Research Article

Learning framework for carbon emissions predictions incorporating a RReliefF driven features selection and an iterative neural network architecture improvement

Antonio Marcio Ferreira Crespo¹ · Chun Wang¹ · Thiago Marques Ferreira Crespo² · Li Weigang³ · Alexandre Barreto⁴

Received: 4 November 2020 / Accepted: 22 February 2021 / Published online: 16 March 2021

© The Author(s) 2021

Abstract

Inaccurate carbon emissions predictions may be one of the root factors leading to the overall ineffectiveness of the European Union environmental regulatory framework. Therefore, the present article aims at introducing a novel computational learning framework for carbon emissions prediction incorporating a RReliefF driven features selection and an iterative neural network architecture improvement. Our learning framework algorithmic architecture iteratively chains the features selection process and the backpropagation artificial neural network architecture design based on the data assessment accomplished by the RReliefF algorithm. Thus a better features set / neural network architecture combination is obtained for each specific prediction target. The implemented framework was trained and tested with real world data obtained from the European Union, International Energy Agency, Organisation for Economic Co-operation and Development, and World Bank, for the period 1990–2017. The framework evaluation against current mainstream machine learning models, and its benchmarking comparing to recent published researches on carbon emissions prediction indicates that our research contribution is relevant and capable of supporting the improvement of environmental policies.

✉ Antonio Marcio Ferreira Crespo, antonio.crespo@ieee.org | ¹Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Canada. ²Technology Development and Implementation, DreamerGate Research and Development Inc, Montreal, Canada. ³Computer Science Department, University of Brasilia, Brasilia, Brazil. ⁴C4I and Cyber Center, George Mason University, Fairfax, USA.

Graphic abstract

Keywords CO₂ Emissions prediction · Backpropagation neural network · RReliefF · Features selection · Explainable AI

Abbreviations

ANN	Artificial Neural Network	MtCO ₂	Millions of tons of CO ₂
BP	Backpropagation	MtCO ₂ -eq	Millions of tons of CO ₂ equivalent
CDD	Cooling Degree Days	NN	Neural Network
EEA	European Environment Agency	NN/BP-BFE	NN/BP combined with backward feature elimination (BFE)
EED	Energy Efficiency Directive	NN/BP Control	Direct application of Backpropagation Neural Network (NN/BP)
EU	European Union	NN/BP-CT	NN/BP combined with Chaos Theory (CT).
EU28	European Union—28 States	NN/BP-IPSO	NN/BP combined with Improved Particle Swarm Optimization (IPSO).
EU-ETS	European Union Emissions Trading System	NN/BP-PCA	NN/BP combined with Principal Component Analysis (PCA).
FEM	Features Engineering Module	NN/BP-RF	NN/BP combined with Random Forest (RF).
FGMVM	Grey Multivariable Verhulst model improved by Fourier series.	NN/BP-RReliefF/BFE	NN/BP combined with RReliefF driven BFE
GBM-BFE	Gradient Boosting Machine (GBM) combined with backward feature elimination (BFE).	OECD	Organisation for Economic Co-operation and Development
GHG	Greenhouse Gases	PEM	Prediction Explanation Module
GM	Grey Model.	RED	Renewable Energy Directive
HDD	Heating Degree Days		
IEA	International Energy Agency		
MAPE	Mean Absolute Percentage Error		
MEM	Model Evaluation Module		
MGM	Model Generation Module		
ML	Machine Learning		

RF-BFE	Random Forest (RF) combined with backward feature elimination (BFE).
RMSE	Root Mean Square Error
SVM-BFE	Support Vector Machine (SVM) combined with backward feature elimination (BFE).
SVM-RF	Support Vector Machine (SVM) combined with Random Forest (RF).
XAI	Explainable Artificial Intelligence

1 Introduction

After the Kyoto Protocol signature in 1997, governments, industry stakeholders and academia began to work on the development of effective and efficient environmentally driven policies and economic mechanisms. In such context, the European Union (EU) current efforts in support to sustainable development and climate change avoidance comprises three main challenges, i.e. greenhouse gases (GHG) emissions reduction, consistent increment on energy production from renewables (Renewable Energy Directive—RED), and increase in energy efficiency (Energy Efficiency Directive—EED).

Within the EU environment protection framework, the European Union Emissions Trading System (EU-ETS) was launched in 2005, and it currently covers approximately 45% of EU28 polluting emissions. The EU-ETS implementation observed a staggered approach, and 2020 is the last year of the third phase, as presented below.

- EU-ETS phase 1 (2005–2007) - the absence of reliable data on actual emissions and consequent wrong estimations led to allowances surplus (in million metric tonnes of CO₂ equivalents - MtCO₂-eq):
 - Allocated allowances: 6370 MtCO₂-eq;
 - CO₂ emissions: 6215 MtCO₂.
- EU-ETS phase 2 (2008–2012):
 - Allocated allowances: 11373 MtCO₂-eq;
 - Carbon emissions: 9613 MtCO₂.
- EU-ETS phase 3 (2013–2020):
 - Allowances surplus in 2017: 1.6 billion.

Moreover, in 2019 the EU had to artificially remove 700 million allowances from circulation. Hence, the observed

evolution of the EU trading system indicates it sustained and substantial structural supply–demand imbalance that kept distorting the market and compromising the scheme effectiveness as an emissions reduction driver [1].

Moreover, notwithstanding the systematic and continued EU efforts and policies addressing the reduction of carbon emissions, the EU member States' projections converge to an EU-wide total GHG emissions reduction of at most 32%, which falls short of the 40% target for 2030. EU-ETS specific projections indicate that the stationary installations could reach a 10% reduction on emissions between 2020 and 2030, which is insufficient for the accomplishment of the 2030 reduction target of 43% compared to 2005 levels [2].

Thus, inaccurate carbon emissions predictions may be one of the root factors leading to the overall ineffectiveness of the EU environmental regulatory framework. Therefore the achievement of the European Union ambitious targets will require additional policies resulting from new holistic and creative approaches targeting carbon emissions trends prediction.

In such scenario, and considering the findings related to the EU-ETS experience, our contribution explored the following research opportunities: (a) market based climate change avoidance policies efficiency could be relevantly improved by a better accuracy on carbon emissions trends prediction; (b) there is a crucial need for more accurate carbon emissions predictions better supporting each climate change avoidance initiative (i.e. EU-ETS, EED, RED) by considering the particularities of industry / economy sectors under their coverage; (c) machine learning (ML) methods and techniques have the potential capacity to grasp such particularities from economic and energy consumption indicators, what might lead to more realistic carbon emissions forecasts.

Therefore, the present article introduces a computational learning framework for carbon emissions predictions incorporating a sophisticated and effective algorithm for the assessment and selection of the most relevant environmental impacting factors. Such algorithm was innovatively engineered with an artificial neural network (ANN), for the conformation of a forecast framework able to improve its architecture according to the selected impacting factors (predicting features). The framework evaluation against current mainstream machine learning models, and its benchmarking to recent published researches on carbon emissions prediction indicates that our research contribution is relevant and capable of supporting the improvement of environmental policies effectiveness.

