovation and policy changes.

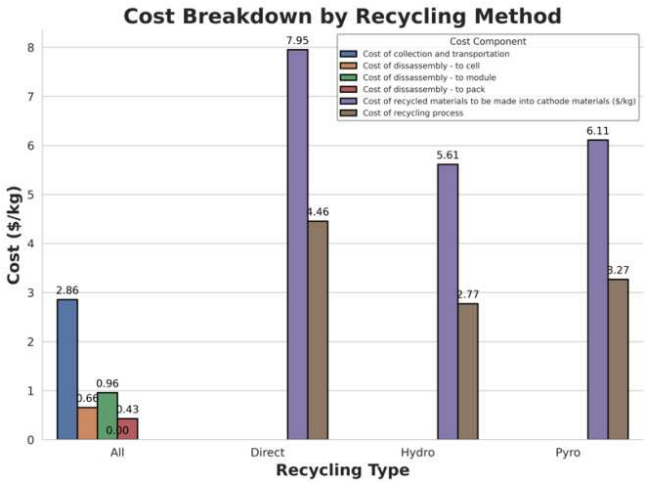
Figure 9b presents a graph showing the relationship between profit and carbon footprint in lithium-ion battery recycling. The carbon footprint values for three recycling methods—Pyro, Hydro, and Direct—are calculated by combining base emissions with transportation emissions. Base emissions are based on literature values for carbon footprints per kWh: 5.11 kg CO_2-eq/kWh for Pyro, 2.68 kg CO_2-eq/kWh for Hydro, and 1.0 kg CO_2-eq/kWh for Direct[49–51]. These values are then converted to per kg of battery using an energy density of 0.25 kWh/kg and a cathode mass fraction of 0.45[52, 53]. This results in carbon footprints of 2.84 kg CO_2-eq/kg for Pyro, 1.49 kg CO_2-eq/kg for Hydro, and 0.56 kg CO_2-eq/kg for Direct.

Transportation emissions depend on location and mode: US-truck (0.15 kg CO_2-eq/kg, from 150 g CO_2/ton-km over 1000 km), US-train (0.03 kg CO_2-eq/kg, from 30 g CO_2/ton-km over 1000 km), and China-truck & tanker (0.225 kg CO_2-eq/kg, from combined truck and sea factors over 6000 km)[54–56]. Total carbon footprint is the sum of base and transportation emissions for each scenario. Profit data is sourced directly from the dataset's 'Profit ($)' column.

In Figure 9b, the scatter plot shows that Direct recycling in the China-truck & tanker scenario stands out with the highest profit of 2.41 and a relatively low carbon footprint of 0.785 kg CO_2-eq/kg. Hydro shows moderate profits, with China-truck & tanker still being the most profitable (1.44), but with a much higher carbon footprint (1.715 kg CO_2-eq/kg). Pyro recycling, on the other hand, has the lowest profits across all scenarios, especially in the US-train (-0.03) and US-truck (-0.28), and its carbon footprint is the highest, reaching 3.065 kg CO_2-eq/kg in the China-truck



(a) Recovered Material Yields by Recycling Method.



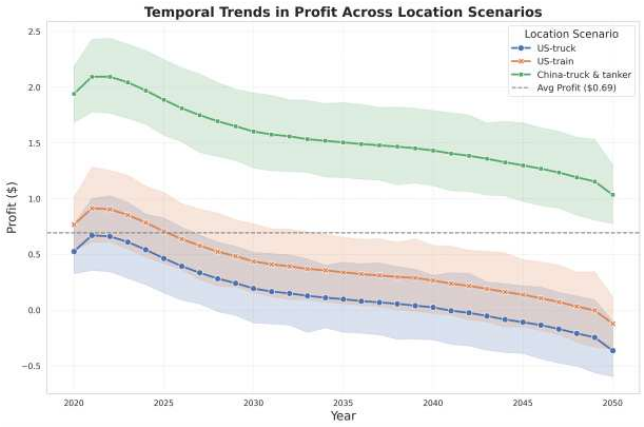
(b) Cost Breakdown by Recycling Method.

Fig. 8: Comparative Analysis of Recycling Methods: Material Yields and Cost Breakdown.

