Journal of Hazardous Materials 424 (2022) 127330

Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

Forecasting plastic waste generation and interventions for environmental hazard mitigation

Yee Van Fan [a,*], Peng Jiang [b], Raymond R. Tan [c], Kathleen B. Aviso [c], Fengqi You [d], Xiang Zhao [d], Chew Tin Lee [e], Jiří Jaromír Klemeš [a]

[a] Sustainable Process Integration Laboratory – SPIL, NETME Centre, Faculty of Mechanical Engineering, Brno University of Technology - VUT Brno, Technická 2896/2, 616 69 Brno, Czech Republic

[b] Department of Industrial Engineering and Engineering Management, Business School, Sichuan University, Chengdu 610064, China

[c] Chemical Engineering Department, De La Salle University, 2401 Taft Avenue, 0922 Manila, Philippines

[d] Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853, USA

[e] Department of Bioprocess Engineering, School of Chemical and Energy Engineering, Universiti Teknologi Malaysia (UTM), 81310 UTM Johor Bahru, Johor, Malaysia

ARTICLE INFO

Editor: Jörg Rinklebe

Keywords:
Plastic pollution
Machine learning
Environmental hazard mitigation
Microplastic
Clustering analysis
Scenario analysis

ABSTRACT

Plastic waste and its environmental hazards have been attracting public attention as a global sustainability issue. This study builds a neural network model to forecast plastic waste generation of the EU-27 in 2030 and evaluates how the interventions could mitigate the adverse impact of plastic waste on the environment. The black-box model is interpreted using SHapley Additive exPlanations (SHAP) for managerial insights. The dependence on predictors (i.e., energy consumption, circular material use rate, economic complexity index, population, and real gross domestic product) and their interactions are discussed. The projected plastic waste generation of the EU-27 is estimated to reach 17 Mt/y in 2030. With an EU targeted recycling rate (55%) in 2030, the environmental impacts would still be higher than in 2018, especially global warming potential and plastic marine pollution. This result highlights the importance of plastic waste reduction, especially for the clustering algorithm-based grouped countries with a high amount of untreated plastic waste per capita. Compared to the other assessed scenarios, Scenario 4 with waste reduction (50% recycling, 47.6% energy recovery, 2.4% landfill) shows the lowest impact in acidification, eutrophication, marine aquatic toxicity, plastic marine pollution, and abiotic depletion. However, the global warming potential (8.78 Gt CO_2eq) is higher than that in 2018, while Scenario 3 (55% recycling, 42.6% energy recovery, 2.4% landfill) is better in this aspect than Scenario 4. This comprehensive analysis provides pertinent insights into policy interventions towards environmental hazard mitigation.

1. Introduction

Plastic pollution is an emerging global environmental issue (Borrelle et al., 2020); the problem is serious enough that there have been recent calls for an international policy framework to manage it (Nordic Council of Ministers, 2020). Plastic import ban and single-use plastic ban have also been advocated in developing countries (Wen et al., 2021) and developed countries (Charitou et al., 2021). More than 90% of plastic is petroleum-based and non-biodegradable (Zhao et al., 2020). The low degree of circularity and leakage to the environment are among the issues of concern (Klemeš et al., 2021). Lau et al. (2020) suggested that even with immediate and concerted action, it is estimated that 710 Mt of plastic waste would enter aquatic and terrestrial ecosystems around the year 2040. There are growing concerns about microplastics (i.e., polymer particles of diameter under 5 mm) and even nanoplastics (Wong et al., 2020) as a threat to the health of humans (Cox et al., 2019), as well as to aquatic (de Sá et al., 2018) and terrestrial (de Souza Machado et al., 2018) ecosystems. The primary and secondary microplastics are both problematic. The latter consists of particles from the gradual breakdown of plastic products during the use and waste disposal of plastic products in the environment through chemical and physical mechanisms. Microplastics can be found in different environmental compartments; they can be transported through air and water (Wang et al., 2020) and accumulate in food chains (Wong et al., 2020) when ingested by organisms. A recent review of the state of scientific literature on microplastic pollution can be found in Petersen and Hubbart (2021), and

* Corresponding author.
E-mail address: fan@fme.vutbr.cz (Y.V. Fan).

https://doi.org/10.1016/j.jhazmat.2021.127330
Received 27 June 2021; Received in revised form 19 September 2021; Accepted 21 September 2021
Available online 23 September 2021

analysis of bibliometric trends is reported by Zhang et al. (2020). Shen et al. (2020) highlighted that the contribution of plastic to global greenhouse gas emissions (GHG) has been noticeable and growing and could reach up to 1.34 Gt/y by the year 2030.

Plastic pollution, like other major environmental issues such as climate change, has the characteristics of a "wicked problem" that resists clear resolution. Entrenched socio-economic factors and business practices often hinder efforts to maximise recycling (Carey, 2017). A vicious cycle occurs that tends to favour unsustainable practices – company decisions on new product choices are constrained by market preferences, while public behaviour is in turn influenced by currently available products (Chiu et al., 2020). More recently, the disruptions caused by the COVID-19 pandemic have led to a global escalation of plastic pollution (Klemeš et al., 2020a). The complex interdependencies that influence plastic use and waste generation lead to "vicious networks", resulting in persistent unsustainable behaviour (Tan et al., 2021). Given the capability to handle highly complex systems, data-driven machine learning approaches can draw insights into how to curb the plastic pollution problem and support future plastic pollution mitigation policies (Nordic Council of Ministers, 2020). However, current literature only sheds light on applying machine learning to mitigate plastic pollution through micro-level technologies (e.g., using artificial intelligence to identify plastic waste for sorting Ozdemir et al., 2021). Studies that focus on the macro-level application of machine learning to analyse country-level trends from public statistics are comparatively scarce. An exception is a macro-level analysis considering five different global scenarios conducted by Lau et al. (2020); this work considered the big picture but lacked the regional resolution to give useful insights for local policy development. Their estimation of pollution is based on the Monte Carlo simulation. This notable research gap provides opportunities for developing a data-driven approach to reducing plastic pollution and its environmental hazards.

As plastics are ubiquitous in daily life, reducing the hazards of plastic requires multiple actions from different stakeholders in designing effective policy interventions (Jia et al., 2019). It is crucial to understand plastic consumption quantitatively and to recognise causalities to develop effective countermeasures. However, these data are not always available as the management is done by different channels, complicating

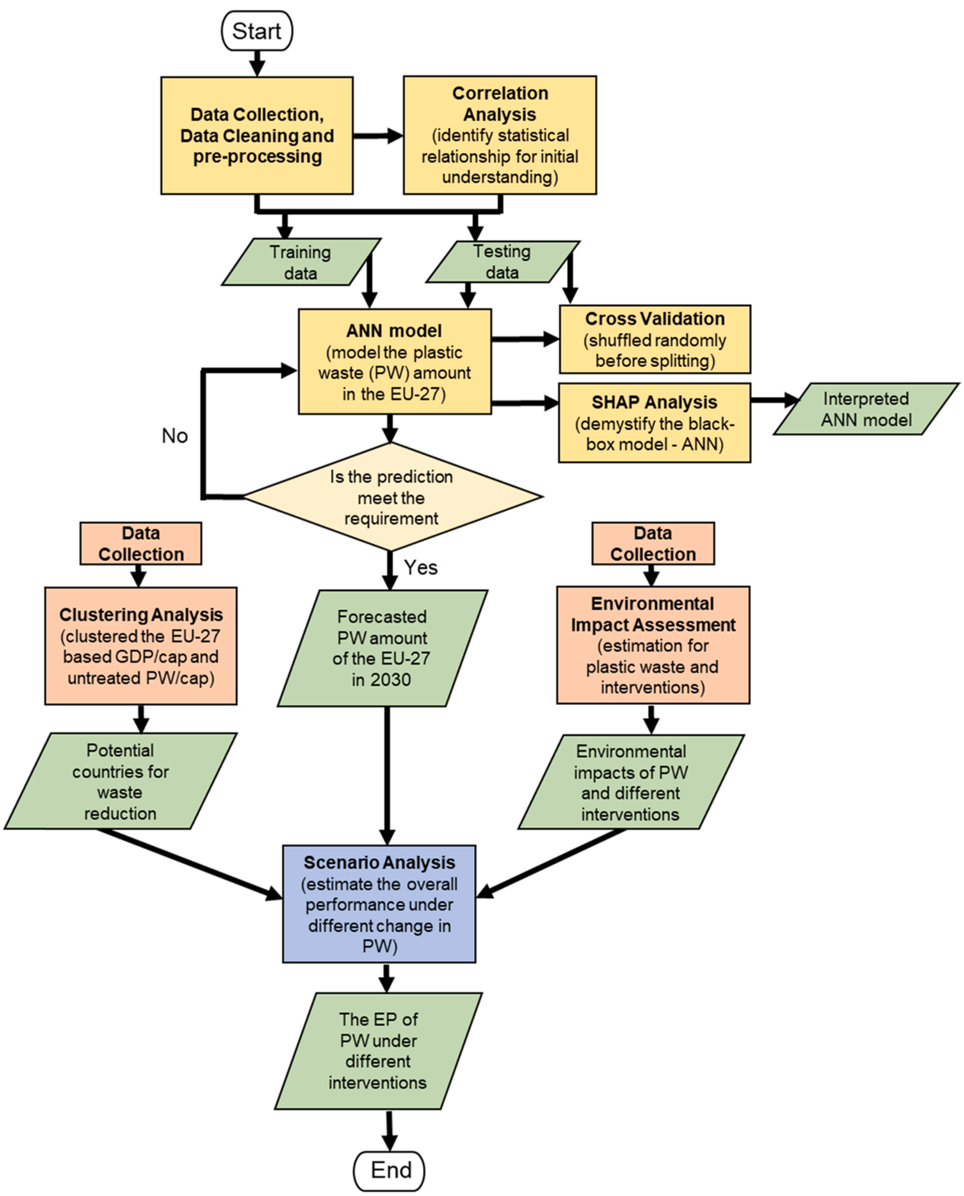

Fig. 1. The overall framework in assessing plastic waste generation and management for environmental hazard mitigation. (Note: PW = plastic waste, EP = Environmental Performance).

the data collection. Estimation and forecasting are usually conducted to inform decision–making and develop data-driven strategies. For the estimation process, one straightforward method is the "top-down" method, which multiplies solid waste generation data by the average proportion of plastic. This method is widely applied for plastic waste trade analysis at a regional scale, as illustrated by Liang et al. (2021). Various studies focused on developing models to forecast waste generation and assess the corresponding pattern. The waste generation modelling approaches range from statistical learning methods (Abdulredha et al., 2018) to machine learning approaches (Abbasi and El Hanandeh, 2016). Ceylan (2020) compared the performance of the Bayesian Gaussian process regression, multiple linear regression, and Bayesian support vector regression models in estimating municipal waste generation. Jiang and Liu (2016) forecasted the municipal waste generation under uncertainties based on data patterns by the hidden Markov model. Pavlas et al. (2020) integrated reconciliation techniques and regression models in forecasting municipal solid waste and its fraction, such as plastic waste in the Czech Republic. Ghayebzadeh et al. (2020) forecasted the plastic waste inputs from land into the Gulf using a regression model to identify the mismanaged plastic and highlighted the necessity for pollution control. However, the forecasting was mainly dependent on the gross domestic product. Machine learning approaches are generally better for identifying trends and patterns (Hao and Ho, 2019) for prediction than statistical learning, especially when dealing with big and complex data (e.g. nonlinear) commonly in macro-level work. Table A1 summarises the studies assessing plastic waste issues under different machine learning approaches where the strength and differences to the present work are highlighted, supporting the research gap.

