Neural Computing and Applications (2021) 33:7073–7087
https://doi.org/10.1007/s00521-020-05476-4

ORIGINAL ARTICLE

Machine learning models for ecological footprint prediction based on energy parameters

Radmila Janković[1] · Ivan Mihajlović[2] · Nada Štrbac[2] · Alessia Amelio[3]

Received: 17 November 2019 / Accepted: 26 October 2020 / Published online: 17 November 2020

Abstract
The ecological footprint is an excellent tool to better understand the consequences of the human behavior on the environment. The growing need for natural resources emphasizes the necessity of their accurate observation, calculation, and prediction. This paper develops and compares four hybrid machine learning models for predicting the total ecological footprint of consumption based on a set of hyper-parameters predefined by the Bayesian optimization algorithm. In particular, K-nearest neighbor regression (KNNReg), random forest regression (RFR) with 93 trees, and two artificial neural networks (ANNs) with two hidden layers were developed and later compared in terms of their performance. As energy inputs, the primary energy consumption from (1) natural gas sources, (2) coal sources, (3) oil sources, (4) wind sources, (5) solar photovoltaic sources, (6) hydropower sources, (7) nuclear sources, and (8) other renewable sources was used. Additionally, population number has also been used as an input. The models were developed using a set of data that include 1804 instances. The ANNs were modeled using two different activation functions in the hidden layers: ReLU and SPOCU. The performance was evaluated using the mean absolute percentage error (MAPE), mean absolute scaled error (MASE), normalized root-mean-squared error (NRMSE), and symmetric mean absolute percentage error (SMAPE). The results show that KNNReg performs the best with MASE of 0.029, followed by the RFR (0.032), ReLU ANN (0.064), and SPOCU ANN (0.089). Moreover, SMOGN was utilized to produce a synthetic test set which was used to additionally test the best performed model. The performance on the SMOGN set demonstrates good performance (MASE=0.022). Lastly, the best performed model was implemented into a GUI that calculates the ecological footprint based on user inputs.

Keywords Ecological footprint · Prediction · Energy · Modeling · Machine learning

1 Introduction

The sustainability is a very important concept of future development. As the industry develops, the comfort of the everyday life increases based on the increase in the national gross domestic product (GDP). Higher GDP leads to consumption and, consequently, waste increase and decrease in the natural capital. All human activities leave consequences for the environment. For instance, the transportation industry is one of the biggest users of fuel and burning; it has many negative implications, such as the increase in carbon in the atmosphere and, as a result, the increase in global warming, the deterioration of the environment, and the negative impact on the public health.

Many studies show that the amounts of some of the natural resources are close to a very low level. According to the British Petroleum [1], all coal reserves will disappear

✉ Radmila Janković
rjankovic@mi.sanu.ac.rs

Ivan Mihajlović
imihajlovic@tfbor.bg.ac.rs

Nada Štrbac
nstrbac@tfbor.bg.ac.rs

Alessia Amelio
a.amelio@dimes.unical.it

[1] Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, Serbia

[2] Technical Faculty in Bor, University of Belgrade, Bor, Serbia

[3] DIMES, University of Calabria, Rende, Cosenza, Italy

in the next 132 years, the natural gas reserves in 52 years, while the oil reserves are going to disappear in 50 years. As fossil fuels represent the main cause of greenhouse emissions in the atmosphere, it is no wonder that in 2018 the level of global CO_2 emissions increased by 2% [1]. Despite many environmental policies, the levels of pollution and nature degradation increase each year, hence amplifying the need to transition to renewables. Furthermore, the development of the industry highly depends on natural resources, and the upcoming Fourth Industrial Revolution may result in an increase in energy production and consumption because it will include the integration of systems and smart technologies.

The global environment protection is one of the main goals of many organizations and entities, such as the United Nations, which listed 17 sustainable goals that have to be reached by 2030 [2]. In these terms, the protection of the natural resources of our planet includes production and consumption in accordance with the sustainable policies, a responsible use of natural resources, and lowering the effects of the climate change [2]. Furthermore, in 2015, 195 countries signed the Paris Agreement by which they legally agreed to take actions in order to limit the global warming [3]. In particular, the aim is to keep the increase in the global average temperatures below 2°C comparing to the pre-industrial levels and to limit the global emissions [3]. In order to evaluate the sustainability, the amount of available natural capital, and the human demands for the resources accumulated as the natural capital, the ecological footprint (EF) needs to be observed, predicted, and compared.

The EF was developed during the 1990s and represents the measure of the natural capital that is needed for human activities [4–6]. The EF can also be considered as a metric of sustainability as every human activity leaves a "footprint" on the planet. Furthermore, the EF can be perceived as an element of the demand and usage of the natural resources. In order to understand the EF, the biocapacity should also be observed. The biocapacity represents a biologically reproductive area. From the economic perspective, the relation between the EF and the biocapacity is the relation of demand and supply of natural resources. In this sense, the biocapacity represents the supply, while the EF represents the demand.

There are two types of EF: (1) the EF of production and (2) the EF of consumption. The EF of production can be defined as the sum of the footprints for all the products in an observed category [7]. On the contrary, the EF of consumption is the capacity of biologically productive area (biocapacity) used to support the consumption patterns of the defined population. Both biocapacity and the EF are measured in units of global hectares (gha) and should be, in most cases, compared so that the difference between the needs of one population and the available natural resources can be observed.

The EF is calculated based on the six different land use types: (1) cropland, (2) grazing land, (3) forest land, (4) carbon footprint, (5) fishing grounds, and (6) built-up land [7]. For cropland, grazing land, and fishing grounds, the EF is calculated by summing up the impact of variety of products on the environment [7]. The built-up land represents the area of land that is covered by infrastructure, while the forest land presents the forests needed to absorb the carbon dioxide emissions [7] and also includes harvests of wood and timber. Two methods are used for determining the EF: (1) the integral method, which observes the energy and material flows of an observed system, and (2) the component method, which observes the real energy and material consumption of an individual [8].

The natural resources are highly used for producing energy. In particular, the fossil fuels which are non-renewable natural sources and include coal, natural gas, and oil represent the biggest part of energy production and consequently consumption. These three sources of energy make up to 89% of the global carbon dioxide (CO2) emissions [9]. Burning fossil fuels releases big amounts of carbon and other greenhouse gases in the atmosphere which cannot be absorbed by the oceans and forests; hence, these gases accumulate and create the greenhouse effect. This leads to an increase in global temperatures, the melting of polar ice caps, and consequently to the sea level rise, which all together have serious and devastating consequences to the environment. Moreover, resource exhaustion decreases the biocapacity and increases the EF, especially the carbon and forest land footprints. The importance of energy research is even higher considering the fact that in 2018 the global energy demand increased of 2.3%, almost twice the average rate from 2010 [10].

Modeling of ecological and climatic processes is often complicated due to the presence of chaos and uncertainty. As we investigate the EF which is composed of different land area types that can be observed as ecosystems in a way, the EF by itself can behave differently than the processes in each land type separately. The behavior of smaller systems of the biosphere can be different than the behavior of a complete biosphere [11]. For example, carbon land is affected by carbon emissions which are by nature chaotic. The same can be observed for energy and population, which are both chaotic processes. Additionally, many processes are so complicated that they are stochastic and chaotic—such example is the methane release from the soil [11], but other greenhouse gas emissions including carbon emissions can also be considered chaotic [12]. It is assumed that the chaotic behavior of ecological processes is driven by different complicated interactions between the components in the process [11]. Furthermore, the presence

of chaos means that there is an unpredictability of a deterministic system [13]. Since EF, energy consumption, and population highly depend on the human factor, they can be observed as complex chaotic systems.

This paper proposes a hybrid approach composed of Bayesian optimization (BO) using the tree-structured Parzen estimator (TPE) and state-of-the-art machine learning (ML) models for EF prediction. The novelty of the research is explained in detail in the next section, after the presentation of the literature research. The paper is organized as follows. Section 2 presents the literature review, Sect. 3 describes the data and used methodology, while Sect. 4 presents the results. Finally, Sect. 5 makes a discussion and draws conclusions.