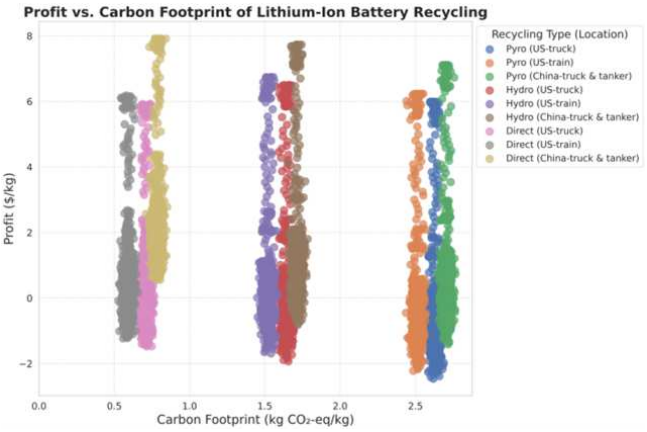
& tanker scenario. Overall, Direct recycling provides the best balance of profit and environmental impact, while Pyro is the least profitable and most carbon-intensive option.

4.6 Predictive Web Interface for LIB Recycling

The web interface enables users to perform simulations and generate predictions across various input scenarios, as demonstrated in Figure 10. It consists of three distinct modules—Circularity, Cost, and Profitability—each with its own set of input fields.



(a) Temporal Trends in Profit Across Location Scenarios.



(b) Profit Vs. Carbon Footprint of LIB Recycling.

Fig. 9: Evaluating LIB Recycling Profitability Over Time and Environmental Impact.

Users can toggle between these modules, enter scenario-specific data, and retrieve a prediction result based on their inputs.

Figure 10a displays the Circularity analysis mode of the LIB Recycling Prediction tool. With the year set to 2025, inputs include 38% repurposed batteries, a 72-year reuse lifespan, and Material 620. Retired battery values are 223 kWh and 41 kg, with 87 kg marked as recycled, and a low 20% collection rate. Recycling scores favor cleaner methods: Direct (89), Hydro (83), and Pyro (3). Sales values are SDS (3), STEPS (599), and cathode scores LFP (48), NCK (43). The tool outputs a low Circularity Score of 0.0309, reflecting limited material recovery due to low collection and unrealistic reuse assumptions.

Figure 10b presents the Cost analysis mode of the LIB Recycling Prediction tool. Set to the year 2025, inputs include Material type 4, a value of 73, Recycling Throughput of 30, Cathode Scenario 22, and Location Scenario 3. Recycling process scores are Direct (3), Hydro (69), and Pyro (41), while location scores are China (71) and the U.S. (62). The predicted cost result is 0.8808, indicating a high recycling cost under these conditions. This suggests a need for optimizing throughput, material selection, or process types to improve cost efficiency.

Figure 10c shows the Profitability prediction interface for LIB recycling. Key inputs include virgin material value and recycling cost both at \$67, recycling method scores—Direct (67), Hydro (64), Pyro (31), and cathode scenarios LFP and NCK (62 each). Transportation and location factors cover China (truck & tanker: 58, recycling: 69) and the U.S. (train: 42, truck: 55, recycling: 45). The tool calculates an extremely low profitability of 1.58×10^{-12}, indicating recycling is not economically viable under these conditions, emphasizing the need for improvements in cost, recovery, or logistics.

