IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 5 December 2023, accepted 22 January 2024, date of publication 30 January 2024, date of current version 6 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3359990

RESEARCH ARTICLE

Estimating Average Vehicle Mileage for Various Vehicle Classes Using Polynomial Models in Deep Classifiers

NAGHMEH NIROOMAND[1] AND CHRISTIAN BACH[2]

[1]School of Management and Law, Zurich University of Applied Sciences, 8400 Winterthur, Switzerland
[2]Automotive Powertrain Technologies Laboratory, Swiss Federal Laboratories for Materials Science and Technology, 8600 Dübendorf, Switzerland

Corresponding author: Naghmeh Niroomand (naghmeh.niroomand@zhaw.ch)

ABSTRACT Accurately measuring vehicle mileage is pivotal in precise CO_2 emission calculations and the development of reliable emission models. Nonetheless, mileage data gathered from surveys relying on self-estimation, garage reports, and other estimation-based sources often yield rough approximations that substantially deviate from the actual mileage. To tackle this issue, we present a comprehensive framework aimed at bolstering the accuracy of CO_2 emission models. This paper harnesses two innovative techniques: the deep learning semi-supervised fuzzy C-means (SSFCM) and polynomial classifier models. By leveraging these sophisticated mathematical techniques, we achieve successful classification of passenger vehicles, enabling more precise evaluations of average mileage. Real data shows that vehicles in Switzerland considerably exceed the estimated mileage in the years following the first registration of the vehicle. The difference lies in the covered mileage after vehicles reach five years of age. Our framework supports segment-based analysis for assessing average mileage and enhancing emission models for better understanding of vehicle-related environmental impact.

INDEX TERMS Average vehicle mileage, mileage model, CO_2 emissions, deep feature learning, polynomial deep classifiers, vehicle classification.

I. INTRODUCTION

The adoption of the Paris agreement over 8 years ago [1], which aimed to mitigate global warming to a level below 1.5°C, has not yielded favorable results. Global greenhouse gas emissions persistently continue to rise, which is a cause for concern. The 2016 EU Reference Scenario indicates that without a determined commitment to decarbonization, carbon dioxide (CO_2) emissions from transportation are forecasted to experience a modest reduction of only 8% between 2010 and 2050, ultimately peaking by 2050 [2], [3]. Various factors contribute to this feeble progress, including a significant proliferation of passenger cars, sluggish uptake of electric vehicles, and a restricted transition to alternative fuels. These factors hinder progress and impede the substantial mitigation of emissions.

According to the International Energy Agency [4], Switzerland's contribution to global anthropogenic CO_2 emissions from fossil fuels is less than 0.2%. However, the transportation sector has a substantial impact on Switzerland's overall carbon footprint, constituting around 30.6% of the nation's CO_2 emissions in the year 2021. Among the various transportation modes, road transport is predominantly responsible, accounting for 97.3% of these emissions. Passenger cars, specifically, constitute a significant portion of Swiss road transport emissions, making up approximately 71.2% of the total emissions [5]. It is worth noting that the normative CO_2 emissions from passenger cars in Switzerland have displayed a fluctuating pattern. After experiencing a continuous decline since 2003 for both gasoline and diesel vehicles, the normative CO_2 emissions witnessed a slight increase in 2017 due to the partial introduction of the new WLTP normative measurement procedure for European type approval and a significant rise in 2021 due to its full introduction. While the introduction of the new normative CO_2 measurement procedure had a significant impact on the normative CO_2 emissions, no impact on the CO_2 emissions on the road are expected; however, the difference

The associate editor coordinating the review of this manuscript and approving it for publication was Jesus Felez.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

between normative and real CO_2 emissions could be reduced significantly [6]. Estimating CO_2 emissions involves employing calculation models that heavily rely on factors such as the vehicle fleet composition, fuel parameters, and average mileage of the vehicles [7], [8], [9], [10].

Due to the lack of standardization in estimating vehicle mileage, which varies greatly between periodic technical inspections (PTI), garage reports, and individual estimations, accurately determining the true CO_2 emissions from road traffic has become increasingly challenging and unreliable. Additionally, the implementation of new carbon dioxide legislation, which includes an EU fleet average normative emission target of 95 g CO_2/km according to the old measurement procedure, has resulted in significant changes in new immatriculated vehicle fleet composition, as well as the technical and dimensional characteristics of vehicles over time [11]. Despite advancements in technology and measures such as purchasing new vehicles and scrapping old or damaged ones, Swiss passenger car fleet continues to have high CO_2 emissions. Therefore, understanding the relationship between estimated and actual mileage of passenger cars and the impact of these differences on CO_2 emissions is of utmost importance in achieving the goal of zero net CO_2 emissions by 2050.

Hence, this study aims to develop a mathematical model to calculate average vehicle mileage for different vehicle segments, thereby improving the accuracy of CO_2 emissions calculations. Given the limited informative value of CO_2 standard values for real emissions, this approach represents an important step towards a new CO_2 assessment of road traffic. The study builds upon previous work focused on developing a machine learning methodology for the segmentation of passenger cars based on technical and dimensional features [12], [13], [14]. Fig. 1 illustrates the core challenge of vehicle segmentation in this context.

Our primary objective was to enhance the accuracy of CO_2 emission calculations and gain a deeper understanding of the impact of variations in vehicle class on the CO_2 footprint of passenger vehicle fleets. To achieve this, we employed a meticulous approach by categorizing passenger vehicles based on their technical and dimensional characteristics [14]. This segmentation allowed for better analysis of the intricate variations within each class (intra-class) as well as comparisons between different classes (inter-class). By doing so, we aimed to comprehend the diverse factors influencing the calculation of accurate average vehicle mileage across the passenger vehicle fleet. In our approach, we conducted a comparative analysis of various semi-supervised clustering algorithms to predict labels obtained from unsupervised clustering algorithms. Our focus was on utilizing a feature learning technique, which effectively learns representations in datasets with high dimensionality and significant uncertainties [15], [16], [17], [18], [19], [20], [21], [22], [23]. Additionally, our research aimed to develop a model for calculating average vehicle mileage for both inter-class and intra-class scenarios, thereby improving the accuracy of CO_2 emission calculations and understanding the impact of vehicle class variations on the CO_2 footprint of passenger vehicle fleets [9]. Ultimately, this study serves a greater purpose by facilitating a better understanding of the impact vehicle class variations have on the overall CO_2 footprint of passenger vehicle fleets. With more precise calculations and deeper insights, we can drive advancements toward reducing emissions.

Section II briefly introduces the Swiss motor vehicles system. Section III presents the related research. Section IV describes the methods. Section V provides concise details on the used datasets, the algorithms, the performed experiments and the discussion of the results and last, section VI provides the majors findings of our work and recommendations for further research.

II. SWISS MOTOR VEHICLES AND CO_2 EMISSIONS

Switzerland registered over 6.6 million motor vehicles in 2023. Out of these, more than 4.7 million were passenger cars. On average, these vehicles are used for nine years. Despite a high rate of the population accepting public transport modes (59%), car travel still accounts for two thirds of the total passenger kilometers [24]. In 2023, the collective distance covered annually by these vehicles amounts to 55 billion kilometers, with an average daily distance of 20.8 kilometers. As reported by the Federal Office for Spatial Development, this is equivalent to a rate of 100,000 kilometers per minute. [25]. Switzerland records vehicle odometer readings during periodic technical inspections (PTI). New cars undergo their first PTI after 5 years, followed by a second test for cars after three more years. Subsequent tests are required every two years. The cantonal road traffic office in Switzerland manages and standardizes PTIs, maintaining an extensive vehicle database with odometer readings. Additionally, there was a consistent decline in the average normative CO_2 emissions for newly registered cars, dropping from around 190 g CO_2/km in 2003 to approximately 134 g CO_2/km in 2016. However, the mean CO_2 emissions of new registrations saw an increase, reaching 137.8 g CO_2/km in 2018. By 2022, the average CO_2 emissions of all new cars were approximately 120.9 g CO_2/km, indicating a decrease of around 9 grams compared to 2021. Despite this reduction, the specified target value of 118 g CO_2/km (measured using the world harmonized light-duty vehicles test procedure (WLTP)) that came into effect in 2022 was not fully achieved. This outcome is primarily attributed to the implementation of the new WLTP measurement method. A real-world factor of 1.4 was applied to NEDC-based CO_2 emissions, while a factor of 1.2 was utilized for WLTP-based CO_2 emissions. During the intermediate period, the factor used was 1.3.

Fig. 2 depicts the monthly progress of CO_2 emissions from newly registered cars between 2012 and 2022. The transition from new European driving cycle (NEDC) to the more accurate WLTP measurement method resulted in higher recorded average CO_2 emissions from new vehicles.

