Data-driven Energy Consumption Modelling for Electric Micromobility using an Open Dataset

Yue Ding, Sen Yan, Maqsood Hussain Shah, Hongyuan Fang, Ji Li and Mingming Liu

***Abstract*—The escalating challenges of traffic congestion and environmental degradation underscore the critical importance of embracing E-Mobility solutions in urban spaces. In particular, micro E-Mobility tools such as E-scooters and E-bikes, play a pivotal role in this transition, offering sustainable alternatives for urban commuters. However, the energy consumption patterns for these tools are a critical aspect that impacts their effectiveness in real-world scenarios and is essential for trip planning and boosting user confidence in using these. To this effect, recent studies have utilised physical models customised for specific mobility tools and conditions, but these models struggle with generalization and effectiveness in real-world scenarios due to a notable absence of open datasets for thorough model evaluation and verification. To fill this gap, our work presents an open dataset, collected in Dublin, Ireland, specifically designed for energy modelling research related to E-Scooters and E-Bikes. Furthermore, we provide a comprehensive analysis of energy consumption modelling based on the dataset using a set of representative machine learning algorithms and compare their performance against the contemporary mathematical models as a baseline. Our results demonstrate a notable advantage for data-driven models in comparison to the corresponding mathematical models for estimating energy consumption. Specifically, data-driven models outperform physical models in accuracy by up to 83.83% for E-Bikes and 82.16% for E-Scooters based on an in-depth analysis of the dataset under certain assumptions.**

***Index Terms*—Electric Micromobility, Sustainability, Machine Learning, Energy Consumption Modelling, Open Dataset**

I. Introduction

Nowadays, micromobility is attracting significant attention worldwide for its potential to help establish a sustainable, cost-effective, and widely accessible modern transportation network [1]. Electric Micromobility (EM), including e-scooters and e-bikes, is set to be a key contributor to effectively mitigate first and last-mile problems by connecting various modes of the transportation network [2].

However, range anxiety remains a significant concern pertaining to the utilization of any e-mobility tool, undermining user confidence and hindering the widespread adoption of e-mobility as a preferred mode of transportation [3]. To mitigate this concern, it is paramount to gain a comprehensive understanding of energy consumption patterns and, consequently, establish reliable methods for estimating energy consumption. This is crucial in facilitating trip planning for users, thereby enhancing their satisfaction with the services, particularly within the context of shared mobility. In this regard, some recent studies have used mathematical modelling techniques. However, these approaches are often challenging to prove their effectiveness in real-world scenarios due to the lack of an open dataset for further model evaluation and verification [4], [5]. Recent advancements in AI have positioned "data" at the forefront of progress across various vertical domains. By harnessing relevant data effectively, we can undertake diverse prediction tasks. For instance, data-driven approaches have become pivotal in energy modelling tasks, however, in the context of e-mobility in general and micro e-mobility in particular, there is an apparent gap which calls for open datasets for energy modelling related studies [6], [7].

As per [8], the manufacturer-reported maximum travel ranges exceed owner-reported averages by about 30%. This discrepancy stems from inadequate energy prediction practices within these companies, leading to ineffective monitoring and control of vehicle battery usage and energy efficiency. Consequently, it creates a disparity between advertised and actual performance, undermining consumer trust and satisfaction as experiences fail to meet expectations. However, current energy models [9], [10] simplify the complex energy dynamics of these systems, often missing crucial details. Moreover, studies [10]–[12] in energy modelling lack detailed insights into travel patterns and data-driven methods which are essential to improve the accuracy and effectiveness of energy consumption predictions. We note that only a limited number of studies have primarily focused on comprehensive datasets pertaining to micromobility. The work [13] indicates a gap in current research dedicated to dataset development and refinement in this specific domain. Even when available, current datasets [10], [14], [15] often lack details such as rider weight, terrain type, State of Charge (SoC), and weather conditions, which can affect the efficacy of energy modelling for various e-mobility tools. Moreover, the specifications provided by electric micromobility manufacturers are often not precise, particularly when it comes to their products' maximum travel range, as this is typically assessed under optimal and ideal conditions, making it challenging for users to estimate the energy needed for real-world journeys.

These aforementioned limitations in current work urgently underscore the need for comprehensive and openly accessible datasets pertaining to Electric Scooter (E-Scooter)s and Electric Bike (E-Bike)s, including crucial factors affecting energy consumption such as rider weight, terrain variations, and weather conditions. Furthermore, exploring a diverse range of

Y. Ding and H. Fang are with the SFI Centre for Research Training in Machine Learning (ML-Labs) at Dublin City University. S. Yan, M. H. Shah and M. Liu are with the SFI Insight Centre for Data Analytics and the School of Electronic Engineering, Dublin City University, Dublin, Ireland. J. Li is with the Department of Mechanical Engineering, University of Birmingham.