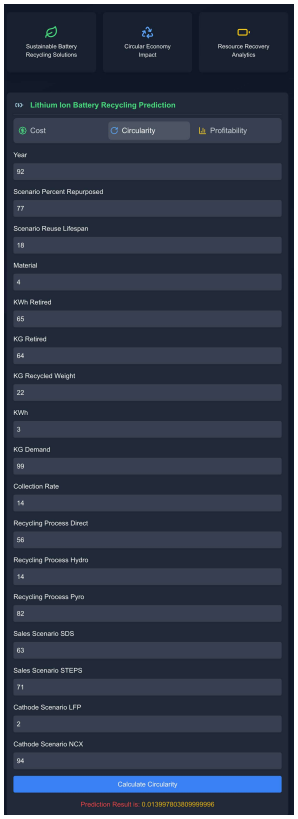
4.7 Benchmarking Model Effectiveness

Several studies have applied ML to battery prediction, showing varied results. Ouyang et al. [31] and Zhu et al. [32] used WOA-LSTM on NASA data, achieving good accuracy but limited generalization. Wang et al. [33] and Li et al. [26] explored hybrid models but lacked real-world factors or broader comparisons. Singaram et al. [25] and Huang et al. [35] applied ensemble and DL methods, requiring complex setups and further feature tuning. Our study outperforms all with PCA-IG and Fidelity Regression, reaching $\mathbf{R^2}$ of 0.99 and low MAE across all recycling tasks. Table 12 presents a comparison of leading models used for LIB prediction, highlighting their methods, performance, and key limitations.

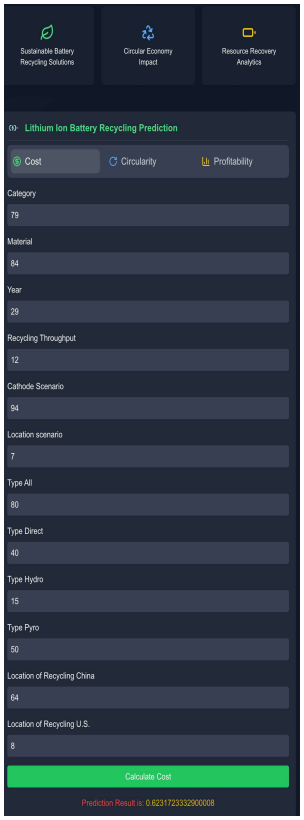
5 Discussion

The experimental results confirm that the proposed fidelity-based BRR framework delivers highly accurate and consistent predictions across LIB recycling objectives, including circularity, cost efficiency, and profitability. Following PCA-IG feature selection, the model achieved an $\mathbf{R^2}$ of 0.99 across all datasets, with low MAE values of 0.01, 0.04, and 0.11 for Circularity, Cost, and Profit, respectively, substantially outperforming conventional models. In contrast, LR and RFR showed limited predictive capability, while SVR failed to generalize effectively, highlighting the complex nonlinear and heteroscedastic nature of LIB recycling policy and economic data. The observed performance improvements are architectural rather than incremental. The proposed fidelity-based framework separates global material-flow trends through a low-fidelity component while refining complex local non-linearities using a high-fidelity BRR model with probabilistic regularization. Through optimized α-weighting and L2 regularization, the framework effectively reduces overfitting and improves generalization across 30 years of large-scale LIB supply-chain and policy scenarios, outperforming conventional single-fidelity and laboratory-focused approaches [31, 32, 40].

The proposed fidelity-based architecture improves the accuracy–computational efficiency trade-off, a key requirement for real-time industrial decision systems. Unlike



(a) Circularity Prediction Using Web-Based Interface.



(b) Cost Prediction Using Web-Based Interface.



(c) Profit Prediction Using Web-Based Interface.

Fig. 10: Online Interface for Predicting the Future of LIB Recycling.

Table 12: Comparison of State-of-the-Art Models for LIB Prediction.

Reference	Dataset	Method	Accuracy (%)	Limitation
[31]	NASA PCoE Data	WOA, VMD, and LSTM	RMSE =0.69 MAPE=0.43	The model was only tested on NASA 18650 lithium-ion batteries, which may limit its applicability to other battery types with different degradation behaviors.
[32]	NASA batteries Data	WOA ,LSTM	RMSE=0.23 MAE=0.02	The study compares WOA-LSTM only with PSO and ABC optimizers. It does not benchmark against other advanced hybrid models.
[33]	Two Lithium-ion Battery Data	ISSA ,VMD and LSTNet	MAE= 1.5, RMSE= 2 and r=92.9	This study does not account for real-world battery health factors such as current, voltage, and temperature, which can significantly influence RUL predictions.
[26]	NASA PCoE Laboratory Data	SOH, SVM, CNN LSTM	RMSE = 0.94, MAPE = 45.82, and R2 = 0.9298	Few models were explored, and more suitable algorithms or model fusion techniques could improve prediction accuracy.
[25]	Monitoring Data	AdaBoost, CNNs Gaussian Processes	Accuracy= 0.94 Precision=0.96 Recall =0.92 and F1 score= 0.94	Complex battery data requires advanced deep learning for effective modeling.
[35]	Sensor Data	ANN,KNN ,GBDT, RF	R2 =0.92 MAE =74.52 RMSE = 110.07	More domain-specific feature engineering is required.
Our Study	LIBs Recycle Data	PCA-IG, Fidelity Regression, LIME	Circularity: R2= 0.99, MAE= 0.01; Cost: R2= 0.99, MAE= 0.04; Profit: R2= 0.99, MAE= 0.09	More domain-specific feature engineering is required.