Artificial Neural Networks (ANN) are a widely utilised machine learning modelling approach for sustainability issues (Gue et al., 2020). They have already been used for various applications for tens of years (Klemeš and Ponton, 1992). ANNs have also been successfully implemented to predict waste generation, as Coskuner et al. (2021) demonstrated for domestic, commercial, and construction waste. ANN generally has higher predictive power and lower sensitivity to outliers than the regression method in forecasting waste generation. It can model nonlinear and complex relationships and has been successfully applied for various solid waste management modelling (Abdallah et al., 2020). Kumar et al. (2018) suggested that the ANN model has the best prediction accuracy on India's plastic waste generation rate compared to the other nonlinear machine learning models (support vector machine and random forest). However, ANN studies specifically dealing with the niche of plastic waste generation at macro level application and with the aim of hazard mitigation are comparatively scarce. A study by Adeleke et al. (2021) applied ANN for predicting the composition of municipal solid waste (e.g. paper, plastics, textile), with a focus on the impact of seasonal variation (meteorological parameters) in the City of Johannesburg. The priority of this study has been given to developing a robust model (e.g., identify the topology) as illustrated in most of the ANN studies (Wu et al., 2020), and the practical implication is focused on facilitating waste management (Abdallah et al., 2020), such as to improve the collection system. The interaction and underlying insights, which are equally valuable for strategic planning such as plastic pollution reduction, are not their main consideration. This is a limitation of machine learning compared to statistical learning, intended for drawing inferences about the relationships between variables (Bzdok et al., 2018). There have been a few methods developed to overcome these issues to give better insights, including Local Interpretable Model-Agnostic Explanations (LIME) (Ribeiro et al., 2016) and SHapley Additive exPlanation (SHAP) (Lundberg and Lee, 2017). The SHAP method can enhance the interpretability of the trained ANN model and offer good visualisation to facilitate communication. It is one of the recent efforts in demystifying the black-box model, based on game theory and capable of providing a complete explanation between the global average and the model output for a particular explanation (Lundberg and Lee, 2017). In contrast to LIME, SHAP can provide an entire model explanation (Lundberg and Lee, 2017), while LIME is for a single prediction explanation with the advantage of speed (Ribeiro et al., 2016). SHAP can be regarded as an improvement over the original linear LIME and has been commonly applied for medical studies (Lundberg et al., 2018), but not in interpreting the plastic waste forecasting model.

In the presented work, machine learning (ANN, supported by SHAP) is used as a basis for developing a novel data-driven approach to detecting hidden patterns in environmental and socio-economic data in the European Union (EU-27). The aim is to project and assess the effectiveness of interventions or environmental-related initiatives, such as waste reduction, recycling, and energy recovery, in reducing the environmental hazard of plastic waste in the EU-27. The EU-27 is chosen to represent the situation and projection of developed regions actively involved in developing strategies for plastics management, including forming a circular economy (EC, 2021a). Based on the machine learning model and SHAP interpretation, five scenarios are assessed in this study to analyse the effectiveness of interventions. Scenario analysis could serve as a process of estimating the expected plastic waste generation after a given change in the values of key factors. Integrating interpretation and scenario analysis-supported modelling could improve understanding and decision-making, allowing optimised responses to future events.

2. Method

Fig. 1 shows the overall framework in assessing plastic waste generation and environmental hazard mitigation through different interventions in the EU-27. The accounting of plastic waste generation includes the imported plastic waste handled in the countries. The primary database utilised in this study is retrieved from the Eurostat Statistic Database (Eurostat, 2021). The database consists of the amount of plastic waste handled within the countries (PW-handled), plastic recycling rate, real gross domestic product per capita (GDP), population, final energy consumption (Energy), Circular Material Rate (CMU), Environmental Tax Revenue (ETR), Environmental Protection Investment (EPI), persons employed related to circular economy sectors (Employed-CE), and Gross Value Added in environmental goods and services sector (GVA-E). The statistics for imported (PW – import) and exported plastic waste (PW – export) are extracted from United Nations Commodity Trade Statistics Database (UN Comtrade, 2021). The economic complexity index (ECI-MIT) applied is developed at the Massachusetts Institute of Technology (The Observatory of Economic Complexity (OEC), 2021a). Table S1 in the Supplementary Material summarises the definition of the data inputs, units, and their physical meaning to plastic waste generation.

The data from 2010 to the latest available (2018) for the EU-27, which includes Austria, Belgium, Bulgaria, Croatia, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, is employed, as provided in the supplementary material (Table S2). Data cleaning or data pre-processing is conducted and discussed in detail in the respective sections. Section 2.1 describes the correlation analysis followed by the ANN modelling. The trained ANN model based on the selected predictors is validated, demystified, and applied to plastic waste prediction as described in Section 2.2. Predictors are also commonly known as independent variables, features, or conditional attributes. In this study, the term "predictors" is applied when referring to the ANN model developed for plastic waste prediction. "Features" is quoted in SHapley Additive exPlanations (SHAP) and correlation analysis, which is mainly an individual measurable property or characteristic of a phenomenon being observed (Lundberg and Lee, 2017). SHAP is used to interpret the ANN model. Finally, clustering analysis is conducted according to the description in Section 2.3. The main purpose of clustering analysis is to group the EU-27 countries based on the plastic waste management and economic status for appropriate discussion and interventions (e.g., waste reduction, recycling) to mitigate the environmental hazard in the EU-27. The analyses and model are built-in Jupyter notebook, Version 6.3.0 (Jupyter, 2021), using Python programming language (Python Software Foundation, 2021). TensorFlow Keras 2.3.0 (TensorFlow, 2021), a neural network library, is applied in training the ANN model. Scikit-learn library (Scikit Learn, 2021) is implemented for clustering analysis. Environmental performance assessment and scenario analysis are performed as described in Section 2.4 and Section 2.5 using an Excel spreadsheet.

2.1. Correlation analysis

Pearson correlation, as in Eq. (1), is conducted to understand the degree of the relationship between various features. The assessed features are stated in Section 2, and the complete definitions are provided in Table S1 in the Supplementary Material. The correlation analysis serves as an initial step to enhancing the understanding of the relationship between plastic waste generation and multiple features prior to predictors selection and later development of the ANN model. The closer the values are to − 1 or 1, the stronger the linear correlation. A value closer to 0 suggests a weak linear relationship. The thresholds of strong and weak linearity are usually at > 0.7 (+,−) and < 0.4 (+,−), depending on the field of study as discussed by Akoglu (2018).

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\ \sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}} \tag{1}$$

where r is the Pearson correlation coefficient between x (features) and y (dependent variable), n is the number of observations. x_i is the value of x for ith observation, y_i is the value of y for the ith observation.

2.2. Artificial Neural Network (ANN)

The ANN approach is selected due to several advantages, as discussed in the introduction, including the ability to deal with nonlinear relationships between problem parameters. Four components (data, model, objective function, and optimisation algorithm) are needed to train the algorithm for this study with the amount of plastic waste generated as the intended response variable or output of the model. It is assumed that the amount of plastic waste generated (decision attribute) by countries can be associated with GDP, population, CMU, energy consumption, and ECI (conditional attributes), as summarised in Table S1 in the Supplementary Material. GDP, population, CMU, and energy consumption have been utilised as predictors of waste generation by Fan et al. (2021a, 2021b). The basis for predictor selection is further discussed in Section 3.1 and Table S1. The selection is primarily based on the physical meaning, potential relationship, and the identified correlation coefficient. The information or input data is extracted as described in Section 2. The data is split into training and testing set with the predictors normalised by MinMaxScaler scaling techniques in SciKitLearn tools (Scikit Learn, 2021), following Eq. (2). A 9:1 ratio between training and testing data is selected due to model performance reflected in learning curves. This ratio is adopted as more data could be used for training. The ANN model can be further improved with updated data. The performance of the learning curve is assessed, followed by cross-validation to prevent underfitting and overfitting (Brownlee, 2019).

$$Nor = \frac{O - min}{max - min} \tag{2}$$

where Nor and O are the normalised and original values, max and min are the maximal and minimal values in the data series.

The selected predictors for the model are judged by several factors, as stated in Section 2.1 and further discussed in Section 3.2. The model type of this study is Sequential (TensorFlow, 2021), with a commonly used three-layer structure and an activation function of ReLU (Rectified Linear Activation). It consists of 5 input parameters (predictors) with 30 nodes in the first hidden layer and 15 nodes in the second hidden layer, which were determined according to the performance of the learning curve. The optimisation algorithm that controls the learning rate is Adam, an extension of stochastic gradient descent, and the lost function used is Mean Squared Error (MSE), as shown in Eq. (3). The detailed settings are summarised in Table S3.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2 \tag{3}$$

where N is the number of data samples points, f_i is the value simulated by the model, y_i is the actual value for the data sample i.

After the validation and assessment of MSE, the developed ANN model is utilised to estimate the absolute amount of plastic waste generated in each EU-27 country and the average plastic waste in the EU-27. The population projection in 2030 is based on the hypothetical "What if" exercises by Eurostat (Eurostat, 2021); refer to Table S4. The GDP growth in EUR/cap (Eurostat, 2021) is estimated according to a linear model grounded on a series of data (years 2000–2020) of the respective EU-27 members. A 10% increment in CMU and ECI and a 5% increment in energy consumption to that of 2018 are fixed as a baseline. The output of the model (plastic waste generation) is inverted from normalisation to obtain the exact values.

2.2.1. SHapley Additive exPlanations (SHAP) analysis

SHAP is an approach proposed by Lundberg and Lee (2017) to interpret a black box or machine learning model. The SHAP values could increase the model transparency and interpretability. It measures the impacts of features by considering the interaction with other variables, providing insight into why such output is suggested or predicted by the model. SHAP is based on cooperative game theory, where the "game" reproduces the outcome of the model and the "player" is the incorporated feature. The contribution or marginal effect that each feature brings to the prediction from the model is exemplified. The main equation specifies the SHAP explanation as in Eq. (4), where more comprehensive information can be obtained in the original work (Lundberg and Lee, 2017). The computation of SHAP values is similar to that of the Shapley value used for determining the fair allocation of profits or costs within a coalition (Shapley, 1951). In this study, the latest SHAP library by Lundberg (2021), the key developer of SHAP, is applied to combine and approximate insights from current additive feature attribution methods. The coding for SHAP analysis is written in Python programming language. The marginal effects are illustrated through the SHAP summary plot and dependence scatter plot (two features – one by values and another with colour indication) (Lundberg, 2021). The SHAP summary plot shows the SHAP values of all features for each sample. It sorts features by the sum of SHAP value magnitudes over all samples and uses SHAP values to show the distribution of each feature's impacts on the model output. The importance of these features is ranked in descending order. In this study, negative SHAP values suggest the feature associated with a lower prediction. In a SHAP summary plot, the feature in red represents high, and the feature in blue represents low. For example, when using population as a feature, red represents a high population, and blue represents a low population.

$$g\left(z'\right) = \varnothing_0 + \sum_{i=1}^{M}\varnothing_i z_i' \tag{4}$$

where g is the explanation model, $z' \in \{0,1\}^M$ is the coalition factor, M is the maximum coalition size or number of simplified input features. $\varnothing_i \in R$ is the feature attribution for a feature i (the Shapley values), $\in$ denotes that an element is in a set, R is the set of real numbers.

