

Generative Counterfactual Explanations for Electric Vehicle Eco-Driving

Leonardo Carneiro Feltran¹[0000-0002-2309-0542] (✉), Paulo Rogério^{1,2}[0000-0002-7367-4335], Luan L. Santos³[0009-0001-5598-0945], Lucas A. C. Souza³[0000-0002-8941-3121], Marcelo Luiz Drumond Lanza³[0009-0006-4412-6707], Robson Francisco Dias³[0000-0002-5281-9179], Luís Henrique Maciel Kosmowski Costa³[0000-0002-7789-3359], Miguel Elias Mitre Campista³[0000-0002-8752-9382], and Andre Carlos Ponce de Leon Ferreira de Carvalho¹[0000-0002-4765-6459]

¹ University of São Paulo (USP), São Paulo, Brazil

² Federal University of Maranhão (UFMA), São Luís, Brazil

³ Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil
leonardo.feltran@usp.br

Abstract. Driver behavior is a critical factor impacting the battery consumption of electric vehicles (EVs), presenting significant opportunities for energy savings through behavioral modification. Current eco-driving systems can identify high-consumption patterns, but they rarely provide prescriptive guidance on how to improve driving. This paper proposes a framework that generates counterfactual speed profiles showing how a driver could complete a given route with lower energy consumption. For such, three feature extraction methods are evaluated on two datasets: a proprietary dataset and the publicly available eVED dataset. Two wavelet-based approaches (Haar and Daubechies) and a Variational Autoencoder (VAE) are compared as latent representations for counterfactual search. The VAE consistently provides superior counterfactuals across both datasets, producing speed profiles that remain geometrically closer to the original trajectory. Simulation of the VAE-generated profiles in the Autonomie® vehicle simulator confirms an average reduction of 12.64% in energy consumption, indicating that latent-space counterfactual search is a practical approach for generating prescriptive speed profiles that generalizes across two independent datasets.

Keywords: Eco-driving, Counterfactual Explanations · Explainable AI (XAI) · Electric Vehicles · Machine Learning · Driver Behavior · Time-series Analysis

1 Introduction

The relationship between driving behavior and electric vehicle (EV) energy consumption has been extensively investigated in the literature [12, 16, 21]. Some

Generative AI tools were used for language revision. The authors take full responsibility for the content of this work.

studies primarily analyze the impact of driver actions on vehicle energy consumption, while others adopt a proactive approach, recommending driving patterns to save energy. These studies show that the air-conditioning system and speed patterns are highly correlated with high energy consumption [12, 16], and that improving driver behavior can significantly extend the vehicle’s battery range.

Machine Learning and eXplainable Artificial Intelligence (XAI) techniques have been applied to model these complex relationships. For instance, neural networks have been used to estimate vehicle consumption, identifying high-traffic urban areas as the most energy-intensive factor [19], while frameworks integrating statistical, predictive, and causal perspectives have shown that relying on a single analytical method may conceal important insights [12]. Post-hoc explanation techniques have also been explored to improve model transparency and facilitate eco-driving [6]. However, the most used XAI techniques are model-agnostic approaches that quantify feature importance at the instance level, being inherently diagnostic: they explain *why* a particular outcome occurred, but do not prescribe *how* to change the inputs to achieve a different result [1].

To translate consumption analysis into actionable tools, several studies have explored driver feedback systems. Real-time frameworks have been proposed to guide drivers toward energy-efficient behavior through corrective suggestions [15, 24]; however, the delivery mechanism can negatively affect user experience and adoption [5]. Gamification techniques, such as leaderboards, offer an alternative motivational approach [8], but still do not inform drivers about how to modify specific driving maneuvers.

A concrete example of prescriptive reasoning in the automotive domain is “EV Life” [23], an application that shows owners of internal combustion engine (ICE) vehicles what each trip would have looked like in an EV, covering cost savings and CO₂ reduction. While effective at the *strategic* level of vehicle adoption, it does not address the *operational* level of daily driving optimization, answering “what if I had an EV?” but not “how should I drive my EV to consume less energy?”