GPU-intensive DL surrogates requiring temporal backpropagation [16, 17], the LF HF cascade uses lightweight algebraic inference, enabling rapid "what-if" analysis at the low-fidelity stage (e.g., circularity under policy shifts), while the high-fidelity BRR is selectively activated for high-stakes, precision-critical decisions. The hierarchical design is crucial for the web-based evaluation system. The Flask backend deploys three pre-trained fidelity models via joblib, enabling sub-second JSON inference, which is difficult to achieve with ensemble LSTM or CNN pipelines. The fidelity-based BRR model uses a compact hyperparameter grid for α and λ $(10^{-6}, 10^{-3}, 10^{-1})$, optimized via 5-fold cross-validation on the training data only. This controlled search strategy reduces computational cost and carbon footprint compared to PSO-driven DL models [26, 31], while improving deployment feasibility for small and medium recycling enterprises.

The integration of LIME transforms the fidelity-based BRR framework into an interpretable decision-support system for LIB circular economy analysis. Across circularity, cost, and profitability datasets, LIME consistently identifies key drivers

such as recycled material volume and material type (positive impact on circularity), recycling throughput and process category (cost sensitivity, with pyro increasing and direct recycling reducing costs), and virgin material value alongside direct recycling (profit enhancement), while logistics and recycling cost act as major negative factors. These explanations convert high R^2 predictions into actionable policy and operational insights, enabling transparent model auditing and supporting data-driven interventions in sustainable battery recycling systems.

The multi-dimensional scenario analysis (Figs. 7, 8, 9) highlights clear implications for U.S. Recycled Content Standards (RCS) [9, 40]. The Sustainable Development Scenario (SDS) consistently improves circularity and material recovery compared to STEPS, confirming that stricter policy targets enhance system efficiency. Cost analysis shows hydrometallurgical recycling as the most economical pathway ($8.38/kg), while direct recycling, though costlier ($12.41/kg), provides higher material value for advanced applications. The profit–carbon trade-off reveals direct recycling as the most favorable option, achieving the highest profit ($2.41/kg) with moderate emissions, whereas pyrometallurgical routes are both economically and environmentally inefficient under U.S. logistics conditions. The results underscore the proposed fidelity-based framework as a robust decision-support tool for assessing economic–environmental trade-offs and long-term competitiveness under RCS policy evolution.

Despite strong predictive performance, the proposed fidelity-based framework has several limitations. It is trained on 2020–2050 scenario projections, which limits its applicability to dynamic market and operational variations, positioning it primarily as a strategic foresight tool. The current configuration is also tailored to existing LIB chemistries (LFP, NCA, NMC), requiring retraining or domain adaptation for emerging technologies. However, the web-based decision-support system is still a proof-of-concept and would require robust cybersecurity, user authentication, and scalable API infrastructure for industrial deployment.

6 Conclusion and Future Work

This study applies advanced ML techniques, including fidelity regression, to enhance the accuracy and efficiency of LIB recycling predictions. The models achieve strong performance, with high R^2 scores and low MAE across three datasets, identifying key factors such as recycling throughput and material recovery rates as primary drivers of cost and profitability. By leveraging a comprehensive public dataset and addressing real-world challenges like diverse battery chemistries and evolving market conditions, the research demonstrates the potential of AI-driven approaches to improve both the sustainability and economic viability of LIB recycling. The findings offer a scalable solution to tackle battery waste and support the advancement of a circular economy in energy storage.