2 Literature review

Lu and Bin [14] proposed a Markov chain-based prediction model of Beijing's urban EF. The system dynamic model was used for estimating the EF in the period 2001–2020, and the results showed that the energy parameters strongly affect the EF, while the predictions showed stable values of the EF due to a rational consumption [14]. The EF of Beijing was also predicted in [15] using a support vector machine (SVM) model. The authors first calculated the EF of Beijing in the period 1996–2015, where the EF showed a rising trend until 2013, when the government issued a policy affecting polluting businesses which decreased the total EF. The inputs were then evaluated for further analysis using the partial least squares method, after which six inputs were chosen. In these terms, population, GDP, total retail sales of consumer goods, gross industrial production, total foreign trade, and total energy consumption were used as inputs into the back-propagation neural network (BPNN) model, while the EF was used as an output. The selection of hyper-parameters was performed based on a number of trials. The BPNN predicted the values of the EF with 2% and 3% of error for 2014 and 2015, respectively. The application of the SVM regression model predicted the EF for 2014 and 2015 with 1% and 0.64% of error, respectively [15]. Hence, the authors indicated the appropriateness of using the SVM model for the prediction of the EF, especially with a limited sample size [15]. Moreover, the EF of Beijing is having a new rising trend from 2016 to 2020, especially due to the energy consumption and land use [15].

A study by Wang et al. [16] investigated and predicted the China's environmental sustainability based on a modified EF model which observes the freshwater EF, the net primary productivity, and the energy EF. The authors used the linear ARIMA and nonlinear ANN (in particular BPNN) models to forecast the EF and ecological biocapacity per capita. The optimization of hyper-parameters for the BPNN model included experimentation in a number of trials. The results show that the combination of ARIMA and ANN can be more enhanced for future predictions. Moreover, the findings of this research revealed that the EF in China is going to further increase [16].

Considering the impact of the energy on the EF, Bello et al. [17] investigated the impact of the electricity consumption on CO_2 emissions, carbon footprint, water footprint, and the EF. The results showed that the hydroelectricity positively affects the environment through a decrease in the environmental degradation. On the contrary, it was found that fossil fuels increase the water footprint [17].

The research of a relationship between the economic growth, energy consumption, financial development, and the EF was performed in a study by Destek and Sarkodie [18]. The study analyzed the data of 11 countries in the period 1977–2013 and tested the environmental Kuznets curve hypothesis. The results showed that a high use of energy increases the EF of some of observed countries and, consequently, that a high energy consumption increases the environmental degradation in the observed countries [18].

Lu et al. [19] performed the analysis on the influential factors of the EF and the prediction of the EF. Based on six selected socioeconomic factors, a partial least squares (PLS) method and a grey model first-order one variable (GM(1,1)) method were developed to predict the EF based on the time series data including energy consumption, total population, GDP per capita, annual per capita disposable income of urban households, total retail sales of consumer goods, and cultivated areas. The results of the simulation showed high ability of the PLS regression to predict the EF.

A study by Yao [20] considers the creation of a simulation model of the total EF of Suzhou by developing a BPNN model, based on seven factors: the GDP, secondary industrial products, tertiary industrial products, urban population, rural population, annual income of city dwellers per capita, and annual income of rural residents per capita. The developed model performed well in terms of the simulation ability.

Huimin [21] performed a grey relational analysis of EF, energy consumption, and environmental protection. It was found that the negative behavior of energy consumption affects the ecosystem more than the positive behavior of the environmental protection. Moreover, it was shown that the electricity represents the most important part of the energy consumption, but is also more affected by the levels of the environmental capital [21]. Furthermore, in a study by Wu et al. [29] the authors used the autoregressive integrated moving average and BPNN to investigate several sustainability factors such as the EF, ecological capacity, ecological deficit/surplus, optimum population

size, EF of 104-yuan Gross National Product, and EF diversity. The hyper-parameters were selected based on a number of trials and comparison. The obtained results showed that both models achieved good precision in predicting the EF and the ecological capacity for the Tianjin City in China. Moreover, it was found that the deterioration of the environment will continue in the future [29].

The wide scope of application of ML algorithms is demonstrated by the usage of ML to tackle problems in many different areas of application. In [22], new deep genetic hierarchical network of learners (DGHNL) was proposed for prediction of credit scoring on the Statlog German credit approval data. The proposed DGHNL has 29 layers and uses five types of learners (two types of SVM algorithms, k-nearest neighbors (kNN), probabilistic neural networks (PNN), and fuzzy systems), two types of normalization techniques, two types of feature extracting methods, three types of kernel functions, and three types of parameter optimizations. The model obtained a prediction accuracy of 94.6% which is the best performance obtained for this dataset in the literature [22]. Credit scoring of Australia was predicted with the use of deep genetic cascade ensembles of classifiers (DGCEC) in [23]. The classifier comprises 16 layers and includes two types of SVM classifiers, normalization methods, feature extraction methods, three types of kernel functions, parameter optimizations, and stratified tenfold cross-validation. The model obtained a high accuracy of 97.39%, confirming the appropriateness of machine learning models in credit score decision-making [23].

A study by [24] presented a novel ResNet model for signal recognition, which uses ResNet18, ResNet50, and ResNet101 as feature extractors. The dataset included sensor data obtained from daily and sports activities such as sitting, standing, climbing stars, moving, and standing in elevator. The networks obtained 1000 features each, comprising a total dataset of 3000 features from which 1000 features were selected using ReliefF. The selected features represent inputs into the SVM with a cubic activation function. The proposed method obtained high accuracy of 99.96% and 99.61% for gender and activity recognitions, respectively [24].

In [25], the authors proposed a face recognition architecture based on fuzzy-based discrete wavelet transform and fuzzy with two novel local graph descriptors—vertical local cross-pattern (VLCP) and horizontal local cross-pattern (HLCP) that together compose the local cross-pattern. For feature classification, linear discriminant analysis, quadratic discriminant analysis, quadratic kernel-based SVM, and KNN were used. The model performance was evaluated on AT&T, CIE, Face94, and FERET databases. The proposed system obtained an accuracy of classification from 96% to 100%, depending on the dataset [25].

One of the most prospective applications of ML is in the area of medicine for disease classification and prediction. In [26], the authors proposed an evolutionary-based computer-aided diagnosis (CAD) system that uses machine learning to classify the wart disease treatment response. The study applied a combined approach of improved adaptive particle swarm optimization (IAPSO) algorithm and artificial immune recognition system (AIRS) to classify 180 records. Additionally, the authors used K2, K3, K4, K5, and K10 cross-validation to test the system. The best results indicated an accuracy of 90% for K10 [26].

Another use of machine learning techniques is for the heartbeat classification based on nonlinear morphological features and voting [27]. The proposed method used the MIT-BIH arrhythmia database comprising records of Holter monitoring where each record is 30 minutes long and contains 15 types of normal and abnormal heartbeats. Five classifiers were used, in particular naïve Bayes, LDA, QDA, J48, and J48 consolidated. A voting scheme with various combinations of these classifiers was used. The overall accuracy of each individual classifier suggests that the best performance was obtained for the J48 (94.32%), followed by LDA (92.60%) and naïve Bayes (81.47%), while the worst performance in terms of accuracy was obtained by QDA (16.20%) and J48 consolidated (8.86) algorithm, on a set with five classes. Using a voting scheme with five classes, a combination of J48, LDA, and naïve Bayes obtained an accuracy of 90.71%, while a combination of J48, naïve Bayes, and QDA obtained an overall accuracy of 83.6% [27]. Combining J48 consolidated, naïve Bayes, and QDA an accuracy of 71.51% was obtained, while a combination of J48 consolidated, naïve Bayes, and LDA obtained an accuracy of 87.91%. The overall accuracy of the investigated algorithms on three classes suggests that the best performing algorithm is the LDA (95.19%), followed by J48 (93.71%), naïve Bayes (85.15%), and J48 consolidated (79.76%), while QDA obtained a low accuracy of 19.9%. In terms of Voting scheme, on three classes the best performance was obtained for the J48, LDA, and naïve Bayes combination of classifiers (94.08%) [27]. The authentication based on heartbeat patterns obtained from electrocardiogram (ECG) was performed in [28] using two deep neural networks (DNNs). The first DNN is a convolutional neural network (CNN), while the second DNN is a residual convolutional neural network (ResNet) with attention mechanism (ResNet-Attention) [28]. The authors developed the DNNs based on two ECG datasets—PTB and CYBHi. Five different CNN architectures were tested, and the best performance was obtained for the CNN comprising of four convolution layers, two max pooling layers, three dropout layers, and two fully connected layers. A rectified linear unit (ReLU) was used as an activation function, while dropout was set to

0.2. The CNN obtained the accuracy of 98.6% (PTB database) and 99.7% (CYBHi database). The ResNet-Attention network obtained an accuracy of 98.8% (PTB database) and 99.2% (CYBHi database) [28].