Current literature thus reveals a gap in providing operational, instance-level guidance: no existing approach prescribes *how* a driver should modify a specific speed profile to reduce energy consumption. Counterfactual explanations [7] are well-suited to address this limitation, as they describe the minimal input changes that lead to a different outcome: a speed profile associated with lower energy consumption.

This paper addresses this gap by proposing a framework that generates counterfactual (CF) speed profiles to guide drivers toward more energy-efficient behaviors. Three feature extraction methods are evaluated for their suitability as latent representations for CF search. Two real-world datasets are used to validate the methodology, the first collected by the authors, and the publicly available eVED dataset. Class imbalance arising from the categorization process is addressed by evaluating four balancing strategies. Finally, the Autonomie® [17] high-fidelity vehicle simulator is used to quantify the resulting energy savings and validate the proposed concept.

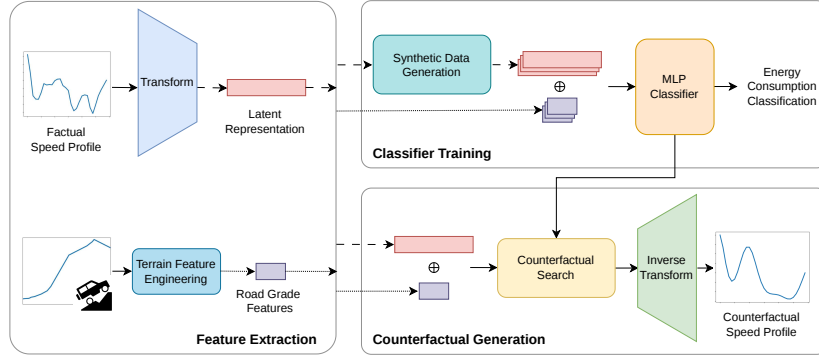


Fig. 1: Overview of the proposed eco-driving framework. The architecture encompasses three stages: (i) *Feature Extraction*, which processes speed profiles into a latent representation and extracts terrain features; (ii) *Classifier Training*, which uses synthetic data generation to address class imbalance and trains an MLP classifier for energy consumption classification; and (iii) *Counterfactual Generation*, which leverages the trained classifier to guide a search that produces an alternative speed profile with lower energy consumption.

2 Methodology

The proposed methodology analyzes vehicle speed profiles and suggests alternative driving behaviors that reduce EV energy consumption. Figure 1 illustrates the complete framework, which consists of three interconnected stages. The following subsections describe the datasets, followed by the target variable construction, the description of each stage of the framework, and metrics used to evaluate the proposed framework.

2.1 Datasets

The proposed pipeline is tested on two datasets: one proprietary and one public. Table 1 summarizes general information about each, and the text below describes them in detail.

Proprietary Dataset : The first dataset was collected under a proprietary campaign and is not publicly available due to contractual confidentiality constraints. It comprises 102 real-world urban trips completed by a single EV along the same route (mean distance 11.84 km) between August 2024 and February 2025, with signals captured via a CANEdge OBD-II logger at 1 Hz. GPS coordinates were queried against OpenStreetMap to retrieve road elevation, enabling identification of uphill and downhill segments that are key to distinguishing consumption from regeneration patterns.

Table 1: Window counts and class distribution for each dataset split after segmentation (30 s windows, 1 s stride).

Dataset	Split	Windows	Class 0	Class 1	Minority (%)
Proprietary	Train+Val	98783	79028	19755	20.0
	Test	23381	19470	3911	16.7
	Total	122164	98498	23666	19.4
eVED (BEV)	Train+Val	83796	67037	16759	20.0
	Test	21993	17206	4787	21.8
	Total	105789	84243	21546	20.4

Extended Vehicle Energy Dataset (eVED) : The second dataset is the Extended Vehicle Energy Dataset (eVED) [25], a publicly available collection of OBD-II signals enriched with map-matched GPS traces and road elevation retrieved via the Google Maps Elevation API. Data were collected from privately owned vehicles in Ann Arbor, Michigan, USA, with GPS traces de-identified prior to release. We use the battery electric vehicle (BEV) subset, comprising 3 vehicles and 491 trips.