Table 1
The conversion factors for environmental performance assessment of plastic waste (units are expressed per kg of plastic waste). Complete calculation and data sources are presented in Table S6.

Per kg of plastic waste	GWP kg CO_2eq	Acidification kg SO_2eq	Eutrophication kg PO_4^{3-}eq	Marine aquatic toxicity kg DCBeq	Abiotic depletion MJ	Plastic marine pollution kg
● Without EOL	1.49	3.32×10^{-3}	3.68×10^{-4}	97.2	55.5	4.67×10^{-3}
● Incineration (energy recovered)	1.19×10^{3}	- 2.62	$- 2.64 \times 10^{-1}$	$- 8.33 \times 10^{4}$	$- 2.73 \times 10^{4}$	NA
● Landfill	0.0717	1.98×10^{-4}	1.96×10^{-4}	7.07	1.03	NA
● Recycling (material recovered)	- 1.61	$- 3.74 \times 10^{-3}$	$- 3.98 \times 10^{-4}$	- 24.62	- 71.21	NA

Note: EOL = end of life cycle, DCB = dichlorobenzene, GWP = global warming potential, NA = not applicable.

Table 2
The assessed scenarios and descriptions, where I = interventions.

Scenarios	Description
Scenario 1 = 2018	• Plastic waste of EU-27 in 2018 = 16.77 Mt (Eurostat, 2021) • Recycling rate (average of EU-27) = 32.5% (Eurostat, 2021), as in 2018 • Energy recovery (average of EU-27) = 42.6% (Eurostat, 2021), as in 2018 • Landfill (average of EU-27) = 24.9% (Eurostat, 2021), as in 2018
Scenario 2 = 2030 (I0), T (baseline)	• Plastic waste of EU-27 in 2030 = 17.00 Mt (Predicted in this study based on ANN model, further discussed in Section 3.2) • Recycling rate = 55% (EC, 2020a) (Average target set by EU-27) • Energy recovery rate = 42.6% • Landfill = 2.4%
Scenario 3 = 2030 (I1)	• Plastic waste of EU-27 in 2030 = 15.51 Mt (Waste reduction enforced in clustered group 1 (yellow) and 2 (green), see Fig. 5) • Recycling rate = 55% (EC, 2020a) (Average target set by EU-27) • Energy recovery rate = 42.6% • Landfill = 2.4%
Scenario 4 = 2030 (I2)	• Plastic waste of EU-27 in 2030 = 15.51 Mt (Waste reduction enforced in clustered group 1 (yellow) and 2 (green), see Fig. 5) • Recycling rate = 50% (5% less than targeted) • Energy recovery rate = 47.6% (increased) • Landfill = 2.4%
Scenario 5 = 2030 (I3)	• Plastic waste of EU-27 in 2030 = 15.51 Mt (Waste reduction enforced in clustered group 1 (yellow) and 2 (green), see Fig. 5) • Recycling rate = 50% (5% less than targeted) • Energy recovery rate = 42.6% • Landfill = 7.4%

2.3. Clustering analysis

In this study, the clustering is performed based on the k-means clustering method (Bholowalia and Kumar, 2014), where each data sample is assigned to the closest centroid to form a cluster. The number of clusters defined is the key of k-means clustering. It is decided based on Elbow Method (Bholowalia and Kumar, 2014) by running several k-means, the sum of squared distance (SSE) of each iteration is plotted, and the point where the SSE curve starts to bend (form an elbow and flatten out) suggests the appropriate cluster number. Non-recycled plastic waste per capita is selected as the x-axis and GDP/capita as the y-axis. The rationale and the applicability are discussed in Section 3.3. The coding for clustering analysis is written in the Python programming language. Scikit-learn library (Scikit Learn, 2021) is implemented, referring to Table S5.

2.4. Environmental performance assessment

The environmental performance assessed in this study includes global warming potential, acidification, eutrophication, marine aquatic toxicity, abiotic depletion, and plastic marine pollution. Table 1 summarises the conversion factors applied and converted using a basis of 1 kg of plastic waste. The underlying assumptions for plastic marine pollution are, as stated by Herberz et al. (2020), who reported that out of the 2.5% mismanaged plastic in the EU, 25% ends up in the ocean. The environmental performance of plastic manufacturing stages before the end of the life (EOL) cycle (without EOL) is evaluated based on the assessment by Herberz et al. (2020) for the EU. Based on 11 observations (different plastic products and types), the average performance is extracted and converted to the same functional unit of 1 kg of plastic. The data for the environmental performance of the end-of-life stages, including incineration, landfill, and recycling, is mainly extracted from the GaBi database (Thinkstep, 2017). Avoided energy of incineration is evaluated based on the energy recovered (electricity and steam). Avoided material (unburdening impact) is considered in the environmental performance assessment of plastic waste recycling. The detailed information on the calculation is given in Table S6 in the Supplementary Materials.

2.5. Scenario analysis

Five different scenarios are assessed in this study, as listed in Table 2. Scenario 1 is based on the data of 2018 (Eurostat, 2021). Data including the pandemic impact are still going to be published. Four other scenarios with different interventions are compared to Scenario 1. Scenarios 2, 3, 4, and 5 are for 2030, where the amount of plastic waste is predicted based on the developed model using population, GDP, CMU, ECI, and energy consumption as predictors. The population forecast for the EU-27 (Table S4) is obtained from the Eurostat Statistic Database (Eurostat, 2021). The GDP growth is calculated based on a linear regression model of the time series data from 2002 to 2020 (Table S7). A 5% increase in energy consumption and a 10% increment in CMU and ECI are assumed. Two EU-27 members, Luxembourg and Malta, are among the countries which do not have reported ECI. An average for EU-25 is applied. The data of all the input features are supplemented and explained in Tables S4 and S7.

The recycling rate in Scenario 2 is increased to 55%, while the energy recovery is unchanged (42.6%) and landfill is reduced to 2.4%. 55% of recycling is selected in this scenario analysis as it is an EU-wide average target to be achieved by 2030 (EC, 2020a). Scenarios 3, 4 and 5 represent the scenarios that consider waste reduction interventions on top of a varying recycling rate. Waste reduction is implemented in countries categorised under Cluster 1 and Cluster 2, which show a high plastic waste per capita and untreated plastic waste amount (Fig. 5). The deduction is to the level of the EU-27 average at 15.51 Mt/y, referring to Table S8. However, Scenario 4 (50% recycling, 47.6% energy recovery, 2.4% landfill) and Scenario 5 (50% recycling, 42.6% energy recovery, 7.4% landfill) are in the situation where the EU-wide target of 55% recycling rate is not achieved and with the capacity of energy recovery increase (Scenario 4) and without (Scenario 5).

3. Results and discussion

Four main sections are included in the results and discussion section. Section 3.1 explores the correlation among various variables

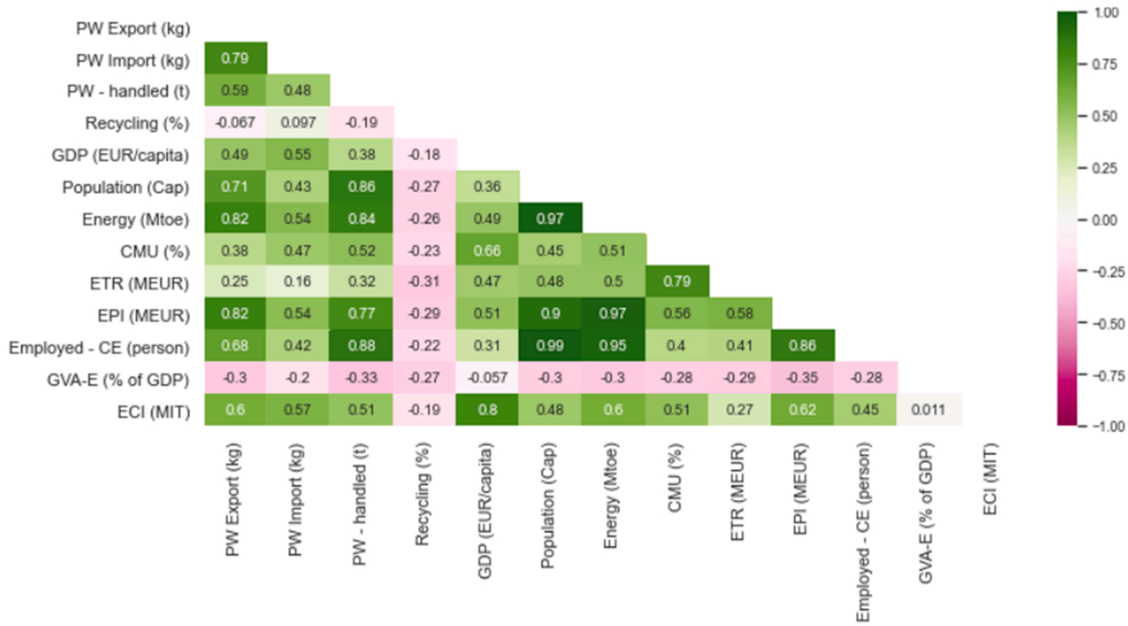

Fig. 2. The correlation coefficients between potential model predictors. PW = plastic; plastic – handled = plastic manage within a country; CMU = circular material use rate; EPI = Environmental Protection Investment; Employed – CE = person employed in Circular Economy sector; GVA – E = Gross value added in environmental goods and services sector; ECI = Economic Complexity Index.

incorporated into the ANN model. Section 3.2 discusses the developed ANN model for plastic waste generation in the EU-27, including validation and insight of this black-box method. Section 3.3 discusses the EU-27 clustering results. Section 3.4 analyses the environmental performance of the different scenarios under different interventions for hazard mitigation, where the forecasted plastic waste in 2030 is based on the ANN model.