Table 1 Methodologies, techniques and predicting variables for carbon emissions prediction

AUTHOR	PROBLEM SCOPE	PROBLEM ANALYSIS METHODOLOGY	TECHNIQUE
Chang [22]	Carbon emissions prediction (predictors: crude oil consumption, coal consumption, natural gas consumption)	Statistical Analysis	Multivariate co-integration Granger causality / Vector Error Correction Modeling
Hong et al. [23]	Carbon allowance price prediction (predictors: oil, brent, natural gas, coal)	Statistical Learning	Predictive Regression Bagged (bootstrap aggregation) / Decision Tree
Liu et al. [24]	Carbon emissions predictions	Computational Learning Theory / Statistical Learning Theory / Chaos Theory	Neural Networks
Zhou et al. [25]	Carbon emissions prediction (predictors: thermal power capacity, thermal power generation, urbanization rate, GDP)	Computational Learning Theory / Statistical Learning Theory	Neural Networks / Particle Swarm Optimization (PSO)
Sun et al. [26]	Carbon emissions prediction (predictors: coal consumption, primary industry GDP, secondary industry GDP, tertiary industry GDP, population, urbanization level, transport, power generation, steel production, total investments, final consumption)	Computational Learning Theory / Statistical Learning Theory / Grey Systems Theory	Neural Networks / Principal Component Analysis (PCA) / Grey Model
Zhou et al. [27]	Carbon emissions prediction (predictors: GDP, urbanization rate, electricity consumption, coal consumption, thermal power capacity)	Statistical Learning / Grey Systems Theory	Support Vector Machine (SVM) / Particle Swarm Optimization / Grey Relation Analysis
Li et al. [28]	Carbon emissions prediction (predictors: energy consumption, i.e. coal, oil, and natural gas)	Computational Learning Theory / Statistical Learning	Neural Networks
Support Vector Machine (SVM)			
Song et al. [29]	Carbon allowance price prediction (predictors: demand-related policies)	Stochastic Processes	Fuzzy Stochastic Differential Model
Sun et al. [30]	Carbon emissions prediction (predictors: coal consumption, primary sector GDP, secondary industry GDP, tertiary industry GDP, final consumption, population, power generation, exports, urbanization level (%), investments, transportation, fuel and power purchase price index, cement production, urban green areas, total retail sales of consumer goods, finished steel production)	Computational Learning Theory / Statistical Learning	Support Vector Machine / Particle Swarm Optimization / Principal Component Analysis / Neural Networks
Wen et al. [31]	Carbon emissions prediction: total population, household consumption level, urbanization level, GDP per capita, primary industry GDP, secondary industry GDP, tertiary industry GDP, commercial department GDP, total energy consumption, and total retail sales of social consumer goods	Computational Learning Theory / Statistical Learning Theory / Evolutionary Computation	Neural Networks / Particle Swarm Optimization / Random Forest / Support Vector Machine / Genetic Algorithm

2 Relevant related researches

A myriad of researches analyzed carbon emissions behavior, and attempted to predict it by means of several different approaches, methods and techniques, and considering different emissions impacting factors (predictors). Econometric approaches were used by Guan et al. [3], Anger [4], Li and Lu [5], Robalino-Lopez et al. [6], Scott et al. [7], Mi et al. [8]. And Game Theory emerged as one of the preferred approaches to address decision-making and supply chain challenges linked to carbon emissions and carbon policies, as in Chang and Chang [9], Yang et al. [10], Yang et al. [11], and Xu et al. [12].

Whereas General Equilibrium Theory (Wang and Wang [13], Gavard et al. [14], Zhang et al. [15]), Operational Research (Cui et al. [16], Hong et al. [17]), Index Number Theory (Wang et al. [18], Solaymani et al. [19]), Variational Inequality Theory (Allewi et al. [20]), and Grey Systems Theory (Jiang et al. [21]) also played an important role in support of such studies, a significant number of researches opted for the application of techniques within the Statistical and Computational Learning domain, as presented in Table 1.

The literature review provided fundamental insights on the existing carbon emissions impacting factors and how to apply them in emissions prediction models. It was noted that the researches benefiting from Computational Learning and Statistical Learning theories were the ones providing more information regarding how the predictors (or the availability / choice of different predictors) impacts prediction confidence level and accuracy.

The literature review also allowed us to identify some very important challenges related to carbon emissions prediction, when considering the amount and diversity of potential predictors. Firstly, a particular predictor correlation to a specific target varies depending on the scenario / region. Secondly the systematic generation of trustworthy carbon emissions information started in the 1990 decade, and its availability is restricted to some parts of the world.

Thirdly, carbon emissions can be characterized as a worldwide multisectoral interconnected phenomenon, e.g. the pollution outcomes international flights may have contributing components spread in all continents if we consider the airline headquarters location, the flight route (origin-overfly area-destination), the aircraft manufacturer (engines manufacturer, fuselage manufacturer, tires manufacturer, etc.), the fuel producer (petroleum, biofuels).

As an additional example, consider, a huge transnational enterprise may move its heavily polluting production to regions where carbon policies are less strict or even inexistent (carbon leakage). In such scenario, carbon emissions prediction models should be scalable in order

to progressively cover broader scopes and process more data.

However, such required scalability would lead to the use of an increasing number predictors (predicting model features space dimension), what would not be accompanied by additional instances, once the availability of trustworthy data is limited. Thus, any intended predicting model should be capable of addressing data related problems such as endogeneity, heteroskedasticity, multicollinearity, and dimensionality.

The overall insights accrued from the literature review drove us to choose a computational learning / statistical Learning approach for the design of our prediction framework. It was also noted that, considering the nature of carbon emissions related data, it would be generally impossible to work with any parametric learning method. Additionally, fitting high-dimensional statistical models often requires the use of non-linear parameter estimation procedures [32]. Therefore, we opted for Neural Networks as the core learning model of our proposed predicting framework.

3 Methods and framework evaluation

Within our research we designed and implemented a computational learning framework for carbon emissions predictions incorporating a RReliefF driven features selection method, and an iterative neural network architecture improvement.

The Relief family of algorithms [33, 34] incorporates the ability to qualify attributes in a dataset as a function of the Euclidean distance computed between neighbouring instances. Such non-parametric and non-myopic algorithms are able to capture non-linear relationships, and run in low-order polynomial time.

The algorithms outcomes are attributes weights that feature a probabilistic interpretation, i.e. the weights are proportional to the difference between two conditional probabilities: the probability of the attribute's value being different conditioned on the given nearest hit, and on the nearest given miss [34].

Figure 1 graphically presents the designed framework, composed by four modules: the Features Engineering Module (FEM), the Model Generation Module (MGM), the Model Evaluation Module (MEM), and the Prediction Explanation Module (PEM).

3.1 Data analysis

Our research focused on the European Union—28 States (EU28), and our dataset comprises data obtained from the European Union (Eurostats), International Energy

Fig. 1 Computational learning framework modules

Table 2 Prediction targets—MtCO₂ / IEA

T1	Total CO ₂ emissions
T2	CO ₂ emissions from energy industries
T3	CO ₂ emissions from energy intensive industries
T4	CO ₂ emissions from commercial and public services
T5	CO ₂ emissions from transport
T6	CO ₂ emissions from residences
T7	CO ₂ emissions from aviation

Agency (IEA), Organisation for Economic Co-operation and Development (OECD), and World Bank (WB) databases, as described below.

- Sources: Eurostats, IEA, OECD, World Bank.
- Scope:
 - EU28;
 - 1990 – 2017;
 - Total CO₂ emissions / sectoral CO₂ emissions;

- 26 economic / energy indicators (candidate predictors).

• Data Aggregation Levels:

- Regional;
- Annual;
- Total Emissions / Energy Industries / Industries / Commerce – Public Services / Transport / Residential / Aviation.

Tables 2 and 3 introduce the prediction targets and potential predictors explored in our research.

The characteristics of our research data led us to the application of neural networks (NN) as the core learning model in our proposed network. In this section we provide a deeper analysis of our data, which corroborates our choice and provides additional insights on how to deal with the data issues mentioned in the aforementioned section.

