Artificial Intelligence (AI)-driven approach to climate action and sustainable development

Haein Cho [1,4] ✉ & Emmanuel Ackom[2,3]

Countries have pledged commitment to the 2030 Sustainable Development Goal (SDGs) and the Paris Agreement to combat climate change. To maximize synergies between SDGs and climate actions (CAs), we evaluate the alignment of national commitment to SDGs and emissions reduction targets by comparing action plans embodied in Voluntary National Review (VNR) reports and the Nationally Determined Contributions (NDCs) across 67 countries. An Artificial Intelligence (AI)-based approach is proposed in this study to explore the interconnectedness by applying machine learning classifier and natural language processing. Middle- and low-income countries with high emissions tend to have low NDC targets and contain similar information in VNR reports. High-income countries show less alignment between their NDCs and VNRs. The economic status of countries is found to be connected to their climate actions and SDGs alignment. Here, we demonstrate utility and promise in using AI techniques to unravel interactions between CA and SDG.

The two most important international treaties signed in 2015, the 2030 Sustainable Development Goals (SDGs) Agenda and the Paris Agreement, have geared global policies to combat climate change and achieve sustainable development. The SDG agenda and the Paris Agreement have many actions and priorities in common. They 1) are based on international agreements involving all countries to take action, 2) intend to benefit everyone globally, 3) achieve inclusive society, 4) encourage multilateral partnerships, and 5) adopt a bottom-up approach where participating countries design their implementation plans[1]. For the SDGs, countries are encouraged to report the progress against SDG targets through Voluntary National Review (VNRs) every four or five years. At the same time, the Paris Agreement asks the Parties to set national emission reduction targets and describe national efforts to achieve the target in Nationally Determined Contributions (NDCs).

The 2030 Sustainable Development Goals Agenda was adopted by setting out 17 Goals with 169 targets to encourage countries to achieve sustainable development. In recognition of the fundamental importance of the interdependence of social, environmental and economic dimensions of development, the Agenda requests to design and implement a coherent and integrated policy framework. For example,

Article 7 of the Paris Agreement recommends that the Parties (countries) design climate actions to enhance adaptive capacity and reduce vulnerability to climate change, contributing to sustainable development[2]. Although the main idea of NDCs is to reduce emissions, countries describe national efforts to additionally adapt to climate change by strengthening resilience. Building resilience and adaptive capacity to natural disasters exacerbated by climate change is now a prerequisite for sustainable development[3].

Moreover, emission reduction actions are undertaken in the context of sustainable development to improve access to clean water and energy and enhance countries' socio-economic development[4-7]. For example, SDG 2 (Zero Hunger) would not be possible without securing food systems from the negative impacts of climate change. SDG 7 (Affordable and Clean Energy) and SDG 8 (Decent Work and Economic Growth) would be interrupted without transforming to low-carbon energy sources.

Despite the strong interconnection between SDGs and NDCs, both have been separately treated. Different ministerial departments are responsible for the implementation of SDGs and NDCs, and they often work in silos[8]. The main government units in charge of the NDCs are the Departments of Environment and Climate Change, while SDGs

[1]National Assembly Futures Institute, Seoul, Republic of Korea. [2]Department of Geosciences, College of Arts, Sciences and Engineering, University of North Alabama, Florence, Alabama, USA. [3]Department of Forest Resources Management, Faculty of Forestry, University of British Columbia, Vancouver, Canada. [4]Present address: Samsung Electronics, Gyeonggi-do, Republic of Korea. ✉e-mail: haein.cho123@gmail.com

are mainly led by the Department of Planning, Finance, Economy, and Development in 90 countries[9]. Coordination mechanisms across leading ministries would facilitate a higher level of coordination between the SDGs and NDCs[10]. The conventional structure for decision-making and policy implementation does not suffice to address the complexity of the interconnections between the two[11–13].

Many studies have investigated the alignment and interactions between NDCs and SDGs. Janetschek et al.[9] reveal that climate activities in NDCs are closely related to the 17 SDGs through content analysis of climate activities indicated in 164 NDCs. The extent to which the NDCs are aligned with the targets of the SDGs was analyzed by estimating the visibility of NDCs in the SDG targets. For example, by applying the most conservative visibility level, 21 out of 169 SDG targets were found to be mentioned explicitly in South Africa's NDC[14]. Previous study applies conventional methods to provide concrete examples of jointly implementing SDGs and NDCs, focusing on five aspects, including institutional coordination and alignment between SDGs and NDC targets. From the perspective of climate financing, the amount and trends of climate-related development assistance assigned to implement the SDGs were evaluated before and after the Paris Agreement[15]. Here, we attempted to apply Artificial Intelligence (AI) to examine the interactions between NDCs and SDGs and provide actionable insights toward policy alignments.

AI is reported to have demonstrated an impressive capacity to understand and solve complex problems in many academic fields and industries[16]. The algorithms that learn patterns and make predictions or decisions have been applied in many fields[17–19], proving superior performance in pattern recognition and prediction compared with traditional methods[20,21]. Advances in AI have substantially affected social science research and revolutionized how we understand and interpret human behavior, society, and culture[22–24]. Surveys and interviews have been widely used in social science research[25,26], including SDG-NDC interaction studies[27,28]. In addition to that, modeling and a combination of modeling results and expert surveys have been applied to explore the interactions[29]. These studies discuss analytical challenges and technical limitations due to the computational capabilities of modeling and the lack of a quantitative approach to assess and map interconnections[27–29]. AI offers a new research direction that enables us to interpret complex phenomena by analyzing large databases containing diverse information from various sources and identifying interactions and causalities. As previous research has shown, text mining technique was applied in SDG research to quantitatively analyze information contained in VNR[30]. Given that data comes from an extensive range of sources in myriad forms (e.g., text, numbers) at many different levels (e.g., the level of country, sector) through advanced measurement and regulated reporting systems[31,32], advanced analytical technologies are in need for handling the complexity of the database.