As the energy and the EF are closely related, which was confirmed in many studies [8, 14–17], the objective of this research is to develop machine learning models for predicting the total EF of consumption, based on the population and energy parameters, and compare the obtained results in order to estimate the models' performance. The motivation behind this research lies in the fact that machine learning models can be used for modeling complex processes associated with the environment. As climate change progresses, it is certain that humanity needs to shift from fossil fuel energy sources to using renewable sources. In order to track the progress of that shift, machine learning can be employed to model the dependencies and predict the future of energy and its impact on sustainability, which is in this study represented as the total ecological footprint of consumption. Furthermore, the current models found in the literature mainly focus on assessing the environmental Kuznets curve or they use time-series models such as ARIMA, and complex ML models such as neural networks. Furthermore, the selection of hyper-parameters of EF ML models is mostly performed using trials, not parameter optimization algorithms. To address these shortcomings, we propose four ML models based on Bayesian optimization: (i) K-nearest neighbor regression (KNNReg), (ii) random forest regression (RFR), (iii) artificial neural network with ReLU activation function (ANN ReLU), and (iv) ANN with the Scaled POlynomial Constant Unit (SPOCU) activation function (ANN SPOCU).

There are several novelties in our study. First, the energy consumption and population number are used as the main inputs for modeling the EF using the ML algorithms, as these input parameters have not yet been modeled together to predict the total EF of consumption. Second, as environmental data are usually very complex, for successful ML modeling it is important to carefully choose the best set of model hyper-parameters; thus, this study proposes a hybrid approach for EF prediction in which the hyper-parameters were optimized using the Bayesian optimization (BO), particularly the tree-structured Parzen estimator (TPE). To the authors' knowledge, the BO has not been utilized in the field of ecological footprint prediction, as studies on the EF mostly employ trials to select the hyper-parameters of the model [15, 16, 29]. Third, we develop an ANN model using the newly developed SPOCU function [30] in its hidden layers which demonstrated good performance for dynamic data prediction. Fourth, we used RFR and KNN algorithms to model the EF, which has not been done before in this area of study, as to the authors' knowledge. Fifth, we used SMOGN, a synthetic minority over-sampling technique for regression with Gaussian noise, to develop a synthetic dataset that was fed into the best model to predict the EF. As environmental modeling frequently includes limited and/or missing data, it is important to investigate the performance on synthetic datasets that might be used as a substitute of the real-world data. Finally, the model with the best performance was used to build the graphical user interface (GUI) which automates the process of predicting the values of the total EF of consumption, based on user inputs. The developed models (as well as the GUI) can be implemented in many aspects of the industry in order to calculate the EF of a company, region, or country. Moreover, the models may be used to plan the resource allocation, create new policies and goals to preserve the environment, perform energy optimization, and understand the demographic and economic trends.

3 Data and methodology

3.1 Data

Data analysis and modeling, presented in this manuscript, was performed in several stages and included modeling of the dependency between the EF values and the data on energy consumption. The dataset contained one output and nine input variables, in particular: (1) the total EF of consumption (in gha), (2) population number, primary energy consumption (in terawatt-hours—TWh) from (3) oil sources, (4) gas sources, (5) coal sources, (6) solar photovoltaic (PV) sources, (7) other renewables, (8) wind sources, (9) nuclear sources, and (10) hydropower. The data regarding the primary energy consumption from different sources were provided by British Petroleum and obtained from Our World in Data Web site [31], the EF data were provided by the Global Footprint Network [32] and obtained from Data World Web site [33] (https://data.world/footprint/nfa-2018-edition), and the population data were obtained from the World Bank open data [34] (https://data.worldbank.org/indicator/SP.POP.TOTL). The final dataset included 1804 datalines from 41 countries in the period 1971–2014. The descriptive statistics of the dataset for each of the variables is presented in Table 1. The following can be observed. Among fossil fuels, oil is a predominant source of energy with a mean value of 667.26 TWh, followed by coal with a mean value of 324.86 TWh. Considering the renewable energy sources, hydro sources have the highest mean of 39.37 TWh, followed by other renewable sources, wind, and finally solar PV. It should be noted that energy produced from nuclear sources has a mean of 35.19 TWh, almost the same as hydro sources which are more frequently utilized worldwide.

Table 1 Descriptive statistics of the dataset

Statistics	Total EF	Population	Oil	Gas	Coal	Solar	Other renew.	Wind	Nuclear	Hydro
Mean	2.13E+08	7.10E+07	667.26	46.45	324.86	0.29	5.35	1.80	35.19	39.37
St. dev.	4.16E+08	1.50E+08	1503.87	991.00	908.75	2.07	18.14	9.53	112.07	73.69
Minimum	2.20E+06	9.56E+05	7.86	0	0	0	0	0	0	0
Maximum	3.08E+09	1.29E+09	10913.54	8025.12	6681.01	36.06	296.78	183.49	849.44	428.33

3.2 Methodology

In this paper, four ML algorithms were used jointly with the Bayesian optimization to develop the prediction models of the EF. Moreover, the best performing model was additionally tested on a set of synthetically generated data using the SMOGN algorithm.

The proposed models were developed on a computer with 2.3GHz quad-core processor, NT-based OS, 8 GB DDR4 RAM, Ryzen 7 processor, and Radeon RX Vega 10 GPU. The experiments were performed in Python 3.7.

3.2.1 Data preparation and correlation

To test if there is a relationship between the observed input and output variables, the correlation test was utilized. Correlation was examined using Spearman's correlation test which is a nonparametric test of association between two variables. The values of the Spearman's correlation coefficient can range between -1 and 1, where values closer to |1| indicate a strong correlation between the observed variables [35]. Furthermore, a correlation value of zero indicates that there is a nonexisting relationship, a correlation of up to 0.4 indicates a weak relationship, while a correlation of up to 0.7 shows a moderate relationship between the variables [35]. The values of correlation of above 0.7 show strong correlation. Depending on the sign of the correlation coefficient, the relationship between the variables can be positive or negative.

Before the model development, the data were scaled using the StandardScaler from *sklearn* library. StandardScaler standardizes each feature column of the dataset independently so that it has a mean of zero and standard deviation of 1 [36]. This preprocessing tool is based on the following equation of standardization:

$$Z = (x - \mu)/\sigma \tag{1}$$

where μ represents the mean and σ is the standard deviation of the feature column of x, while x is the input value.

3.2.2 ML models

The KNN regression maps the regression function f between input and output variables by calculating the similarity of features. Similarity in a sample of continuous variables is calculated using a distance function such as Euclidean, Minkowski, or Manhattan. The performance of the algorithm depends on the value of k where a large value of k minimizes the loss [37], while the small value of k is more efficient in terms of computation costs. To obtain the best possible performance, model parameters should be a subject of careful selection and evaluation. The parameter optimization technique used in this research (see Sect. 3.3) suggested using Manhattan distance, which can be calculated as:

$$\sum_{i=1}^{k} |x_i - y_i| . \tag{2}$$

The algorithm makes predictions $\hat{y_i}$ of x by taking the average of the outputs y_i of k-nearest neighbors. Therefore, predictions can be calculated as:

$$\hat{y} = \frac{1}{K} \sum_{i=1}^{k} y_i \tag{3}$$

The RF is an ensemble learning algorithm that is based on bootstrap aggregation, otherwise known as bagging and was first proposed in [38]. In RF, bagging consecutive trees are constructed regardless of each other [39] using a different bootstrap data sample. Sample split is performed using a predefined number of randomly selected k predictor. Furthermore, RF avoids the problem of bias and variance by constructing the trees in parallel [38]. The RF algorithm is proved to perform well compared to other state-of-the-art techniques such as SVMs and is also robust to overfitting [38]. A RF regression predictor is calculated as [40]:

$$f^{C}(x) = \frac{1}{C} \sum_{i=1}^{C} T_i(x) \tag{4}$$

where x is the input, C represents the number of trees, and $T_i(x)$ is a single regression tree.

The ANN presents an analogy to the biological neural system and consists of interconnected neurons (nodes)

[41]. The ANN performs tasks based on given examples; hence, it learns from the experience and detects relationships between the variables even when no obvious relationship exists [41]. No apriori assumptions are made when utilizing neural networks [42], as opposed to the multiple regression models and other parametric techniques. Therefore, the ANN can be applied to various types of data and for different purposes.