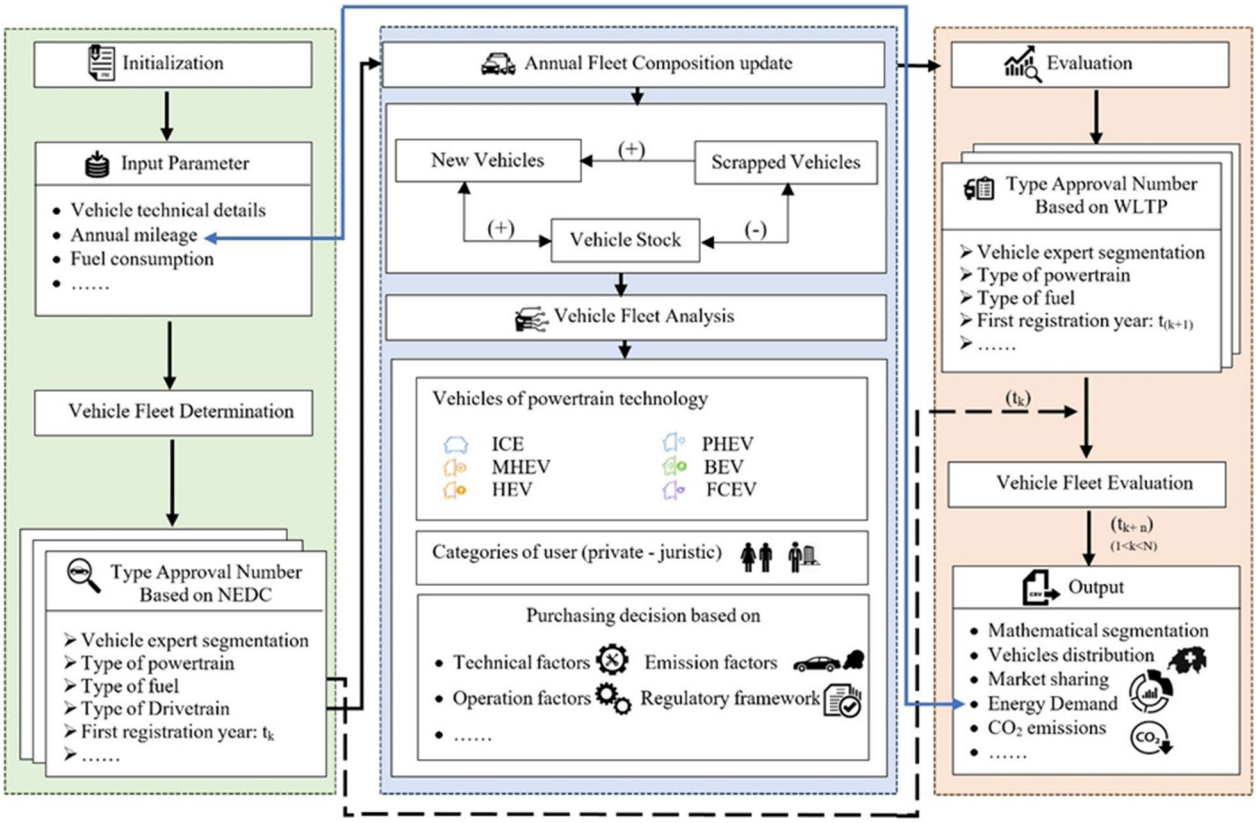

FIGURE 1. Characterizing vehicle fleet composition structure-type data input framework. Internal combustion engine (ICE), mild hybrid electric vehicle (MHEV), full hybrid electric vehicle (HEV), plug-in hybrid electric vehicle (PHEV), battery electric vehicle (BEV), fuel cell electric vehicle (FCEV), new European driving cycle (NEDC) and world harmonized light-duty vehicles test procedure (WLTP).

To prevent a sudden and drastic tightening of the CO_2 target, adjustments were made to align the CO_2 target value with the EU standards [26]. While road traffic in Switzerland has previously operated on its own energy system, which was relatively simple to evaluate in terms of CO_2 emissions, the growing adoption of electric vehicles will complicate the differentiation between energy consumption from road traffic and other stationary energy sources. The development of a precise mathematical methodology to accurately estimate the mileage of passenger vehicles is crucial for determining the actual CO_2 emissions from road traffic in the future.

III. RELATED WORK

Over the last decades, despite achieving partial success in meeting the normative CO_2 emission targets, actual CO_2 emissions in real-world conditions have only experienced a modest decrease of approximately 10% [27]. However, a notable difference of 42% now exists between the estimated and real-world emissions, resulting in a significant discrepancy of 31 g CO_2/km in supposedly saved emissions [28], [29]. One crucial aspect in accurately calculating emissions is determining the average mileage of vehicles, which can be challenging to obtain precise values for or often rely on estimations. Researchers implemented advanced simulation programs to construct comprehensive emission inventories, enhancing the accuracy and reliability of their findings [30], [31], [32], [33], [34], [35], [36]. Simulation programs play a crucial role in bridging the gap between the two primary estimation techniques. Top-down approaches focus on market dynamics, such as fuel consumption patterns and economic factors, to estimate CO_2 emissions on a broader scale. Conversely, bottom-up approaches concentrate on intricate technological details, taking into account factors such as vehicle class, vehicle mileage, and engine efficiency. By employing simulation programs, researchers are able to integrate these complex factors and interactions, specifically in the case of vehicle class and average mileage of vehicle, leading to more precise estimates of CO_2 emissions. These programs simulate real-world scenarios and consider a wide range of parameters, enabling a comprehensive assessment of the environmental impact of different activities and technologies. Consequently, the compilation of emission inventories becomes more reliable and comprehensive. Simulations also prove particularly valuable in compensating for the limitations of laboratory test methods. Traditional lab tests are conducted under controlled conditions, which may not fully capture the diverse and dynamic factors that influence real-world emissions. In contrast, simulation programs enable more realistic and dynamic simulations by considering a broader range of variables and scenarios.

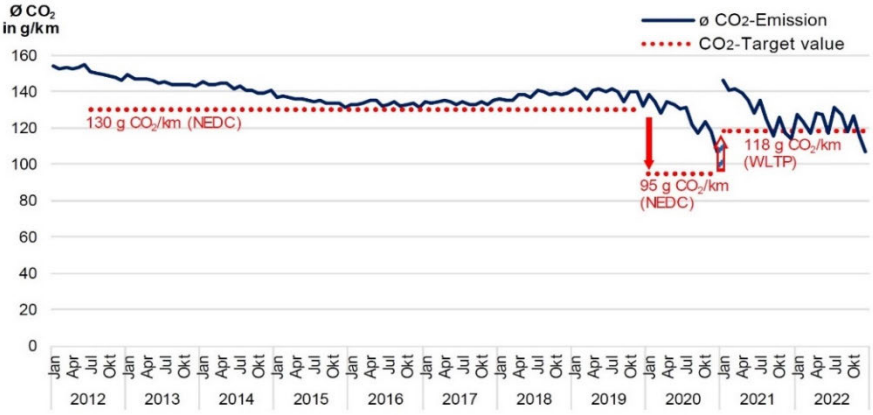

FIGURE 2. Monthly normative CO_2 emissions 2012-2022. Data source: ASTRA (IVZ/TARGA), BFE (CO_2 enforcement data).

Jimenez et al. [37] conducted a review focusing on the influence of vehicle classification, vehicle characteristics, vehicle brand, and registration year on real-world CO_2 emissions. The researchers utilized a database consisting of 650 passenger cars. Their study aimed to elucidate how these factors contribute to the disparity between real-world emissions and type-approval emission values. Hiselius et al. [38] suggested targeting CO_2 emission reduction in the upper quintiles to have a more significant impact compared to uniform reductions across all quintiles. However, eliminating passenger mileage in the sustainable category contributes only minimally to achieving the required one-third reduction. Pejić et al. [39] devised a model that utilizes the age of vehicles and their population size to determine the average mileage. The model assumes an annual reduction in mileage of 5% for passenger cars and small delivery vehicles, 5% for medium trucks, 9.1% for large trucks, and 9% for buses.

However, limitations exist in simulation techniques when it comes to considering variations in emissions within vehicle classes and conducting detailed analyses. Feature learning techniques show promise in addressing uncertainties and improving classification but have been underutilized in predicting vehicle CO_2 emissions on high-dimensional datasets [40], [41]. Ghahramani and Pilla [42] employed a combination of deep learning and support vector machine (SVM) model to forecast CO_2 emissions through energy consumption and mileage monitoring. The model demonstrated a high level of accuracy in its predictions, as evidenced by the low value of the Root Mean Square Error. Pei et al. [43] introduced a method to estimate emissions and mileage using driving cycle data. Their approach incorporates temporal features and a clustering method, leading to improved accuracy. The proposed driving cycle construction technique eliminates the need for manual parameters and is evaluated using visualizations and the COPERT emission model. Experimental results demonstrate significant enhancements in accuracy and robustness. Chrysos et al. [44] provided a principled approach to study state-of-the-art classifiers as polynomial expansions. The research highlighted the prevalence of polynomial functions in various classifiers and elucidated their underlying design principles within a unified framework. The suggested framework can be applied to compress models or enhance model performance.

In this research, our primary aim was to address the challenges posed by diverse methodologies used to estimate average mileage and CO_2 emissions. To achieve this, we developed simulation programs with the goal of enhancing the accuracy of emission estimations. Among the various simulation-based approaches, we utilized a combination of feature extraction methods and deep learning techniques. This approach proved effective in overcoming the limitations associated with conventional laboratory test methods and significantly improving the accuracy of emission models.

IV. MATERIALS AND METHODS

A. SEMI-SUPERVISED CLUSTERING

Semi-supervised clustering endeavors to optimize cluster accuracy by identifying superior clusters in comparison to those obtained through unsupervised learning algorithms [18], [45], [46], [47]. Traditionally, semi-supervised clustering techniques yield subpar results when represented in the original feature space. To enhance the effectiveness of semi-supervised clustering, integrating deep feature learning [15], [48], [49], [50] is rational. The framework of the suggested clustering approach is depicted in Fig. 3.

In contrast to commonly employed methodologies in semi-supervised clustering that rely on feature extraction techniques, our approach integrates three different types of information (diffusion labels, extracted core data, and extracted feature vectors) in order to improve classification accuracy and tackle challenges such as imbalanced class distribution and overlapping among multiple classes.