2.2 Target Variable Construction

The target variable is motor energy efficiency per window. This value for the proprietary dataset is calculated using the formula

$$E_c = \frac{E}{d} \left[\frac{\text{kWh}}{\text{km}} \right] \quad (1)$$

where E is the energy consumed and d the distance traveled in the window. Stationary windows ($d \leq 10^{-5}$) are discarded, as no meaningful efficiency can be defined.

The eVED dataset does not provide motor power consumption, however the battery current I_{bat} and voltage V_{bat} are provided, with the addition of air-conditioning system P_{HVAC} and heater system P_{heater} power consumption. Therefore, motor power consumption is approximated as:

$$P_{motor} = V_{bat}I_{bat} - P_{HVAC} - P_{heater} \quad (2)$$

For both datasets a binary label is assigned using the 80th percentile of the efficiency distribution as threshold η : windows with $E_c > \eta$ are labeled as high consumption (class 1), and the remainder as moderate consumption (class 0). A percentile-based threshold is used instead of a fixed value to avoid systematic bias across vehicles with different baseline consumption profiles, yielding a consistent 80/20 class imbalance ratio that requires and motivates the balancing strategies evaluated in this work.

2.3 Feature Extraction

This stage has two steps, the first extracts features from speed profiles using a Variational Autoencoder (VAE) [14], while the second processes terrain features variations (TFV)s.

VAE is used to encode speed profiles into a continuous, probabilistic latent space. By constraining the encoder to output distributions rather than point estimates, the VAE ensures a smooth topology where similar instances remain proximal, a property that facilitates CF search. The encoder-decoder is implemented with LSTM cells [11]. As an alternative, the Discrete Wavelet Transform (DWT) decomposes each speed window into one approximation and four detail coefficient sets. Five statistics (mean, standard deviation, energy, maximum, and median) are computed per set and concatenated into a 25-dimensional feature vector. Haar and Daubechies (DB) [4] wavelets are evaluated alongside the VAE.

TFVs, which are not controllable by the driver, directly influence energy consumption through gravitational load and is therefore treated as a static contextual input: CF generation must find a new speed profile that reduces consumption under the same terrain conditions. Four terrain features are computed per window from the altitude signal: uphill and downhill distance ($d_{\uparrow}$, $d_{\downarrow}$) and intensity ($\iota_{\uparrow} = \Delta h_{\uparrow}/d_{\uparrow}$, $\iota_{\downarrow} = \Delta h_{\downarrow}/d_{\downarrow}$), where Δh denotes the cumulative absolute altitude change. Each trip is segmented into overlapping 30-second windows with a stride of 1 second.

2.4 Classifier Training

The classifier guides CF search by identifying latent-space instances belonging to the low-consumption class closest to the query. Gradient-based search requires a differentiable model, so an MLP is used: two hidden layers of 64 and 32 neurons respectively (ReLU, dropout 0.1) and a single output neuron for binary classification. Training includes L_2 regularization, early stopping, and a learning rate scheduler to mitigate overfitting.

The resulting 80/20 class imbalance, whose origin is detailed in Section 2.2, can produce a classifier that fails to detect high-consumption instances, which may degrade the CF search. This motivates the use of synthetic oversampling strategies.

SMOTE [3] generates synthetic minority-class instances by interpolating between a sample and its k -nearest neighbors. Despite its effectiveness, interpolation without boundary awareness may introduce noise in overlapping regions. Three variants address this limitation. Borderline-SMOTE [9] restricts generation to samples near the decision boundary. ADASYN [10] adapts the number of synthetic samples per region according to local class density, concentrating oversampling where the boundary is hardest to learn. SMOTE-Tomek [2] combines SMOTE with Tomek-link removal, which discards majority-class instances closest to the boundary after oversampling, producing a cleaner class separation.

2.5 Counterfactual Generation

A CF explanation describes the minimal input change that would lead to a different model outcome [20]. Following [13], this objective is formulated as the following optimization problem:

$$x^* = \arg \min_{x'} \text{dist}(x, x') \quad \text{s.t.} \quad h(x) \neq h(x'), \quad x' \in \mathcal{P} \quad (3)$$

where $x, x' \in \mathcal{X}$ are the factual and CF instances, $h(\cdot)$ is the classifier, $\text{dist}(\cdot, \cdot)$ a distance metric, and $\mathcal{P}$ the set of plausible constraints. In this framework, $\mathbf{T}(\cdot)$ extracts features from a speed series s as $x = \mathbf{T}(s)$; after solving Equation 3, the result is mapped back via $s^* = \mathbf{T}^{-1}(x^*)$. This search is performed by DiCE [18], which supports differentiable and genetic-algorithm approaches.