Table 3 Candidate predictors

A1—GDP (constant million 2010 USD / World Bank)	A14—Final Consumption (expenditure of households current prices million Euro / Eurostat)
A2—GDP(current prices million Euro / Eurostat)	A15—Final Consumption (expenditure current prices million Euro / Eurostat)
A3—Population (Eurostat)	A16—Primary Energy Consumption (MTOE / IEA)
A4—Temperature (HDD / Eurostat)	A17—Final Energy Consumption (MTOE / IEA)
A5—Temperature (CDD / Eurostat)	A18—Energy Use (KGOE per capita / World Bank)
A6—Technology Innovation (GERD per GDP / OECD)	A19—Energy Intensity Level of Primary Energy (MJ / per GDP PPP 2011 USD / World Bank)
A7—Technology Innovation (GERD million PPP 2010 USD / OECD)	A20—Energy Intensity Level of Primary Energy (KGOE per GDP PPP 2011 USD / World Bank)
A8—Technology Innovation (GERD per GDP / World Bank)	A21—Energy Intensity of GDP (KGOE per thousand Euro Chain Linked Volumes 2010 / Eurostat)
A9—Technology Innovation (GERD million 2010 USD / World Bank)	A22—Coal Total Primary Energy Supply (KTOE / IEA)
A10—Technology Innovation (GERD per GDP / Eurostat)	A23—Oil Total Primary Energy Supply (KTOE / IEA)
A11—Technology Innovation (GDP per direct material input PPP 2010 USD per kg / OECD)	A24—Natural Gas Total Primary Energy Supply (KTOE / IEA)
A12—Technology Innovation (GDP per direct material input PPP current USD per kg / OECD)	A25—Biofuels and Waste Total Primary Energy Supply (KTOE / IEA)
A13—Final Consumption (gross national expenditure current USD / World Bank)	A26—Non Combustion Electricity Generation (KTOE / IEA)

Table 4 Pearson correlation test—target T1 / Total CO₂ (MtCO₂)

Order	Predictor	Coefficient	p-value	Order	Predictor	Coefficient	p-value
1	A18 (WB)	0.9668816	1.07E-09	14	A7 (OECD)	−0.8692267	1.23E-05
2	A6 (OECD)	−0.9513449	1.52E-08	15	A9 (WB)	−0.8683733	1.29E-05
3	A10 (EU)	−0.9453401	3.38E-08	16	A25 (IEA)	−0.8499821	3.05E-05
4	A23 (IEA)	0.9381434	7.88E-08	17	A3 (EU)	−0.8264887	7.92E-05
5	A16 (IEA)	0.9357116	1.03E-07	18	A24 (IEA)	0.7928254	2.49E-04
6	A8 (WB)	−0.9335975	1.28E-07	19	A15 (EU)	−0.7707846	4.74E-04
7	A22 (IEA)	0.9290128	2.02E-07	20	A14 (EU)	−0.7617238	6.06E-04
8	A17 (IEA)	0.9266759	2.51E-07	21	A2 (EU)	−0.7524735	7.70E-04
9	A11 (WB)	−0.9137785	7.54E-07	22	A1 (WB)	−0.6586972	5.52E-03
10	A12 (OECD)	−0.910324	9.84E-07	23	A13 (WB)	−0.5127864	4.22E-02
11	A21 (EU)	0.9017806	1.82E-06	24	A4 (EEA)	0.3722717	1.56E-01
12	A19 (WB)	0.8834384	5.73E-06	25	A5 (EEA)	−0.1650201	5.41E-01
13	A20 (WB)	0.8834384	5.73E-06	26	A26 (IEA)	−0.0055333	9.84E-01

Pearson's coefficient is a test that measures the statistical association between two continuous variables as a function of the covariance observed between them. It provides information about the magnitude of the association, or correlation, as well as the direction of the relationship.

The results of the Pearson correlation test are bound by some important assumptions regarding the tested data, i.e. the variables should be normally distributed, a feature linearity and homoskedasticity. The test outcomes are values ranging -1 to $+1$, where $+1$ indicates a perfect positive relationship, -1 indicates a perfect negative relationship, and a 0 indicates no relationship exists; strong correlations are indicated by values between ± 0.5 and ± 1 . The research

data was submitted to the test, and Table 4 presents the potential predictors listed in order of correlation strength.

The analysis of the Pearson's test outcomes flags down some important insights. The predictor A18 (Energy Use) shows the strongest correlation with the total CO₂ emissions, what is in accordance with the knowledge accrued from the literature review. However, as regards to the predictor A3 (population), the test result is counterintuitive, as it indicates a strong negative relationship between population and CO₂ emissions; the literature review also contradicts such negative correlation.

Figure 2 provides the visualization of the predictors A18 and A3, and corroborates the Pearson's test outcome. A

Fig. 2 Exploratory visualization for T1, A18, A3 and A4

deeper analysis of the test outcome raises another important flag, i.e. the predictor A4 (Temperature HDD) shows an irrelevant correlation with total CO₂ emissions, and considering the research scope (EU28), such outcome seems inconsistent with the real-world energy consumption dynamics.

The combined analysis of Table 4 and Fig. 2 indicates potential non-linearity between the variables, as well as eventual additional violations of the Pearson's test assumptions. Thus there is a need for a more sophisticated correlation analysis of the data, and as such, we submitted it to the Spearman correlation test and to the Hoeffding's D Statistic test.

Spearman correlation uses ranks instead of assumptions about the distributions of the two variables and, as such, it analyzes the association between variables according to ordinal measurement levels. Thus, the test does not assume that the variables are normally distributed, and it can be applied to the cases in which the Pearson's assumptions (continuous-level variables, heteroskedasticity, and normality) are not fulfilled.

Similarly to the Pearson's test, Spearman analysis outcomes are values between -1 and $+1$, and the test results are representative once the data can be ordinally arranged and the scores on one variable can be monotonically related to the other variable. The Hoeffding's D test [35] measures the independence of the data sets by computing the distance between the product of the marginal distributions under the null hypothesis and the empirical bivariate distribution. The test is able to identify linear / non-linear, monotonic / non-monotonic functions, and also non-functional relationships. The test outcome is a

value between -0.5 and $+1$ where larger values indicate stronger relationships, and there is no information about the variables correlation direction.

We then continued the analysis of the potential abnormal results related to the predictors A3 and A4. Table 5 presents the results of the aforementioned tests, and it is possible to observe that A3 increases its relative correlation level as the test sophistication is improved. Moreover, Spearman test result corroborates the abnormal negative relationship between population and total carbon emissions.

As regards to A4, the joint results imply that none of the tests were able to properly analyze the correlation level between temperature and CO₂ emissions. In such scenario, the next data analysis step comprised the use of a state-of-the-art computationally efficient, non-myopic, and non-parametric algorithm able to indicate and weight complex patterns of association, i.e. the RReliefF algorithm. Once a target variable is defined, RReliefF scores the correlated variables with values ranging from -1 (worst) to $+1$ (best).

Table 6 presents the outcomes of the aforementioned analysis. The Hoeffding's D test and the RReliefF algorithm do not provide information about the variables correlation direction, thus Pearson's and Spearman's outcomes are presented in terms of absolute value.