Our proposed framework includes four primary layers, where the first three layers have been previously discussed in a prior study [14]. In the initial layer, we partition the labeled data into separate training and testing sets which are used for constructing and evaluating classifiers, respectively. In the second layer, the training set is utilized along with

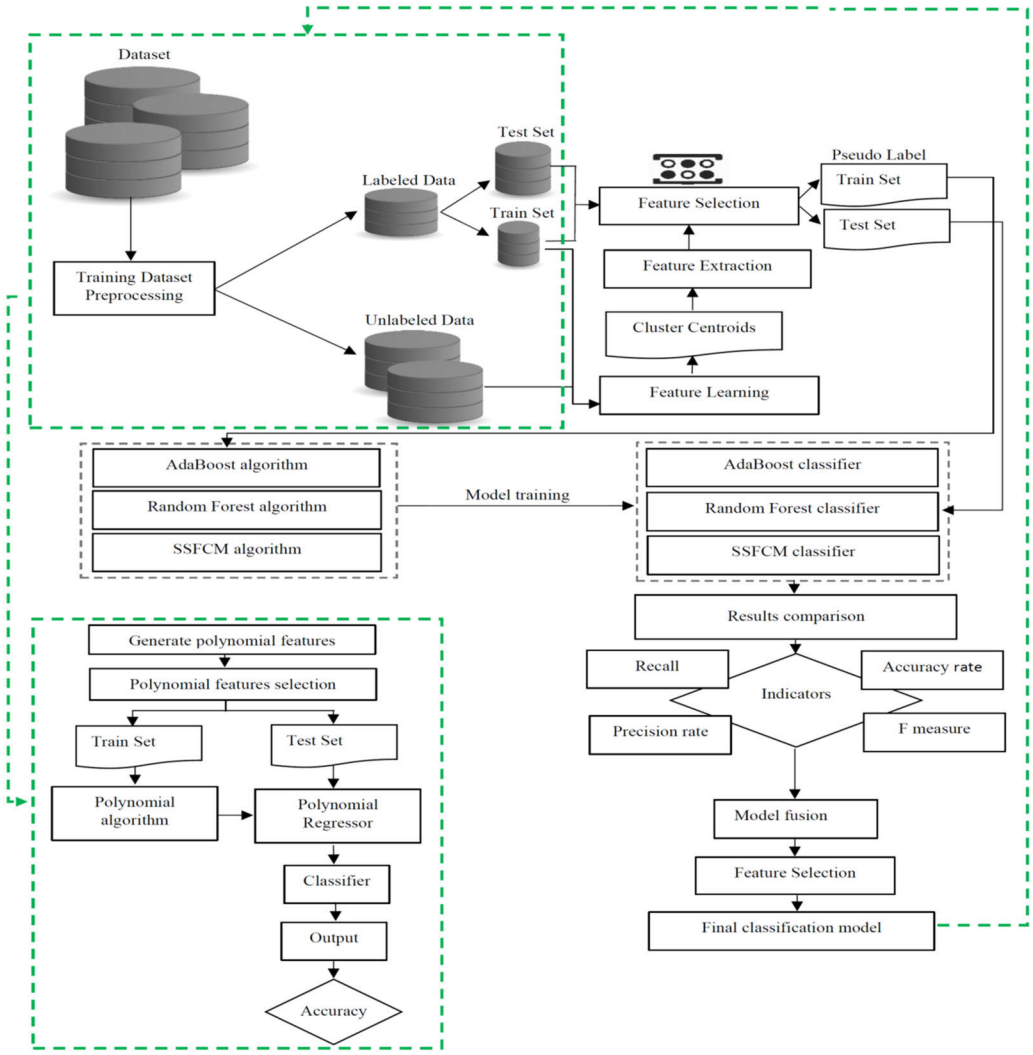

FIGURE 3. The structure of the proposed semi-supervised deep learning and Polynomial regression approach.

unlabeled data as input for the feature learning process. The output of this step yields cluster centroids, which serve as a basis for projecting data from both the training and testing sets into a newly learned space. Furthermore, this projection allows for the extraction of feature vectors during the subsequent feature extraction step. In the classification step, we construct AdaBoost [51], Random Forest [52], and semi-supervised fuzzy C-means clustering (SSFCM) models using the feature vectors derived from the training set. These models are then utilized to predict labels for the corresponding feature vectors within the testing set. The third layer involves the comparison of performance parameters among the three individual models and a fusion model, with the aim of evaluating their effectiveness in terms of data classification and prediction. Lastly, the experimental outcomes from the third layer are applied to a dataset concerning used cars. In this context, we independently employ the polynomial regression algorithm for each vehicle class, with the objective of establishing a model that accurately calculates the average mileage of a vehicle belonging to a specific class. To validate the coefficients obtained from the experimental model, a representative subset is randomly selected from each class and compared with a real dataset corresponding to the given year.

B. SEMI-SUPERVISED FUZZY C-MEAN CLUSTERING

A semi supervised fuzzy C-means clustering incorporates deep feature learning to further improve its effectiveness and eliminate redundant information [21], [46], [53]. Let u_{ki} be a weighted squared errors function known as membership function and can be defined as follow:

$$u_{ki} = \frac{1}{\sum_{j=1}^{C} \left(\frac{D_{kiA}}{D_{kjA}}\right)^{2/(m-1)}} \tag{1}$$

where C is the number of clusters; m is a weighting exponent that determines the degree of fuzziness and that was set to 2 in order to ensure high membership values for each data point to its closest cluster; A is a positive and symmetric (n $\times$ n) weight matrix. The calculation for the updated cluster center

is as follows:

$$v_i = \frac{\sum_{k=1}^{N} u_{ki}^m X_k}{\sum_{k=1}^{N} u_{ki}^m} \tag{2}$$

This method aims to minimize the objective function (J) as follows:

$$Min\ J\ (X; U, V) = \sum_{k=1}^{N} \sum_{i=1}^{C} u_{ki}^m \|X_k - v_i\|_A^2 \quad (1 \le m < \infty) \tag{3}$$

$$s.t.\ \sum_{i=1}^{c} u_{ki} = 1 \quad (0 \le u_{ki} \le 1) \tag{4}$$

where N is number of data elements, X_k represents the data k of X = {X1,X2,X3,…,XN} in the i^{th} cluster; U is the fuzzy partition matrix of the dataset X into c cluster; v_i is vectors of center in ith cluster; K denotes the features, and $\|x_k - v_i\|_A^2$ denotes to the Euclidean distance function and it is computed in the A norm between j^{th} data and i^{th} cluster center.

C. STEPS OF DEEP SEMI-SUPERVISED FUZZY C-MEAN CLUSTERING ALGORITHMS

The SSFCM algorithm comprises the following steps:

Algorithm 1 Membership and Centroid of FCM

Input: N data elements X = {X1, X2,…,XN}, weight matrix (A), number of clusters (C), degree of fuzziness (m=2), max iteration number (T), error threshold (ε)
Output: u_{ki}, v_i
Set t = 0
1. Initialize centroid vectors v_i
2. Update t = t + 1
3. Calculate membership degrees u_{ki}
4. Calculate updated centroid vectors v_i
5. Until $||u_t - u_{t-1}|| < \varepsilon$ is satisfied, then stop
6. Otherwise repeat from step 3.

Subsequently, algorithm 2 is employed to compute the memberships and centroids of deep FCM.

Algorithm 2 Training Strategies for Deep FCM

Input: N data elements X= {X_1, X_2,…,X_N}, number of clusters (C), clusters feature (K), labeled dataset (L), unlabeled dataset (UN), membership degree (U), max iteration number (T), error threshold (ε)
Output: u_{iL}, u_{iUNL}, v_{iL}^k, v_{iUNL}^k
Set t = 0
1. Initialize v_i^k (random for labeled data)
2. Update t = t + 1
   a) Calculate u_{iL}, u_{iUNL}
   b) Calculate v_{iL}^{k+1}, v_{iUNL}^{k+1}
   c) If the stopping criterion, until $||J_t - J_{t-1}|| < \varepsilon$,is fulfilled for all labeled and unlabeled objective functions, then stop
3. Otherwise repeat from step 2

Then, employing algorithm 3, we select the features (s⊂K) through the utilization of the random oversampling (ROS) technique. The aim of employing the ROS technique is to maintain a balance between the feature subsets of labeled classes and unlabeled data elements [14].

Algorithm 3 Feature Extraction of Deep FCM

Input: N data elements X = {X_1, X_2,…,X_N}, clusters feature (K), labeled dataset (X_L), unlabeled dataset (X_{UNL}), $\mu\ (D)$ mean of the elements of D, set of the centroids (v_{iL}^k, v_{iUNL}^k)
Output: Set of extract features of labeled and unlabeled dataset
Set Q = ∅
1. Calculate $D_{Lk} = \left\| x_{iL} - v_{iL}^k \right\|$
2. Calculate $D_{UNLk} = \left\| x_{iUNL} - v_{iUNL}^k \right\|$
3. Calculate means D_{Lk}&D_{UNLk} of elements $\mu_i\ (D_{iL})$, $\mu_i\ (D_{iUNL})$
4. feature extraction ($f_k(x) = \max(0, \mu\ (D) - D_k)$
   a) for all L and UNL features do
5. Return the set Q

In the following step, we utilize the Euclidean distance technique, which is widely used as a metric to measure similarity or distance between labeled and unlabeled feature vectors. The result is determined by finding the maximum average of the maximum relevant and minimum redundant features between each selected feature of unlabeled data and labeled classes:

$$\max\ Sim_i(X_j, V_L^s) = \min d_{jiL} = \min \left| X_j - V_{iL}^s \right| \quad (1 \le i \le c), X_j \epsilon X_{UNL} \tag{5}$$

Finally, in algorithm 4 the maximum average of the maximum similarity between the selected features are estimated, which is then utilized in the classifiers.

Algorithm 4 SSFCM Classifier

Input: N data elements X={X_1,X_2,…,X_N} with minimum features in any subset (s), set of the centroid (V_{iL}^s, V_{UNL}^s) of selected features
Output: Predicted labeled data (Q= {q_{L+1}, q_{L+2},…, q_{L+N}})
Set Q = ∅
1. For each centroid index i ϵ {1, …, c} do
2. For each data element index j ϵ {1, …, N}, do the following steps:
   a) Employ V_{iL}^s to calculate $max\ Sim_i$
   b) If maximum average of $max\ Sim_i \epsilon\ i^{th}$ labeled class, then
   c) Append X_j to i^{th} labeled class
   d) Update the set Q if a labeled class is achieved
   e) For all $V_{iL}^s \epsilon V_L^s$ do
3. Return the set Q

D. STATE-OF-THE-ART METHODS

To improve the accuracy and performance of classification, two ensemble learning methods, namely Random Forest and

AdaBoost, are utilized [54], [55]. The Random Forest technique employs parallel learning and utilizes bagging for data training. Its purpose is to minimize variance and bias in the model by creating multiple decision trees (sets) from the original data. Importantly, in the parallel process, these decision trees are independent of one another.