Y. Ding and S. Yan are joint first authors. *Corresponding author: Mingming Liu. Email:* mingming.liu@dcu.ie.

modelling approaches beyond regression-based methods [10], [16] would contribute to a better understanding of energy consumption in the context of e-mobility frameworks. To bridge the identified research gaps, our work presents an openly accessible dataset covering both E-Scooters and E-Bikes, with meticulous consideration for critical factors in energy consumption. Our analysis extends beyond conventional regression-based models, incorporating diverse modelling techniques, including both Machine Learning (ML) and deep learning, as well as its comparison with contemporary physical models.

The rest of the paper is organised as follows. In Section II, we provide a literature review of current datasets and related methods. In Section III, we introduce the data collection and processing pipeline and provide an overview of the E-Bike and E-Scooter datasets. In Section IV, we present the models and describe the experimental results and analysis. The experiment results and relevant discussion are illustrated in Section V. Finally, we sum up the discussions and limitations with a future pathway in Section VI and Section VII.

II. Related Works

In this section, we review multiple datasets pertaining to E-Bike and E-Scooter in detail. We identify and discuss some limitations of these datasets and highlight how our work compares and contributes to the existing body of knowledge. A summary of the comparison is presented in Table II.

The existing datasets for EM studies are limited in both quantity and quality. One notable dataset [14] originates from the Chicago Pilot Programme, documenting E-Scooter trips capturing the commencement and conclusion of each journey during the 2019 initiative. This dataset was enhanced in 2020 with additional data columns pertaining to the vendor names. Despite its breadth, the dataset exhibits certain shortcomings including instances of missing data, non-descriptive trip identifiers, and errors such as identical start and end times. In effect, this dataset resembles an unprocessed data file, lacking in-depth trip details and omitting crucial energy consumption metrics. The Moby Bikes Dataset [18], centred on E-Bikes in Dublin, offers insights into trip details but is marred by Global Positioning System (GPS) data error issues and the absence of energy consumption information. Similarly, the Micromobility dataset [17], encompassing E-Scooters and E-Bikes from July 2019 onwards, is updated weekly but echoes the shortcomings of the Chicago dataset with a notable absence of energy consumption or State of Charge (SoC) information.

Regarding energy consumption modelling, the recent works [11], [12] primarily revolve around mathematical models that account for factors such as the braking system, energy load, etc. However, these models overlook the travel patterns of users. Moreover, there seems to be a gap in the integration of data-driven methods into these studies. This oversight could potentially limit the accuracy and applicability of the models, as user travel patterns can significantly influence energy consumption. The study in [10] is relevant to our focus on energy consumption modelling which utilises Swedish E-Scooter data on timestamps, locations, and SoC. However, its dataset is not publicly available, which constrains broader research utility. The study's data collection lacks direct E-Scooter trajectory information, instead estimating travel metrics from origin-destination points with a minor time error due to data update frequency. Unlike our research that includes both E-Scooters and E-Bikes, it only considers E-Bike data and also lacks a thorough comparison for different data-driven methods. It also overlooks key energy consumption factors like riders' physical attributes and weather conditions.

Comparing with relevant works, a summary of our major contributions is outlined in the following:

- Performed thorough real-world data collection for both E-Scooters and E-Bikes, using a detailed and transparent methodology, resulting in the creation of an accessible General Data Protection Regulation (GDPR)-compliant open dataset.
- Performed an extensive analysis using various Artificial Intelligence (AI) techniques to model energy consumption and compared them with traditional physical models.

III. Electric Micromobility Dataset

We introduce the data collection procedure and present our data processing flow for the dataset. The dataset will be used for further analysis in Section IV.

A. Pipeline of the E-mobility Trip Data Collection

This section details the data collection process and encapsulates the process for the datasets. The complete process of data collection is wrapped up as a standard pipeline presented in Fig. 1 and introduced in detail in the following sections.

1) E-Bike: The data collection pipeline is listed below.
Experiment Object: For E-Bike, we experimented with the model *Electric Trekking Bike T1*[1], a mountain bike equipped with a 250 W motor and capable of speeds up to 25 km/h. Powered by a 450 Wh battery, the E-Bike has a range of 100 km and features a power assist module that offers five levels of assist/electric mode. These modes correspond to speeds of 12, 16, 20, 23, and 25 km/h. The dataset includes trip attributes for various pedal assist levels to mirror real-world scenarios.
Data Collecting Device: We employed an *iGPSPORT iGS630 GPS bike computer*[2] with *LivLov V2 Bike Cadence and Speed Sensors*[3] and the smart cycling application of the E-Bike to gather journey attributes such as timestamps, GPS coordinates, altitude, speed, assistance level, and travel distance. In terms of SoC data, unfortunately, the E-Bike's built-in SoC display provides insufficient granularity (only 5 levels). This limitation is common in contemporary E-Bikes. To address this, we employed a custom *Voltage Logger* sensor[4] to continuously monitor real-time battery voltage. This sensor consists of a voltage-divider circuit, an *ESP8266* microcontroller, and a

[1] https://eleglide.com/products/removable-battery-100km-range-electric-trekking-touring-bike-t1
[2] https://www.igpsport.com/igs630-highlights
[3] https://amzn.eu/d/j2OiRMi
[4] https://ie.rs-online.com/web/p/data-loggers/1799537