Future research should aim to integrate real-time IoT and sensor-based monitoring to enable dynamic, online predictions. Expanding the model to include more battery chemistries, emerging recycling technologies—like direct lithium recovery and green leaching—and optimizing for both environmental and economic outcomes will further strengthen its impact. Additionally, real-world validation through pilot testing with

industry and policy stakeholders will help refine the web-based tool for practical use. Implementing digital twins on large datasets and establishing industry benchmarks will be key steps in making these AI-driven solutions truly effective. These advancements are crucial for turning scalable analytics into actionable strategies that drive a fully circular LIB economy.

Data and Code Availability

The dataset utilized in this research is available on Dryad, and all associated code can be accessed through the GitHub repository linked below. Dryad : https://datadryad.org/dataset/doi:10.25338/B8792H#citations GitHub: https://github.com/ShahriarAyon63/lithium_battery_recycle

Conflict of Interest

The authors declare no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics, Consent to Participate, and Consent to Publish declarations

Not applicable

Authors Contribution

Shahriar Siddique Ayon: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft. Md. Akbar Hossain: Supervision, Project administration, Methodology, Writing – review & editing. Md. Jisun Abedin Aurnob: Methodology, Investigation, Formal analysis, Resources, Validation, & Writing – original draft. Majharul Islam Shanto: Methodology, Validation, Formal analysis, Writing – review & editing. Md. Ebrahim Hossain: Data Curation, Validation, Writing – review & editing.

References

[1] Sheth, R.P., Ranawat, N.S., Chakraborty, A., Mishra, R.P., Khandelwal, M.: The lithium-ion battery recycling process from a circular economy perspective—a review and future directions. Energies **16**(7), 3228 (2023) https://doi.org/10.3390/en16073228

[2] Ralls, A.M., Leong, K., Clayton, J., Fuelling, P., Mercer, C., Navarro, V., Menezes, P.L.: The role of lithium-ion batteries in the growing trend of electric vehicles. Materials **16**(17), 6063 (2023) https://doi.org/10.3390/ma16176063

[3] Siddiqua, A., Hahladakis, J.N., Al-Attiya, W.A.K.: An overview of the environmental pollution and health effects associated with waste landfilling and open dumping. Environmental Science and Pollution Research **29**(39), 58514–58536 (2022) https://doi.org/10.1007/s11356-022-21578-z

[4] Huang, Y., Li, J.: Key challenges for grid-scale lithium-ion battery energy storage. Advanced Energy Materials **12**(48), 2202197 (2022) https://doi.org/10.1002/aenm.202202197

[5] Mayyas, A., Steward, D., Mann, M.: The case for recycling: Overview and challenges in the material supply chain for automotive li-ion batteries. Sustainable Materials and Technologies **19**, 00087 (2019) https://doi.org/10.1016/j.susmat.2018.e00087

[6] Yang, Z., Huang, H., Lin, F.: Sustainable electric vehicle batteries for a sustainable world: perspectives on battery cathodes, environment, supply chain, manufacturing, life cycle, and policy. Advanced Energy Materials **12**(26), 2200383 (2022) https://doi.org/10.1002/aenm.202200383

[7] Global Market Insights Inc.: U.S. Electric Vehicle Market Size & Share: Forecast Report, 2032. Accessed: 2024-05-14 (2023). https://www.gminsights.com/industry-analysis/us-electric-vehicle-market

[8] Meegoda, J.N., Malladi, S., Zayas, I.C.: End-of-life management of electric vehicle lithium-ion batteries in the united states. Clean Technologies **4**(4), 1162–1174 (2022) https://doi.org/10.3390/cleantechnol4040071

[9] Maeder, M., Fröhling, M.: Conceptualizing circular economy policy instruments: The case of recycled content standards. Sustainable Production and Consumption **52**, 333–346 (2024) https://doi.org/10.1016/j.spc.2024.11.009

[10] Su, D., Mei, Y., Liu, T., Amine, K.: Global regulations for sustainable battery recycling: Challenges and opportunities. Sustainability **17**(7), 3045 (2025) https://doi.org/10.3390/su17073045

[11] CAS Insights: Innovations in Lithium-ion Battery Recycling. Accessed: 2024-05-14 (2025). https://www.cas.org/resources/cas-insights/innovations-in-lithium-ion-battery-recycling