Algorithm 5 Random Forests Classifier

Input: Training set (S), number of decision trees in the forest (B), subsample size (μ), maximum iteration number (T)
Output: Set K = $\emptyset$
1. Initialize the iteration number t ϵ {1, …, T} do
2. For each decision tree index b ϵ {1, …, B} do the following steps:
   a) Sample μ instances from S with replacement, creating a subsample set S_t
   b) construct a decision tree K_t using decision tree b on the subsample set St
   c) Add the trained decision tree classifier K_t to set K
3. Return the set K

Conversely, AdaBoost functions as a sequential learning approach that builds decision stumps based on the training data. Each subsequent decision stump in this sequential process depends on the previous one. Specifically, any errors made by the initial decision stump, such as misclassifying a few datasets, impact the subsequent decision stump by assigning higher weights to those particular training data.

Algorithm 6 AdaBoost Classifier

Input: Data X whose number of elements N, training set (S), decision tree in forest (B), subsample size (μ), max iteration number (T)
1. Initialize data weights $\{D_n\}$ to 1/N
2. for t ϵ {1, …, T} do
   a) find best weak classifier $y_m(x)$ by minimizing weighted error function J_m:
   $J_m = \sum_{n=1}^{N} D_n^{(m)} 1[y_m(x_n) \neq t_n]$
   b) compute
   $err_m = \sum_{n=1}^{N} D_n^{(m)} 1[y_m(x_n) \neq t_n] / \sum_{n=1}^{N} D_n^{(m)}$
   c) assign weight $\alpha_m = \log(\frac{1-\varepsilon_m}{\varepsilon_m})$ to classifier $y_m(x)$
   d) update the data weights:
   $D_n^{(m+1)} = D_n^{(m)} \exp\{\alpha_m 1[y_m(x_n) \neq t_n]\}$
   e) Normalize $D_n^{(m+1)}$ to be proper distribution
Output: Make prediction using the final model:
$Y_M(x) = sign(\sum_{m=1}^{M} \alpha_m y_m(x))$

E. PERFORMANCE MEASURE

To evaluate the effectiveness of the various algorithms, we analyze the confusion matrix to calculate metrics. These metrics are used to assess the performance of the algorithms and are outlined below:

		Predicted Value (class *i*)		
		Positive	**Negative**	
Actual Value	**Positive**	*True Positive Prediction (TP)*	*False negative prediction (FN)*	Recall (R_i) $\frac{TP_i}{TP_i+FN_i}$
	Negative	*False Positive Prediction (FP)*	*True negative prediction (TN)*	Specificity $\frac{TN_i}{TN_i+FP_i}$
		Precision (P_i) $\frac{TP_i}{TP_i+FP_i}$	Negative Predictive $\frac{TN_i}{TN_i+FN_i}$	F-Measure $\frac{2P_iR_i}{P_i+R_i}$
Rand Index		$\frac{TP+TN}{TP+FN+TN+FP}$		$(0 \leq RI \leq 1)$
Adjusted Rand Index		$\frac{RI-E[RI]}{\max(RI)-E[RI]}$		$(-1 \leq ARI \leq 1)$

F. MODEL FUSION

The Model fusion method is a deep learning technique that combines multiple classification predictive models with individual weights to improve the final estimation. This approach serves as a more robust meta-classifier by leveraging a majority voting classifier estimator, which helps overcome the limitations of individual classifiers and results in higher classification accuracy. The two commonly used types of voting classifiers are the hard voting classifier and soft voting classifier. The hard voting classifier determines the majority vote by giving equal weights to each classifier (selecting the mode of all predicted labels), while the soft voting classifier calculates the majority vote by assigning different weights to each classifier (considering the probability of all predicted labels). The predictions of the voting classifier can be defined as:

$$H_{vote}(x) = \max\left\{\sum_j lab(x, j, 1), \ldots, \sum_j lab(x, j, c)\right\} \quad (1 \leq j \leq T)(1 \leq c \leq K) \tag{6}$$

$$S_{vote}(x) = \max\left\{\frac{\sum_i p(x, j, 1)}{n_T}, \frac{\sum_i p(x, j, 2)}{n_T}, \ldots, \frac{\sum_i p(x, j, c)}{n_T}\right\} \tag{7}$$

where $H_{vote}(x)$ represent the outcome of the hard voting process. The function lab (x, j, c) acts as an indicator, determining whether x belongs to the label c as calculated by the j^{th} classifier, $S_{vot}e(x)$ represents the result of the soft voting process. The probability $p(x, j, c)$ is associated with the likelihood of the j^{th} classifier surpassing certain threshold values. Here, n_T denotes the total number of classifiers, while k signifies the number of labels.

G. POLYNOMIALS AND DEEP CLASSIFIERS

Polynomials are mathematical expressions that establish a connection between an input variable and coefficients. In the context of regression analysis, polynomial regression is employed to handle data that deviates from the assumptions

of basic models [57], [58]. When combined with ensemble methods, polynomial regression can improve the overall model's generalization performance. This combination has the potential to decrease both bias and variance, resulting in improved predictions for unseen data. A principled approach is adopted to investigate advanced classifiers as polynomial expansions. It is observed that polynomials play a recurring role in various classifiers, and their design choices can be interpreted under a unified framework. Building upon existing methods, we introduce extensions that lead to enhanced classification accuracy. Specifically, we represent state-of-the-art ensemble learning methods as polynomials, allowing us to gain insights into the inductive bias of each vehicle class. This allows for evaluating performance under different changes in the training distribution, such as limited samples per class or a long-tailed distribution.

Algorithm 7 Third-Degree Polynomials

Input: Data X whose number of elements *N,* training set (*S*), polynomial coefficients (C), degree of polynomial (t)

Output:

1. Set t = 3
2. Update t = t - 1
2. Initialize data weights $W^{[n]}$
3. $\Phi_{iL}^{[t]}(X) = CX_{iL}, \Phi_{iL}^{[t-1]}(X) = CX_{iL}, \Phi_{iL}^{[t-2]}(X) = CX_{iL}$
4. $Y_{St} = (\Phi_{iL}^{[t]}(X)\,\Phi_{iL}^{[t-1]}(X))\Phi_{iL}^{[t-2]}(X) + \beta$
5. $Y = \sum_{n=1}^{N}(w^{[n]}\Phi_{iL}^{t} * X_i) + \beta$

V. EXPERIMENTS

A. DATA PREPARATION

In this study, the primary dataset is the Swiss Motor Vehicle Information System (MOFIS) [59]. It contains information about more than 4.7 million passenger vehicles. This information includes various details such as type approval numbers, physical characteristics, weight properties, ownership information, technical specifications, and registration dates. Additionally, we have also incorporated data on vehicle technical specifications and periodic technical inspections from the Technical Type Approval Information provided by the Federal Roads Office (ASTRA) [60] and the Vehicles Expert Partner [61] respectively.

To align with the goal of the paper, we divided the dataset into two parts: a training set and a testing set. The training set consisted of 308,824 newly registered passenger cars in 2018. Initially, a filtering process was applied to remove vehicles that didn't fit the conventional definitions of passenger cars, such as small pickup trucks, standard pickup trucks, vans, special purpose vehicles (SPVs), sports cars, and multi-purpose vehicles (MPVs). These cars were then categorized into various types based on their make, model, and manufacturer code, resulting in 366 unique passenger car types. These types were further classified into classes: 18 in the micro class, 50 in the small class, 110 in the middle class, 84 in the upper middle class, and 104 in the large class and luxury class. Due to limitations of the unsupervised FCM clustering algorithm, only labeled data with true labels and a membership degree higher than 0.95 were used as the core dataset. This core dataset was utilized to extract accurate classifications and serve as the foundation for subsequent training steps. Furthermore, 10% of the data from each class was randomly selected as training labeled samples. Lastly, the used cars dataset [62], consisting of 1,880,417 entries, was utilized. This comprehensive dataset contains essential information about the mileage covered by each car and their estimated age. Its purpose is to facilitate precise predictions concerning the mileage associated with different passenger car types.

B. EXPERIMENTAL SETUP AND RESULTS

The initial analysis revealed a strong correlation between emissions, vehicle segments, sub-segments, and influencing factors. To process the data, a combination of labeled and unlabeled data was used, along with the core dataset, and principal component analysis was applied to address multicollinearity. New features were extracted to reduce the number of features, and a selection process involving resampling and Euclidean distance was used to identify the best features (algorithm 2-4). Pseudo labels were assigned to unlabeled data for pre-training different classification algorithms (algorithm 5-6). Model fusion was performed using labeled data to improve accuracy. The results indicated that the soft voting fusion model and SSFCM algorithm achieved the highest accuracy (Table 1). The final features extracted from the model fusion were used to re-evaluate the single algorithms and select the ultimate classification model. These experimental results demonstrate that the SSFCM algorithm is capable of extracting more valuable information from the

TABLE 1. Evaluation of model performance on a dataset with labeled rate of 10% from each class.

Techniques	Method	Feature Learning Techniques		Feature Extraction Techniques	
		Accuracy Rate	Precision Rate	Training Accuracy	Test Accuracy
Algorithm 4	SSFCM	0.954	0.953	0.952	0.904
Algorithm 5	Random Forest	0.902	0.89	0.903	0.837
Algorithm 6	AdaBoost	0.891	0.871	0.781	0.715
Model Fusion	Hard Voting	0.921	0.935		
	Soft Voting	0.942	0.956		

TABLE 2. Inter-class and intra-class classification of passenger cars using SSFCM in the year 2018.