The ANN consists of the input, output, and hidden layers which are composed of neurons. The number of neurons in the input and the output layers is determined by the number of input variables (for the input layer) and the type of output variable (for the output layer). The number of hidden layers and hidden nodes is a crucial part of modeling with ANNs. The inappropriate selection of these parameters impacts the model's ability to generalize. In these cases, two situations can occur. The first one is known as underfitting when the network has a poor performance on the training data because it is unable to fit the function to the data, thus making it unable to perform well on the test data, while overfitting occurs when the network has an excellent performance on the training data, thus fitting the function too closely to the data, but it fails to generalize on unseen data.

3.2.3 Parameter optimization

Machine learning models require a predefined set of parameters; hence, a careful parameter selection is needed before training the models. In this study, parameter estimation was performed using the tree-structured Parzen estimator (TPE)—a Bayesian optimization (BO) algorithm for hyper-parameter optimization.

The BO is a sequential optimization approach that puts a prior probability distribution over an objective function, evaluates the function, and then changes the prior probability distribution to posterior distribution. It is a probabilistic algorithm that can perform optimization using Gaussian processes (GP), random forests, and TPE. The TPE aims to optimize objective functions that are costly to evaluate. Here, the parameter space is tree-structured and sequentially constructed; thus, parameters preserve their conditional dependence. It supports different types of variables in the configuration space such as uniform, log-uniform, quantized log-uniform, and categorical [43].

The algorithm simultaneously chooses which variables to optimize, and instead of modeling $p(y|x)$ directly, it rather models $p(x|y)$ and $p(y)$ [43]. In fact, TPE models $p(x|y)$ by performing the Parzen window density estimation which generates two densities, $\ell(x)$ and $g(x)$, as:

$$p(y|x) = \begin{cases} \ell(x) & if \quad y < y^* \\ g(x) & if \quad y > y^* \end{cases}$$

where y^* is some quantile of the observed y value, thus $p(y < y^{\gamma}) = \gamma$ [43].

The expected improvement (EI) is maximized when points with high probability belong to $\ell(x)$ and those with low probability to $g(x)$ [43]. The optimization of the TPE algorithm is expressed as: $EI_{y^*}(x)\alpha\left(\gamma + \frac{g(x)}{\ell(x)}(1-\gamma)\right)^{-1}$.

Parameter search was performed for all the models developed in this study to find the best combination that will yield the highest performance. The set of parameters that were evaluated by the TPE for each of the models is given in Table 2.

Table 2 TPE parameters

Algorithm	Parameter grid	Proposed configuration
KNNReg	Number of neighbors [2, 200]	Number of neighbors = 2
	Ttypes of algorithms [auto, ball tree, kd tree, brute]	Type of algorithm = brute
	Power parameter p [1, 2]	p = 1
	Leaf size [1, 80]	Leaf size = 69
RFR	Number of estimators [1, 300]	Number of estimators = 93
	Bootstrap [True, False]	Boostrap = False
	Minimum samples split [1, 80]	Minimum samples split = 2
	Maximum depth [1, 50]	Maximum depth = 33
	Maximum features [auto, sqrt, log2]	Maximum features = sqrt
ANN ReLU	Batch size [16, 32, 64, 128, 256]	Batch size = 256
	Number of neurons [3, 200]	Number of neurons (120, 136)
	Number of hidden layers [1, 5]	Number of hidden layers = 2
	Dropout [0.3, 0.4, 0.5, 0.6]	Dropout = 0.3
ANN SPOCU	Batch size [16, 32, 64, 128, 256]	Batch size = 32
	Number of neurons [3, 200]	Number of neurons (14, 168)
	Number of hidden layers [1, 5]	Number of hidden layers = 2
	Dropout [0.3, 0.4, 0.5, 0.6]	Dropout = 0.3

After selecting the optimal parameters, the ML models were specified.

3.2.4 Performance evaluation

To objectively determine the model performance, ML models need to be evaluated on new data. As ML usually requires large datasets, when dealing with a limited number of samples one can utilize a re-sampling technique to perform evaluation on an unseen data in order to preserve the whole set for training. Cross-validation (CV) uses a portion of the data that are set aside and not used during training. In this study, the KNNReg and RFR were evaluated using K-fold cross-validation with ten folds. K-fold CV splits the dataset into a k number of folds where each fold is used to validate the model. The process starts by shuffling and then splitting the dataset into k-fold. In the process, onefold is used to validate the model, while the remaining k-1 folds are used for training [44], until all k-subsamples are used for validation exactly one time.

The overall model performance was evaluated using four statistical measures. In particular, we have obtained and compared the MAPE, MASE, NRMSE, and SMAPE.

MAPE presents a measure of forecasting accuracy and is represented in percentages which makes the interpretation of the results easier than with other metrics. It is calculated as in:

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{e_i}{x_i}\right| \tag{5}$$

where e_i is the error between actual and predicted values, while x_i are the actual values. MAPE can be interpreted as follows: Up to 10% indicates a highly accurate forecasting, from 10 to 20% MAPE indicates a good forecasting, 20 to 50% MAPE shows reasonable forecasting, while MAPE above 50% indicates a weak model [45]. MAPE is widely used for energy and environmental modeling, although it has some drawbacks including the inability to handle zero values in which case MAPE is undefined.

The NRMSE presents the normalized square root of the MSE (i.e., root-mean-squared error—RMSE), as in Eq. 6. It shows the normalized standard deviation of prediction errors, and it should be close to zero to indicate good prediction. It is expressed as:

$$\text{NRMSE} = \frac{1}{\bar{x}}\sqrt{\sum_{i=1}^{n} e_i^2}\,. \tag{6}$$

SMAPE is a modification of the MAPE developed to overcome the disadvantages of MAPE, especially the inability to process zero values. SMAPE is defined as:

$$\text{SMAPE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|x_i - \hat{x_i}|}{(x_i - \hat{x_i})/2}\right) * 100\% \tag{7}$$

where x_i is the actual value and $\hat{x_i}$ is the predicted value.

The final metric that was considered is the MASE. MASE is a metric that was also developed to overcome the disadvantages of other measures and is robust and stable. It is represented as:

$$\text{MASE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - \hat{x_i}}{\frac{1}{n-1}\sum_{i=1}^{n}|x_i - x_{i-1}|}\right|. \tag{8}$$

A MASE lower than 1 indicates that the model performs better than the naive benchmark model [46].

Additionally, the losses of the developed ANNs considered minimization of mean squared error (MSE). The MSE calculates the average squared difference between the predicted and real values, as in:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} e_i^2\,. \tag{9}$$

3.2.5 Generation of a new synthetic test set

The performance of the above-mentioned models was evaluated on the test set composed of 30% of data for RFR and KNNReg models and 20% of data for the ANN models. However, considering the complexity of the data, we aimed to create a new test set composed of synthetic data generated using the SMOGN algorithm. SMOGN is developed with the aim to balance imbalanced regression datasets [47]. It works by combining under-sampling with two over-sampling techniques—SmoteR and the introduction of Gaussian noise [47]. By combining these techniques, SMOGN creates a diverse set of examples taking into account the real-world variable distribution.

The algorithm starts by processing a predefined rarity threshold based on which it draws partitions of data that are clustered in two groups—one containing samples with a relevance score below the threshold (i.e., the less relevant samples) and the other containing partitions with higher relevance samples [47]. Random under-sampling is then applied to samples with lower relevance score; while depending on the distance between the case in a partition and the k-nearest neighbor, the higher relevance samples are processed using SmoteR or the introduction of Gaussian noise [47]. In this study, the rarity threshold was set to 0.5, the over-sampling interpolation was performed based on 7 neighbors, while the amount of perturbation was set to 0.05 indicating the amount of noise that will be introduced to synthetic data. The minority values are automatically determined by the box plot extremes. Using SMOGN, we generated a set of 1461 new cases from which we randomly

resampled 1000 cases that are used for testing the best performed algorithm. In Fig. 1, the distribution of the actual and the SMOGN modified values of the EF is represented.