3.1. Correlation analysis

Fig. 2 shows the correlation coefficient of different variables as potential predictors for plastic waste generation in the EU-27. The correlation coefficients at the 3rd row and 3rd column show the correlation with the amount of plastic waste generated and is represented as "PW – handled (t)." Population (0.86) and energy consumption (0.84) show a relatively strong linear correlation with plastic waste generation. Demographic factors such as population size have been suggested as one of the important factors in determining waste generation in several studies, including municipal solid waste (Kaza et al., 2018), e-waste (Duman et al., 2019), and other hazardous waste (Ndanguza et al., 2020). An increasing population usually leads to an increasing amount of waste, but socio-economic factors and other potential interventions could influence the output and must be considered. The waste generation pattern can be impacted by the income level or urbanisation rate (Kaza et al., 2018). Low-income countries usually do not have the luxury of wastage or consumption compared to high-income countries. However, higher-income countries have a higher capability in mitigating and introducing interventions for waste reduction. This phenomenon fits the theory of the Environmental Kuznets Curve (Boubellouta and Kusch-Brandt, 2020). It is a complex question involving the interaction of different factors where there is no single factor that can thoroughly explain waste generation and management observation. Although real GDP/capita is applied, this study did not show a strong linear correlation (0.38). GDP is a critical predictor for plastic waste generation, as discussed by Lebreton and Andrady (2019). Statistics is capable of highlighting the candidate predictors. However, knowledge of the subject area and plausible causality are also required.

Energy consumption could be an effective predictor for plastic waste generation (0.84) as most manufacturing and production processes consume energy, and energy could be recovered from waste. It is expected to have a close relationship with plastic waste generation, which is also highly associated with economic development. The Circular Material Use (CMU) rate measures the share of material recovered and fed back into the economy. A higher rate value suggests that more secondary materials substitute the primary raw materials. It has a moderately linear relationship with plastic waste generation (0.52). However, the CMU value can reflect the effort of countries in material recovery, a better replacement for the recycling rate indicator. Another potential predictor with moderate linearity (0.51) is ECI (Economic Complexity Index) (0.51). ECI reflects the knowledge intensity of a country's economic activities. It is estimated using diverse data sources, including employment, stock market, trade, and patent (The Observatory of Economic Complexity (OEC), 2021b). Other variables or features that have a high linear correlation with plastic waste generation are EPI (Environmental Protection Investment) and Employed – CE (persons employed in Circular Economy sector), which are 0.77 and 0.88. These features are potentially useful in reflecting the interventions effort of different countries. However, they are highly correlated with population and energy consumption (> 0.9, see Fig. 2), suggesting it can be from the same origin and linearly predicted from the other predictors or features with a substantial degree of accuracy. Although the multicollinearity issues would not influence those over-parameterised ANN models (De Veaux and Ungar, 1994), sparse models with fewer predictors, but without compromising the predictive power, are preferable and interpretable (Scheinost et al., 2019). The interventions of countries on the plastic demand and waste generation could be reflected by GDP, CMU, and ECI, where the data is relatively available even in the non-EU countries.

The identified correlation coefficients suggest that despite the linear regression model having higher simplicity and interpretability, it could not forecast the plastic waste generation accurately in the EU-27. Some of the features which expected to correlate with predictability for plastic waste generation are not linearly related. Plastic waste generation is affected by the local demand and the intervention effort or policy in a specific region. There is a complex relationship between the factors or variables and economic development. ANN does not have a priori

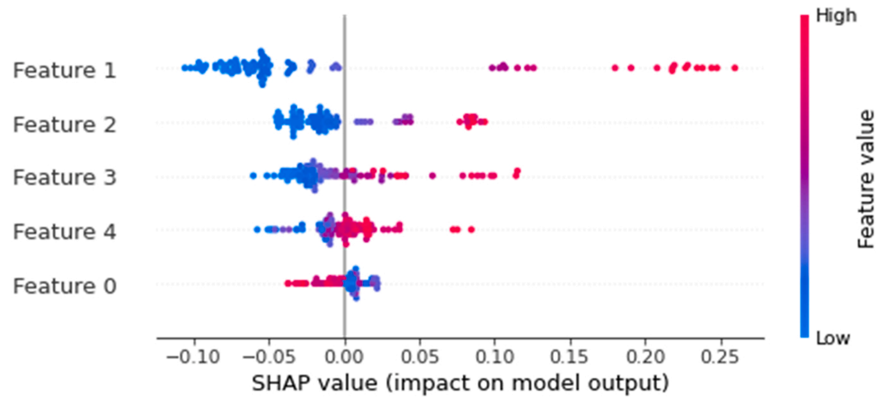

Fig. 3. The summary plots of the SHAP permutation importance. Feature 0 = GDP (EUR/capita), Feature 1 = Population (capita), Feature 2 = Final Energy Consumption (Mtoe), Feature 3 = CMU (%), Feature 4 = ECI.

assumptions on the relationship between predictors and output, and their nonlinear modelling capability is one of the appropriate approaches in such a case. In this study, population, GDP, CMU, ECI, and energy consumptions are selected as the predictors, where SHAP further assesses and interpret the interactions between the predictors. Another correlation method that helps understand the association between different variables is Spearman correlation, evaluating monotonic relationships instead of a linear relationship. As suggested by Jiang et al. (2021), Spearman correlation is applied to identify the appropriate fit, to determine whether a linear model is sufficient or if a polynomial, exponential, or radial basis function is needed. However, as fitting identification is not the primary concern of ANN and SHAP, Pearson correlation is sufficient for a preliminary understanding of our modelling.

3.2. The ANN model

The learning and validation (based on the holdout dataset) curves of the developed ANN model to forecast plastic waste generation under the settings as in Table S3 are shown in Fig. S1, where the epoch reflects the network's learning trend. The fitness of the model serves as one of the indicators to show the appropriateness of the model. In general, the learning curves do not show a significant trend of overfitting (training loss continues to decrease with experience, validation loss decrease to a point and increase) and underfitting (training loss remains flat, training loss continues to decrease) (Brownlee, 2019). Fig. S1(b) shows the performance of the model under cross-validation, where 32 observations are assessed. The grey dot labelled as "Actual" shows the real amount of plastic waste, and the pink dot labelled as "Predicted" shows the estimated plastic waste generated. The vertical distance between the pink and grey dots indicates the error of this predictive model. The brown dots indicate where the actual and predicted values are the same. In general, the differences between actual and predicted values are acceptable, with a mean squared error of 0.0028 generated by normalised input. ANN is a black box method where the visible items are the inputs and outputs. The knowledge of its internal mechanism of causality can be elucidated via SHAP analysis, as shown in Fig. 3. They are essential for policy actions.

Fig. 3 summarises SHAP permutation importance, showing what is driving the predicted plastic waste amount. It provides an overview for feature comparison. The colour shows whether that feature was high or low, and the horizontal location (SHAP value) shows whether the feature (in a different colour) caused a higher (positive) or lower (negative) prediction. Feature 1, population, is suggested as the top influential feature in determining the plastic waste where the SHAP value ranges from − 0.10 to 0.25. Population (high) has a positive impact on absolute plastic waste generation, as shown in Fig. 3. It indirectly suggests the hardly replaceable status and usefulness of plastic. There have been many interventions in the EU-27 to reduce the plastic waste demand, including the introduction of plastic replacement, banning the utilisation of single-use products, and plastic recovery. However, the influence of population is still dominant, suggesting that the current interventions barely could override the plastic demand, and the substitutive material is yet as feasible as plastic. Plastics are not an enemy, which has to be replaced at any cost. Their proper use can, in some applications, decrease environmental footprints. The issue lies in excessive or unsustainable consumption and mismanaged plastic, where Plastic Waste Footprint (Klemeš et al., 2021) could be conducted to facilitate sustainable decision-making. This development is consistent with several studies suggesting that substitute or alternative materials and products still need to improve to achieve competitive environmental performance and economic feasibility (Klemeš et al., 2021). The issue is apparent during the current COVID-19 pandemic, where plastic in food packaging and PPE (Klemeš et al., 2020b) still dominates. Reducing plastic waste, including curtailing plastic consumption, is a prioritised initiative to minimise the environmental footprints in most policies. However, it is undeniable that there is a threshold on the effectiveness of this effort. The plastic demand is inevitable to support economic development and human need. End-of-life management is an important strategy to mitigate the environmental footprint, forming a circular economy under such constraints. However, there is a limit for reduction without compromising the development.

Feature 0 (GDP reported in EUR/capita) is suggested as the least important among all the features of interest. This is explainable, especially when the dataset is focused on the EU-27. The modelling is not at the global scale and does not exhibit huge income variations among countries (e.g., developed and developing countries). The GDP in the EU-27 is generally higher than the world's average. Based on Fig. 3, an observable trend is that a higher GDP contributes to a lower SHAP value. This encouraging trend supports the Environmental Kuznets Curve hypothesis, where the EU-27 generally reaches the turning point of inverted U (Boubellouta and Kusch-Brandt, 2020). More effort is invested in protecting the environment (minimising plastic consumption or improving recovery rates) and stabilising the economy. Fig. 3 shows the summary plots for the first indications of the relationship between a feature value and the impact of the prediction. However, to understand the exact form of the relationship, SHAP dependence plots (Fig. 4) could provide further insights, including the distribution.

The SHAP dependence can be interpreted first by focusing on the shape (distribution). By referring to Fig. 4(a–d), the slope shows that a higher (value in x-axis) Feature 1 (population) contributes to higher plastic waste generation as reflected in a higher SHAP value. By referring to the colour indicators, it does not show a consistent trend of, for example, from blue (low) to red/pink (high) or red/pink (high) to blue (low). As shown in Fig. 4(a), when the Feature 0 (GDP) is high (pink), it can cause the predicted plastic waste amount in the EU-27 to be either