2.6 Evaluation Metrics

Classification The classification is highly impacted by class imbalance, which motivates the use of metrics other than accuracy. Area Under the ROC Curve (AUC-ROC) measures class discrimination, which is particularly relevant in the CF context, where class separability is sought. Area Under the Precision-Recall Curve (AUC-PR) provides a threshold-independent summary of the precision-recall trade-off, and captures better performance on the minority class than AUC-ROC alone. The last metric used in this work is F1-score, which is the harmonic mean of recall and precision.

Counterfactual evaluation Evaluating CF for time series requires assessing not only the class flip but also plausibility and proximity. All metrics used to evaluate CF are listed below:

Dynamic Time Warping (DTW) [22] measures the similarity between the factual speed profile (x) and the CF (x^*). Unlike Euclidean distance, DTW is robust to temporal shifts and distortions, providing a more reliable alignment cost.

Boundary adherence (B_a): To be effective, a CF must share the same start and end points as the factual speed profile. This metric measures the proximity of the CF's boundary speed values to those of the original curve.

$$B_a = \sqrt{(x_0^* - x_0)^2 + (x_t^* - x_t)^2} \quad (4)$$

Physical Plausibility: A CF speed profile is considered physically valid if it satisfies $0 \leq v_t \leq V_{max}$ and $a_{min} \leq a_t \leq a_{max}$ for all t , where the bounds $V_{max} = 42$ m/s, $a_{min} = -5$ m/s², and $a_{max} = 3.2$ m/s² are derived from the tested vehicles' specifications. Profiles that violate any of these constraints are discarded before CF selection.

Relative Traveled Distance (E_d): To ensure the CF represents a useful alternative trajectory, the total distance traveled should remain comparable to the original trip; the relative deviation is quantified as:

$$E_d = \frac{|\sum x \cdot \Delta t - \sum x^* \cdot \Delta t|}{\sum x \cdot \Delta t} \quad (5)$$

Composite score (C_{score}): This metric aggregates CF metrics into one to facilitate the comparison. For this purpose, each metric is first rescaled via min-max normalization within each evaluation fold, yielding values in $[0, 1]$ and physically implausible CF are discarded prior to this step. The composite score is computed as:

$$C_{score} = W_{Ed} \hat{E}_d + W_{BA} \hat{B}_a + W_{DTW} \widehat{DTW} \quad (6)$$

where $\hat{\cdot}$ denotes min-max normalization and weights $W_{Ed} = 8$, $W_{BA} = 4$, $W_{DTW} = 1$ reflect the relative importance of distance preservation, boundary continuity, and shape similarity, respectively. The candidate with the lowest C_{score} is selected as the best CF for each query.

2.7 Energy Consumption with Autonomie[®]

Autonomie[®] uses the speed profile to estimate motor consumption. For each window, the factual speed is applied as an input to yield its power consumption; a cumulative sum is then computed and its final value at $t = 29$ s is recorded. The same procedure is applied to each CF window. The following formula quantifies the energy savings achieved by a CF:

$$Ec_{gain} = \frac{Ec(x) - Ec(x^*)}{Ec(x)} \quad (7)$$

2.8 Experimental Setup

To evaluate the framework using both datasets, their trips are divided into 80% for training and 20% for testing using a fixed random seed. The training set is evaluated with a 5-fold cross-validation strategy to ensure robust evaluation and reduce variance. A final model is then trained on the full training set and validated using the hold-out test data. To ensure physical relevance, a minimum speed threshold of 2 km/h is established for the 30-second time windows, as data analysis showed that high energy consumption during vehicle launch (standstill to motion) occurs regardless of smooth acceleration patterns.

3 Results and Discussion

The experimental evaluation is conducted in two steps. The first step evaluates classifier performance across feature extraction methods and oversampling

Table 2: MLP classifier performance marginalized over balancing strategies (mean $\pm$ std across all folds and balancing configurations). Best extractor per dataset in **bold**.