3.3 Model configuration

The Scikit library [48] in Python was used for developing the KNNReg and RFR models, while Keras library [49] was used to develop the ANN models. The BO-TPE was performed using the Hyperopt library [50]. After standardizing the data based on Eq. 1, the dataset was split into 70% for training and 30% of the instances for testing the KNNReg and RFR, while the ANN models used 70% of instances for training, 20% for testing, and 10% for validation. The data were then used to estimate the optimal parameters for each of the models using the TPE. The best configuration derived by the BO-TPE is presented in Table 2. For the KNNReg algorithm, the number of neighbors was set to 2, as suggested by the optimization algorithm. The nearest neighbors were computed using the brute algorithm, with a leaf size of 69. The chosen power parameter p for the Minkowski metric is 1, thus making it equivalent to using the Manhattan distance.

The developed RFR has 93 trees in the forest, while the minimum number of samples that is needed to split the nodes is 2. The TPE also suggested that the maximum depth of a tree should be 33, while the number of features that should be considered when searching for the best split is the sqrt of the total number of features.

Two sequential ANN models were built: one using the ReLU activation function in the hidden layers and the other using the SPOCU activation function. A sequential model presents a linear stack of layers [49], in this case fully connected dense layers. The ReLU ANN was built using a batch size of 256 samples, as suggested by the TPE. The model includes one input layer with 9 neurons, two hidden layers with 120 and 136 neurons, respectively, and an output layer with 1 neuron. The last hidden layer is followed by a dropout layer with a dropout rate of 0.3. A normal weight initialization was used, where the weights represent small Gaussian random values [49]. Moreover, in order to fit the neural network, a rectified linear unit (ReLU) activation function was used for the input and hidden layers, while the linear activation function was chosen for the output layer. ReLU is defined as [36]:

$$f(x) = \max(x, O). \tag{10}$$

In order to ensure that the developed model is not overfitting, a regularization technique was employed. In these terms, the early stopping regularization method was applied on the validation set with patience parameter set to 10. Patience represents the number of epochs the algorithm will wait before early stopping if there is no progress in the error on the validation set during the training [51].

The weights are updated to the network by using an optimizer, and here the adaptive moment estimation (Adam) optimizer was applied. Adam is a stochastic gradient descent optimization algorithm that calculates the adaptive learning rates for every parameter [51]. This method has many advantages, such as the ease of use, efficiency, and small memory requirements [52]. The objective function (i.e., the loss function) used for the evaluation of the model is the mean squared error (MSE). The model was trained for 300 epochs.

The same configuration was used with the SPOCU activation function in the hidden layers, with the following exceptions: The batch size, as suggested by the parameter optimization algorithm, was set to 32 samples, and the number of neurons in the hidden layers is 14 and 168, respectively. SPOCU is given by [30]:

$$s(x) = \alpha h\left(\frac{x}{\gamma} + \beta\right) - \alpha h(\beta) \tag{11}$$

where $\beta \in (0, 1); \alpha, \gamma > 0$ and $h(x)$:

$$\begin{cases} r(c), & x \geq (c) \\ r(c), & x \in [0, c) \\ 0, & x < 0 \end{cases}$$

Furthermore, the $r(x)$ is expressed as [30]:

$$r(x) = x^3(x^5 - 2x^4 + 2). \tag{12}$$

where $1 \leq c < \infty$.

Lastly, in order to build the GUI, the *tkinter* library in Python was used.

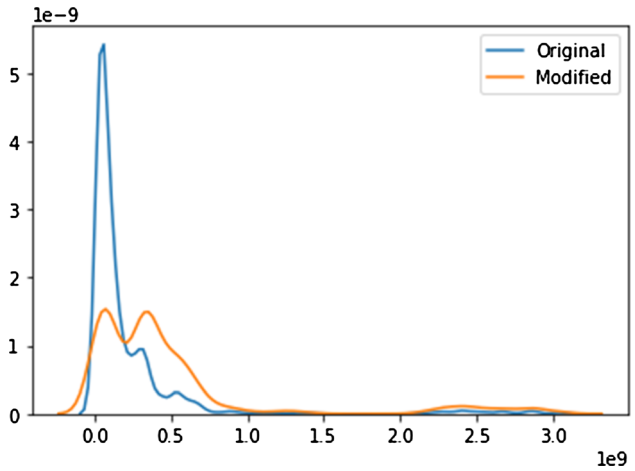

Fig. 1 The distribution of SMOGN generated and real-world EF value

4 Results and discussion

The first step of the modeling included the correlation analysis in order to observe the relationship between the variables. Even though correlation does not indicate causation, and causation is not the focus of this study, it is useful to estimate the potential relationships between the parameters, especially between the independent and the dependent variables. The obtained values of the Spearman's correlation are all positive, and the strongest correlations are observed between the fossil fuel energy sources and the dependent variable—the total EF of consumption, while renewable sources do not correlate with the total EF that strong. The correlation heating map is presented in Fig. 2.

After the initial analysis, the data were modeled using ML algorithms and the results of the modelings are presented in Table 3, with the best obtained values bolded. In terms of all of the observed performance measures, the KNNReg performed the best with the lowest obtained errors and lowest computation time. The MASE and NRMSE values of the KNNReg model are 0.029 and 0.006, respectively, while MAPE and SMAPE obtained values of 5.136 and 5.214, respectively. The RFR obtained MASE of 0.032, NRMSE of 0.007, and MAPE and SMAPE of 5.688 and 5.52, respectively. The ANN model with ReLU activation function obtained MASE of 0.064, NRMSE of 0.015, MAPE of 13.79, and SMAPE of 13.43. Lastly, the ANN with SPOCU activation function obtained MASE of 0.089, NRMSE of 0.011, MAPE of 22.454, and SMAPE of 18.31. The KNNReg took less time to train and validate, but it should be noted that the computation time highly depends on the used libraries and hardware; thus, it cannot be used as performance measure, though it can be observed as a measure of computational cost. The model with the best obtained performance, in this case the KNNReg, was used to test the SMOGN generated data. The performance on the synthetic dataset is good with low values of MASE (0.022), MAPE (3.243), and SMAPE (3.318), confirming the effectiveness of the proposed approach. These predictions are shown in Fig. 6.

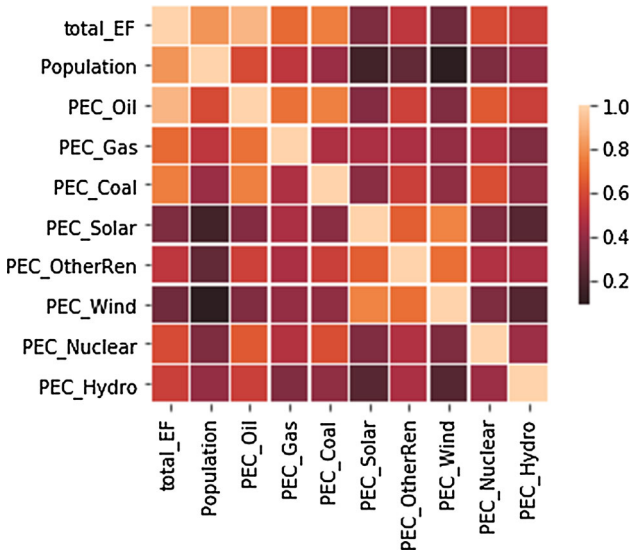

Fig. 2 The correlation heating map of the dataset

The RF algorithm calculates the importance of each feature in the model as the normalized total reduction in Gini importance. Feature importance is shown in Fig. 3. The highest importance is observed for the fossil fuels, i.e., primary energy production from coal, oil, and gas sources. These three variables account for more than 70% of importance in the model. This is no surprise as EF represents a measure of sustainability, and fossil fuels negatively impact the environment. Interestingly, population holds less than 10% of importance, suggesting that it is not a main driver of the EF prediction in the RFR model.

Additionally, the developed ANNs were compared in terms of the obtained loss (Table 4). The ANN with ReLU activation function obtained higher values of the training loss and lower values of the validation loss than the ANN with SPOCU activation function in the hidden layers. Both models took similar time to converge—45 epochs versus 49 epochs for ReLU and SPOCU, respectively.