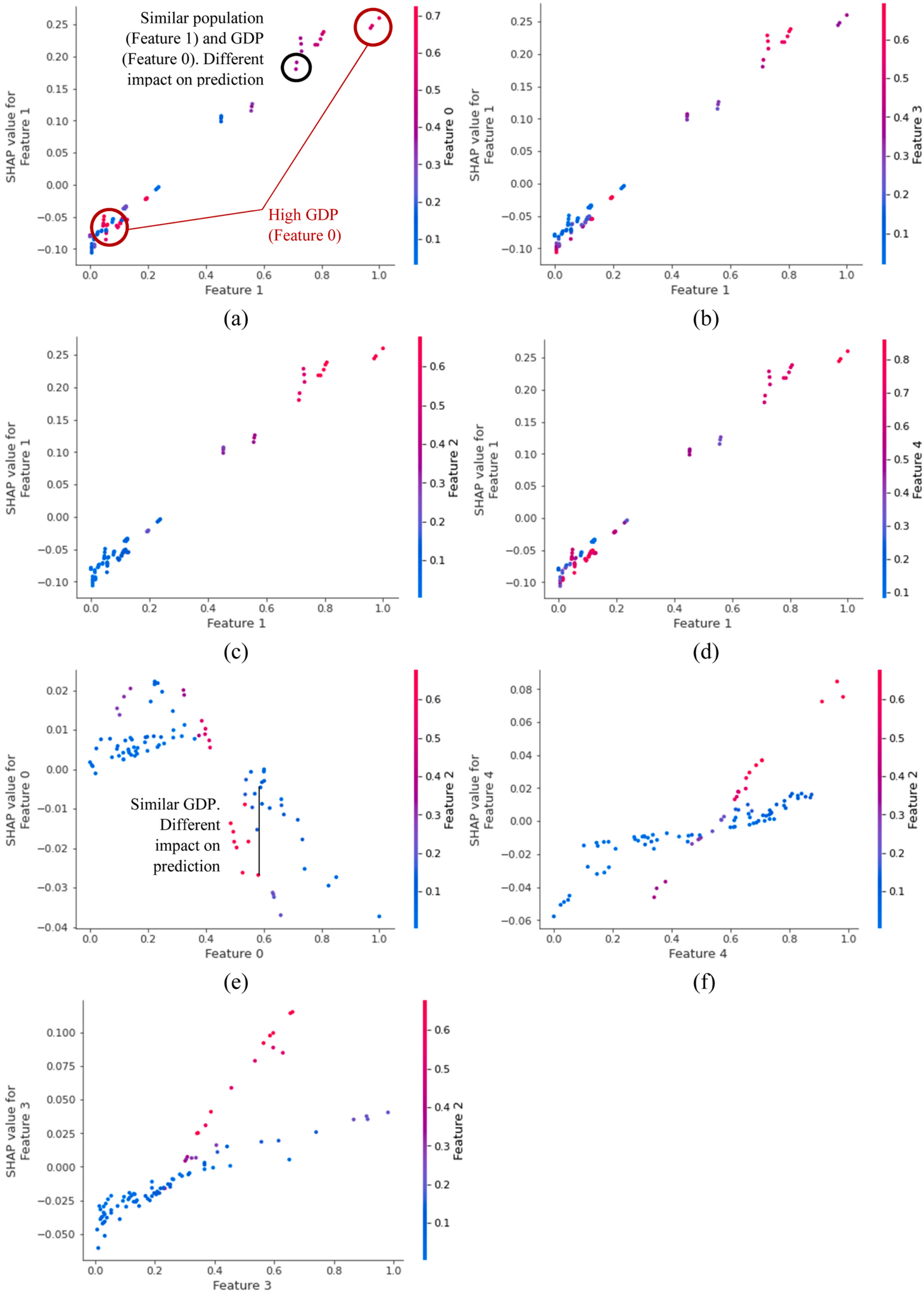

Fig. 4. SHAP dependence contribution plot. Feature 0 = GDP (EUR/capita), Feature 1 = Population (capita), Feature 2 = Final Energy Consumption (Mtoe), Feature 3 = CMU (%), Feature 4 = ECI. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

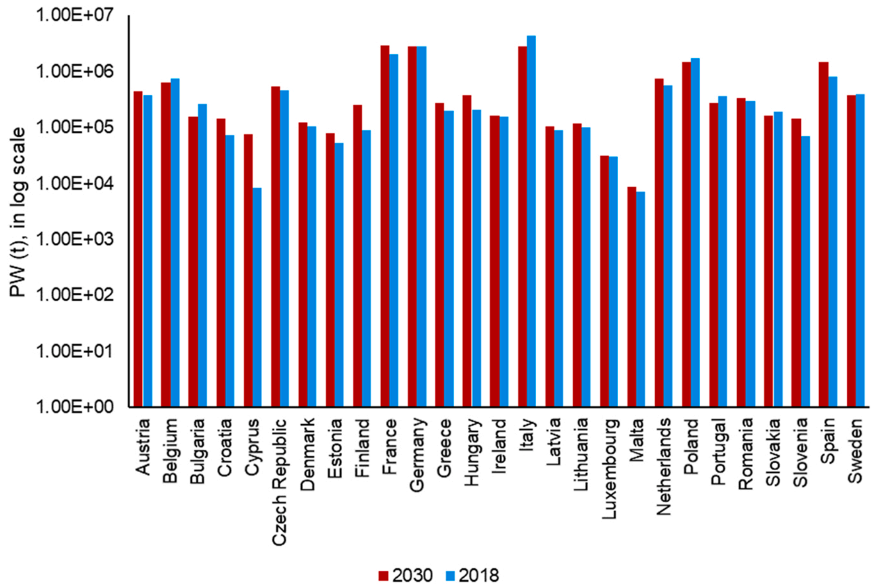

Fig. 5. The predicted amount of waste based on the baseline condition stated in Section 2.5 for 2030 compared to 2018 using the ANN model. Referring to Table S8 for tabulated data. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

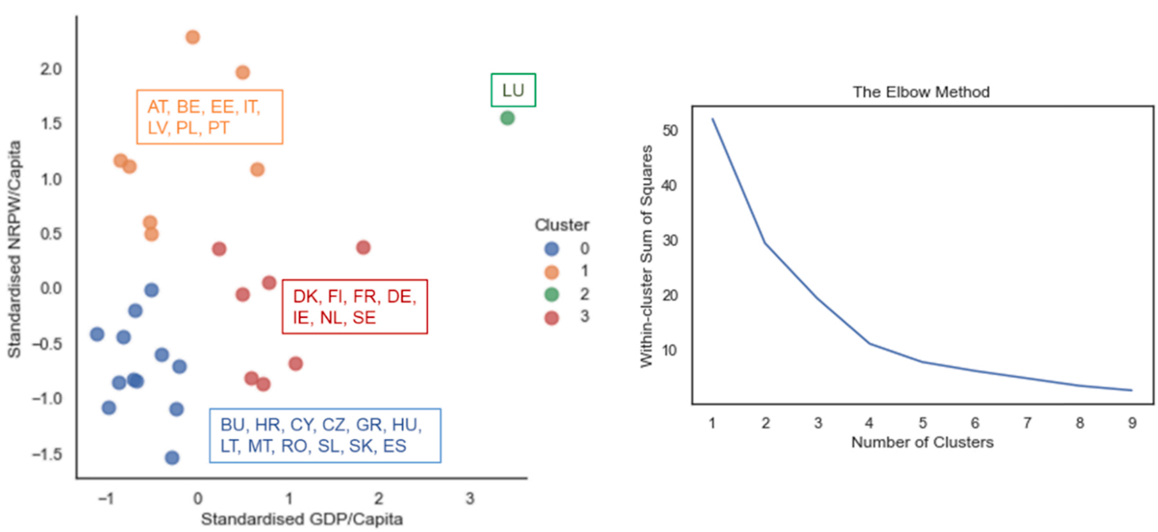

Fig. 6. The identified clusters in the EU-27 based on K-means clustering with the support of the Elbow method. NRPW = non-recycled plastic waste. GDP = gross domestic product. The abbreviation stated in the boxes is the two-letter country code. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

higher or lower (inconsistent trend), and Feature 1 (population) has a more prevailing influence. When referring to the dot located at a similar population (Feature 1) and under similar GDP (Feature 0 is in the same colour) in Fig. 4(a), there is a different impact on the prediction (SHAP value). This observation suggests that other features also interact with the population. The interaction of other features to plastic waste generation is relatively distributed, as shown in Fig. 4(e–g). Similar to the relationship identified in Fig. 3, higher GDP generally contributes to a lower plastic waste generation (Fig. 4e). A similar GDP (a normalised value, see dots connected by a straight line in Fig. 4e) shows a different SHAP value contributed by the impact of Feature 2 (Final energy consumption).

Fig. 5 shows the predicted amount of plastic waste for 27 EU countries in 2030 using the developed ANN model. The overall plastic waste generation in the EU-27 is expected to increase by 0.2 Mt compared to 2018. It represents an incremental change in per capita consumption from 37.5 kg/cap/y to 37.8 kg/cap/y. It is lower than the projection by Kaza et al. (2018), which mainly used GDP as a predictor. Kaza et al. (2018) predicted 56.94 kg/cap/y (1.3 kg/cap/d where 12% is plastic waste) of plastic waste in 2030. However, the estimation is not for direct comparison as the estimation is categorised for Europe and Central Asia, while in this study, the target is for the EU-27. Most countries expect a typical waste management trend, an increment in plastic waste, except Belgium, Bulgaria, Germany, Italy, Poland, Portugal, Slovakia and Sweden. They are estimated to have a slight decrease in plastic waste generation. By referring only to the individual variables (e.g., population), no apparent conclusion can be drawn whether it contributes to a lower or higher plastic waste generation. For example, the decrease in population in Belgium and the increase in Bulgaria both show a reduction in plastic waste. This trend further supports the associated interrelated relationship between a range of variables, as illustrated in Fig. 4, and support the application of the machine learning method to capture a complex trend. Scenario analysis is conducted to understand the environmental performance if different interventions are implemented to