Dataset	Extractor	F1-score	AUC-ROC	AUC-PR
Proprietary	VAE	0.7476 ± 0.0419	0.9475 ± 0.0205	0.8533 ± 0.0384
	Haar	0.7334 ± 0.0421	0.9447 ± 0.0177	0.8411 ± 0.0355
	DB	0.6900 ± 0.0462	0.9272 ± 0.0214	0.7979 ± 0.0427
eVED	VAE	0.7318 ± 0.0776	0.9410 ± 0.0325	0.8394 ± 0.0748
	Haar	0.7324 ± 0.0710	0.9414 ± 0.0291	0.8366 ± 0.0693
	DB	0.6748 ± 0.0874	0.9209 ± 0.0367	0.7822 ± 0.0800

strategies. The second step analyzes how both factors impact CF generation. The two best-performing CF models are then simulated in Autonomie[®] to quantify energy consumption reduction.

3.1 Classifier Performance

The classifier serves as the guide for the CF search process; therefore, high class separability is essential. Table 2 summarizes the MLP performance across different feature extraction methods for both datasets, with results averaged over all balancing strategies.

This table shows that VAE leads on the proprietary dataset across all metrics. On eVED, Haar and VAE are statistically equivalent, with VAE leading on AUC-PR. DB consistently performs worst.

Figure 2 compares the impact of balancing strategy on classifier performance. SMOTE-Tomek performed significantly worse than all pure oversampling methods (SMOTE, ADASYN, Borderline) and the no-treatment baseline in both

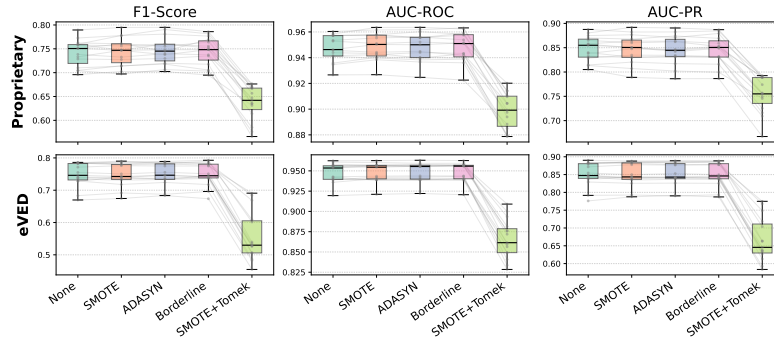


Fig. 2: Effect of balancing strategy on classification performance. Each box pools all extractor $\times$ fold observations. Block lines connect individual (extractor, fold) pairs across strategies. Results shown for both datasets.

datasets. SMOTE, ADASYN, Borderline, and the no-treatment baseline were statistically indistinguishable across all metrics.

3.2 Counterfactual Analysis

The subsection examines how feature extraction impacts CF generation. For each dataset, 10 CF candidates were generated per factual instance, and the best was selected using the composite score defined in Section 2.6. Any CF exhibiting physically implausible speeds or accelerations was discarded, though this occurred in only a minority of cases. In our tests, at least 99.96% of the counterfactuals generated with the VAE extractor were valid in both datasets, with a single query for which no counterfactual could be found, while both wavelet extractors produced physically plausible counterfactuals in 100% of cases.

Figure 3 shows that the VAE outperforms both wavelet methods across CF metrics. Haar and VAE had similar classification performance; however, Haar performed worse in CF generation, demonstrating that classification performance does not guarantee CF search quality. A visual comparison is shown in Figure 4, where four curves illustrate the differences in CF found using each latent representation. Boundary adherence and DTW differences are both visually apparent: DTW is worse for both wavelet approaches, meaning CFs generated in their latent space have shapes more different from the originals; boundary adherence is also worse, as the CF and factual curves diverge more than in the VAE case.

Contrary to expectations, DB (which yielded the worst classification performance) outperforms Haar on DTW and relative distance error in both datasets. On boundary adherence, DB and Haar have similar performance on both datasets. These results suggest that classification and CF search quality do not share the same latent space requirements.