Vehicle Classification	Power (kW)	Fuel Type
Class representatives: SMART Fortwo 451, TOYOTA Aygo AB1, FIAT 500 312, SUZUKI Jimny FJ		
Micro class	(Q1<54),	
- Not-SUV	(54=<Q2<=66),	Benzin, Diesel
- SUV	(54=<Q2<=66)	
Class representatives: VW Polo AW, AUDI S1 8X, ALFA ROMEO MiTo 955, SUZUKI Vitara LY		
Small class	(Q1<70),	
- Not-SUV	(70=<Q2<=103),	Diesel, Benzin
- SUV	(Q3>103)	
Class representatives: DACIA Duster SR, TOYOTA C-HR AX1, ALFA ROMEO Giulietta 940, VW Tiguan 5N		
Middle class	(Q1<103),	
- Not-SUV	(103=<Q2<=135),	Diesel, Benzin
- SUV	(Q3>135)	
Class representatives: SKODA Octavia 5E, ALFA ROMEO Giulia 952, AUDI A4 B8, HYUNDAI Santa TM		
Upper middle class	(Q1<114),	
- Not-SUV	(114=<Q2<=185),	Diesel, Benzin
- SUV	(Q3>185)	
Class representatives: AUDI A6 4G, BMW 5er G5L, MERCEDES-BENZ AMG 212, AUDI Q7 4L		
Large & luxury class	(Q1<135),	
- Not-SUV	(135=<Q2<=240),	Diesel, Benzin
- SUV	(Q3>240)	

Inter-class (namely micro, small, middle, upper middle, large and luxury class) and intra-class (namely sport utility vehicle (SUV) and non-sport utility vehicle (Not-SUV)), Interquartile power range (Q)

TABLE 3. Accuracy of polynomial model coefficients validated on 10% randomly chosen SSFCM labeled samples within the vehicle classes.

Vehicle classes and sub-classes	polynomial regression (k=3)	R^2 value	Accuracy
Micro class	$y = -1.5608x^3 + 44.105x^2 - 554.94x + 13382$	0.875	0.952
Small class	$y = -1.0057x^3 + 30.805x^2 - 466.24x + 14674$	0.739	0.903
Middle class	$y = 0.2663x^3 - 22.979x^2 - 70.834x + 17908$	0.910	0.964
SUV	$y = 1.128x^3 - 60x^2 + 533.35x + 14567$	0.787	0.926
Upper middle class	y = 1.2903x3 - 58.968x2 + 109.1x + 20252	0.850	0.937
SUV	$y = 0.9501x^3 - 29.58x^2 - 270.75x + 19330$	0.841	0.931
Large & luxury class	y = 6.0331x3 - 234.5x2 + 1931.3x + 18118	0.877	0.943
SUV	y = 3.7257x3 - 140.56x2 + 1063.2x + 16550	0.836	0.904
MPVs	y = -1.9785x3 + 78.987x2 - 1290.5x + 20673	0.824	0.918
Sport Cars	y = -2.489x3 + 100.31x2 - 1226.2x + 17701	0.715	0.817

vehicle dataset, resulting in improved recognition rates compared to other classifiers.

The underlying assumption of feature extraction is that it leads to improved classification results in comparison to the initial classifier's predictions with the original features. In algorithm 7, particularly during the Polynomial features selection step, the inter-class and intra-class classification results obtained from the SSFCM approach are employed on used cars dataset. These results encompass a total of five classes, each accompanied by their respective sub-classes, as described in Table 2.

The extraction of average mileage data has been conducted specifically for used cars within the age range of up to 20 years, focusing on data obtained in the year 2018. Further-

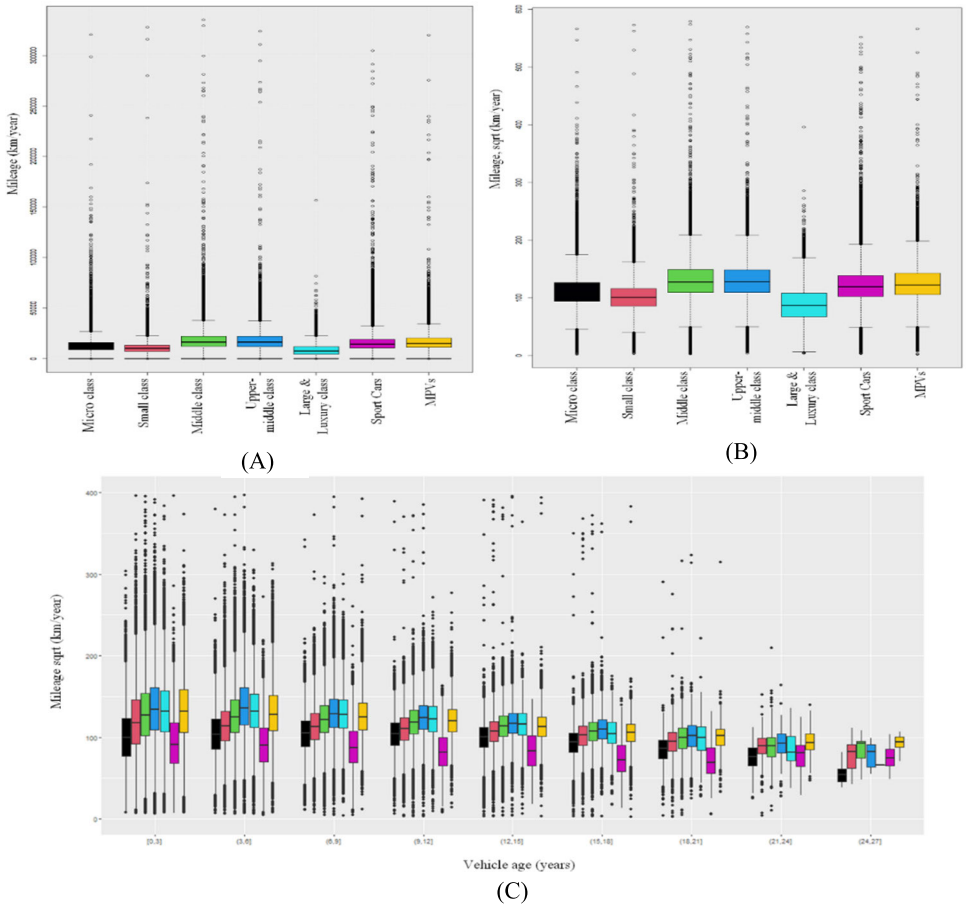

FIGURE 4. Overall comparison of inter-class differences (Boxplots A and B) and mileage-age relationship in each segment (Boxplot C).

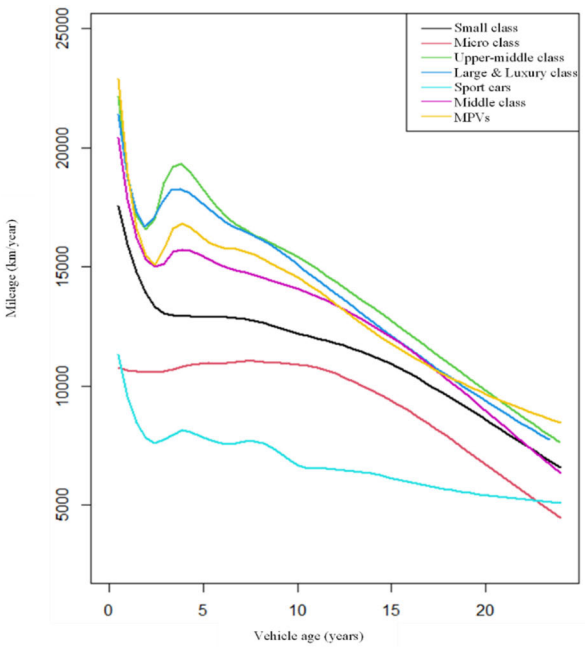

FIGURE 5. SSFCM classifier and polynomial regression performed on each segment.

more, in-depth analysis of the dataset from the year 2015 has been carried out to examine the average mileage data for each vehicle class. Additionally, the dataset has been expanded to include sport cars and MPVs. As a result, there are now seven distinct car segments available for mileage analysis. Rigorous data quality checks are performed to eliminate mileage records with unrealistic values, such as zero mileage or a negative mileage difference between consecutive years for a given vehicle. In Fig. 4, an encompassing comparison of inter-class differences is depicted by employing the utilization of boxplots. Furthermore, it offers a comprehensive overview of the relationship between mileage and age within each distinct class.

Following data refinement, a third-degree polynomial analysis is conducted on the average mileage and age data, Fig. 5. This analysis takes into consideration the life cycle pattern of vehicles, where the highest annual mileage is typically observed at the initial stage, followed by a period of stabilization and gradual decline. Consequently, the utilization of a third-degree polynomial analysis provides a more accurate representation of the actual vehicle operation. To validate the coefficients obtained from the resulting model, a stratified sampling approach is employed based on the number of unique vehicles in some intra-classes.