TABLE I
COMPARISON OF DATASETS IN E-MOBILITY RESEARCH

Dataset	**Coverage**	**Quality**	**Opensource**	**Focus**
Chicago Pilot Programme [14]	E-Scooter trips, Chicago	Missing data, errors	✓	Trip details, lacks energy metrics
Micromobility Dataset [17]	E-scooters and E-bikes	Lacks energy data	✓	Trip details
Gothenburg Dataset [10]	E-Scooter, Gothenburg	Trajectories, energy data	✗	Energy consumption
Moby Bikes Dataset [18]	E-bikes, Dublin	Errors, lacks energy data	✓	Trip details
Our work	E-scooters and E-bikes, Dublin	Trajectories, energy data	✓	Trip details, energy consumption

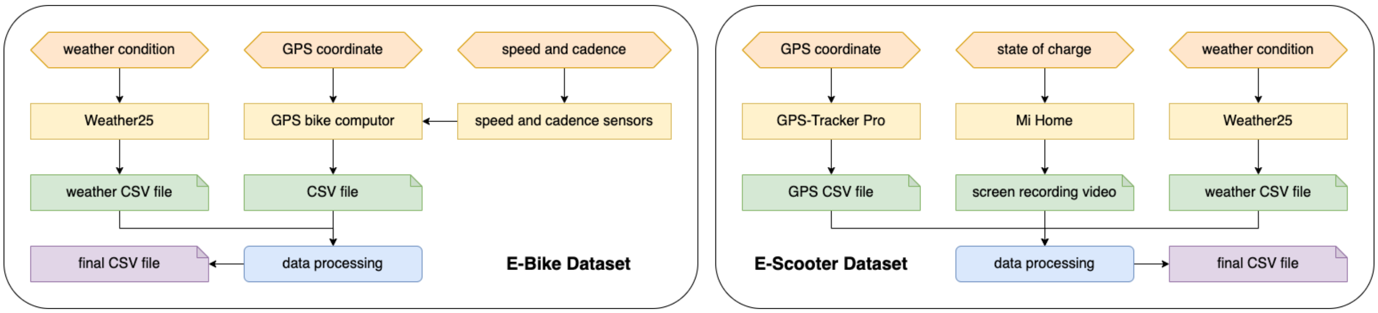


Fig. 1. The process of data collection and dataset creation.

relay for isolation, as shown in the upper green box in Fig. 2. The microcontroller applies the voltage divider equation to calculate battery voltage and stores the data in a CSV file.

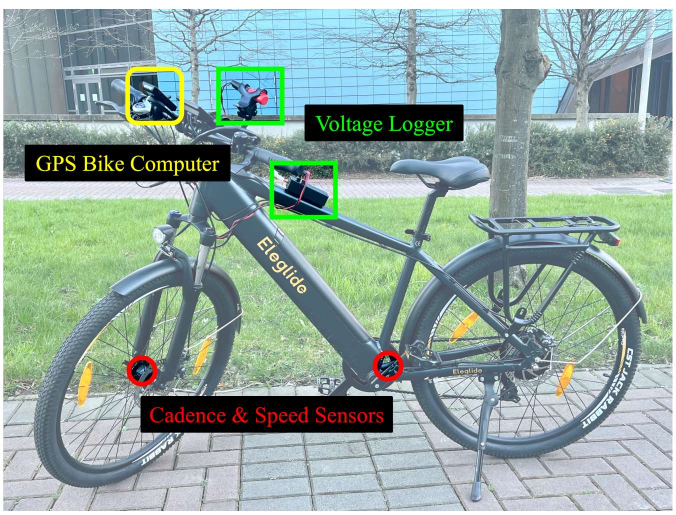
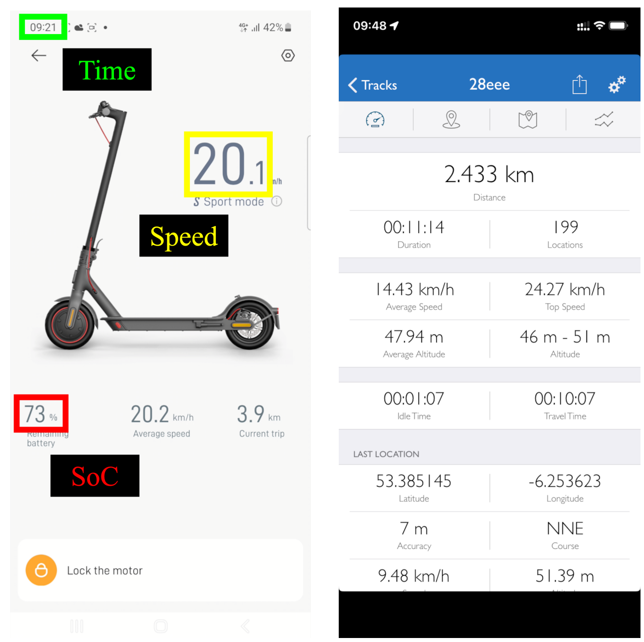


Fig. 2. E-Bike and sensors for experimentation.