Although the RReliefF score for feature A3 does not contradict the information obtained by the other tests, the analysis of the score attributed to feature A4 indicates an important level of correlation between temperature (A4) and CO₂ emissions, whilst such condition was not

Table 5 Potential predictors ranked according to the specified testes outcomes

Pearson product-moment correlation target: total CO ₂ emissions (MtCO ₂)			Spearman rank-order correlation target: total CO ₂ emissions (MtCO ₂)			Hoeffding's D correlation target: total CO ₂ emissions (MtCO ₂)		
Predictor	Coefficient	p-value	Predictor	Coefficient	p-value	Predictor	Coefficient	p-value
A18 (WB)	0.9668816	1.07E-09	A18 (WB)	0.970588	4.72E-10	A18 (WB)	0.7218721	1.00E-08
A6 (OECD)	−0.9513449	1.52E-08	A23 (IEA)	0.917647	5.53E-07	A10 (EU)	0.5781834	1.00E-08
A10 (EU)	−0.9453401	3.38E-08	A16 (IEA)	0.914706	7.01E-07	A6 (OECD)	0.573296	1.00E-08
A23 (IEA)	0.9381434	7.88E-08	A17 (IEA)	0.9	2.05E-06	A16 (IEA)	0.5263772	1.00E-08
A16 (IEA)	0.9357116	1.03E-07	A6 (OECD)	−0.894118	3.01E-06	A23 (IEA)	0.5053688	1.03E-08
A8 (WB)	−0.9335975	1.28E-07	A10 (EU)	−0.888072	4.37E-06	A21 (EU)	0.4782913	2.59E-08
A22 (IEA)	0.9290128	2.02E-07	A22 (IEA)	0.879412	7.19E-06	A8 (WB)	0.4779412	2.62E-08
A17 (IEA)	0.9266759	2.51E-07	A8 (WB)	−0.797059	2.18E-04	A17 (IEA)	0.4456116	7.88E-08
A11 (WB)	−0.9137785	7.54E-07	A21 (EU)	0.788235	2.86E-04	A22 (IEA)	0.4393091	9.76E-08
A12 (OECD)	−0.910324	9.84E-07	A19 (WB)	0.782353	3.41E-04	A9 (WB)	0.4337068	1.18E-07
A21 (EU)	0.9017806	1.82E-06	A20 (WB)	0.782353	3.41E-04	A19 (WB)	0.4325397	1.23E-07
A19 (WB)	0.8834384	5.73E-06	A9 (WB)	−0.755882	7.06E-04	A20 (WB)	0.4219188	1.76E-07
A20 (WB)	0.8834384	5.73E-06	A11 (WB)	−0.752941	7.61E-04	A7 (OECD)	0.3960084	4.26E-07
A7 (OECD)	−0.8692267	1.23E-05	A12 (OECD)	−0.752941	7.61E-04	A3 (EU)	0.3927404	4.77E-07
A9 (WB)	−0.8683733	1.29E-05	A3 (EU)	−0.75	8.20E-04	A25 (IEA)	0.3927404	4.77E-07
A25 (IEA)	−0.8499821	3.05E-05	A25 (IEA)	−0.75	8.20E-04	A15 (EU)	0.377451	8.02E-07
A3 (EU)	−0.8264887	7.92E-05	A7 (OECD)	−0.747059	8.82E-04	A11 (WB)	0.3601774	1.44E-06
A24 (IEA)	0.7928254	2.49E-04	A15 (EU)	−0.726471	1.44E-03	A12 (OECD)	0.3601774	1.44E-06
A15 (EU)	−0.7707846	4.74E-04	A2 (EU)	−0.705882	2.25E-03	A2 (EU)	0.3122082	7.39E-06
A14 (EU)	−0.7617238	6.06E-04	A14 (EU)	−0.705882	2.25E-03	A14 (EU)	0.3122082	7.39E-06
A2 (EU)	−0.7524735	7.70E-04	A24 (IEA)	0.688235	3.20E-03	A24 (IEA)	0.2361111	1.49E-04
A1 (WB)	−0.6586972	5.52E-03	A1 (WB)	−0.597059	1.46E-02	A1 (WB)	0.1973623	4.52E-04
A13 (WB)	−0.5127864	4.22E-02	A13 (WB)	−0.505882	4.56E-02	A4 (EEA)	0.076564	2.07E-02
A4 (EEA)	0.3722717	1.56E-01	A4 (EEA)	0.4	1.25E-01	A26 (IEA)	0.0151727	1.92E-01
A5 (EEA)	−0.1650201	5.41E-01	A26 (IEA)	0.229412	3.93E-01	A13 (WB)	0.0140056	2.01E-01
A26 (IEA)	−0.0055333	9.84E-01	A5 (EEA)	−0.091176	7.37E-01	A5 (EEA)	−0.0085201	5.27E-01

Table 6 Data tests results / comparative perspective for Target T1

Feature	Pearson correlation	Spearman correlation	Hoeffding's D correlation	RReliefF	Feature	Pearson correlation	Spearman correlation	Hoeffding's D correlation	RReliefF
A1	0.6587	0.5971	0.1974	0.0021	A14	0.7617	0.7059	0.3122	0.0103
A2	0.7525	0.7059	0.3122	0.0187	A15	0.7708	0.7265	0.3775	0.0086
A3	0.8265	0.7500	0.3927	0.0217	A16	0.9357	0.9147	0.5264	0.0801
A4	0.3723	0.4000	0.0766	0.1016	A17	0.9267	0.9000	0.4456	0.0728
A5	0.1650	0.0912	−0.0085	−0.0181	A18	0.9669	0.9706	0.7219	0.0698
A6	0.9513	0.8941	0.5733	0.0520	A19	0.8834	0.7824	0.4325	0.0525
A7	0.8692	0.7471	0.3960	0.0295	A20	0.8834	0.7824	0.4219	0.0525
A8	0.9336	0.7971	0.4779	0.0392	A21	0.9018	0.7882	0.4783	0.0636
A9	0.8684	0.7559	0.4337	0.0319	A22	0.9290	0.8794	0.4393	0.0558
A10	0.9453	0.8881	0.5782	0.0281	A23	0.9381	0.9176	0.5054	0.0578
A11	0.9138	0.7529	0.3602	0.0553	A24	0.7928	0.6882	0.2361	0.1191
A12	0.9103	0.7529	0.3602	0.0609	A25	0.8500	0.7500	0.3927	0.0104
A13	0.5128	0.5059	0.0140	−0.0295	A26	0.0055	0.2294	0.0152	0.0771

Fig. 3 Exploratory visualization: T4 and A4

Table 7 Hoeffding's D statistic test—prediction target T4

Order	Feature	Co-efficient	p-value	Order	Feature	Coefficient	p-value
1	A4	0.2464	6.94E-05	14	A1	0.0716	2.46E-02
2	A16	0.2369	1.45E-04	15	A9	0.0670	2.87E-02
3	A17	0.2055	3.57E-04	16	A11	0.0663	2.94E-02
4	A24	0.1967	4.60E-04	17	A12	0.0663	2.94E-02
5	A18	0.1406	2.63E-03	18	A3	0.0587	3.82E-02
6	A20	0.1228	4.53E-03	19	A25	0.0587	3.82E-02
7	A19	0.1094	7.01E-03	20	A23	0.0579	3.93E-02
8	A6	0.1083	7.27E-03	21	A15	0.0497	5.23E-02
9	A10	0.0957	1.10E-02	22	A7	0.0409	7.18E-02
10	A8	0.0909	1.29E-02	23	A22	0.0236	1.37E-01
11	A21	0.0847	1.58E-02	24	A13	0.0018	3.36E-01
12	A2	0.0758	2.12E-02	25	A26	0.0000	3.64E-01
13	A14	0.0758	2.12E-02	26	A5	-0.0412	1.00E+00

apparent in the other tests outcomes. Consequently, such discrepancy between the tests was further investigated.

The aggregation level of the data is a characteristic that may adversely bound the effectiveness of such tests. Therefore, the next research step comprised the analysis of the CO₂ emissions in a lower level of aggregation, i.e. the test of emissions data of specific industry / economy sectors. Figure 3 shows the comparison between the emissions from commercial and public services (T4) and temperature (A4).