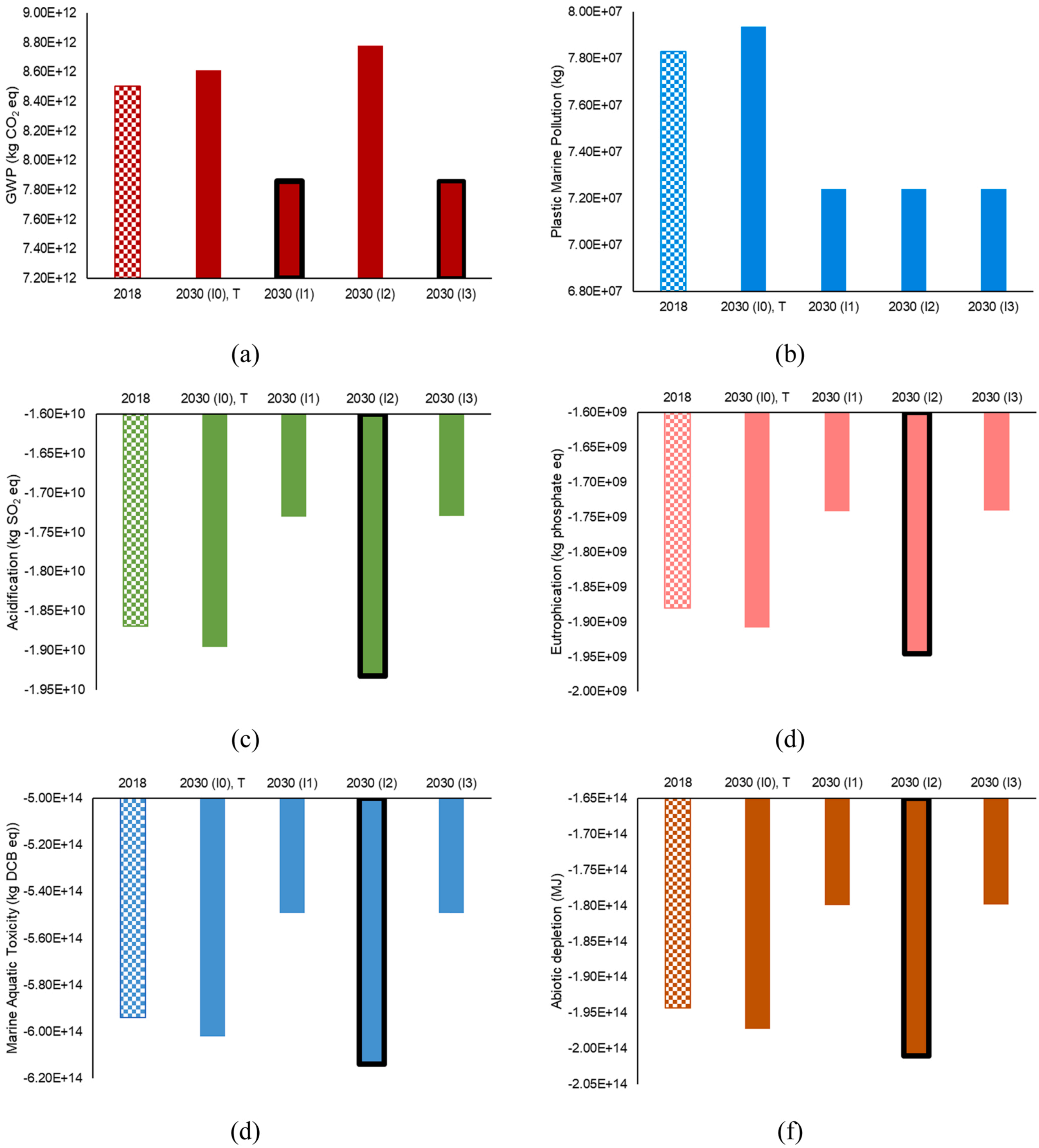

Fig. 7. The environmental performance of plastic waste management in 2018 and 4 other scenarios (a) Global Warming Potential, (b) Plastic Marine Pollution, (c) Acidification, (d) Eutrophication, (e) Marine Aquatic Toxicity, (f) Abiotic Depletion. The checkered bar represents the situation in 2018 (Scenario 1). The bar in solid colour represents the situation in 2030 under different interventions see Table 2. The bars with a thick black border represent the scenarios with the lowest impacts of the assessed environmental categories. Scenario 2 = 2030 (I0),T; Scenario 3 = 2030 (I1), Scenario 4 = 2030 (I2), Scenario 5 – 2030 (I3). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mitigate plastic waste generation impacts. Cluster analysis is first conducted to strengthen the implementation feasibility of the interventions based on the current performance and capability.

3.3. Identified clusters of the EU-27

Fig. 6 shows the four clusters formed in the EU-27 based on the GDP and plastic waste recovered performance where plastic waste generated is also deliberated. Four clusters are selected, determined based on the elbow method, picking the elbow of the curve as the number of clusters to apply. Cluster 0 (blue) has low GDP/capita and low NRPW/capita, suggesting a relatively smaller room for improvement. High NRPW/capita (Cluster 1 and Cluster 2) shows the potential in plastic waste reduction, including recycling and reducing the demand. High GDP/

capita (Cluster 2 and Cluster 3) reflects the economic capability to support and invest in environmental protection initiatives. In this study, the countries categorised under Cluster 1 (orange) and Cluster 2 (green), including Austria, Belgium, Estonia, Italy, Latvia, Luxembourg, Poland, and Portugal, are targeted for interventions. It is plausible that they have the capability to reduce the amount of waste to the EU-27 average plastic waste amount in 2018, as stated in Table 2.

A colour-coded map to determine whether the identified clusters in Fig. 6 correlate with geographic proximity is included in Fig. S2. In general, it does not fully follow the scale division of Europe according to spatial criteria, such as Western Europe, Southern Europe, Central Europe, Northern Europe, and Southeastern Europe. However, all the countries in Southeastern Europe (Bulgaria, Cyprus, Greece, Romania) of the EU-27 are in Cluster 0, and all Northern Europe countries of the EU-27 (Denmark, Finland, Estonia, Sweden) are grouped under Cluster 3. They show a better performance in non-recycled plastic waste per capita, where the generated plastic waste and the non-recovered portion are lower.

3.4. *Environmental performance*

Fig. 7 shows the environmental performance, including global warming potential, acidification, eutrophication, marine aquatic toxicity, abiotic depletion, and plastic marine pollution, in 2018 and under different scenarios in 2030. The negative values are contributed by the environmental impacts avoided or savings, for example, from the recovered energy or the avoided material consumption. A negative value in the GHG footprint is commonly reported in the literature representing the avoided emissions, such as Wicke et al. (2020). The bar chart in the checker pattern shows the performance of the baseline (2018 – Scenario 1) for comparison.

In 2018, the plastic waste generation in the EU-27 was 16.77 Mt, with a recycling rate of 32.5%. Based on the current policy, the recycling rate is targeted to increase up to 55%. However, even assuming the target is achieved, the environmental performance is not as anticipated (Scenario 2 - second bar in Fig. 7). It is not lower than in 2018, especially in global warming potential and plastic marine pollution. The global warming potential in 2018 is 1 Gt CO_2eq lower than in 2030 (Scenario 2 - second bar in Fig. 7). This observation is identified in contrast to the other organic waste. Plastic is difficult to degrade and less likely to emit GHG when it ends up in the landfill. The main issue of plastic waste disposal is not GHG emissions but the land footprint (Law et al., 2020) and, more importantly, pollution of aquatic and terrestrial environments (Wong et al., 2020). Recycling plastic waste reduces the acidification potential, eutrophication potential, marine aquatic toxicity, and abiotic depletion. Plastic waste reduction through reducing consumption plays a more dominant role in decreasing GHG emissions. Different strategies can be used, such as introducing alternative materials (Kabir et al., 2020), driving innovation and investment (EC, 2021a), and developing alternatives to single-use plastics (EC, 2021b). Other potential policy measures include extended producer responsibility (Watkins et al., 2017), development of the European market for recycled plastics (EU, 2021b), cradle-to-cradle product design (Helms and Russell, 2016), and various related strategies in the Circular Economy Action Plan (EC, 2020b). These options need to be further strengthened to reduce the region's environmental footprints from plastic pollution.

Scenario 3 (3rd bar in Fig. 7) and Scenario 5 (5th bar in Fig. 7) show the lowest global warming potential, representing the intervention where the recycling rate increases and the generated plastic waste is reduced. However, Scenario 4 (4th bar in Fig. 7) did not perform better than in 2018. It suggests that, for plastic waste, energy recovery has a higher global warming potential than landfilling. This highlights the importance of considering different indicators or footprints in defining environmental performance. Misleading notions (e.g., "landfilling is environmentally sustainable than energy recovery or vice versa") could be obtained from the analysis of a single environmental impact/footprint. The global warming potential of energy recovery depends on the current energy grid mix at a place. In this study, the EU average is applied. For a more effective policy implication, allocating recovered energy to replace the energy demand supplied by the source with higher carbon emission intensity (grid mix with a lower share of renewable energy) could mitigate the global warming potential. The benefit of integrated regional waste management, which adequately matches the sources, demand, treatment capacity and resources, in further reducing the environmental footprints of waste management has been previously demonstrated by Fan et al. (2021a, 2021b) in Central Europe.

In general, a conclusive observation cannot be identified to answer which is the best scenario due to the existing trade-off. Scenario 3 (waste reduction, recycling rate = 55%, energy recovery = 42.6%) and Scenario 5 (waste reduction, recycling rate = 50%, energy recovery = 42.6%) have the best performance in terms of global warming potential. Scenario 4 (waste reduction, recycling rate = 50%, energy recovery = 47.6%) has the best performance in acidification potential, eutrophication potential, abiotic depletion and marine aquatic toxicity. The results serve as a quantitative reference for decision-making. A localised assessment is needed based on stakeholders' concerns, existing infrastructure, and resources at a place. In the case that environmental sustainability is the main concern without restricting the economic and social aspects, environmental price (CE Delft, 2017) and eco-cost (TU Delft, 2021) could be used as a medium to accumulate the impacts and determine the overall best scenarios. Eco-cost has been applied by Wen et al. (2021) in assessing the environmental impact of plastic waste trade flow at a global scale. For example, Scenario 4 is expected to have a better overall environmental performance as the savings from eco-costs of acidification (8.75 EUR/kg SO_2eq) and eutrophication (4.7 EUR/kg PO_4^{3-}-eq) are more significant than global warming (0.116 EUR/kg CO_2eq) (TU Delft, 2021).

4. Conclusion

The classification of plastic waste as hazardous has been commonly discussed and raised in Basel Convention. Despite the importance of plastics in the modern world, the environmental impacts of plastic waste are undeniable. This study utilised machine learning to detect hidden patterns in socio-economic and environmental data. Scenario analysis is conducted to assess the effectiveness of interventions in reducing the environmental hazards of plastic waste in the EU-27. By assessing the environmental impacts of the identified plastic waste amount in 2030 using the ANN model, it suggests that a 55% recycling rate as targeted in the EU-27 alone is insufficient to reduce the environmental impacts of plastic waste. It works towards mitigation; however, to reduce the environmental impacts, especially global warming potential, a hand-in-hand effort and reduction in plastic consumption are needed.

A feasible solution is to target the countries including Austria, Belgium, Estonia, Italy, Latvia, Poland, and Portugal, which have a larger potential for improvement (high non-recycled plastic waste per capita), and Luxembourg is economically capable of improving their performance. The environmental impacts contributed by plastic waste could be reduced compared to that in 2018 by implementing integrated management among the EU-27 members and reducing waste generation, even if the 55% recycling rate could not be fully achieved. Scenario 3 (waste reduction, recycling rate = 55%, energy recovery = 42.6%) and Scenario 5 (waste reduction, recycling rate = 50%, energy recovery = 47.6%) offer the lowest global warming potential (7.86×10^{12} kg CO_2eq). Scenario 4 (waste reduction, recycling rate = 50%, energy recovery = 47.6%) has the best performance in acidification potential (-1.93×10^{10} kg SO_2eq), eutrophication potential (-1.95×10^{9} kg PO_4^{3-}-eq), abiotic depletion (-2.01×10^{14} MJ) and marine aquatic toxicity (-6.14×10^{14} kg DCB eq).

This study indirectly underpins the Environmental Kuznets Curve hypothesis. It suggests the EU-27 is generally after the turning point of the inverted U-shaped (the higher the GDP, the lower the plastic waste

Table A1
Machine learning methods applied in plastic waste studies and the assessed scope.