Data Collection Process: Before each E-Bike trip, we activated the GPS bike computer, initialised the voltage logger, set the assistance level to a constant value, fixed the gear ratio, and commenced the ride. We stopped data collection at the trip's conclusion and switched off all the devices.

2) *E-Scooter:* The data collection pipeline is listed below.
Experiment Object: For E-Scooter studies, we chose the popular *Mi Electric Scooter Pro 2* model [5] due to its widespread availability and representative features. This model is equipped with a powerful 600-watt motor and offers three speed modes, each with its own speed range. During our data collection, we selected the sports mode, limiting the speed range to 0-25 km/h. The scooter is outfitted with a high-capacity 446 Wh



(a) Mi Home interface. (b) GPS-Tracker Pro interface.

Fig. 3. Sample user interfaces of data collecting devices.

lithium battery, providing a maximum range of 45 km on a full charge and a top speed of 25 km/h.

Device Specification: For the data collection device, two mobile applications were employed to gather attributes of E-Scooter trips, containing timestamp, GPS coordinates, altitude, speed, and SoC. The SoC was extracted from screen recordings using *Xiaomi*'s official mobile application, *Mi Home*[6], as shown in the red box in Fig. 3(a), on an Android device (*SAMSUNG Galaxy A53*), while other attributes were directly acquired via a GPS tracking mobile application, *GPS-Tracker*

[5]https://www.xiaomi.ie/mi-electric-scooter-pro-2

[6]https://play.google.com/store/apps/details?id=com.xiaomi.smarthome

Pro[7], as shown in Fig. 3(b), on an Apple device (*iPhone 11*).

Data Collection Process: Before commencing each E-Scooter trip, we initiated the data collection applications on two mobile devices and concluded data collection upon trip completion. After multiple iterations of data collection, we proceeded to process and conduct a thorough analysis of the data gathered.

B. Data Processing

1) E-Bike Data: The key steps for data processing are:

Data Integration: Data from both sources (the GPS bike computer and sensors) is merged into a unified file.

SoC Calculation: We exported the recorded voltage data for offline analysis. Although voltage was measured throughout the trip with a 1-second sampling rate, only the first and last data points were used to calculate the total SoC drop, representing the overall energy consumption for the trip. A nonlinear equation (1), customised for lithium polymer battery characteristics [19], [20], is then used to establish a precise correlation between voltage V and SoC. The E-Bike used in our study is equipped with a lithium polymer battery. We illustrate the characteristics of one battery cell, i.e., *Tenergy 18650 battery cell*[8], in Fig. 4, where the blue and yellow lines represent the actual and fitted curves, respectively. The SoC parameters in the yellow curve are obtained by least squares fitting with the *Tenergy 18650 battery cell* data specification.

$$V = k_0 + k_1 \cdot SoC + \frac{k_2}{SoC} + k_3 \cdot \ln(SoC) + k_4 \cdot \ln(1 - SoC) \quad (1)$$

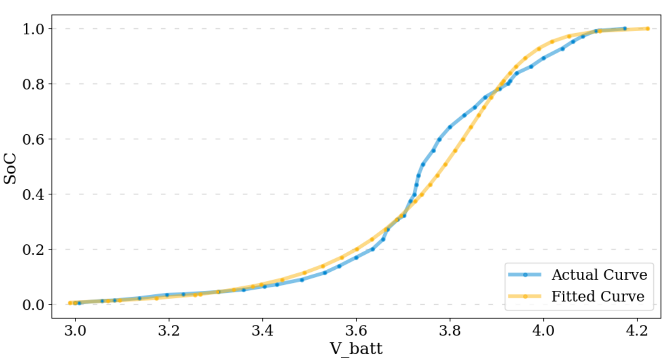


Fig. 4. Lithium polymer battery characteristics (SoC against voltage).

Energy Efficiency Calculations: Energy consumption is measured by analysing the decrease in SoC relative to battery capacity. Energy efficiency (Wh/km) is calculated by dividing the consumed energy by the travel distance collected from the odometer, providing insights into battery usage.

Weather Data Incorporation: Historical weather data (wind speed, direction, and descriptions) from Weather25[9] is integrated to enrich the dataset.

Synthetic Data Generation: To overcome the limitations of our small real-world dataset, we employed synthetic data generation to augment our training set. Python's Synthetic Data Vault library[10], which has been applied in many studies [21]–[23], is used in our work to generate a 10,000-record supplement to the established real-world dataset. This library uses a variety of ML algorithms, such as Gaussian Copula and CTGAN, to learn data patterns to generate synthetic data, enhancing the robustness and generalisation of data-driven models. Through rigorous validation to ensure consistency with the original dataset, we found that the quality of the generated data was 80.34%.

2) E-Scooter Data: The key steps for data processing are:

Digit Extraction: As the position of the digits representing SoC in the screen recording video is fixed, we selected this area as the Region of Interest (RoI), as shown in the yellow box in Fig. 3(a). When pixels in RoI change, we mark the previous and next frames that changed as keyframes and extract the digits by pytesseract[11], an optical character recognition tool for Python. We then conduct manual validation on the extracted data to guarantee its accuracy. The extracted digits are stored along with the corresponding video progress for data alignment.