Table 3 Results of the KNNReg, ANN, and RFR algorithms

	KNNReg	RFR	ANN ReLU	ANN SPOCU	KNN SMOGN
MASE	**0.029**	0.032	0.064	0.089	**0.022**
NRMSE	**0.006**	0.007	0.015	0.011	0.009
MAPE	**5.136**	5.688	13.794	22.454	**3.243**
SMAPE	**5.214**	5.520	13.428	18.311	**3.318**
Training time (s)	0.129	0.319	1.743	5.743	N/A*
Validation time (s)	1.878	3.222	7.767	18.009	N/A*
TPE time (s)	7.314	03.47 min	01.46 min	03.38 min	N/A*

* *Note* The KNN SMOGN was tested on the KNNReg model, so there were no training, validation, and optimization steps involved

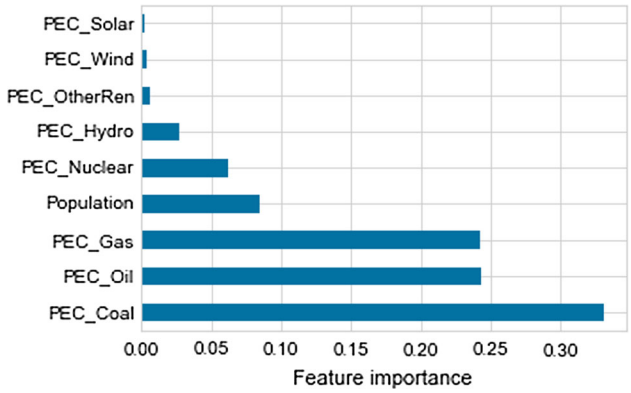

Fig. 3 RFR feature importance plot

Table 4 Losses in the last epoch

Model	Epochs	Training loss	Validation loss
ANN ReLU	45	0.0127	0.0035
ANN SPOCU	49	0.0114	0.0079

To be able to better understand the prediction accuracy, the actual and predicted values were plotted and are shown in Fig. 4. The plots confirm the results in Table 3, with the most accurate prediction obtained by the KNNReg, followed by the RFR and the ANNs.

Additionally, Fig. 5 presents the plot of losses through epochs of SPOCU and ReLU ANNs during training and validation. The ANN with the SPOCU function has a smooth trend, with continuous nature of errors. The gap between the training and validation loss is small, with training losses being slightly higher than the validation losses which indicates that the model did not overfit/underfit. The ANN with the ReLU function performs good on both sets, but the difference between training and validation losses is bigger and the loss is slightly volatile with a small decrease through epochs. Lastly, the predictions of the KNNReg on the synthetic dataset are presented in Fig. 6 confirming the results from Table 3.

The obtained results support some of the findings in the literature. In particular, the ANN prediction algorithm

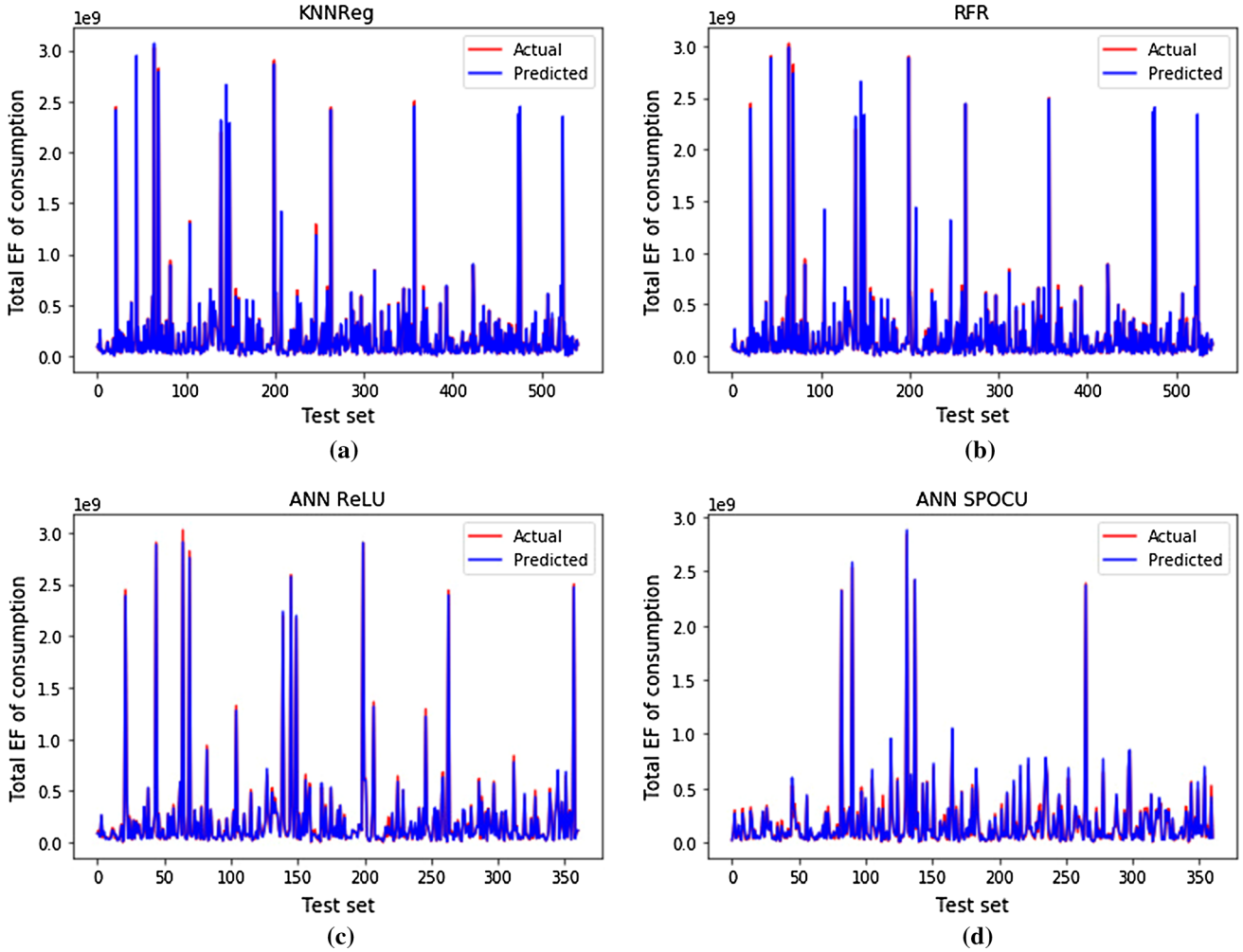

Fig. 4 Actual vs predicted values of the total EF of consumption for **a** KNNReg, **b** RFR, **c** ANN ReLU, and **d** ANN SPOCU