The similarities among the curves are relevant, and Table 7 confirms such observation by presenting the results of the Hoeffding's D correlation analysis, where one

may observe the feature A4 listed as the most relevant one when T4 is the target variable. The application of the Hoeffding's D test to a carbon emissions dataset featuring a lower level of aggregation confirmed the outcome of the RRelief analysis.

Based on such conclusions, the next research step consisted of the expansion of the RRelief analysis to our research dataset in its entirety, while taking the carbon emissions split into sectoral emissions (Table 2). Table 8 presents the results of the RRelief scoring for our research data.

Table 8 RReliefF scores for the research dataset

Predictor	Predictor importance on target variable						
	T1	T2	T3	T4	T5	T6	T7
A1	0.0269	0.0441	-0.0015	-0.0334	0.0348	-0.0018	0.0147
A2	0.0264	0.0397	0.0078	-0.0228	0.0316	0.0048	0.0164
A3	0.0373	0.0543	0.02	-0.0283	0.0383	-0.0078	0.0248
A4	0.0469	0.0149	-0.002	0.0891	-0.0086	0.124	-0.0563
A5	-0.1044	-0.0813	-0.0378	-0.065	-0.0236	-0.0839	0.0176
A6	0.0528	0.0355	0.0437	0.0239	0.0241	0.0325	-0.007
A7	0.0425	0.0447	0.0078	-0.0093	0.0278	0.0223	0.0003
A8	0.0568	0.039	0.0532	0.0241	0.0257	0.035	-0.0069
A9	0.0512	0.0533	0.0046	-0.0024	0.0305	0.0271	0.0003
A10	0.0538	0.0396	0.0503	0.0171	0.0333	0.0294	-0.0095
A11	0.067	0.0664	0.0336	0.0025	0.0379	0.0246	-0.0038
A12	0.0656	0.069	0.0349	0.0052	0.0397	0.0275	-0.0049
A13	-0.0161	-0.0136	0.0239	0.0113	0.0154	-0.0284	0.072
A14	0.017	0.0272	0.0149	-0.0204	0.0272	0.0027	0.0131
A15	0.0185	0.0292	0.0158	-0.0204	0.0275	-0.003	0.0219
A16	0.0582	0.0417	0.0455	0.0612	0.0288	0.0486	0.0303
A17	0.0283	-0.0017	0.0298	0.1019	0.0182	0.064	0.013
A18	0.0403	0.0183	0.0279	0.0773	0.0255	0.0571	0.0058
A19	0.0615	0.0582	0.0199	-0.0009	0.0287	0.0424	0.005
A20	0.0197	0.0231	0.0012	-0.0125	0.0159	0.0181	-0.0148
A21	0.0685	0.0622	0.0254	0.0042	0.0283	0.0496	0.0028
A22	0.1227	0.1367	-0.017	-0.0373	0.0652	0.0334	0.0038
A23	0.0476	0.0231	0.0483	0.055	0.0237	0.056	-0.0251
A24	0.0018	-0.0023	0.0091	0.0759	0.0279	0.0519	0.0039
A25	0.0243	0.0326	0.0342	-0.0114	0.0271	-0.0008	0.0117
A26	-0.0316	-0.0168	-0.0262	0.0437	-0.0138	0.0143	-0.0057

Fig. 4 Exploratory visualization: T6 and A4

Table 9 RReliefF algorithm notation

Number of training instances: n (1)	Number of features (attributes): a (2)
Number of training instances (user defined; $m < n$) used to update W : m (3)	
Vector of Attributes (features): $A_1, A_2, \dots, A_a$ (4)	
Instance Space (examples): $I_1, I_2, \dots, I_n$ (5)	
$R_i \in I$: randomly selected target instance (6)	
$\tau(\cdot)$: prediction value (7)	k : nearest instances I_j (8)
N_{dC} : weights for different $\tau(\cdot)$ (9)	N_{dA} : weights for different attribute (10)
$N_{dC \& dA}[A]$: weights for different prediction and different attribute (11)	
$d(i, j) = \frac{d_1(i, j)}{\sum_{l=1}^k d_1(i, l)}$ (12)	$d_1(i, j) = e^{-\left(\frac{\text{rank}(R_i, I_j)}{\sigma}\right)}$ (13)
$\text{Rank}(R_i, I_j)$: rank of instance I_j in a sequence of instances ordered by the distance from R_i (14)	
σ : user defined parameter controlling the influence of the distance (15)	
$\text{Diff}(A, I_1, I_2) = \frac{ \text{value}(A, I_1) - \text{value}(A, I_2) }{\max(A) - \min(A)}$ (16)	

Fig. 5 RReliefF algorithm pseudocode (Robnik-Šikonja and Kononenko, 2003)

```

1.  set all  $N_{dC}, N_{dA}[A], N_{dC\&dA}[A], W[A]$  to 0;
2.  for  $i := 1$  to  $m$  do begin
3.      randomly select instance  $R_i$ ;
4.      select  $k$  instances  $I_j$  nearest to  $R_i$ ;
5.      for  $j := 1$  to  $k$  do begin
6.           $N_{dC} := N_{dC} + \text{diff}(\tau(\cdot), R_i, I_j) \cdot d(i, j)$ ;
7.          for  $A := 1$  to  $a$  do begin
8.               $N_{dA}[A] := N_{dA}[A] + \text{diff}(A, R_i, I_j) \cdot d(i, j)$ ;
9.               $N_{dC\&dA}[A] := N_{dC\&dA}[A] + \text{diff}(\tau(\cdot), R_i, I_j) \cdot$ 
10.                   $\text{diff}(A, R_i, I_j) \cdot d(i, j)$ ;
11.          end;
12.      end;
13.  end;
14.  for  $A := 1$  to  $a$  do
15.       $W[A] := N_{dC\&dA}[A] / N_{dC} - (N_{dA}[A] - N_{dC\&dA}[A]) / (m - N_{dC})$ ;

```

Still analyzing the feature A4, it is possible to observe a strong correlation towards the target T6 (residential emissions), what is confirmed by the exploratory visualization in Fig. 4.

Such findings confirmed the applicability of the RReliefF algorithm to assess (score and rank) the correlation level of our research data, and qualified its use in our learning framework features engineering process.

3.2 Features engineering and neural network architecture iterative design

In the proposed learning framework, the features engineering and the model generation (i.e. NN architecture design) are iteratively accomplished by two modules, i.e.

the Features Engineering Module (FEM) and the Model Generation Module (MGM), as can be observed in Fig. 1.

The FEM accomplishes the data dimensionality and quality treatment. Such combined treatment is done by a RReliefF driven Backwards Feature Elimination (BFE) aiming at: a) selecting relevant predictors, in order to reduce the dataset features space and avoid the dimensionality curse; b) reducing the computational complexity of the learning algorithm featured in the MGM; c) improving the accuracy of predictions; d) facilitating the interpretation of results, and; e) reducing the data storage space. The feature selection process observes the notation presented in Table 9

The RReliefF Algorithm, as presented in Fig. 5, uses as input a vector of attribute values $[A]$ and predicted value

τ for each training instance I , and provide as outcome a vector W containing the score of the attributes.

As observed in the algorithm steps 8 and 9, RReliefF uses Eq. 16 for the iterative update of features weights according to their probabilistic relevance for the predictions. The intuition behind such weights computation as an expression of probabilistic relevance is conveyed by Eq. 17.

$$W[A] = \frac{P_{diffC|diffA} \cdot P_{diffA}}{P_{diffC}} - \frac{(1 - P_{diffC|diffA}) \cdot P_{diffA}}{1 - P_{diffC}} \quad (17)$$

In the formulation of Eq. 17, P_{diffA} represents the probability of having different values of A within the nearest instances, P_{diffC} represents the probability of having a different prediction within the nearest instances, and $P_{diffC|diffA}$ represents the probability of having a different prediction given a different value of A within the nearest instances.