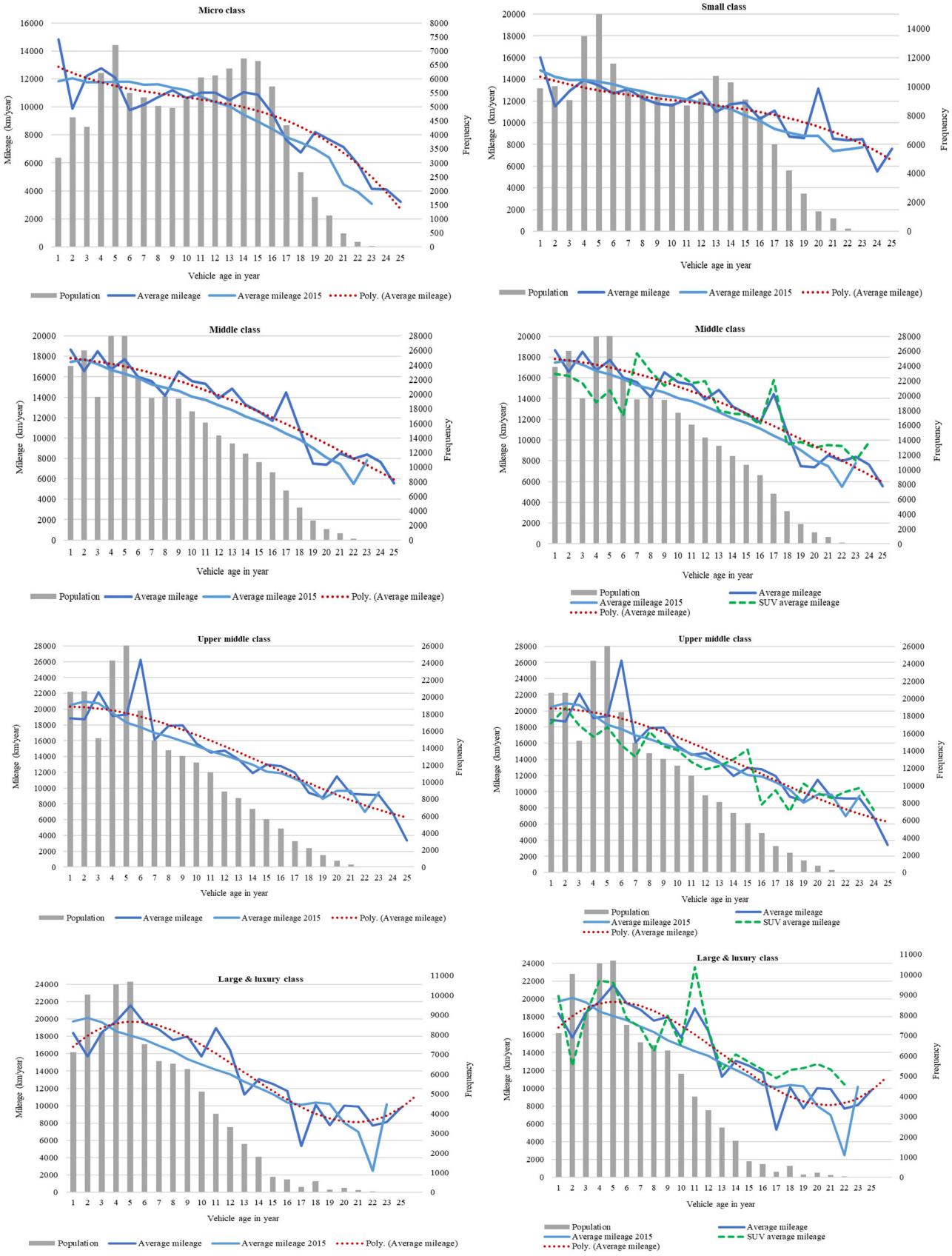

FIGURE 6. Applying a polynomial regression of the third order for each vehicle segment, along with 10% sample of average mileage in some intra-classes as well as the average mileage for the year 2015.

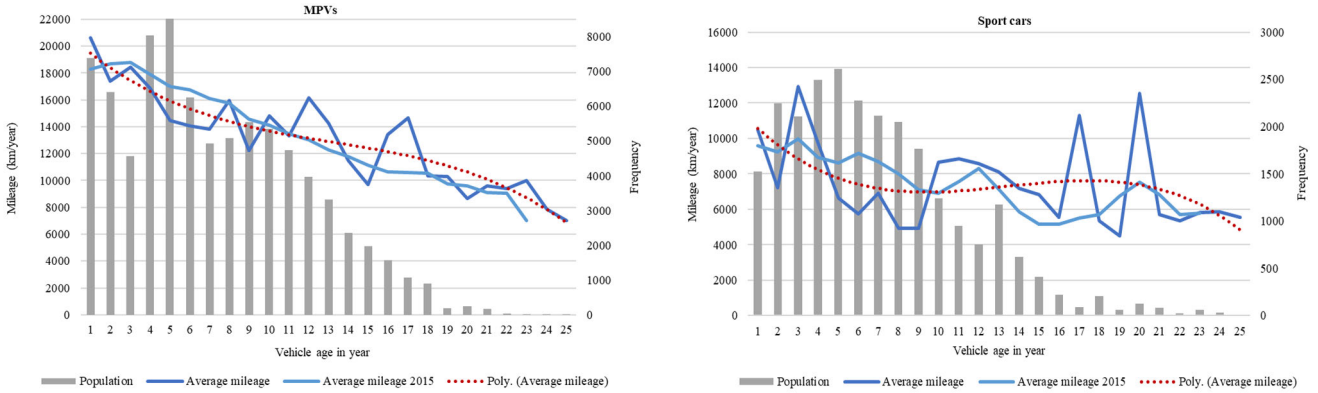

FIGURE 6. *(Continued.)* **Applying a polynomial regression of the third order for each vehicle segment, along with 10% sample of average mileage in some intra-classes as well as the average mileage for the year 2015.**

Specifically, 10% of the data from each class is randomly selected as training labeled samples from SSFCM classifiers, representing their respective classes Fig. 6. Finally, the resulting model is compared to an existing one from 2015 for evaluation and comparison purposes as presented in Table 3.

C. DISCUSSIONS

The experiment results have demonstrated that there is a significant decrease in the overall fleet size for each vehicle class within the age range of up to three years. This reduction in fleet size can be attributed to the ongoing scarcity of used cars that are specifically three years old or younger. These vehicles are consistently 17% less available compared to other age ranges that have slightly higher supply. However, it is important to note that despite this decline in fleet size, the average age of passenger cars in Switzerland has continued to increase throughout the study period. Specifically, the average age of passenger cars has risen from 9 years in 2018 to 9.3 years by the end of 2021. This upward trend suggests that older vehicles are remaining in use for longer periods of time. It could also indicate a growing interest in electric vehicles among some individuals. Furthermore, based on observations, a newly purchased vehicle was found to cover an average distance of 17,935 km annually. However, after 5 years, this annual distance reduced by 25%, and after 10 years, it decreased by 40%. Despite the majority of passenger kilometers being covered by cars in Switzerland, there is a notable variation in mileage between rural and urban areas, particularly for older vehicles. For instance, 10-year-old vehicles in cities travel approximately 20% fewer kilometers on average compared to their rural counterparts.

The distribution of mileage in various segments tends to shift towards higher values. The range of driving performance is also quite extensive, with some vehicles only traveling a few thousand kilometers per year, while others cover several tens of thousands of kilometers. Moreover, the mileage of vehicles is not constant throughout their lifespan. It generally decreases over time, although the decrease is not linear during the first ten years but becomes more linear thereafter. Across all segments, the mileage is halved over a span of 20 years. To estimate the average mileage, we considered the entire operational period. We used a polynomial model that takes into account the vehicle age and population size as input features for each vehicle class. Experimental results demonstrate discrepancies between the estimated data and the actual vehicle data. However, we validated the model by comparing it with the actual data for 2015, as shown in Fig. 7. It is worth noting that the difference mainly arises in the accumulated mileage after vehicles reach five years of age, indicating that used cars generally accumulate more mileage than initially predicted. This underscores the significance of updating the model coefficients every three to five years, leading to recommendations for regular updates. Furthermore, the accuracy of the chosen model coefficients was validated by applying them to a randomly selected sample from within the vehicle class. This test demonstrated their applicability and reliability. Additionally, except for sports cars, we observed a strong positive correlation ($R^2 > 0.90$) between the proposed estimated mileage and the data provided by the federal vehicle control authority for all vehicle classes. Hence, we used distinct approaches to assess the mileage in both cases, and the results exhibit a high level of correlation.

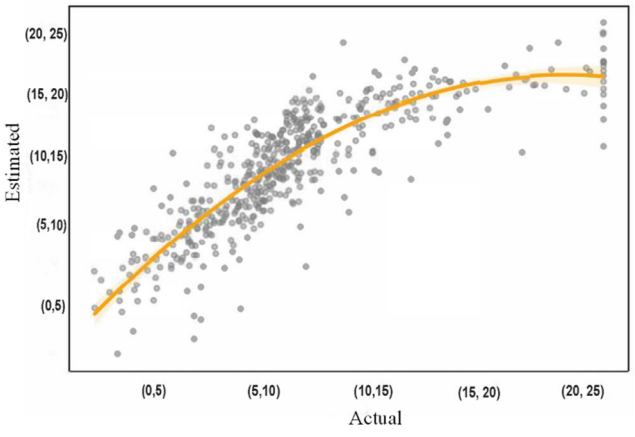

FIGURE 7. **Comparison of actual average mileage and estimated values.**

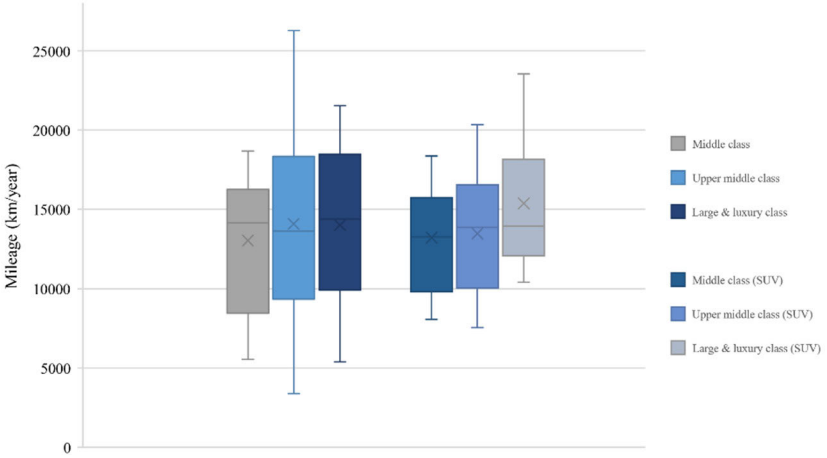

FIGURE 8. Distribution of mileage within selected passenger car segments. Additionally, a 10% sample of average mileage in specific intra-classes is included. Boxplot representation with median and 25/75% quartiles and mean (×) of the mileage of the passenger car segments.