[12] Collis, G.E., Dai, Q., Loh, J.S., Lipson, A., Gaines, L., Zhao, Y., Spangenberger, J.: Closing the loop on lib waste: a comparison of the current challenges and opportunities for the us and australia towards a sustainable energy future. Recycling **8**(5), 78 (2023) https://doi.org/10.3390/recycling8050078

[13] Roy, J.J., Rarotra, S., Krikstolaityte, V., Zhuoran, K.W., Cindy, Y.D.-I., Tan, X.Y., Carboni, M., Meyer, D., Yan, Q., Srinivasan, M.: Green recycling methods to treat lithium-ion batteries e-waste: a circular approach to sustainability. Advanced Materials **34**(25), 2103346 (2022)

[14] Yang, Y.: A machine-learning prediction method of lithium-ion battery life based on charge process for different applications. Applied Energy **292**, 116897 (2021)

[15] Baum, Z.J., Bird, R.E., Yu, X., Ma, J.: Lithium-ion battery recycling overview of techniques and trends. ACS Publications (2022)

[16] Iglesia, D.H., Corbacho, C.C., Dib, J.Z., Alonso-Secades, V., López Rivero, A.J.: Advanced machine learning and deep learning approaches for estimating the remaining life of ev batteries—a review. Batteries **11**(1), 17 (2025) https://doi.org/10.3390/batteries11010017

[17] Khan, B., Ali, S.M., Ullah, Z.: Deep learning based digital twins augmented reality: Model predictive control for battery and storage optimization in renewable energy prosumers districts. Journal of Energy Storage **131**, 117565 (2025) https://doi.org/10.1016/j.est.2024.117565

[18] Pagliaro, M., Meneguzzo, F.: Recycling of lithium batteries. Sustainable Separation Engineering: Materials, Techniques and Process Development, 591–603 (2022) https://doi.org/10.1002/9783527830185.ch20

[19] Li, Z.: A cost-effective lithium-ion battery direct recycling process. In: Electrochemical Society Meeting Abstracts 233, pp. 606–606 (2018). https://doi.org/10.1149/MA2018-01/4/606 . The Electrochemical Society, Inc.

[20] Velázquez-Martínez, O., Valio, J., Santasalo-Aarnio, A., Reuter, M., Serna-Guerrero, R.: A critical review of lithium-ion battery recycling processes from a circular economy perspective. Batteries **5**(4), 68 (2019) https://doi.org/10.3390/batteries5040068

[21] Dos Santos, M., Garde, I.A.A., Ronchini, C.M.B., Cardozo Filho, L., De Souza, G.B.M., Abbade, M.L.F., Regone, N.N., Jegatheesan, V.J., De Oliveira, J.A.: A technology for recycling lithium-ion batteries promoting the circular economy: The recyclib. Resources, Conservation and Recycling **175**, 105863 (2021) https://doi.org/10.1016/j.resconrec.2021.105863

[22] Kousar, S., Jalal, T., Kausar, N., Agarwal, P., Sakinah Syed Ahmad, S., Kasmin, F.: Enhancing efficiency and sustainability in lithium battery recycling: A multi-objective goal programming approach. INTERCIENCIA **260**(10) (2024) https://doi.org/10.59671/xsk3q

[23] Neumann, J., Petranikova, M., Meeus, M., Gamarra, J.D., Younesi, R., Winter,

M., Nowak, S.: Recycling of lithium-ion batteries—current state of the art, circular economy, and next generation recycling. Advanced energy materials **12**(17), 2102917 (2022) https://doi.org/10.1002/aenm.202102917

[24] Valizadeh, A., Amirhosseini, M.H., Ghorbani, Y.: Predictive precision in battery recycling: unveiling lithium battery recycling potential through machine learning. Computers & Chemical Engineering **183**, 108623 (2024)

[25] Singaram, B., Vinmathi, M.S., Rajan, H.B.M., H, J., Manikandan, T.: Data-driven estimation of lithium-ion battery state-of-health prediction approach using machine learning algorithm for enhanced battery management systems. Nanotechnology Perceptions **20**(s7) (2024) https://doi.org/10.62441/nano-ntp.v20is7.7