Thus, within the FEM, the initial features $A_1 \dots A_{26}$ (potential predictors) are scored by the RReliefF algorithm, the features then are indexed and ranked based on the attributed scores, what leads to a rearranged feature set. Next step, the interaction with the MGM starts, i.e. the rearranged features set is fed to the Backpropagation Neural Network Architecture (NN/BP), the network is trained and the vector containing the current learning framework status (features subset, NN/BP architecture, NN/BP prediction accuracy) is stored. Subsequently, new features subsets are created by backward feature elimination, the NN/BP is trained with the new subset, and the learning framework status vector is updated.

The NN/BP featured in the MGM has the following characteristics:

Feed-forward network $v(x)$ defined as follows:

$$v(x) := f^L(W^L f^{L-1}(W^{L-1} \dots f^1(W^1 x) \dots)) \quad (18)$$

Number of layers: 3; hidden layer activation function (transfer function): logistic (sigmoid), defined as follows:

$$f^{L-1}(x_i) = \frac{1}{1 + e^{-x_i}} \quad (19)$$

Training method: backpropagation;

Normalization method: unit interval;

Training cost (loss) function: residual sum of squares, defined as follows:

$$RSS = \sum_{i=1}^n (y_i - v(x_i))^2 \quad (20)$$

where y_i is the i th value of the target variable; x_i is the i th value of the predicting variable; $v(x_i)$ is the predicted value of y_i .

As previously mentioned, the FEM and the MGM interact in an iterative way, and such interaction allows for an innovative BFE/RReliefF driven improvement for the NN/BP, i.e. the number of neurons in the hidden layer is changed along with the features subsets, and the learning framework status vector is updated accordingly. Once framework stop conditions are achieved, the status vector encloses the best features subset and the best NN/BP architecture in terms of prediction accuracy.

3.3 Learning framework evaluation

As observed in Fig. 1, the Model Evaluation Module (MEM) built in the implemented learning framework features the best NN/BP architecture (MGM outcome) and feeds it with the evaluation dataset. The module also features three additional ML models (Support Vector Machine—SVM, Gradient Boosting Machine—GBM, and Random Forest—RF), which are used to complement (benchmarking) the learning framework performance evaluation. The results of the benchmarking are presented in the next subsection, along with the overall assessment of the proposed framework performance.

3.3.1 Original contribution

To the best of our knowledge, our proposed learning framework is the first to implement an iterative neural network architecture improvement supported by a Backward Feature Elimination search method driven by the RReliefF algorithm.

The framework iteratively learns (NN/BP architecture, features subset) on ad hoc basis, i.e. specifically for each economy / industry sector. The implemented framework evaluation / validation processes benefited from real world data accrued by the EU, OECD, IEA, and the World Bank. The whole dataset covers the period 1990–2017, and the training and test datasets were determined in accordance to the Pareto principle for data sampling.

Table 10 presents the accuracy (Root Mean Square Error—RMSE for the test dataset) of the proposed learning framework for the totality of the EU28 CO2 emissions as well as for sectoral emissions. The table also presents the accuracy for the experiment control NN/BP, the framework featuring NN/BP supported by plain BFE (NN/BP-BFE), and the framework featuring NN/BP supported by RReliefF driven BFE (NN/BP-RReliefF/BFE).

The accuracy figures in Table 10 demonstrate the improved performance of the proposed learning framework when compared to the control NN/BP, as well as to other possible framework designs. The table also presents the number of predictors in the learned features subset, and the number of neurons in the hidden layer.

Table 10 Computation learning framework evaluation figures

Industry / economy sector	Model parameters / prediction accuracy metric	Methods			
		NN control	NN-BFE	NN-RReliefF/BFE	Proposed framework
Total emissions (T1)	#Predictors	26	15	17	21
	# Neurons HL	3	3	3	7
	RMSE (MtCO ₂)	692.5226	191.9399	145.2302	143.2545
Energy industries emissions (T2)	#Predictors	26	9	24	18
	# Neurons HL	3	3	3	7
	RMSE (MtCO ₂)	264.1396	55.5251	80.3972	53.0494
Industry emissions (T3)	#Predictors	26	8	12	20
	# Neurons HL	3	3	3	8
	RMSE (MtCO ₂)	139.3530	5.6810	7.6481	5.6749
Commercial and public services emissions (T4)	#Predictors	26	11	5	18
	# Neurons HL	3	3	3	5
	RMSE (MtCO ₂)	15.8050	10.9382	9.0678	6.4312
Transport emissions (T5)	#Predictors	26	21	14	21
	# Neurons HL	3	3	3	5
	RMSE (MtCO ₂)	63.3391	29.1528	40.6110	26.5271
Residential (T6)	#Predictors	26	12	12	16
	# Neurons HL	3	3	3	5
	RMSE (MtCO ₂)	99.7645	24.2505	22.3417	21.9795
Aviation (T7)	#Predictors	26	19	13	12
	# Neurons HL	3	3	3	4
	RMSE (MtCO ₂)	6.9858	6.9283	6.9406	6.9285

As observed in Table 10, the proposed approach combining backward feature elimination, RReliefF feature qualification, and iterative improvement of the NN/BP architecture effectively boosted the carbon emissions prediction accuracy for the EU28 scope dataset.

The computational complexity of the neural network is $O(h^a)$, where h is the number of hidden layers and a is the number of features (predictors), and the training process converged in less than 100 epochs, with a learning rate of 0.1. The RReliefF algorithm computational complexity is $O(n.m.a)$, where n is the number of training instances,

and m is the number of training instances used by the algorithm to update the weights. The computational environment featured the CPU Intel i9-9900 k supported by the GPU NVIDIA RTX 2080 (8 Gb) and 32 Gb DDR4 RAM, and the whole prediction process took approximately 1 h (average) per prediction target.

3.3.2 Learning framework validation

The validation of the original contribution provided by the proposed learning framework consisted of two different

Table 11 Computational learning framework benchmarking figures (MAPE) for prediction target T1—Total EU28 Carbon Emissions (NN/BP specific)

Method	Comparison parameters								
	Proposed framework	NN/BP [25]	NN/BP [26]	NN/BP-CT [24]	NN/BP-IPSO [25]	NN/BP [27]	NN/BP [30]	NN/BP-PCA [30]	NN/BP-RF [31]
Scope	EU28	China	China	China	China	China	China	China	China
	1990	1993	1980	1960	1993	1995	1990	1990	1997
	2017	2014	2014	2011	2014	2014	2016	2016	2017
Test instances (years)	5	5	5	5	5	5	5	5	4
# Predictors	16 (26)	5	11	3	5	5	24	24	7 (17)
MAPE	2.28	5.07	5.49	3.43	2.53	3.72	4.03	2.87	8.09

Table 12 Computational learning framework benchmarking figures (MAPE) for prediction target T1—Total EU28 Carbon Emissions (mainstream ML models)

Method	Comparison parameters							
	Proposed framework	GBM-BFE	RF-BFE	SVM-BFE	Econometric / system dynamics [6]	GM [26]	SVM-RF [31]	FGMVM [21]
Scope	EU28	EU28	EU28	EU28	Ecuador	China	China	China
	1990	1990	1990	1990	1980	1980	1997	2008
	2017	2017	2017	2017	2010	2014	2017	2016
Test instances (years)	5	5	5	5	–	5	4	2
# Predictors	16 (26)	17 (26)	19 (26)	6 (26)	25	11	7 (17)	3
MAPE	2.28	7.05	7.98	17.58	15.96	5.77	13.37	15.14

analysis. Firstly, its outcomes were compared to three current mainstream ML models (i.e. SVM, GBM, RF) using our research dataset. Secondly, our accuracy figures were benchmarked against the results of recently published researches targeting carbon emissions prediction.