Reference	Method[c]	Remarks
Gruber et al. (2019)[a]	• Linear discriminant analysis (LDA) • k-nearest neighbour • Support vector machines (SVM) • Ensemble models with a decision tree • Convolutional neural networks (CNN)	• **Aim:** increase and compare the accuracy of classification models • **Type of waste:** plastic • **Main result(s):** CNN has the highest overall classification accuracy • **Differences**[d]**:** classify plastics waste for recycling instead of predicting plastic waste generation for interventions and hazard mitigation
Wang et al. (2019)[a]	• Support vector machines (SVM)	• **Aim:** classify plastic based on colour into 7 categories • **Type of waste:** plastic bottle • **Main result(s): The** SVM classification model has an accuracy of 94.7% for colour recognition • **Differences**[d]**:** classify plastics waste for recycling instead of predicting plastic waste generation for interventions and hazard mitigation
de Medeiros Back et al. (2021)[a]	• Support vector machines (SVM) • Random Forests (RF) • Decision Trees • K-nearest Neighbours • Others: Logistic regression, Gaussian Naïve Bayes	• **Aim:** identify the best method to automatically classify microplastic spectra • **Type of waste:** Microplastics collected at sea • **Main result(s):** SVM is the best-suited model in characterising microplastics to evaluate potential impacts and target specific mitigation actions • **Differences**[d]**:** classify plastics waste for recycling instead of predicting plastic waste generation for interventions and hazard mitigation
Zhu et al. (2019)[a]	• Support vector machines (SVM)	• **Aim:** Identify plastic solid waste according to the plastics group (polypropylene, polystyrene, polyethylene etc.) • **Type of waste:** Plastic waste • **Main result(s):** identification accuracy up to 97.5%. • **Differences**[d]**:** classify plastics waste for recycling instead of predicting plastic waste generation for interventions and hazard mitigation
Wolf et al. (2020)[a,b]	• Convolutional neural networks	• **Aim:** present novel machine learning algorithm to detect, classify and quantify floating and washed plastic litter ashore in Cambodia • **Type of waste:** aquatic plastic litter • **Main result(s):** An accuracy of 83% can be achieved • **Differences**[d]**:** predictors, assessed countries/regions, intention: environmental impacts are not assessed, clustering and SHAP is not performed
Kannangara et al. (2018)[b]	• Neural networks • Decision tree	• **Aim:** develop and identify the prediction of MSW in Canada based on two machine learning algorithms • **Type of waste:** MSW • **Main result(s):** The prediction error is 16–23% • **Differences**[d]**:** predictors, assessed countries/regions, intention: environmental impacts are not assessed, clustering and SHAP is not performed
Wu et al. (2020)[b]	• Artificial Neural Network (ANN)	• **Aim:** develop and optimise municipal solid waste (MSW) prediction model for mainland China • **Type of waste:** MSW • **Main result(s):** Models for MSW prediction in the southern and northern region of mainland China share many similarities in dependency on predictors • **Differences**[d]**:** predictors, assessed countries/regions, intention: environmental impacts are not assessed, clustering and SHAP is not performed
Adeleke et al. (2021)[b]	• Artificial Neural Network (ANN)	• **Aim:** Investigate the optimal ANN settings for predicting the physical composition of MSW in Johannesburg using seasonal variation related predictors • **Type of waste:** MSW • **Main result(s):** Single hidden layer is identified as the optimal architecture network • **Differences**[d]**:** predictors, assessed countries/regions, intention: environmental impacts are not assessed, clustering and SHAP is not performed
Kumar et al. (2018)[b]	• Artificial Neural Network (ANN) • Support Vector Machine (SVM) • Random Forest (RF)	• **Aim:** Predict the generation rate of different plastic wastes and the possible revenue from informal recycling in India where three nonlinear machine learning models are compared • **Type of waste:** Plastic waste • **Main result(s):** the plastic waste generation rate of a higher socio-economic group is 51 g/c/d and 8 g/c/d for a lower socio-economic group. ANN performed best for the prediction accuracy of the plastic waste generation rate. • **Differences**[d]**:** predictors, assessed countries/regions, intention: environmental impacts are not assessed; instead, the potential revenue is estimated, clustering based on socio-economic, and SHAP is not performed

[a] Examples of study with the main purpose of plastic waste classification (microanalysis) or polymerisation design (Rizkin et al., 2020) rather than waste generation forecasting at a macro level for environmental hazard mitigation strategies.

[b] Studies focus on waste generation forecasting, however, they mainly compared the accuracy of the models, rather than provide insights.

[c] The comparison of model characteristics of the different machine learning methods (Decision trees, neural networks, Naïve Bayes, k-nearest neighbours (KNN) algorithm, support vector classification) is discussed in de Medeiros Back et al. (2021) based on 13 characteristics. The advantages of ANN compared to decision trees are discussed in Kannangara et al. (2018), followed by a review conducted by Guo et al. (2020) summarising the suitable application fields of different machine learning models.

[d] "Differences" focus on the main differences between the conducted studies by the cited authors and the presented work.

amount). The other complex interdependencies that influence plastic interactions are also discussed to facilitate communication with the final decision-makers. One of the limitations of this study is the exclusion of recent input data affected by the COVID-19. It is still unclear if these effects will persist after the pandemic ends. However, as machine learning algorithms are capable of learning based on the provided data, an updated assessment could be conducted based on the changes in the selected predictors when the data impacted by COVID-19 are available. Future work could also be conducted at a global scale for a more comprehensive mapping. Such an extension can uncover patterns unique to developing countries and variations based on different socio-cultural norms. Optimal recycling and energy recovery rates and the capability of each country towards contributing to environmental impact reductions could be assessed. Such investigations can provide insights for the effective reduction of hazards brought by plastic use and disposal by offering customised strategies.

CRediT authorship contribution statement

Yee Van Fan: Conceptualization, Data collection, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Peng Jiang:** Conceptualization, Writing – review & editing, Validation. **Raymond R. Tan:** Conceptualization, Writing – review & editing, Validation. **Kathleen Aviso:** Conceptualization, Writing – review & editing, Data collection. **Fengqi You:** Conceptualization, Writing – review & editing. **Xiang Zhao:** Conceptualization, Writing – review & editing. **Chew Tin Lee:** Conceptualization, Writing – review & editing, Supervision. **Jiří Jaromír Klemeš:** Conceptualization, Writing – review & editing, Funding acquisition, Project administration, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The financial support from the EU supported project Sustainable Process Integration Laboratory – SPIL funded as project No. CZ.02.1.01/0.0/0.0/15_003/0000456, by Czech Republic Operational Programme Research and Development, Education, Priority 1: Strengthening capacity for quality research under the collaboration agreement with De La Salle University and Universiti Teknologi Malaysia is acknowledged. This work was also supported by the National Natural Science Foundation of China (Grant No. 72061127004) related to "AI + Waste to Resources Management toward Sustainable City".

Appendix A

See Table A1 here.

Appendix B. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jhazmat.2021.127330.

References

Abbasi, M., El Hanandeh, A., 2016. Forecasting municipal solid waste generation using artificial intelligence modelling approaches. Waste Manag. 56, 13–22.

Abdulredha, M., Al Khaddar, R., Jordan, D., Kot, P., Abdulridha, A., Hashim, K., 2018. Estimating solid waste generation by hospitality industry during major festivals: a quantification model based on multiple regression. Waste Manag. 77, 388–400.

Abdallah, M., Talib, M.A., Feroz, S., Nasir, Q., Abdalla, H., Mahfood, B., 2020. Artificial intelligence applications in solid waste management: a systematic research review. Waste Manag. 109, 231–246.

Adeleke, O., Akinlabi, S.A., Jen, T.C., Dunmade, I., 2021. Application of artificial neural networks for predicting the physical composition of municipal solid waste: an assessment of the impact of seasonal variation. Waste Manag. Res. 0734242×21991642.

Akoglu, H., 2018. User's guide to correlation coefficients. Turk. J. Emerg. Med. 18 (3), 91–93.

Borrelle, S.B., Ringma, J., Law, K.L., Monnahan, C.C., Lebreton, L., McGivern, A., Rochman, C.M., 2020. Predicted growth in plastic waste exceeds efforts to mitigate plastic pollution. Science 369 (6510), 1515–1518.

Boubellouta, B., Kusch-Brandt, S., 2020. Testing the environmental Kuznets Curve hypothesis for E-waste in the EU28+ 2 countries. J. Clean. Prod. 277, 123371.

Bholowalia, P., Kumar, A., 2014. EBK-means: a clustering technique based on elbow method and k-means in WSN. Int. J. Comput. Appl. 105 (9) https://doi.org/10.5120/18405-9674.

Brownlee J., 2019. How to use learning curves to diagnose machine learning model performance. ⟨machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/⟩ (Accessed 8 June 2021).

Bzdok, D., Altman, N., Krzywinski, M., 2018. Points of significance: statistics versus machine learning. Nat. Methods 15, 233–234.

Carey, J., 2017. On the brink of a recycling revolution? Proc. Natl. Acad. Sci. USA 114, 612–616.

CE Delft, 2017. Environmental prices handbook ⟨www.cedelft.eu/en/environmental-prices⟩ (accessed 24 June 2021).

Ceylan, Z., 2020. Estimation of municipal waste generation of Turkey using socio-economic indicators by Bayesian optimization tuned Gaussian process regression. Waste Manag. Res. 38 (8), 840–850.

Charitou, A., Aga-Spyridopoulou, R.N., Mylona, Z., Beck, R., McLellan, F., Addamo, A.M., 2021. Investigating the knowledge and attitude of the Greek public towards marine plastic pollution and the EU Single-Use Plastics Directive. Mar. Pollut. Bull. 166, 112182.

Chiu, A.S.F., Aviso, K.B., Baquillas, J., Tan, R.R., 2020. Can disruptive events trigger transitions towards sustainable consumption? Clean. Responsible Consum. 1. Article 100001.

Coskuner, G., Jassim, M.S., Zontul, M., Karateke, S., 2021. Application of artificial intelligence neural network modeling to predict the generation of domestic, commercial and construction wastes. Waste Manag. Res. 39 (3), 499–507.

Cox, K.D., Covernton, G.A., Davies, H.L., Dower, J.F., Juanes, F., Dudas, S.E., 2019. Human consumption of microplastics. Environ. Sci. Technol. 53, 7068–7074.

de Medeiros Back, H., Junior, E.C.V., Alarcon, O.E., Pottmaier, D., 2021. Training and evaluating machine learning algorithms for ocean microplastics classification through vibrational spectroscopy. Chemosphere 287 (1), 131903.

de Sá, L.C., Oliveira, M., Ribeiro, F., Rocha, T.L., Futter, M.N., 2018. Studies of the effects of microplastics on aquatic organisms: what do we know and where should we focus our efforts in the future? Sci. Total Environ. 645, 1029–1039.

de Souza Machado, A.A., Kloas, W., Zarfl, C., Hempel, S., Rillig, M.C., 2018. Microplastics as an emerging threat to terrestrial ecosystems. Glob. Change Biol. 24, 1405–1411.

De Veaux, R.D., Ungar, L.H., 1994. Multicollinearity: a tale of two nonparametric regressions. Selecting Models from Data. Springer, New York, NY, USA, pp. 393–402.

Duman, G.M., Kongar, E., Gupta, S.M., 2019. Estimation of electronic waste using optimized multivariate grey models. Waste Manag. 95, 241–249.

EC (European Commission), 2020a. ⟨ec.europa.eu/info/sites/default/files/plastics-factsheet-challenges-opportunities_en.pdf⟩ (Accessed 24 June 2021).

EC (European Commission), 2020b. Circular economy action plan for a cleaner and more competitive Europe. ⟨ec.europa.eu/environment/circular-economy/pdf/new_circular_economy_action_plan.pdf⟩ (Accessed 25 August 2020).

EC (European Commission), 2021a. ⟨ec.europa.eu/environment/strategy/plastics-strategy_en⟩ (Accessed 24 June 2021).

EC (European Commission), A European strategy for plastics in a circular economy. ⟨ec.europa.eu/environment/pdf/circular-economy/plastics-strategy-brochure.pdf⟩ (Accessed 5 September 2021).

Eurostat, 2021. ⟨ec.europa.eu/eurostat/en/⟩ (Accessed 24 June 2021).