Our previous findings indicated significant variations in average CO_2 emissions among different vehicle classes [14]. This underscores the importance of considering both average mileage within and between vehicle classes to effectively address emission reductions. Additionally, our observations revealed that the average mileage of SUVs tends to increase as vehicles age. This notable finding highlights that the SUV fleet in Switzerland covered an extensive distance of 12.6 billion kilometers in 2018, resulting in the unnecessary production of CO_2 emissions with each kilometer traveled, Fig. 8. Therefore, the integration of inter-class and intra-class classification offers crucial insights for developing strategies to transform the passenger vehicle fleet and promote decarbonization. Utilizing an existing estimation-based model from another country [63], a comparative analysis was conducted using real data from Switzerland. It is important to acknowledge that direct comparisons between two countries with diverse driving fleets, driving behaviors, road infrastructures, and vehicle lifespans may not be straightforward. Nevertheless, these comparisons can provide valuable insights into the key differences. The findings indicate that vehicles in Switzerland greatly surpass the estimated annual mileage in the years following their initial registration.

VI. CONCLUSION

The accurate estimation of average annual vehicle mileage holds immense importance in conducting effective emission analyses and making informed decisions in sustainable transport planning. Incorrect or unreliable mileage values can result in misguided incentives and long-term consequences. Therefore, this study aimed to establish a precise model for calculating average vehicle mileage, enabling a better understanding of the influence of vehicle segments on real CO_2 emissions. To develop the model, extensive analysis of mileage data was conducted for vehicles up to 20 years of age in 2018. Utilizing technical and dimensional features, vehicles were classified based on a mathematical model. Additionally, the model considered population size and vehicle age as inputs for calculating average mileage within each vehicle class. The results demonstrated that the actual mileage covered by vehicles in Switzerland exceeded the estimated mileage, particularly after five years of vehicle age. The model's validity was assessed by comparing it with actual data from 2015, leading to recommendations for updating the model coefficients every three to five years. Additionally, the accuracy of selected model coefficients was affirmed by applying them to a randomly selected sample within the vehicle class, exemplifying their applicability and reliability.

Overall, this study successfully developed a model for accurately calculating average vehicle mileage. The proposed approach offers several advantages, including automated vehicle classification of vast databases, facilitating fleet analysis. The adoption of clustering-based mathematical segmentation also allows for standardized comparisons of databases across different regions. Furthermore, as mileage varies over the age of vehicles, it was observed that the average mileage of SUVs tends to increase over time. As a result, combining inter-class and intra-class classification is essential for gaining valuable insights to formulate fleet transformation strategies aimed at decarbonizing the passenger vehicle fleet. An area that holds promise for future research involves utilizing CO_2 estimates derived from real-world measurements instead of relying solely on type approval values.

This approach would enable a more precise evaluation of fleet CO_2 emissions and further enhance our understanding of the environmental impact of vehicles. Our results emphasize the importance of adjusting the vehicle composition and size to reduce CO_2 emissions. This study's comprehensive analysis and the development of an accurate model for calculating average vehicle mileage contribute to advancing CO_2 emission analysis, informing sustainable transport planning, and paving the way for effective fleet transformation strategies to reduce CO_2 emissions in the passenger vehicle sector.

ACKNOWLEDGMENT

The authors would like to thank the Federal Roads Office (FEDRO) for providing the Swiss Vehicle Information System (MOFIS) data and the vehicle technical dataset and the Vehicle Expert Partners for providing the expert segmentation data.

REFERENCES

[1] UNFCCC. *The Paris Agreement*. Accessed: Dec. 2023. [Online]. Available: https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

[2] Eur. Commission. (2016). *EU Reference Scenario 2016, Energy, Transport and GHG Emissions. Trends to 2050*. [Online]. Available: https://ec.europa.eu/ener-gy/sites/ener/files/documents/ref2016_report_final-web.pdf

[3] P. Capros et al., "EU reference scenario 2016—Energy, transport and GHG emissions—Trends to 2050," Eur. Commission Directorate, Gen. Climate Action Directorate, Gen. Mobility Transp., Luxembourg, 2016, U.K., Tech. Rep., 2050, doi: 10.2833/001137.

[4] *World Energy Outlook 2018*, Int. Energy Agency, Paris, France, 2018.

[5] *Federal Office for the Environment (FOEN)*. Accessed: Oct. 2023. [Online]. Available: https://www.bafu.admin.ch/bafu/en/home/topics/climate/state/data/greenhouse-gasinventory/transport.html

[6] (2016). *International Council on Clean Transportation (ICCT)*. Accessed: Dec. 2023. [Online]. Available: https://theicct.org/wp-content/uploads/2022/01/FactSheet_FromLabToRoad_ICCT_2016_EN.pdf

[7] R. E. Wilson, J. Anable, S. Cairns, T. Chatterton, S. Notley, and J. D. Lees-Miller, "On the estimation of temporal mileage rates," *Proc. Social Behav. Sci.*, vol. 80, pp. 139–156, Jun. 2013.

[8] L. Fridstrøm, V. Østli, and K. W. Johansen, "A stock-flow cohort model of the national car fleet," *Eur. Transp. Res. Rev.*, vol. 8, no. 3, p. 22, Sep. 2016, doi: 10.1007/s12544-016-0210-z.

[9] S. Caserini, C. Pastorello, P. Gaifami, and L. Ntziachristos, "Impact of the dropping activity with vehicle age on air pollutant emissions," *Atmos. Pollut. Res.*, vol. 4, no. 3, pp. 282–289, Jul. 2013, doi: 10.5094/APR.2013.031.

[10] V. Williams, S. McLaughlin, R. McCall, and T. Buche, "Motorcyclists' self-reported riding mileage versus actual riding mileage in the following year," *J. Saf. Res.*, vol. 63, pp. 121–126, Dec. 2017, doi: 10.1016/j.jsr.2017.10.004.

[11] Eur. Commission. *Reducing CO_2 Emissions From Passenger Cars*. Accessed: Dec. 2023. [Online]. Available: https://ec.europa.eu/clima/policies/transport/vehicles/cars_en

[12] N. Niroomand, C. Bach, and M. Elser, "Vehicle dimensions based passenger car classification using fuzzy and non-fuzzy clustering methods," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2675, no. 10, pp. 184–194, Oct. 2021, doi: 10.1177/03611981211010795.

[13] N. Niroomand, C. Bach, and M. Elser, "Robust vehicle classification based on deep features learning," *IEEE Access*, vol. 9, pp. 95675–95685, 2021, doi: 10.1109/ACCESS.2021.3094366.

[14] N. Niroomand, C. Bach, and M. Elser, "Segment-based CO_2 emission evaluations from passenger cars based on deep learning techniques," *IEEE Access*, vol. 9, pp. 166314–166327, 2021, doi: 10.1109/ACCESS.2021.3135604.

[15] W. Shi, Y. Gong, C. Ding, Z. Ma, X. Tao, and N. Zheng, "Transductive semi-supervised deep learning using min-max features," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 11209. Cham, Switzerland: Springer, 2018, pp. 299–315.

[16] X. Zhu. (2008). *Semi-Supervised Learning Literature Survey*. [Online]. Available: http://pages.cs.wisc.edu/~jerryzhu/research/ssl/semireview.html

[17] L. Zhuo, L. Jiang, Z. Zhu, J. Li, J. Zhang, and H. Long, "Vehicle classification for large-scale traffic surveillance videos using convolutional neural networks," *Mach. Vis. Appl.*, vol. 28, no. 7, pp. 793–802, Oct. 2017.

[18] G. Forestier and C. Wemmert, "Semi-supervised learning using multiple clusterings with limited labeled data," *Inf. Sci.*, vols. 361–362, pp. 48–65, Sep. 2016.

[19] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 2006.

[20] S. Melacci and M. Belkin, "Laplacian support vector machines trained in the primal," *J. Mach. Learn. Res.*, vol. 12, no. 3, pp. 1149–1184, Jul. 2011.

[21] A. Arshad, S. Riaz, and L. Jiao, "Semi-supervised deep fuzzy C-mean clustering for imbalanced multi-class classification," *IEEE Access*, vol. 7, pp. 28100–28112, 2019, doi: 10.1109/ACCESS.2019.2901860.

[22] H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1259–1270, Mar. 2018.

[23] Y.-Z. Ren, G.-J. Zhang, and G.-X. Yu, "Random subspace based semi-supervised feature selection," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Jul. 2011, pp. 113–118.

[24] Swiss Federal Office Energy (SFOE). *CO_2 Emission Regulations for New Cars and Light Commercial Vehicles*. Accessed: Sep. 2023. [Online]. Available: https://www.bfe.admin.ch/bfe/en/home/efficiency/mobility/co2-emission-regulations-for-new-cars-and-light-commercial-vehicles.html

[25] Swiss Federal Office of Energy (SFOE). *Mobility Behavior of the Population*. Accessed: Nov. 2023. [Online]. Available: https://www.bfs.admin.ch/bfs/de/home/statistiken/mobilitaetverkehr/personenverkehr/

[26] Swiss Federal Office of Energy (SFOE). (2022). *Vollzug Der CO_2-Emissionsvorschriften Für Personenwagen*. Accessed: Nov. 2023. [Online]. Available: https://www.bfe.admin.ch/bfe/de/home/effizienz/mobilitaet/co2-emissionsvorschriften-fuer-neue-personen-und-lieferwagen/personenwagen-pw.exturl.html/aHR0cHM6Ly9wdWJkYi5iZmUuYWRtaW4uY2gvZGUvc3VjaGU_a2/V5d29yZHM9NDcw.html

[27] Stanford Earth Matters Magazine. (2020). *COVID Lockdown Causes Record Drop in Carbon Emissions for 2020*. Accessed: Dec. 2023. [Online]. Available: https://earth.stanford.edu/news

[28] F. Grelier, "CO_2 emissions from cars: The facts," Eur. Fed. Transp. Environ., AISBL, Brussels, Belgium, Apr. 2018.