Data Alignment: Data from two mobile applications is integrated using timestamps, a common attribute in GPS data and screen recordings. By adjusting the RoI and labelling keyframes with video progress, we can accurately correspond video to timestamps, ensuring precise SoC changes and effective data integration.

Weather Data Incorporation & Data Generation: We employ the same methods described for E-Bike data to enrich the dataset with weather information and generate a 10,000-record supplement to the established real-world dataset. The overall quality score of the generated data reaches 81.23%.

The dataset is primarily composed of two parts: specific driving data for each journey and a comprehensive summary table for all trips. Our study results in a dataset consisting of two parts: 36 E-Bike trips and 30 E-Scooter trips. Table II and Table III provide an overview of some features from this collected data. A sample E-Scooter trip (Trip 27) is analysed and visualised in Fig. 5, in which Fig. 5(a) shows the speed changes along the trip trajectory, while Fig. 5(b) presents the speed (green line) and SoC (orange line) trends over time. For access to the complete dataset, we suggest interested readers visit our GitHub repository[12].

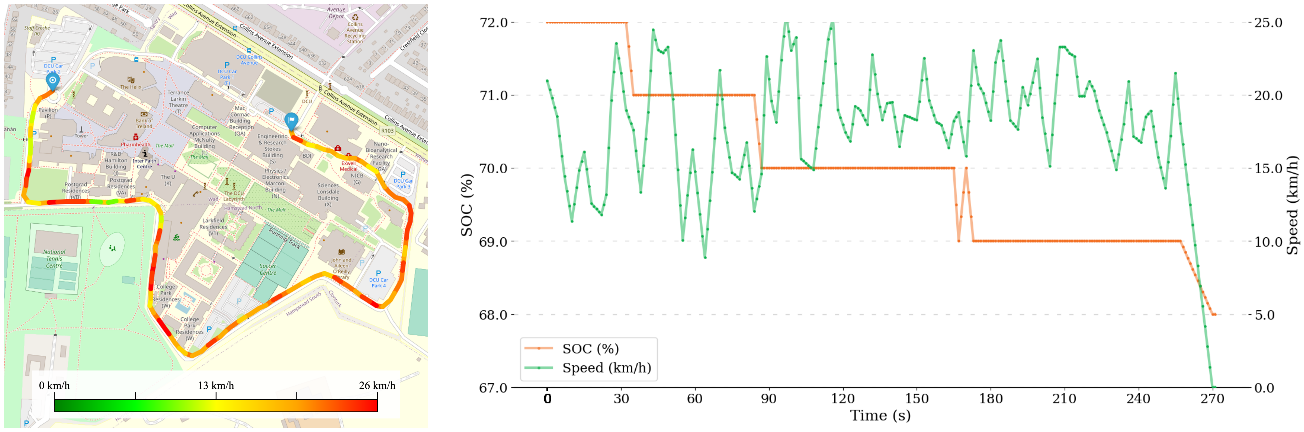
IV. Models and Experiments

In this section, we introduced the experiments using mathematical and data-driven models to forecast energy consumption, discuss the results and analyse the implications for energy management in EM devices.

A. Mathematical Model

In our study, we adopted the mathematical energy consumption model initially proposed in [9], customising it to align with real-world application scenarios specific to our research.

[7] https://apps.apple.com/us/app/gps-tracker-pro/id984920064
[8] https://www.tenergy.com/30005_datasheet.pdf
[9] https://www.weather25.com/europe/ireland/leinster/dublin
[10] https://docs.sdv.dev/sdv
[11] https://github.com/madmaze/pytesseract
[12] https://github.com/SFIEssential/DualEMobilityData-datasets



(a) One sample E-Scooter trip

(b) E-Scooter data analysis

Fig. 5. Data analysis and visualisation of partial trajectory (starting from the 50th sample point) in E-Scooter Trip 27.

TABLE II
E-BIKE DATASET

timestamp	**latitude**	**longitude**	**altitude** (m)	**speed** (km/h)	**wind speed** (km/h)	**wind direction**	**weather**	**temperature** (C)
06/07/2023 09:04:23	53.3854	-6.2564	56.8	18	16.9	S	Cloudy	16
06/07/2023 09:04:24	53.3854	-6.2564	56.8	18	16.9	S	Cloudy	16
06/07/2023 09:04:25	53.3854	-6.2565	56.8	18	16.9	S	Cloudy	16
06/07/2023 09:04:26	53.3854	-6.2565	56.8	18	16.9	S	Cloudy	16
06/07/2023 09:04:27	53.3855	-6.2566	56.8	18	16.9	S	Cloudy	16