Fan, Y.V., Jiang, P., Klemeš, J.J., Liew, P.Y., Lee, C.T., 2021a. Integrated regional waste management to minimise the environmental footprints in circular economy transition. Resour. Conserv. Recycl. 168, 105292.

Fan, Y.V., Klemeš, J.J., Lee, C.T., Tan, R.R., 2021b. Demographic and socio-economic factors including sustainability related indexes in waste generation. Energy Sources Part A: Recovery Util. Environ. Eff. https://doi.org/10.1080/15567036.2021.1974610.

Ghayebzadeh, M., Taghipour, H., Aslani, H., 2020. Estimation of plastic waste inputs from land into the Persian Gulf and the Gulf of Oman: an environmental disaster, scientific and social concerns. Sci. Total Environ. 733, 138942.

Gruber, F., Grählert, W., Wollmann, P., Kaskel, S., 2019. Classification of black plastics waste using fluorescence imaging and machine learning. Recycling 4 (4), 40.

Gue, I.H.V., Ubando, A.T., Tseng, M.L., Tan, R.R., 2020. Artificial neural networks for sustainable development: a critical review. Clean Technol. Environ. Policy 1–17.

Guo, H.N., Wu, S.B., Tian, Y.J., Zhang, J., Liu, H.T., 2020. Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: a review. Bioresour. Technol., 124114

Hao, J., Ho, T.K., 2019. Machine learning made easy: a review of scikit-learn package in python programming language. J. Educ. Behav. Stat. 44 (3), 348–361.

Helms, B.A., Russell, T.P., 2016. Reaction: polymer chemistries enabling cradle-to-cradle life cycles for plastics. Chem 1 (6), 816–818.

Herberz, T., Barlow, C.Y., Finkbeiner, M., 2020. Sustainability assessment of a single-use plastics ban. Sustainability 12 (9), 3746.
Jia, L., Evans, S., van der Linden, S., 2019. Motivating actions to mitigate plastic pollution. Nat. Commun. 10 (1), 1–3.
Jiang, P., Liu, X., 2016. Hidden Markov model for municipal waste generation forecasting under uncertainties. Eur. J. Oper. Res. 250 (2), 639–651.
Jiang, P., Fan, Y.V., Klemeš, J.J., 2021. Data analytics of social media publicity to enhance household waste management. Resour. Conserv. Recycl. 164, 105146.
Jupyter, 2021. ⟨jupyter.org/⟩ (Accessed 24 June 2021).
Kabir, E., Kaur, R., Lee, J., Kim, K.H., Kwon, E.E., 2020. Prospects of biopolymer technology as an alternative option for non-degradable plastics and sustainable management of plastic wastes. J. Clean. Prod. 258, 120536.
Kannangara, M., Dua, R., Ahmadi, L., Bensebaa, F., 2018. Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. Waste Manag. 74, 3–15.
Kaza, S., Yao, L., Bhada-Tata, P., Van Woerden, F., 2018. What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050. Urban Development. World Bank, USA, Washington, DC. ⟨openknowledge.worldbank.org/handle/10986/30317⟩ License: CC BY 3.0 IGO.
Klemeš, J.J., Fan, Y.V., Jiang, P., 2021. Plastics: friends or foes? The circularity and plastic waste footprint. Energy Sources Part A: Recovery Util. Environ. Eff. 43 (13), 1549–1565.
Klemeš, J.J., Fan, Y.V., Tan, R.R., Jiang, P., 2020a. Minimising the present and future plastic waste, energy and environmental footprints related to COVID-19. Renew. Sustain. Energy Rev. 127, 109883.
Klemeš, J.J., Fan, Y.V., Jiang, P., 2020b. The energy and environmental footprints of COVID-19 fighting measures–PPE, disinfection, supply chains. Energy 211, 118701.
Klemeš, J.J., Ponton, J.W., 1992. An analysis of the efficiency of neural nets application for simulation of distillation processes. Theor. Found. Chem. Eng. 26/3, 412–424, 1992 (In Russian, English translation published by Plenum Publishing Corp, NY).
Kumar, A., Samadder, S.R., Kumar, N., Singh, C., 2018. Estimation of the generation rate of different types of plastic wastes and possible revenue recovery from informal recycling. Waste Manag. 79, 781–790.
Law, K.L., Starr, N., Siegler, T.R., Jambeck, J.R., Mallos, N.J., Leonard, G.H., 2020. The United States' contribution of plastic waste to land and ocean. Sci. Adv. 6 (44), eabd0288.
Lau, W.W., Shiran, Y., Bailey, R.M., Cook, E., Stuchtey, M.R., Koskella, J., Palardy, J.E., 2020. Evaluating scenarios toward zero plastic pollution. Science 369 (6510), 1455–1461.
Lebreton, L., Andrady, A., 2019. Future jscenarios of global plastic waste generation and disposal. Palgrave Commun. 5 (1), 1–11.
Liang, Y., Tan, Q., Song, Q., Li, J., 2021. An analysis of the plastic waste trade and management in Asia. Waste Manag. 119, 242–253.
Lundberg, S., Lee, S.I., 2017. A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874.
Lundberg, S.M., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Lee, S.I., 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat. Biomed. Eng. 2 (10), 749–760.
Lundberg, 2021. ⟨github.com/slundberg/shap⟩ (Accessed 24 June 2021).
Ndanguza, D., Nyirahabinshuti, A., Sibosiko, C., 2020. Modeling the effects of toxic wastes on population dynamics. Alex. Eng. J. 59 (4), 2713–2723.
Nordic Council of Ministers, 2020, Possible elements of a new global agreement to prevent plastic pollution. ⟨www.nordicreport2020.com⟩ (Accessed 7 June 2021).
OEC (The Observatory of Economic Complexity), 2021a. ⟨oec.world/en/rankings/eci/hs6/hs96⟩ (Accessed 24 June 2021).
OEC (the Observatory of Economic Complexity), 2021b. Methods. ⟨oec.world/en/resources/methods⟩ (Accessed 7 June 2021).
Ozdemir, M.E., Ali, Z., Subeshan, B., Asmatulu, E., 2021. Applying machine learning approach in recycling. J. Mater. Cycles Waste Manag. 1–17.
Pavlas, M., Šomplák, R., Smejkalová, V., Stehlík, P., 2020. Municipal solid waste fractions and their source separation: forecasting for large geographical area and its subregions. Waste Biomass Valoriz. 11 (2), 725–742.
Petersen, F., Hubbart, J.A., 2021. The occurrence and transport of microplastics: the state of the science. Sci. Total Environ. 758. Article 143936.
Python Software Foundation, 2021. Python Language Reference. ⟨www.python.org⟩ (Accessed 24 June 2021).
Ribeiro, M.T., Singh, S., Guestrin, C., 2016, Why should I trust you?: explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.
Rizkin, B.A., Shkolnik, A.S., Ferraro, N.J., Hartman, R.L., 2020. Combining automated microfluidic experimentation with machine learning for efficient polymerization design. Nat. Mach. Intell. 2 (4), 200–209.
Scheinost, D., Noble, S., Horien, C., Greene, A.S., Lake, E.M., Salehi, M., Constable, R.T., 2019. Ten simple rules for predictive modeling of individual differences in neuroimaging. NeuroImage 193, 35–45.
Scikit Learn, 2021. ⟨scikit-learn.org/stable/⟩ (Accessed 24 June 2021).
Shen, M., Huang, W., Chen, M., Song, B., Zeng, G., Zhang, Y., 2020. Micro) plastic crisis: un-ignorable contribution to global greenhouse gas emissions and climate change. J. Clean. Prod. 254, 120138.
Shapley, L.S., 1951. Notes on the n-Person Game – II: The Value of an n-Person Game. RAND Corporation, Santa Monica, CA, USA.
Tan, R.R., Aviso, K.B., Lao, A.R., Promentilla, M.A.B., 2021. Modelling vicious networks with P-graph causality maps. Clean Technol. Environ. Policy. https://doi.org/10.1007/s10098-021-02096-x.
TensorFlow, 2021. ⟨www.tensorflow.org/⟩ (Accessed 24 June 2021).
Thinkstep, 2017, GaBi Software System and Database for Life Cycle Engineering GaBi ts version, Stuttgart, Germany.
TU Delft, 2021. ⟨www.ecocostsvalue.com/eco-costs⟩ (Accessed 24 June 2021).
UN Comtrade, 2021. United Nations Commodity Trade Statistics Database, 2021. ⟨comtrade.un.org/⟩ (Accessed 24 June 2021).
Wang, Z., Peng, B., Huang, Y., Sun, G., 2019. Classification for plastic bottles recycling based on image recognition. Waste Manag. 88, 170–181.
Wang, C., Xiao, J., Xing, B., 2020. Environmental source, fate, and toxicity of microplastics. J. Hazard. Mater. 407. Article 124357.
Watkins E., Gionfra S., Schweitzer J., Pantzar M., Janssens C., ten Brink P., 2017. EPR in the EU plastics strategy and the circular economy: a focus on plastic packaging. ⟨ieep.eu/uploads/articles/attachments/95369718-a733–473b-aa6b-153c1341f581/EPR%20and% 20plastics%20report%20IEEP%209%20Nov%202017%20final.pdf⟩ (Accessed 5 September 2021).
Wen, Z., Xie, Y., Chen, M., Dinga, C.D., 2021. China's plastic import ban increases prospects of environmental impact mitigation of plastic waste trade flow worldwide. Nat. Commun. 12 (1), 1–9.
Wicke, B., Kluts, I., Lesschen, J.P., 2020. Bioenergy potential and greenhouse gas emissions from intensifying European temporary grasslands. Land 9 (11), 457.
Wolf, M., van den Berg, K., Garaba, S.P., Gnann, N., Sattler, K., Stahl, F., Zielinski, O., 2020. Machine learning for aquatic plastic litter detection, classification and quantification (APLASTIC-Q). Environ. Res. Lett. 15 (11), 114042.
Wong, S.L., Nyakuma, B.B., Wong, K.Y., Lee, C.T., Lee, T.H., Lee, C.H., 2020. Microplastics and nanoplastics in global food webs: a bibliometric analysis (2009–2019). Mar. Pollut. Bull. 158, 111432.
Wu, F., Niu, D., Dai, S., Wu, B., 2020. New insights into regional differences of the predictions of municipal solid waste generation rates using artificial neural networks. Waste Manag. 107, 182–190.
Zhang, Y., Pu, S., Lv, X., Gao, Y., Ge, L., 2020. Global trends and prospects in microplastics research: a bibliometric analysis. J. Hazard. Mater. 400. Article 123110.
Zhao, X., Cornish, K., Vodovotz, Y., 2020. Narrowing the gap for bioplastic use in food packaging: an update. Environ. Sci. Technol. 54, 4712–4732.
Zhu, S., Chen, H., Wang, M., Guo, X., Lei, Y., Jin, G., 2019. Plastic solid waste identification system based on near infrared spectroscopy in combination with support vector machine. Adv. Ind. Eng. Polym. Res. 2 (2), 77–81.