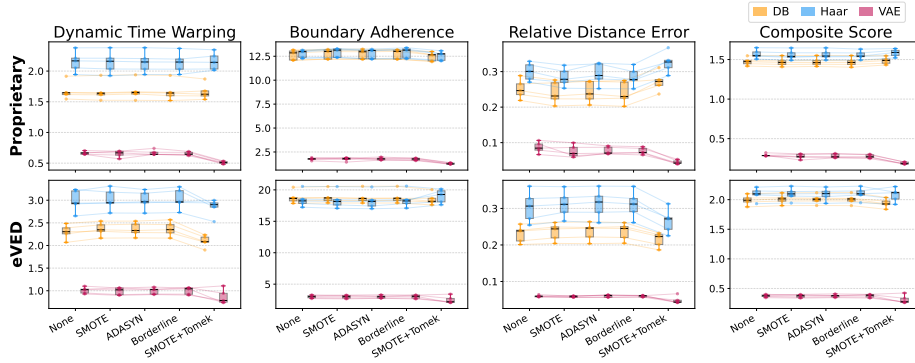


Fig. 3: CF quality metrics by feature extractor and balancing strategy (lower is better). Each group of three boxes corresponds to one oversampling strategy, coloured by extractor. Block lines connect individual fold results within each extractor.

Regarding the impact of balancing strategy on CF quality, one strategy stands apart. SMOTE-Tomek achieves the best mean scores on DTW, boundary adherence, and relative distance error, exactly the opposite of its behavior in classification. This reversal suggests that removing noisy majority-class instances near the decision boundary in the latent space facilitates finding CF that stay geometrically close to the factual profile. The remaining strategies (None, SMOTE, ADASYN, Borderline) are statistically indistinguishable on the composite score in both datasets.

3.3 Consumption Simulation

The Autonomie[®] simulation employed a VAE-based model trained under two conditions: one employing the SMOTE-Tomek oversampling strategy and another without any imbalance treatment. Figure 5 compares the consumption distributions of the factual set and the two resulting CF sets. Over 85.5% of the CF instances exhibit a reduction in energy consumption compared to their corresponding factual instances.

Despite the substantial overlap between all distributions, a one-sided Wilcoxon signed-rank test confirms that factual energy consumption is significantly greater than CF consumption for both strategies ($p < 0.01$). The None strategy achieves a mean reduction of 12.64%, while SMOTE-Tomek achieves 11.93%. A bilateral Wilcoxon test between the two CF sets yields no statistically significant difference ($p > 0.05$).

These results suggest that the classifier quality has limited impact on the energy savings obtained through CF search. Although SMOTE-Tomek substantially degraded classification performance, the classifier still guided the latent-

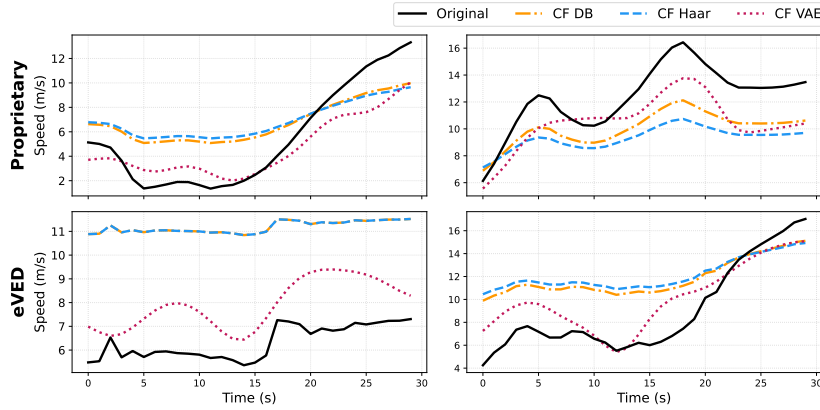


Fig. 4: Representative CF speed profiles for four queries spanning low to high maximum speed. Each panel shows the original profile (black) and the best CF found by each extractor. Queries were selected to be representative of each extractor’s mean CF metrics.