TABLE III
E-SCOOTER DATASET

timestamp	**latitude**	**longitude**	**speed** (km/h)	**altitude** (m)	**SoC** (%)	**wind speed** (km/h)	**wind direction**	**weather**
12/10/2023 13:35:24	53.3916	-6.2637	0.0000	57.2203	97	3	E	Dry
12/10/2023 13:35:31	53.3915	-6.2633	2.1430	56.8687	97	3	E	Dry
12/10/2023 13:35:48	53.3915	-6.2633	0.0000	56.8241	97	3	E	Dry
12/10/2023 13:36:06	53.3915	-6.2634	8.1311	56.7919	97	3	E	Dry
12/10/2023 13:36:11	53.3915	-6.2636	8.0638	55.9300	97	3	E	Dry

According to [24]–[26], it is recognised that motor power is mainly affected by road slope, friction, and air resistance, and we can ascertain some physical parameters. We use m, s, and v to represent the mass of the device and user, the road slope, and the driving speed respectively. The acceleration of gravity g is approximately 9.81 $\mathrm{m/s^2}$. For asphalt surfaces, the rolling resistance coefficient C_r is typically 0.001. The air density ρ is given as 1.29 $\mathrm{kg/m^3}$. Additionally, the combined frontal area A of the device and rider is 0.5 $\mathrm{m^2}$, and the drag coefficient C_d is 0.7. To calculate the energy consumption per unit distance, we further refined the model and arrived at equation (2), which serves as the total energy required for E-Scooter, i.e., $P_{es} = P_d$. For E-Bike, we introduce the power assistance level P_{al}, which ranges from 0 (no assistance) to 1 (full assistance), and formulate our prediction target, i.e., energy consumption demand per unit distance, as $P_{eb} = P_d \cdot P_{al}$.

$$P_d = g \cdot m \cdot s + C_r \cdot m \cdot g + \frac{1}{2} C_d \cdot \rho \cdot A \cdot v^2 \tag{2}$$

B. Data-Driven Models

Nine data-driven algorithms, including Linear Regression (LR), Support Vector Regression (SVR), Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), K Nearest Neighbour (KNN), eXtreme Gradient Boosting (XGB), Light Gradient-Boosting Machine (LGBM), and Multi-Layer Perceptual (MLP), are chosen to develop predictive models for energy consumption in E-Bikes and E-Scooters. Each algorithm is selected based on its unique analytical strengths and attributes, which are essential for addressing the complexities and various factors influencing energy usage.

C. Experiment Setup

Our experimental datasets for both e-bikes and e-scooters contain the following feature categories:

Rider Features: height range (cm), weight range (kg).

Trip Features: *1) E-Bike:* distance (km), average speed (km/h), total ascent (m), average slope (%), power assistance level; *2) E-Scooter:* average speed (km/h), altitude difference (m), average slope (%).

Weather Features: *1) E-Bike:* wind speed (km/h), wind direction, weather description, precipitation (mm), temperature (°C); *2) E-Scooter:* wind speed (m/s), wind direction, weather description.

To process non-numeric features, we employed the following techniques:

Label Encoding: distinct numerical labels were assigned to textual weather descriptions, such as “Dry” and “Wet”.

Vectrisation: wind direction was converted into two numerical columns, *WE* and *NS*, using the Cartesian coordinate system, with the positive directions of the x and y axes representing east and north respectively.

We initially applied the LR, SVR, DT, RF, GB, KNN, XGB and LGBM to implement the prediction on two datasets, which were both divided into training, testing, and validation sets with an 8:1:1 ratio. Then, the MLP model was designed and optimised by the *Adam* optimiser with a default learning rate of $1e^{-3}$. The XGB and LGBM regressors were implemented using their respective Python libraries, while other models were built using *Tensorflow Keras* and *Sklearn* libraries.

We conducted experiments both with and without rider features (height and weight ranges) to assess their impact on prediction results. While these features cannot be removed from the mathematical consumption model due to their physical significance, their influence on ML model performance was analysed. In this study, Watt-hours per kilometre (Wh/km) is the primary metric for assessing energy consumption efficiency in electric vehicles. Lower Wh/km values indicate better efficiency. Model performance was assessed based on their Mean Absolute Error (MAE) in predicting energy consumption, and lower MAE values signify higher accuracy.

V. Results and Discussion

The performance of various predictive models in estimating the energy consumption of E-Bike and E-Scooter is compared in Table IV, measured in Watt-hours per kilometre. In each experiment, the MLP model consistently exhibits superior performance, as evidenced by its MAE values of 4.47, 4.62, 4.77, and 6.32, surpassing other models. This result suggests that while advanced models like MLP can achieve remarkable accuracy, the choice of model should be tailored to the specific characteristics of the data. Following the removal of users' height and weight features, it shows Table IV that the MAE values are consistently higher than those before removal for most models (it slightly dropped for KNN). This increase in MAE indicates a decline in prediction accuracy, underscoring the importance of user-specific features in the model's performance.

VI. Limitations

This section delves into the discussions surrounding our study's findings and critically evaluates the inherent limitations therein. We classify our limitations from the views of Electric Micromobility devices, data sources, contributors to data acquisition, and data collection process.

In our study, we did not compare devices across different models and vendors, hence our work should be regarded as an initial exploration in this domain. Additionally, our devices are relatively new, which may omit considerations for wear-and-tear or ageing effects, commonly referred to as 'friction' in the longevity of device performance.