[26] Li, K., Chen, X.: Machine learning-based lithium battery state of health prediction research. Applied Sciences **15**(2), 516 (2025) https://doi.org/10.3390/app15020516

[27] Antony Jose, S., Cook, C.A.D., Palacios, J., Seo, H., Torres Ramirez, C.E., Wu, J., Menezes, P.L.: Recent advancements in artificial intelligence in battery recycling. Batteries **10**(12) (2024) https://doi.org/10.3390/batteries10120440

[28] Asal, B., Oyucu, S., Aksoz, A.: Ensemble learning in li-ion battery management systems: Focus on voting regression for capacity estimation. In: 2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP), pp. 1–5 (2024). IEEE

[29] Babu, P.S., V, I., S, V.: Enhanced soc estimation of lithium ion batteries with realtime data using machine learning algorithms. Scientific Reports **14**(1), 16036 (2024) https://doi.org/10.1038/s41598-024-66997-9

[30] Roy, A.: Hybrid random forest regression and artificial neural networks for modelling and monitoring the state of health of li-ion battery. J. Electr. Syst **20**(2), 2231–2243 (2024) https://doi.org/10.52783/jes.1990

[31] Ouyang, M., Shen, P.: Prediction of remaining useful life of lithium batteries based on woa-vmd and lstm. Energies **15**(23), 8918 (2022) https://doi.org/10.3390/en15238918

[32] Zhu, Y., Shang, Y., Duan, B., Gu, X., Li, S., Chen, G.: A data-driven method for lithium-ion batteries remaining useful life prediction based on optimal hyperparameters. In: 2022 41st Chinese Control Conference (CCC), pp. 7388–7392 (2022). https://doi.org/10.23919/CCC55661.2022.9847811 . IEEE

[33] Wang, H., Ye, X., Li, Y., Zhu, G.: Remaining useful life prediction for lithium-ion batteries based on improved mode decomposition and time series. Sustainability **15**(12), 9176 (2023) https://doi.org/10.3390/su15129176

[34] Pregowska, A., Osial, M., Urbańska, W.: The application of artificial intelligence in the effective battery life cycle in the closed circular economy model—a perspective. Recycling **7**(6), 81 (2022) https://doi.org/10.3390/recycling7060081

[35] Huang, W., Zhou, T., Ma, J., Chen, X.: An ensemble model based on fusion of multiple machine learning algorithms for remaining useful life prediction of lithium battery in electric vehicles. Innovations in Applied Engineering and Technology, 1–12 (2025)

[36] Rane, N.L., Choudhary, S.P., Rane, J.: Artificial intelligence and machine learning in renewable and sustainable energy strategies: A critical review and future perspectives. Partners Universal International Innovation Journal **2**(3), 80–102 (2024) https://doi.org/10.5281/zenodo.12155847

[37] Nair, P., Vakharia, V., Borade, H., Shah, M., Wankhede, V.: Predicting li-ion battery remaining useful life: an xdfm-driven approach with explainable ai. Energies **16**(15), 5725 (2023) https://doi.org/10.3390/en16155725

[38] Pacios, R., Bekaert, E., Antuñano, N., Reynaud, M., Gandiaga, I.: Recycling of lithium-ion batteries: The black spot of the green deal. DYNA-Ingeniería e Industria **99**(6) (2024) https://doi.org/10.52152/D11287

[39] Oyucu, S., Ersöz, B., Sağıroğlu, Ş., Aksöz, A., Biçer, E.: Optimizing lithium-ion battery performance: Integrating machine learning and explainable ai for enhanced energy management. Sustainability **16**(11), 4755 (2024) https://doi.org/10.3390/su16114755

[40] Dunn, J., Kendall, A., Slattery, M.: Electric vehicle lithium-ion battery recycled content standards for the us – targets, costs, and environmental impacts. Resources, Conservation and Recycling **185**, 106488 (2022) https://doi.org/10.1016/j.resconrec.2022.106488

[41] Dunn, J., Kendall, A., Slattery, M.: United States recycled content standards for lithium-ion batteries. Dryad. Accessed: 2025-07-20 (2022). https://doi.org/10.25338/B8792H . https://datadryad.org/dataset/doi:10.25338/B8792H