[29] G. Fontaras, N.-G. Zacharof, and B. Ciuffo, "Fuel consumption and CO_2 emissions from passenger cars in Europe—Laboratory versus real-world emissions," *Prog. Energy Combustion Sci.*, vol. 60, pp. 97–131, May 2017.

[30] J. Pavlovic, K. Anagnostopoulos, M. Clairotte, V. Arcidiacono, G. Fontaras, I. P. Rujas, V. V. Morales, and B. Ciuffo, "Dealing with the gap between type-approval and in-use light duty vehicles fuel consumption and CO_2 emissions: Present situation and future perspective," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2672, no. 2, pp. 23–32, Dec. 2018.

[31] H. Dai, P. Mischke, X. Xie, Y. Xie, and T. Masui, "Closing the gap? Top-down versus bottom-up projections of China's regional energy use and CO_2 emissions," *Appl. Energy*, vol. 162, pp. 1355–1373, Jan. 2016.

[32] S. D. Tuladhar, M. Yuan, P. Bernstein, W. D. Montgomery, and A. Smith, "A top–down bottom–up modeling approach to climate change policy analysis," *Energy Econ.*, vol. 31, pp. S223–S234, Dec. 2009.

[33] D. P. van Vuuren, M. Hoogwijk, T. Barker, K. Riahi, S. Boeters, J. Chateau, S. Scrieciu, J. van Vliet, T. Masui, K. Blok, E. Blomen, and T. Kram, "Comparison of top-down and bottom-up estimates of sectoral and regional greenhouse gas emission reduction potentials," *Energy Policy*, vol. 37, no. 12, pp. 5125–5139, Dec. 2009.

[34] N. Karali, T. Xu, and J. Sathaye, "Reducing energy consumption and CO_2 emissions by energy efficiency measures and international trading: A bottom-up modeling for the U.S. iron and steel sector," *Appl. Energy*, vol. 120, pp. 133–146, May 2014.

[35] P. Thunis, B. Degraeuwe, K. Cuvelier, M. Guevara, L. Tarrason, and A. Clappier, "A novel approach to screen and compare emission inventories," *Air Qual., Atmos. Health*, vol. 9, no. 4, pp. 325–333, May 2016.

[36] Y. Natarajan, G. Wadhwa, K. R. S. Preethaa, and A. Paul, "Forecasting carbon dioxide emissions of light-duty vehicles with different machine learning algorithms," *Electronics*, vol. 12, no. 10, p. 2288, 2023, doi: 10.3390/electronics12102288.

[37] J. L. Jiménez, J. Valido, and N. Molden, "The drivers behind differences between official and actual vehicle efficiency and CO_2 emissions," *Transp. Res. D, Transp. Environ.*, vol. 67, pp. 628–641, Feb. 2019.

[38] L. W. Hiselius and L. S. Rosqvist, "Segmentation of the current levels of passenger mileage by car in the light of sustainability targets—The Swedish case," *J. Cleaner Prod.*, vol. 182, pp. 331–337, May 2018, doi: 10.1016/j.jclepro.2018.02.072.

[39] G. Pejić, F. Bijelić, G. Zovak, and Z. Lulić, "Model for calculating average vehicle mileage for different vehicle classes based on real data: A case study of Croatia," *PROMET Traffic Transp.*, vol. 31, no. 2, pp. 213–222, Apr. 2019.

[40] Z. He, G. Ye, H. Jiang, and Y. Fu, "Vehicle emission detection in data-driven methods," *Math. Problems Eng.*, vol. 2020, Oct. 2020, Art. no. 4875310, doi: 10.1155/2020/4875310.

[41] C. Saleh, N. R. Dzakiyullah, and J. B. Nugroho, "Carbon dioxide emission prediction using support vector machine," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 114, Feb. 2016, Art. no. 012148, doi: 10.1088/1757-899x/114/1/012148.
[42] M. Ghahramani and F. Pilla, "Analysis of carbon dioxide emissions from road transport using taxi trips," *IEEE Access*, vol. 9, pp. 98573–98580, 2021, doi: 10.1109/ACCESS.2021.3096279.
[43] L. Pei, Y. Cao, Y. Kang, Z. Xu, and Z. Zhao, "UJ-FLAC: Unsupervised joint feature learning and clustering for dynamic driving cycles construction," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 10970–10982, Aug. 2022, doi: 10.1109/TITS.2021.3098353.
[44] G. G. Chrysos, M. Georgopoulos, J. Deng, J. Kossaifi, Y. Panagakis, and A. Anandkumar, "Augmenting deep classifiers with polynomial neural networks," 2021, *arXiv:2104.07916*.
[45] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2247–2256, Aug. 2015.
[46] Y. Ren, X. Hu, K. Shi, G. Yu, D. Yao, and Z. Xu, "Semi-supervised DenPeak clustering with pairwise constraints," in *Proc. 15th Pacific Rim Int. Conf. Artif. Intell.*, 2018, pp. 837–850.
[47] Y. Qin, S. Ding, L. Wang, and Y. Wang, "Research progress on semi-supervised clustering," *Cogn. Comput.*, vol. 11, no. 5, pp. 599–612, Oct. 2019.
[48] A. Arshad, S. Riaz, L. Jiao, and A. Murthy, "Semi-supervised deep fuzzy C-mean clustering for software fault prediction," *IEEE Access*, vol. 6, pp. 25675–25685, 2018.
[49] A. Arshad, S. Riaz, L. Jiao, and A. Murthy, "The empirical study of semi-supervised deep fuzzy C-mean clustering for software fault prediction," *IEEE Access*, vol. 6, pp. 47047–47061, 2018.
[50] G. Chen, "Deep transductive semi-supervised maximum margin clustering," 2015, *arXiv:1501.06237*.
[51] S. Wang and X. Yao, "Multiclass imbalance problems: Analysis and potential solutions," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 4, pp. 1119–1130, Aug. 2012.
[52] A. Verikas, A. Gelzinis, and M. Bacauskiene, "Mining data with random forests: A survey and results of new tests," *Pattern Recognit.*, vol. 44, no. 2, pp. 330–349, Feb. 2011.
[53] S. Riaz, A. Arshad, and L. Jiao, "A semi-supervised CNN with fuzzy rough C-mean for image classification," *IEEE Access*, vol. 7, pp. 49641–49652, 2019, doi: 10.1109/ACCESS.2019.2910406.
[54] T. Hasanin, T. M. Khoshgoftaar, J. Leevy, and N. Seliya, "Investigating random undersampling and feature selection on bioinformatics big data," in *Proc. IEEE 5th Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Apr. 2019, pp. 346–356, doi: 10.1109/BIGDATASERVICE.2019.00063.
[55] R. Kumar and R. Verma, "Classification algorithms for data mining: A survey," *Int. J. Innov. Eng. Technol.*, vol. 1, no. 2, pp. 7–14, 2012.
[56] C. Macdonald and I. Ounis, "Voting for candidates: Adapting data fusion techniques for an expert search task," in *Proc. 15th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, Arlington, VA, USA, Nov. 2006, pp. 387–396.
[57] G. G. Chrysos, S. Moschoglou, G. Bouritsas, J. Deng, Y. Panagakis, and S. Zafeiriou, "Deep polynomial neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4021–4034, Aug. 2022.
[58] Z. Chen, K. Batselier, J. A. K. Suykens, and N. Wong, "Parallelized tensor train learning of polynomial classifiers," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4621–4632, Oct. 2018, doi: 10.1109/TNNLS.2017.2771264.
[59] *MOFIS—Das Motorfahrzeuginformationssystem Der Eidgenössischen Fahrzeugkontrolle (EFKO)*. Accessed: Dec. 2023. [Online]. Available: https://www.experience-online.ch/de/9-case-study/2023-mofis
[60] ASTRA. *Bundesamt Für Strassen*. Accessed: Dec. 2023. [Online]. Available: https://www.astraamin.ch/astra/de/home.html
[61] *Schweizer Partner Für Fahrzeugdaten*. Accessed: Dec. 2023. [Online]. Available: https://www.auto-i-dat.ch
[62] *Autoscout24*. Accessed: Mar. 2023. [Online]. Available: https://www.autoscout24.ch/de
[63] T. Trost, M. Sterner, and T. Bruckner, "Impact of electric vehicles and synthetic gaseous fuels on final energy consumption and carbon dioxide emissions in Germany based on long-term vehicle fleet modelling," *Energy*, vol. 141, pp. 1215–1225, Dec. 2017.

NAGHMEH NIROOMAND received the M.A. and Ph.D. degrees from Eastern Mediterranean University, Cyprus, the IAPM degree from Queen's University, Canada, and the Ph.D. degree from SSPH, University of Lucerne, Switzerland. Since then, she has been a Research Fellow with the Transport and Mobility Laboratory, EPFL Lausanne, and a Senior Scientist with Cambridge Resources International, USA. She is currently a Techno-Energy Economist with the Zurich University of Applied Sciences (ZHAW). Prior to joining ZHAW, she was with the Automotive Powertrain Technologies Laboratory, Swiss Federal Laboratories for Materials Science and Technology (Empa). Her current research interests include vehicle fleet and operational analysis, retro-perspective analyze vehicle specific changes in function of spatial technology and economic frame conditions, and economies of synthetic energy carriers.

CHRISTIAN BACH received the B.Sc. degree in automotive engineering from the Bern University of Applied Sciences, Bern. He performed two internships with the Haagen-Smit Laboratory, California Air Resources Board, El Monte, USA, to study zero and ultra low emission technologies in the transport sector. He is currently the Head of the Automotive Powertrain Technologies Laboratory, Swiss Federal Laboratories for Materials Science and Technology (Empa). He is a Lecturer with ETH Zurich and a member of several expert groups in Switzerland.