Further, our verification of the collected data was limited; for instance, we did not use multiple GPS sensors, speed sensors, or altitude measuring instruments to cross-validate the location, altitude, and speed data. Moreover, we relied on external web sources for climate details, which provided data with coarse granularity (hourly), lacking finer temporal resolution that might affect the accuracy of our analysis. Also, our research was primarily focused on evaluating models utilising average speed as a basis for trip planning. However, we did not investigate detailed aspects of modelling beyond this parameter. This will be part of our future work.

Finally, the data for our study was gathered through voluntary participation, which constrained the size of our dataset. To address this, we employed the Synthetic Data Vault method to augment our data pool. However, this method only simulates the distribution of the original data, which may not accurately reflect the diversity of potential user profiles. Additionally, our approach was not personalised to individual users, mainly due to compliance with the GDPR, limiting our capability to carry out customised energy prediction models and optimisation based on data collected from any specific user for travel behavior analysis. For instance, the height and weight data were collected in ranges, which could affect the precision of personalised modelling. Additionally, our data collection was geographically restricted to Dublin, Ireland, and primarily conducted on university campuses due to health and safety considerations. This limited the diversity of environmental conditions in our data, potentially affecting the applicability of our findings across different terrains and temporal variations.

VII. Conclusion and Future Work

Our research represents a pivotal step in bridging the data shortage in understanding the energy consumption of EMs. We introduce two meticulously collected open-source benchmark datasets for E-Bike and E-Scooter, aiming to capture and analyse the usage patterns of these vehicles. This comprehensive dataset we proposed contributes to the current gap in relevant datasets with detailed trip information, especially energy consumption. Through in-depth experiments and analysis, we found that our MLP model improved prediction accuracy for E-Bikes and E-Scooters. Also, the user-specific features are important for the improvement of models' performance. As part of our future work, we aim to explore the limitations mentioned to enrich our research and practical application. We will use GDPR-compliant historical data to improve predictions and employ federated learning, as introduced in [27], [28], to further enhance accuracy in model performance in a collaborative learning setup. We also plan to diversify our dataset with various terrains for robust, precise energy consumption forecasts and incorporate detailed trajectory data for real-time monitoring and in-depth data analysis.

Acknowledgment

This work has emanated from research supported in part by Science Foundation Ireland under Grant Number *21/FFP-P/10266* and *SFI/12/RC/2289_P2* (Insight the SFI Research Centre for Data Analytics, at Dublin City University). Yue Ding was also supported by the SFI Centre for Research Training in Machine Learning (ML-Labs) at Dublin City

TABLE IV
PERFORMANCE (MAE) OF DISTINCT APPROACHES (WH/KM) BEFORE AND AFTER FEATURE REMOVAL.

Approach	E-Bike		E-Scooter	
	with user features	without user features	with user features	without user features
Mathematical model	29.39	–	27.87	–
LR	4.68	4.84	4.85	6.44
SVR	4.66	4.80	4.87	6.40
DT	6.03	6.49	6.85	9.09
RF	4.63	4.74	4.98	6.57
GB	4.59	4.73	4.83	6.44
KNN	5.47	5.42	5.40	6.95
XGB	4.69	4.86	5.13	6.84
LGBM	4.57	4.70	4.82	6.55
MLP	**4.47**	**4.62**	**4.77**	**6.32**

University. We extend special thanks to our Chief Technical Officer *Paul Wogan* at the School of Electronic Engineering, Dublin City University, for his invaluable support in designing, equipping and integrating the voltage logger into our E-Bike.

The studies involving human participants were reviewed and approved by Data Protection Office, Dublin City University. Written informed consent for participation was required for this study in accordance with the national legislation and the institutional requirements. The project has received ethical approval from Dublin City University Research Ethics Committee with reference number *DCUREC/2023/025*.

REFERENCES

[1] I. D'Adamo, M. Gastaldi, and I. Ozturk, "The sustainable development of mobility in the green transition: Renewable energy, local industrial chain, and battery recycling," *Sustainable Development*, vol. 31, no. 2, pp. 840–852, 2023.

[2] R. J. van Kuijk, G. H. de Almeida Correia, N. van Oort, and B. Van Arem, "Preferences for first and last mile shared mobility between stops and activity locations: A case study of local public transport users in utrecht, the netherlands," *Transportation Research Part A: Policy and Practice*, vol. 166, pp. 285–306, 2022.

[3] D. Pevec, J. Babic, A. Carvalho, Y. Ghiassi-Farrokhfal, W. Ketter, and V. Podobnik, "A survey-based assessment of how existing and potential electric vehicle owners perceive range anxiety," *Journal of cleaner Production*, vol. 276, p. 122779, 2020.

[4] I. Miri, A. Fotouhi, and N. Ewin, "Electric vehicle energy consumption modelling and estimation—a case study," *International Journal of Energy Research*, vol. 45, no. 1, pp. 501–520, 2021.