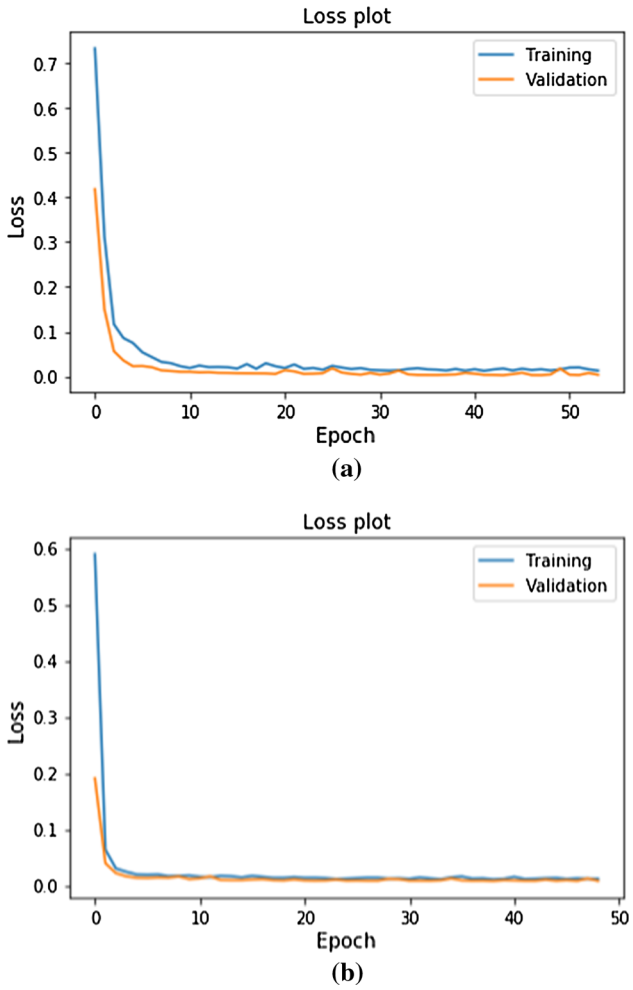

Fig. 5 The loss plots of **a** ANN with ReLU, **b** ANN with SPOCU

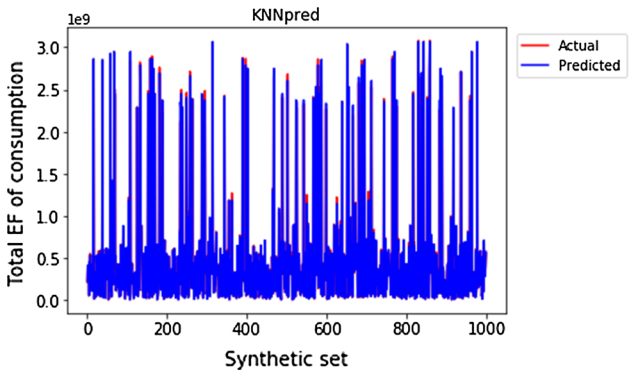

Fig. 6 The predictions of the KNNReg on the synthetic dataset

usually performs good in comparison with other machine learning techniques [53], but in comparison with random forest, its performance results are slightly weaker [54, 55]. In this study, random forest performed better than the ANN, in terms of all observed performance measures. Moreover, the total EF is closely affected by the population and energy consumption [57], particularly from fossil fuels [58], which was demonstrated by the feature importance analysis of RFR. Furthermore, the performed correlation analysis (Fig. 2) indicates a strong correlation between the total EF and the primary energy consumption, showing that the total EF and the energy consumption increase together. The comparison of the results in the present study and the literature in terms of the EF prediction using ML algorithms is presented in Table 5. As most studies focus on studying the environmental Kuznets curve, we compared our results only to the studies that aimed to predict the EF. Our approach obtained the highest MAPE compared to [16, 29, 56], but it is the only study that utilized the BO-TPE contributing to the knowledge on the impact of such methods for the EF prediction based on energy and population parameters.

The developed GUI shows good performance in calculating the total EF of consumption based on user inputs; hence, it can be deployed into the production systems in order to automate the forecasting system. The GUI is presented in Fig. 7.

As observed from the literature, different combinations of inputs have been used for the EF modeling and prediction, mostly based on socioeconomic, industrial, and energy parameters [14, 17–20, 57, 59] such as the population, GDP, trade, agricultural consumption, and retail sales. This research considers only the population and energy parameters for several reasons: (i) Population highly affects the EF of consumption, as with the increase in population number, the needs of the population and, consequently, the consumption are also increasing; (ii) energy consumption, as already observed, is mostly produced from fossil fuels which worsen the environmental quality and increase the EF; hence, these two sets of parameters are inseparable for understanding the dynamic processes of environmental degradation. The results of this research have several implications: (i) The developed models should enable an efficient resource allocation, (ii) the developed models can be applied in various regions and sectors, (iii) the developed models can be deployed into production, and (iv) based on the obtained performance, the KNNReg model was used to build the GUI in order to achieve a better forecasting experience and a more efficient human–computer interaction.

5 Conclusions

The EF shows the human appropriation of the natural resources. In this sense, the human demand for natural resources is constantly increasing, while the biocapacity of the planet decreases. Large amounts of resources are consumed rapidly in order to support different human

Table 5 The comparison of results of the present paper with the literature in the prediction of the EF

References	Output	Model	Optimization	Best model	Evaluation (MAPE)
[16]	EF	ARIMA, ARIMA-BPNN	Trials	ARIMA-BPNN	0.459
[29]	EF	ARIMA, ARIMA-BPNN	Trials	ARIMA-BPNN	1.65
[56]	EF, EFD, EFDC	Grey relational analysis, ANN	N/A	6-4-3 ANN with standardization	1.88
Our approach	EF	KNNReg, RFR, ANN ReLU, ANN SPOCU	BO-TPE	KNNReg	5.136

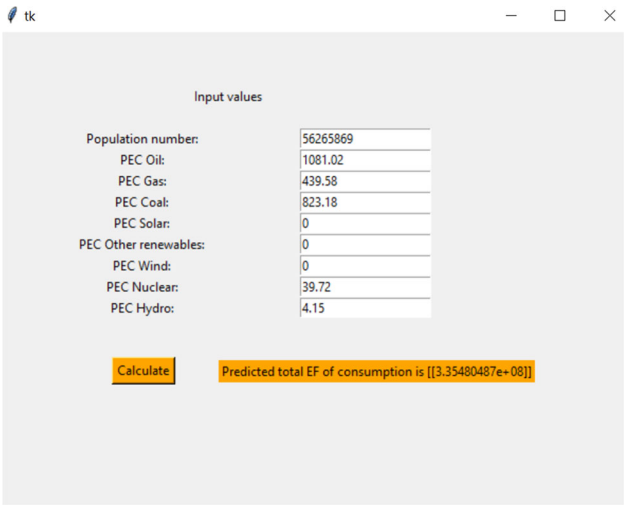

Fig. 7 The developed GUI. The values entered are from the synthetic dataset. The obtained prediction error is 2.95%

activities, while simultaneously the Earth cannot regenerate its resources at that speed.

While coal sources today dominate the energy production, a carefully planned and analyzed use of renewable sources is remarkably important. The future of the industry and technology is highly dependent on the environment and vice versa. In the upcoming industrial revolution, a global decentralization of the industrial processes is expected. Moreover, the Industry 4.0 will require a lot of energy in order to function, because it will integrate new systems and solutions which will be mostly cloud and IT based. These newly utilized devices will require a lot of power to function, hence increasing the energy use. In order to prevent further consequences to the environment, the natural resources should be effectively and efficiently allocated, and in order to allocate them, there is a strong need for accurate prediction models that will imply the rationality of resource consumption.

This paper proposed an approach for predicting the total EF of consumption based on the population and the primary energy consumption from different sources. As environmental and energy data are naturally complex and hard to predict, it is important to carefully select the best set of model hyper-parameters in order to obtain good performance. Therefore, this paper proposed four hybrid machine learning models based on Bayesian parameter estimation: (i) KNNReg, (ii) RFR, (iii) ANN ReLU, and (iv) ANN SPOCU. The model performance was compared using several measures, including MASE, MAPE, SMAPE, and NRMSE. The best model is KNNReg with MASE of 0.029 and MAPE of 5.136%. The KNNReg was additionally tested on a synthetic dataset generated by the SMOGN method. The results demonstrated good performance with MASE of 0.022 and MAPE of 3.243%. Lastly, the KNNReg was selected to develop the GUI for an automated EF calculation based on user input.

The limitations of this study include the following. As the TPE takes into account the specified conditional relationships, it should be noted that the inter-dependencies of other variables may not be captured. Furthermore, we performed correlation analysis which implies there is a relationship between energy, population, and the EF, but this analysis does not address causality; hence, we can only assume the relationship between the aforementioned variables based on theoretical research and experimentation found in the literature. Additionally, we used an over-/under-sampling approach for regression problems—SMOGN—that generated a synthetic dataset based on the actual data, and we implemented the newly developed set to additionally test the KNNReg model. Even though the prediction performance is good, more experiments should be performed. The real-world data are in most cases skewed, and over-/under-sampling techniques model the data by focusing on the extremes above/below the threshold. Hence, more attention needs to be put on synthetic data modeling, as environmental data are often limited, thus limiting the number of approaches that can be used to tackle the problem.

Future work will include modeling the EF while considering the time variable, with a focus on individual countries. Also, the causation relationship will be investigated.