In the validation process we worked with mean absolute percentage error—MAPE as the accuracy metric, and focused on the prediction target T1 (total CO₂ emissions). Thus the best performing model designed by our proposed learning framework features 7 neurons in the hidden layer, and 16 predictors out of the candidates presented in Table 3, i.e.: A18, A4, A23, A16, A8, A17, A10, A6, A22, A12, A11, A25, A9, A7, A24, and A26.

The trained framework achieved 2.28% accuracy performance (MAPE) on the test dataset, and Table 11 allows us to benchmark the result of the proposed learning framework with the results of other NN/BP implementations.

As observed in Table 11, the predictions of the proposed learning framework are significantly better than the ones provided by similar researches.

Table 12 compares the result of our proposed framework with the results of three ML models (GBM, RF, SVM) supported by plain BFE, and using our research dataset, as well as against the results of recently published researches models other than NN/BP.

The figures in Tables 11 and 12 reinforce the relevance of the results accrued by our research, and confirm ANNs as a powerful algorithm capable of processing a large amount of non-linear and non-parametric data. The RReliefF algorithm, in turn, efficiently assess and rank the predicting variables (features) by effectively addressing non-linear relationships, time series, noisy and correlated features, as well as features interactions of high order (complex patterns associations).

The iterative combination of these two algorithms produced a powerful and scalable prediction tool able to process huge datasets featuring complex and incomplete data. The improved ad hoc learning capability of our

framework makes it potentially applicable to any region in the world, and for any level of data aggregation.

Whereas our original contribution represents an important step towards the better design and implementation of environment protection initiatives and policies, its effectiveness would be greatly enhanced once combined with an explainable Artificial Intelligence (XAI) technique, given the black-box nature of ANN algorithms.

3.4 Predictions explanation

Regarding ML outcomes, it is critically important to ensure that its predictions accuracy performance relies on valid features (predictors) computations, i.e. the ML model is providing the right answer for the right reasons. Specifically addressing our research, although ANNs are top performing, non-parametric and scalable algorithms, they lack the required algorithmic transparency to adequately support policy making and decision-making processes targeting the complex environmental challenges. Therefore, we improved our learning framework with an XAI module (i.e. PEM—Prediction Explanation Module) featuring the Local Interpretable Model-agnostic Explanations—LIME technique [36].

LIME is a scalable method that creates local interpretable surrogate models (explanations) around a given instance in order to estimate how data points influence the global model predictions. LIME translates the explanation problem into an optimization problem. The search space comprises explanations generated by the local interpretable surrogate models $g \in G$, where G is a class of interpretable models. Locality is defined by a proximity measure $\pi(x, z)$ expressing the distance between an instance z and x . The interpretability degree of the surrogate model is assessed by means of a complexity measure $\Omega(g)$. Thus, by considering $f(x)$ the model to be explained, the local fidelity measure $\mathcal{L}(f, g, \pi)$ express how unfaithful g is in approximating f in the locality π . Finally the LIME

Table 13 Predictions explanation module outcomes

T1—Total CO2 Emissions / MtCO2				
Year (case)	Actual emissions	Framework prediction	Feature	Feature weight
2013	3,555.66	3,612.61	A18	1.8853
			A4	6.0553
			A23	46.7868
			A16	97.0558
			A8	−65.7551
2016	3,202.18	3,207.78	A18	1.5832
			A4	12.1087
			A23	44.7131
			A16	26.1756
			A8	81.0778

outcome ensuring both interpretability and local fidelity is defined by:

$$\varepsilon(x) = \arg \min_{g \in G} \mathcal{L}(f, g, \pi) + \Omega(g) \quad (21)$$

Within the learning framework XAI module, the model to be explained is defined as $f(x) = v(x)$ (i.e. the ANN). G comprises ridge regression models for the perturbed sample $z' \in Z$ (perturbed samples dataset), such that $g(z') = \beta_g \cdot z'$. The complexity measure $\Omega(g)$ is expressed in terms of non-zero coefficients in the linear model, π is defined by the Euclidian distance, and the local fidelity is computed as square loss. We thus define:

$$\mathcal{L}(f, g, \pi) = \sum_{z, z' \in Z} \pi(x, z) (v(z) - g(z'))^2 \quad (22)$$

From a general perspective, for each prediction to be explained, LIME algorithm permutes (perturbation) the observation n times; the statistics for each variable are extracted and permutations are then sampled from the variable distributions. The model to be explained then predicts the outcome of all permuted observations, and the algorithm calculates the Euclidian distance from all permutations to the original observation, and selects the m features with highest absolute weight in a ridge regression fit of the complex model outcome.

Afterwards a simple model is fitted to the permuted data, explaining the complex model outcome with the m features from the permuted data weighted by its distance to the original observation. And finally, the algorithm extracts the feature weights from the simple model and use them to explain the local behavior of the complex model.

In order to demonstrate the whole outcome of our implemented learning framework, Table 13 presents

predictions and explanations (features weights) for two cases, i.e. predictions for the years 2013 (EU-ETS phase 3 first year) and 2016 (most accurate learning framework prediction). For the sake of simplification, among the 16 selected features, Table 13 presents the explanations for the first 5 as ranked by the RReliefF algorithm.

As observed in Table 13, the XAI module featured with LIME provides insights on how a specific feature contributes for a specific prediction (case). We observe, for instance, that oil based energy supply (predictor A23) consistently and positively contributes for, and induces total CO2 emissions. In such context it is extremely important to differentiate the two different impacts, i.e. contribution and induction. Although it is clear that energy production from oil combustion definitely contributes for CO2 emissions, its use may induce the reduction of total CO2 emissions due to the replacement phenomenon, by which an increment on oil use may imply a reduction on the use of a more polluting source of energy, like coal. Hence, our learning framework outcomes express inductive relationships rather than contributory ones.

Such improvement turns our research contribution much more valuable for the design of better environmental initiatives and policies dependent on CO2 emissions forecasts. Additionally, the outcomes of our proposed learning framework may provide some background for future works addressing carbon emissions causality analysis, as well as potential improvements on both ANNs and XAI techniques.

4 Conclusions

The accurate forecast of CO2 emissions is one of the most important inputs for any decision-making process targeting climate change / global warming avoidance. Therefore, in our attempt to contribute to such global environmental challenge, we implemented a learning computational framework for carbon emissions predictions. Our framework features the capacity to iteratively improve the prediction features set and the backpropagation neural network (NN/BP) architecture according to the data statistical assessment computed by the RReliefF algorithm. Our research counted with real world data obtained from the European Union, International Energy Agency and World Bank, for the period 1990–2017.

The outcomes of the designed prediction framework were successfully evaluated against different NN/BP based solutions (NN/BP, NN/BP-BFE, NN/BP-RReliefF/BFE, NN/BP-CT, NN/BP-IPSO, NN/BP-PCA, NN/BP-RF)), as well as different mainstream machine learning models (GBM-BFE, RF-BFE, SVM-BFE, SVM-RF). Additionally, the featured XAI module provided insights on how different predictors

impacted a specific prediction case. Therefore, our results demonstrated the effectiveness of our approach in terms of increased and explained prediction accuracy, which may adequately support the design and the improvement of environmental initiatives and policies.