[5] Y. Xie, Y. Li, Z. Zhao, H. Dong, S. Wang, J. Liu, J. Guan, and X. Duan, "Microsimulation of electric vehicle energy consumption and driving range," *Applied energy*, vol. 267, p. 115081, 2020.

[6] C. Muli, S. Park, and M. Liu, *A Comparative Study on Energy Consumption Models for Drones*. Springer International Publishing, 2022, p. 199–210.

[7] S. Yan, M. H. Shah, J. Li, N. O'Connor, and M. Liu, "A review on ai algorithms for energy management in e-mobility services," in *2023 7th CAA International Conference on Vehicular Control and Intelligence (CVCI)*. IEEE, Oct. 2023.

[8] Rider Guide, "Do electric scooter companies lie about their stats?" https://riderguide.com/guides/electric-scooter-companies-specifications/, 2023, accessed: 2024-03-07.

[9] E. Burani, G. Cabri, and M. Leoncini, "An algorithm to predict e-bike power consumption based on planned routes," *Electronics*, vol. 11, no. 7, p. 1105, Mar. 2022.

[10] Y. Wang, J. Wu, K. Chen, and P. Liu, "Are shared electric scooters energy efficient?" *Communications in Transportation Research*, vol. 1, p. 100022, Dec. 2021.

[11] M. N. Yuniarto, S. E. Wiratno, Y. U. Nugraha, I. Sidharta, and A. Nasruddin, "Modeling, simulation, and validation of an electric scooter energy consumption model: A case study of indonesian electric scooter," *IEEE Access*, vol. 10, p. 48510–48522, 2022.

[12] K. N. Genikomsakis, G. Mitrentsis, D. Savvidis, and C. S. Ioakimidis, "Energy consumption model of electric scooter for routing applications: Experimental validation," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, Oct. 2017.

[13] H.-H. Schumann, H. Haitao, and M. Quddus, "Passively generated big data for micro-mobility: State-of-the-art and future research directions," *Transportation Research Part D: Transport and Environment*, vol. 121, p. 103795, 2023.

[14] City of Chicago, "E-scooter trips - 2019 pilot," 2019, accessed: 21.11.2023. [Online]. Available: https://data.world/cityofchicago/2kfw-zvte

[15] City of Chicago, "E-scooter trips - 2020," 2020, accessed: 21.11.2023. [Online]. Available: https://data.world/cityofchicago/3rse-fbp6

[16] Y. E. Ayözen, "Statistical optimization of e-scooter micro-mobility utilization in postal service," *Energies*, vol. 16, no. 3, 2023.

[17] City of Norfolk, "Micromobility (electric scooters and bikes) - norfolk," 2020, accessed: 21.11.2023. [Online]. Available: https://data.norfolk.gov/Government/Micromobility-Electric-Scooters-and-Bikes-/wqxq-hhe6

[18] Government of Ireland, "Moby Bikes," 2024, accessed: 12.03.2024. [Online]. Available: https://data.gov.ie/dataset/moby-bikes

[19] G. L. Plett, "Extended kalman filtering for battery management systems of lipb-based hev battery packs," *Journal of Power Sources*, vol. 134, no. 2, p. 262–276, Aug. 2004.

[20] H. Rahimi-Eichi, B. Balagopal, M.-Y. Chow, and T.-J. Yeo, "Sensitivity analysis of lithium-ion battery model to battery parameters," in *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, Nov. 2013.

[21] A. Kiran and S. S. Kumar, "A comparative analysis of gan and vae based synthetic data generators for high dimensional, imbalanced tabular data," in *2023 2nd International Conference for Innovation in Technology (INOCON)*. IEEE, Mar. 2023.

[22] F. Specht, J. Otto, and D. Ratz, "Generation of synthetic data to improve security monitoring for cyber-physical production systems," in *2023 IEEE 21st International Conference on Industrial Informatics (INDIN)*. IEEE, Jul. 2023.

[23] G. Visani, G. Graffi, M. Alfero, E. Bagli, F. Chesani, and D. Capuzzo, "Enabling synthetic data adoption in regulated domains," in *2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, Oct. 2022.

[24] W. J. v. Steyn and J. Warnich, "Comparison of tyre rolling resistance for different mountain bike tyre diameters and surface conditions," *South African Journal for Research in Sport, Physical Education and Recreation*, vol. 36, no. 2, pp. 179–193, 2014.

[25] B. Upadhya, R. Altoumaimi, and T. Altoumaimi, "Characteristics and control of the motor system in e-bikes," 2014.

[26] W. M. Bertucci, S. Rogier, and R. F. Reiser, "Evaluation of aerodynamic and rolling resistances in mountain-bike field conditions," *Journal of sports sciences*, vol. 31, no. 14, pp. 1606–1613, 2013.

[27] M. Liu, "Fed-bev: A federated learning framework for modelling energy consumption of battery electric vehicles," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, 2021, pp. 1–7.

[28] S. Yan, H. Fang, J. Li, T. Ward, N. O'Connor, and M. Liu, "Privacy-aware energy consumption modeling of connected battery electric vehicles using federated learning," *IEEE Transactions on Transportation Electrification*, pp. 1–1, 2023.