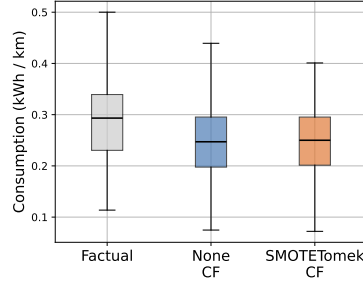


Fig. 5: Comparison of energy consumption between factual and CF instances obtained with Autonomie[®].

space search toward speed profiles with comparable energy reductions. This decoupling between classification accuracy and CF effectiveness is consistent with the finding that SMOTE-Tomek produces a cleaner decision boundary, which may compensate for lower overall accuracy during the search.

4 Conclusions and Future Work

This work demonstrates that latent-space CF search is a viable approach to prescriptive eco-driving feedback. The key finding is that the VAE latent space supports speed profile generation that is physically plausible and energy-efficient across two independent datasets, while wavelet representations produce qualitatively inferior curves despite comparable classification performance. A subtler finding is the decoupling between classifier accuracy and CF effectiveness: SMOTE-Tomek severely degraded classification yet yielded energy savings statistically equivalent to the unbalanced baseline (12% reduction in both cases). This suggests that decision boundary geometry matters more than overall classification accuracy for CF search quality.

This work presents some limitations. First, the framework stops at generating counterfactual speed profiles and still lacks an application layer that turns them into suggestions to the driver. Second, the counterfactual search is purely associational, that is, it does not model the causal dependencies among features, so the recommended maneuvers may not be interventionally valid even when they satisfy the physical-plausibility constraints. Finally, the proprietary dataset comprises trips from a single EV along a single route, which limits the generalization of the findings and calls for broader validation across vehicles, drivers, and routes.

Future work should address these limitations by developing the application layer that delivers actionable, driver-interpretable feedback, and by incorporating causal counterfactuals so that the recommended changes hold under intervention. Broader data collection across vehicles and routes is also needed to confirm generalization.

Acknowledgments. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and Development and Research Foundation - Fundep - Rota 2030.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Asher, N., De Lara, L., Paul, S., Russell, C.: Counterfactual models for fair and adequate explanations. *Machine Learning and Knowledge Extraction* **4**(2), 316–349 (2022). <https://doi.org/10.3390/make4020014>
2. Batista, G.E.A.P.A., Prati, R.C., Monard, M.C.: A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor. Newsl.* **6**(1), 20–29 (2004). <https://doi.org/10.1145/1007730.1007735>
3. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research* **16**, 321–357 (2002). <https://doi.org/10.1613/jair.953>
4. Daubechies, I.: Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics* **41**(7), 909–996 (1988). <https://doi.org/10.1002/cpa.3160410705>
5. Degirmenci, K., Breitner, M.H.: Gamification and sensory stimuli in eco-driving research: A field experiment to reduce energy consumption in electric vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour* **92**, 266–282 (2023). <https://doi.org/10.1016/j.trf.2022.10.014>
6. Farooq, M.S., Saleem, M., Mazhar, T., Anjum, M.A., Shahzad, T., Khan, M.A., Hamam, H.: An interpretable, data-driven framework empowered by explainable ai for fuel consumption and co₂ emission prediction. *Catalysis Today* **465**, 115651 (2026). <https://doi.org/10.1016/j.cattod.2025.115651>
7. Fernández-Loría, C., Provost, F., Han, X.: Explaining data-driven decisions made by ai systems: The counterfactual approach. *Management Information Systems Quarterly* **46**(3), 1635–1660 (2022). <https://doi.org/10.25300/MISQ/2022/16749>
8. Franco, A., Cale, D., Ferreira, J.C.: Gamification system for eco-driving: Enhancing driver motivation and fuel savings through game mechanics. In: *Bio-Inspired Computing*. pp. 129–137. Springer Nature Switzerland (2025). https://doi.org/10.1007/978-3-031-78946-5_12
9. Han, H., Wang, W.Y., Mao, B.H.: Borderline-smote: A new over-sampling method in imbalanced data sets learning. In: *Advances in Intelligent Computing*. pp. 878–887. Springer Berlin Heidelberg (2005). https://doi.org/10.1007/11538059_91
10. He, H., Bai, Y., Garcia, E.A., Li, S.: Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: *2008 IEEE International Joint Conference on Neural Networks*. pp. 1322–1328 (2008). <https://doi.org/10.1109/IJCNN.2008.4633969>
11. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Computation* **9**(8), 1735–1780 (1997). <https://doi.org/10.1162/neco.1997.9.8.1735>
12. Huang, H., Li, B., Wang, Y., Zhang, Z., He, H.: Analysis of factors influencing energy consumption of electric vehicles: Statistical, predictive, and causal perspectives. *Applied Energy* **375**, 124110 (2024). <https://doi.org/10.1016/j.apenergy.2024.124110>