Acknowledgements This work was supported by the Serbian Ministry of Education, Science and Technological Development through Mathematical Institute of the Serbian Academy of Sciences and Arts and through the project of the Ministry of Education, Science and Technological Development of Serbia—TR34023.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. British Petroleum Company (2018) BP statistical review of world energy. https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2018-full-report.pdf. Accessed 10 June 2019
2. United Nations (2015) Transforming our world: the 2030 agenda for sustainable development. https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf Accessed 28 October 2018
3. European Commission (2016) Paris agreement. https://ec.europa.eu/clima/policies/international/negotiations. Accessed 20 June 2019
4. Rees WE (1992) Ecological footprints and appropriated carrying capacity: what urban economics leaves out. Environ Urban. 4(2):121–130
5. Wackernagel M (1993) How big is our ecological footprint?. A handbook for estimating a community's appropriated carrying capacity, Verlag nicht ermittelbar
6. Wackernagel M, Onisto L, Bello P, Linares AC, Falfan ISL, Garcia JM, Guerrero AIS, Guerrero M (1999) National natural capital accounting with the ecological footprint concept. Ecol Econ 29(3):375–390
7. Borucke M, Moore D, Cranston G, Gracey K, Iha K, Larson J, Lazarus E, Morales JC, Wackernagel M (2013) Accounting for demand and supply of the biosphere's regenerative capacity: the national footprint accounts' underlying methodology and framework. Ecol Ind 24:518–533
8. Medved S (2006) Present and future ecological footprint of Slovenia —the influence of energy demand scenarios. Ecol Model 192(1–2):25–36
9. Olivier JG, Schure KM, Peters JAHW (2017) Trends in global CO2 and total greenhouse gas emissions. PBL Netherland Environmental Assessment 5
10. International Energy Agency (2018) Global energy and CO2 status report. https://www.iea.org/geco. Accessed 06 February 2019
11. Stehlık M, Dušek J, Kiseľák J (2016) Missing chaos in global climate change data interpreting? Ecol Complex 25:53–59
12. Liu Y, Tian Y, Chen M (2017) Research on the prediction of carbon emission based on the chaos theory and neural network. Int J Bioautom. 21(4):339
13. Sabolová R, Sečkárová V, Dušek J, Stehlík M (2015) Entropy based statistical inference for methane emissions released from wetland. Chemometr Intell Lab Syst 141:125–133
14. Lu Y, Bin C (2017) Urban ecological footprint prediction based on the Markov chain. J Clean Prod 163:146–153
15. Liu L, Lei Y (2018) An accurate ecological footprint analysis and prediction for Beijing based on SVM model. Ecol Inform 44:33–42
16. Wang Z, Yang L, Yin J, Zhang B (2018) Assessment and prediction of environmental sustainability in China based on a modified ecological footprint model. Resour Conserv Recycl 132:301–313
17. Bello MO, Solarin SA, Yen YY (2018) The impact of electricity consumption on CO2 emission, carbon footprint, water footprint and ecological footprint: the role of hydropower in an emerging economy. J Environ Manag 219:218–230
18. Destek MA, Sarkodie SA (2019) Investigation of environmental Kuznets curve for ecological footprint: the role of energy and financial development. Sci Total Environ 650:2483–2489
19. Lu F, Xu JH, Wang ZY, Hu XF (2010) Qualitatively analysis on influence factors of ecological footprint and dynamic prediction of ecological footprint: a case study in Xinjiang. Geogr Geo-Inf Sci 26(6):70–74
20. Yao H (2012) Simulating the total ecological footprint of Suzhou from 1990 to 2009 by BPANN. Pol J Environ Stud 21(6):1901–1910
21. Huimin L (2013) The impact of human behavior on ecological threshold: positive or negative? grey relational analysis of ecological footprint, energy consumption and environmental protection. Energy Policy 56:711–719
22. Pławiak P, Abdar M, Pławiak J, Makarenkov V, Acharya UR (2020) DGHNL: a new deep genetic hierarchical network of learners for prediction of credit scoring. Inf Sci 516:401–418
23. Pławiak P, Abdar M, Acharya UR (2019) Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring. Appl Soft Comput 84:105740
24. Tuncer T, Ertam F, Dogan S, Aydemir E, Pławiak P (2020) Ensemble residual network-based gender and activity recognition method with signals. J Supercomput 76(3):2119–2138
25. Tuncer T, Dogan S, Abdar M, Ehsan Basiri M, Pławiak P (2019) Face recognition with triangular fuzzy set-based local cross patterns in wavelet domain. Symmetry 11(6):787
26. Abdar M, Wijayaningrum VN, Hussain S, Alizadehsani R, Plawiak P, Acharya UR, Makarenkov V (2019) IAPSO-AIRS: a novel improved machine learning-based system for wart disease treatment. J Med Syst 43(7):220
27. Kandala RN, Dhuli R, Pławiak P, Naik GR, Moeinzadeh H, Gargiulo GD, Gunnam S (2019) Towards real-time heartbeat classification: evaluation of nonlinear morphological features and voting method. Sensors 19(23):5079
28. Hammad M, Pławiak P, Wang K, Acharya UR (2020) ResNet–attention model for human authentication using ECG signals. Expert Systems e12547
29. Wu M, Wei Y, Lam PT, Liu F, Li Y (2019) Is urban development ecologically sustainable? ecological footprint analysis and prediction based on a modified artificial neural network model: A case study of Tianjin in China. J Clean Prod 237:117795
30. Kiseľák J, Lu Y, Švihra J, Szépe P, Stehlík M (2020) "SPOCU": scaled polynomial constant unit activation function. Neural Computing and Applications
31. Ritchie H, Roser M (2019) Energy production and changing energy sources. https://ourworldindata.org/energy-production-and-changing-energy-sources. Accessed 12 February 2019
32. Global Footprint Network (2018) The national footprint accounts, 2018 edition. https://www.footprintnetwork.org. Accessed 12 May 2019
33. Data World (2018) NFA 2018 editon. https://data.world/footprint/nfa-2018-edition. Accessed June 2019
34. World Bank (2019) https://data.worldbank.org/indicator/SP.POP.TOTL. Accessed 09 February 2019
35. Dancey CP, Reidy J (2007) Statistics without maths for psychology. Pearson education, London
36. Ramachandran P, Zoph B, Le QV (2017) Searching for activation functions. arXiv preprint, arXiv:1710.05941
37. Farahnakian F, Pahikkala T, Liljeberg P, Plosila J (2013) Energy aware consolidation algorithm based on k-nearest neighbor regression for cloud data centers. In: IEEE/ACM 6th international conference on utility and cloud computing
38. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
39. Liaw A, Wiener M (2002) Classification and regression by randomForest. R news 2(3):18–22

40. Ahmad MW, Reynolds J, Rezgui Y (2018) Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. J Clean Prod 203:810–821
41. Bandyopadhyay G, Chattopadhyay S (2007) Single hidden layer artificial neural network models versus multiple linear regression model in forecasting the time series of total ozone. Int J Environ Sci Technol 4(1):141–149
42. Zhang G, Patuwo BE, Hu MY (1998) Forecasting with artificial neural networks: the state of the art. Int J Forecast 14(1):35–62
43. Bergstra JS, Bardenet R, Bengio Y, Kégl B (2011) Algorithms for hyper-parameter optimization. Adv Neural Inf Process Syst 24:2546–2554
44. James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. Springer, New York
45. Lewis CD (1982) Industrial and business forecasting methods. Butterworths, London
46. Hyndman Rob J, Koehler Anne B (2006) Another look at measures of forecast accuracy. Int J Forecast 22(4):679–688
47. Branco P, Torgo L, Ribeiro RP (2017) SMOGN: a pre-processing approach for imbalanced regression. Theory and Applications, First International Workshop on Learning with Imbalanced Domains
48. Pedregosa F, Varoquaux G, Gramfort A, Michael V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830
49. Chollet F (2015) Keras documentation. https://keras.io. Accessed 12 February 2019
50. Bergstra J, Yamins D, Cox D (2013) Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. In: International conference on machine learning, pp. 115–123
51. Ruder S (2016) An overview of gradient descent optimization algorithms. ArXiv preprint, arXiv:1609.04747
52. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. Available from: arXiv preprint, arXiv:1412.6980
53. Ahmad MW, Mourshed M, Rezgui Y (2017) Trees vs Neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy Build 147:77–89
54. Dong LJ, Li XB, Kang PENG (2013) Prediction of rockburst classification using random forest. Trans Nonferrous Metals Soc China 23(2):472–477
55. Benali L, Notton G, Fouilloy A, Voyant C, Dizene R (2019) Solar radiation forecasting using artificial neural network and random forest methods: application to normal beam, horizontal diffuse and global components. Renew Energy 132:871–884
56. Wang Q, Wang X, Mao Y (2009) Dynamic assessment of sustainable development based on grey relational analysis and artificial neural network. In: 2009 IEEE International conference on grey systems and intelligent services (GSIS 2009) (pp. 212-217). IEEE
57. Luo J, Zhou L, Li X (2008) Construction and analysis of ecological footprint dynamic prediction model—-a case study of Wuhan. Resources and Environment in the Yangtze Basin 3
58. Wackernagel M, Monfreda C (2004) Ecological footprints and energy. Encycl Energy 2(1):1–11
59. Wu K, Wang L (2006) Partial least square regression model of ecological footprint and its influencing factors. Resour Sci 28(6):182–188