[42] Vinutha, H.P., Poornima, B., Sagar, B.M.: Detection of outliers using interquartile range technique from intrusion dataset. In: Satapathy, S.C., Tavares, J.M.R.S., Bhateja, V., Mohanty, J.R. (eds.) Information and Decision Sciences, pp. 511–518. Springer, Singapore (2018)

[43] Odhiambo Omuya, E., Onyango Okeyo, G., Waema Kimwele, M.: Feature selection for classification using principal component analysis and information gain. Expert Systems with Applications **174**, 114765 (2021) https://doi.org/10.1016/j.eswa.2021.114765

[44] Siddique, A.A., Raza, A., Alshehri, M.S., Alasbali, N., Abbasi, S.F.: Optimizing

tumor classification through transfer learning and particle swarm optimization-driven feature extraction. IEEE Access **12**, 85929–85939 (2024) https://doi.org/10.1109/ACCESS.2024.3412412

[45] Salih, A.M., Raisi-Estabragh, Z., Galazzo, I.B., Radeva, P., Petersen, S.E., Lekadir, K., Menegaz, G.: A perspective on explainable artificial intelligence methods: Shap and lime. Advanced Intelligent Systems **7**(1), 2400304 (2025) https://doi.org/10.1002/aisy.202400304

[46] Siddique Ayon, S., Ebrahim Hossain, M., Ullah Miah, M.S., Rahman, M.M., Mahmud, M.: Explainable ai in feature selection: Improving classification performance on imbalanced datasets. In: International Conference on Neural Information Processing, pp. 303–318 (2024). https://doi.org/10.1007/978-981-96-6606-5_21 . Springer

[47] developers, S.-l.: What's New in Scikit-learn 1.0. https://scikit-learn.org/stable/whats_new/v1.0.html. Accessed: 2025-07-20 (2021)

[48] Foundation, P.S.: Python 3.8.0 Release. https://www.python.org/downloads/release/python-380/. Accessed: 2025-07-20 (2019)

[49] Peiseler, L., Schenker, V., Schatzmann, K., Pfister, S., Wood, V., Schmidt, T.: Carbon footprint distributions of lithium-ion batteries and their materials. Nature Communications **15**(1), 10301 (2024) https://doi.org/10.1038/s41467-024-54634-y

[50] Gopal, S.: Recycling of Lithium Batteries and GHG Emissions. Accessed: 2025-04-22 (2023). https://floodlightglobal.com/recycling-of-lithium-batteries-and-ghg-emissions/

[51] Chen, Q., Lai, X., Gu, H., Tang, X., Gao, F., Han, X., Zheng, Y.: Investigating carbon footprint and carbon reduction potential using a cradle-to-cradle lca approach on lithium-ion batteries for electric vehicles in china. Journal of Cleaner Production **369**, 133342 (2022) https://doi.org/10.1016/j.jclepro.2022.133342

[52] Battery University: BU-1003: Electric Vehicle (EV). Accessed: 2025-04-22 (2021). https://batteryuniversity.com/article/bu-1003-electric-vehicle-ev

[53] Frandsen, H.L., Curran, D.J., Rasmussen, S., Hendriksen, P.V.: High throughput measurement of high temperature strength of ceramics in controlled atmosphere and its use on solid oxide fuel cell anode supports. Journal of Power Sources **258**, 195–203 (2014) https://doi.org/10.1016/j.jpowsour.2014.02.036

[54] Udawattage, W.: Sustainable freight transport – truck vs. train. Accessed: 2025-04-22 (2024). https://www.sustamize.com/blog/sustainable-freight-transport

[55] Nicolai, F.: Rail freight produces 6 times less CO_2 than truck. Accessed:

2025-04-22 (2020). https://www.railfreight.com/policy/2020/01/08/rail-freight-produces-6-times-less-co2-than-truck/

[56] Chaloupy, F.: Ecological transport: boat, train, plane or truck? Accessed: 2025-04-22 (2022). https://kibodo.shop/en/blogs/infos/transport-ecologique-bateau-train-avion-ou-camion