13. Karimi, A.H., Schölkopf, B., Valera, I.: Algorithmic recourse: from counterfactual explanations to interventions. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. p. 353–362. FAccT '21 (2021). <https://doi.org/10.1145/3442188.3445899>
14. Kingma, D.P., Welling, M.: Auto-encoding variational bayes (2013), <https://arxiv.org/abs/1312.6114>
15. McAllister, J., Adams, S.: Eco-routing and driving pattern optimization to minimize ev energy usage. *Journal of Computing and Electronic Information Management* **17**(2), 21–25 (2025). <https://doi.org/10.54097/6vn6nb89>
16. Mobini Seraji, M.H., Shaffiee Haghshenas, S., Shaffiee Haghshenas, S., Simic, V., Pamucar, D., Guido, G., Astarita, V.: A state-of-the-art review on machine learning techniques for driving behavior analysis: clustering and classification approaches. *Complex & Intelligent Systems* **11** (2025). <https://doi.org/10.1007/s40747-025-01988-5>
17. Mohammadnazar, A., Khattak, Z.H., Khattak, A.J.: Assessing driving behavior influence on fuel efficiency using machine-learning and drive-cycle simulations. *Transportation Research Part D: Transport and Environment* **126**, 104025 (2024). <https://doi.org/10.1016/j.trd.2023.104025>
18. Mothilal, R.K., Sharma, A., Tan, C.: Explaining machine learning classifiers through diverse counterfactual explanations. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. pp. 607–617 (2020). <https://doi.org/10.1145/3351095.3372850>
19. Moulik, B., Kaur, S., Ranjan, S., Dixit, S.: Analysing the energy consumption of urban electric vehicles for real driving scenarios. *IETE Journal of Research* **70**(6), 5799–5809 (2024). <https://doi.org/10.1080/03772063.2023.2273295>
20. Pearl, J., Glymour, M., Jewell, N.P.: Causal inference in statistics: A primer. John Wiley & Sons (2016). <https://doi.org/10.1111/biom.13079>
21. Pietra, A., Vazquez Rull, M., Etzi, R., Gallace, A., Scurati, G.W., Ferri, F., Bordegoni, M.: Promoting eco-driving behavior through multisensory stimulation: a preliminary study on the use of visual and haptic feedback in a virtual reality driving simulator. *Virtual Real.* **25**(4), 945–959 (2021). <https://doi.org/10.1007/s10055-021-00499-1>
22. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **26**(1), 43–49 (1978). <https://doi.org/10.1109/TASSP.1978.1163055>
23. Shamma, D.A., Lee, M.L., Filipowicz, A.L.S., Denoue, L., Glazko, K., Murakami, K., Lyons, K.: Ev life: A counterfactual dashboard towards reducing carbon emissions of automotive behaviors. In: *Companion Proceedings of the 27th International Conference on Intelligent User Interfaces*. pp. 46–49. IUI '22 Companion (2022). <https://doi.org/10.1145/3490100.3516451>
24. Wang, J., Zhang, H., Wu, B., Liu, W.: Symmetry-guided electric vehicles energy consumption optimization based on driver behavior and environmental factors: A reinforcement learning approach. *Symmetry* **17**(6), 930 (2025). <https://doi.org/10.3390/sym17060930>
25. Zhang, S., Fatih, D., Abdulqadir, F., Schwarz, T., Ma, X.: Extended vehicle energy dataset (eved): An enhanced large-scale dataset for vehicle energy consumption analysis. In: *2025 IEEE 101st Vehicular Technology Conference (VTC2025-Spring)*. pp. 01–07 (2025). <https://doi.org/10.1109/VTC2025-Spring65109.2025.11174415>