Finally, the outcomes of our implemented learning framework may provide some background for future works addressing carbon emissions causality analysis, as well as potential improvements on both ANNs and XAI techniques.

Data Availability The research used data publicly available on the websites of the mentioned sources.

Compliance with ethical standards

Conflict of interest The author(s) declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Crespo, A. M. F., Wang, C. (2020). "European Union Emissions Trading Scheme: Design Evolution and Effectiveness Analysis." Handbook of Research on Interdisciplinary Approaches to Decision Making for Sustainable Supply Chains, eds. Anjali Awasthi and Katarzyna Grzybowska, IGI Global, 189–210, <https://doi.org/10.4018/978-1-5225-9570-0.ch009>.
- European Environment Agency (2018) "Trends and projections in Europe 2018: Tracking progress towards Europe's climate and energy targets. (No.16/2018). Luxembourg: Publications Office of the European Union. DOI:<https://doi.org/10.2800/931891>.
- Guan D, Hubacek K, Weber CL, Peters GP, Reiner DM (2008) The drivers of Chinese CO2 emissions from 1980 to 2030. Global Environ Change Part A: Human Policy Dimens 18(4):626–634
- Anger A (2010) Including aviation in the European emissions trading scheme: Impacts on the industry, CO2 emissions and macroeconomic activity in the EU. J Air Transport Manage 16(2):100–105
- Li W, Lu C (2015) The research on setting a unified interval of carbon price benchmark in the national carbon trading market of China. Appl Energy 155:728–739
- Robalino-López A, Mena-Nieto A, García-Ramos JE (2016) System dynamics modeling for renewable energy and CO2 emissions: a case study of Ecuador. Energy Sustain Develop 20:11–20. <https://doi.org/10.1016/j.esd.2014.02.001>
- Scott K, Roelich K, Owen A, Barrett J (2017) Extending European energy efficiency standards to include material use: an analysis. Climate Policy 18(5):627–641
- Mi Z, Meng J, Guan D, Shan Y, Liu Z, Wang Y, Feng K, Wei Y-M (2017) Pattern changes in determinants of Chinese emissions. Environ Res Lett. <https://doi.org/10.1088/1748-9326/aa69cf>
- Chang K, Chang H (2016) Cutting CO2 intensity targets of inter-provincial emissions trading in China. Appl Energy 163:211–221
- Yang L, Zhang Q, Ji J (2017) Pricing and carbon emission reduction decisions in supply chains with vertical and horizontal cooperation. Int J Prod Econ 191:286–297
- Yang L, Wang G, Ke C (2018) Remanufacturing and promotion in dual-channel supply chains under cap-and-trade regulation. J Clean Prod 204:939–957
- Xu L, Wang C, Zhao J (2018) Decision and coordination in the dual-channel supply chain considering cap-and-trade regulation. J Clean Prod 197:551–561
- Wang Z, Wang C (2015) How carbon offsetting scheme impacts the duopoly output in production and abatement: analysis in the context of carbon cap-and-trade. J Clean Prod 103:715–723. <https://doi.org/10.1016/j.jclepro.2014.04.069>
- Gavard C, Winchester N, Paltsev S (2016) Limited trading of emissions permits as a climate cooperation mechanism? US–China and EU–China examples. Energy Econ 58:95–104
- Zhang X, Qi T, Ou X, Zhang X (2017) The role of multi-region integrated emissions trading scheme: a computable general equilibrium analysis. Appl Energy 185:1860–1868
- Cui Q, Li Y, Wei Y (2017) Exploring the impacts of EU ETS on the pollution abatement costs of European airlines: an application of network environmental production function. Transp Policy 60:131–142
- Hong J, Su JCP, Chu C, Yen C (2018) Decentralized decision framework to coordinate product design and supply chain decisions: evaluating trade-offs between cost and carbon emission. J Clean Prod 204:107–116
- Wang H, Ang BW, Su B (2017) Assessing drivers of economy-wide energy use and emissions: IDA versus SDA. Energy Policie, 107:585–599
- Solaymani S (2019) CO2 emissions patterns in 7 top carbon emitter economies: the case of transport sector. Energy 168:989–1001
- Allevi E, Gnudi A, Konnov IV, Oggioni G (2018) Evaluating the effects of environmental regulations on a closed-loop supply chain network: a variational inequality approach. Ann Oper Res 261(1–2):1–43
- Jiang H, Kong P, Hu Y-C, Jiang P (2020) Forecasting China's CO2 emissions by considering interaction of bilateral FDI using the improved grey multivariable verhulst model. Environ Dev Sustain. <https://doi.org/10.1007/s10668-019-00575-2>
- Chang C (2010) A multivariate causality test of carbon dioxide emissions, energy consumption and economic growth in China. Appl Energy 87:3533–3537
- Hong K, Jung H, Park M (2017) Predicting European carbon emission price movements. Carbon Manage 8(1):33–44. <https://doi.org/10.1080/17583004.2016.1275813>
- Liu Y, Tian Y, Chen M (2017) Research on the prediction of carbon emission based on the chaos theory and neural network. Int J Bioautomation 21(4):339–348
- Zhou J, Du S, Shi J, Guang F (2017) Carbon emissions scenario prediction of the thermal power industry in the Beijing-Tianjin-Hebei region based on a back propagation neural network optimized by an improved particle swarm optimization algorithm. Polish J Environ Stud 26(4):1895–1904. <https://doi.org/10.15244/pjoes/68881>
- Sun W, Sun J (2017) "Prediction of carbon dioxide emissions based on principal component analysis with regularized

- extreme learning machine: the case of China. *Environ Eng Res* 22(3):302–311
27. Zhou J, Guang F, Tang R (2018) Scenario analysis of carbon emissions of China's power industry based on the improved particle swarm optimization-support vector machine model. *Polish J Environ Stud* 27(1):439–449. <https://doi.org/10.15244/pjoes/74132>
 28. Li M, Wang W, De G, Ji X, Tan Z (2018) forecasting carbon emissions related to energy consumption in Beijing-Tianjin-Hebei region based on grey prediction theory and extreme learning machine optimized by support vector machine algorithm. *Energies*, MDPI, Open Access J 11(9):1–15
 29. Song Y, Liu T, Liang D, Li Y, Song X (2019) A fuzzy stochastic model for carbon price prediction under the effect of demand-related policy in china's carbon market. *Ecol Econ* 157:253–265
 30. Sun W, Jin H, Wang X (2019) Predicting and analyzing CO2 emissions based on an improved least squares support vector machine. *Polish J Environ Stud* 28(6):4391–4401. <https://doi.org/10.15244/pjoes/94619>
 31. Wen L, Yuan X (2020) Forecasting CO2 emissions in chinas commercial department, through BP neural network based on random forest and PSO. *Sci Total Environ*. <https://doi.org/10.1016/j.scitotenv.2020.137194>
 32. Javanmard A, Montanari A (2014) Confidence intervals and hypothesis testing for high-dimensional regression. *J Machine Learn Res* 15:2869–2909
 33. Kira, K., Rendell, L. A. (1992). The feature selection problem: traditional methods and a new algorithm. 10th National Conference on Artificial Intelligence. AAAI-92 Proceedings, San Jose, California
 34. Robnik-Šikonja M, Kononenko I (2003) Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learn* 53:23–69. <https://doi.org/10.1023/A:1025667309714>
 35. Hoeffding W (1948) A non-parametric test of independence. *Ann Math Stat* 19:546–557
 36. Ribeiro, M. T., Singh, S., Guestrin, C (2016) "Why should I trust you? Explaining the predictions of any classifier." Proceedings of the 22nd International Conference on Knowledge Discovery and Data Mining, 1135–1144 <https://doi.org/10.1145/2939672.2939778>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.