Against this background, this work proposes an AI-based approach to identify patterns and evidence linking NDCs and SDGs by applying AI techniques that current traditional approaches have missed. We investigate two research questions. The first is which SDG indicator is the most relevant to NDC targets. For this first question, machine learning algorithms (Extra trees and Random forest models) are used to select important SDG features and map them to NDC targets. The second question explores whether national SDG visions are relevant to NDC targets. For the second question, the Natural Language Processing (NLP) technique is applied to analyze VNRs and examine any relevance between VNRs and NDC targets. Based on the results, we scrutinize the interlinkage between NDCs and SDGs and discuss the usability aspect of AI techniques in assessing global actions of climate change and sustainability and recommending policy measures that would maximize synergies between them.

Results

Classifier-based feature selection

We compare feature importance scores calculated by three different methods: Logistic Regression (LR), Extra Trees (ET), and Random Forest (RF). This score refers to the level of importance of each feature in predicting a target variable, helping select the most relevant features while removing irrelevant ones to enhance the predictive performance of a model. In this case, the target is the NDC goal, and the features are 120 SDG indicators across 17 SDGs defined by the Sustainable Development Report, as countries preparing the VNRs frequently use these. Two tree-based models, ET and RF, calculate the score by measuring the mean decrease in impurity of a dataset. In contrast, LR calculates feature importance by taking the average of the coefficients across classes. Thus, the range of scores differs between tree-based models and LR.

We focus on selecting top-ranked features that contribute most to each model's target variable, NDC, because they enhance both model performance and interpretability of the relationship between SDGs and NDCs. Comparing the best top five features based on feature importance score (Table 1), we find that the most critical feature varies across types of classifiers in that LR selects the unemployment rate

Table 1 | List of top 5 selected indicators across types of classifiers

	Code	Score	Indicator
Logistic Regression (LR)	n_sdg8_unemp	0.015778	Unemployment rate
	n_sdg16_detain	0.013914	Unsentenced detainees
	n_sdg14_cpma	0.013680	Mean area that is protected in marine sites important to biodiversity
	n_sdg17_govex	0.012044	Government spending on health and education
	n_sdg15_forchg	0.011985	Permanent deforestation (3-year average)
Extra Trees (ET)	n_sdg12_explastic	0.050472	Exports of plastic waste
	n_sdg17_govex	0.049064	Government spending on health and education
	n_sdg12_ewaste	0.047990	Electronic waste
	n_sdg13_co2import	0.041699	CO2 emissions embodied in imports
	n_sdg15_cpfa	0.040307	Mean area that is protected in freshwater sites important to biodiversity
Random Forest (RF)	n_sdg16_cpi	0.047178	Corruption Perceptions Index
	n_sdg3_u5mort	0.042754	Mortality rate, under-5
	n_sdg12_explastic	0.041532	Exports of plastic waste
	n_sdg17_govex	0.039496	Government spending on health and education
	n_sdg14_cpma	0.034024	Mean area that is protected in marine sites important to biodiversity

(SDG 8), ET chooses the amount of exported plastic waste (SDG 12) and RF finds corruption perceptions index as the most important.

Interestingly, all three models select government spending on health and education as one of the top 5 critical indicators. For example, Tuvalu, one of the most climate-vulnerable countries worldwide, is committed to spending 22% of its GDP (Gross Domestic Product) on education, health and decarbonizing electricity generation by 100%. Previous work concluded that public investment in education to enhance human awareness and capacity is essential because educated individuals were found to be better prepared for natural disasters, suffer less and recover faster from adverse impacts[33].

Protected area for biodiversity, coded as sdg14_cpma and sdg15_cpfa, is a theme commonly found in three models. Denmark is a leading example, recording above 86% for these two indicators. Biodiversity in Denmark, in general, has declined and experienced changes due to a lack of natural hydrology, drainage of agricultural areas, and oxygen depletion, which might have been caused by climate change[34]. In response, the national legislative framework for biodiversity protection set out rules to prohibit land use conversion. Scaling up of 'nature-based solutions (NBS),' including sustainable use and protection of natural areas, represents a considerable potential for countries to reach their NDC targets cost-effectively. NBS would achieve up to 37% of cost-effective global CO_2 mitigation targets[35]. Not only did the Paris Agreement recognize its importance, but approximately 77% of the NDCs also highlighted the importance of NBS for climate change mitigation and adaptation.

Another critical feature commonly found in both ET and RF is the export of plastic waste. The highest score is found in Slovenia (65), followed by Belgium (29) and Iceland (23). Due to the high rate of reliance on the import of raw materials (77%), Slovenia takes the implementation of a circular economy as a process to create the conditions for sustainable development[36] and strengthen resilience to climate change threats and uncertainty. According to previous studies, SDG12 is found to improve emission reductions by advancing climate actions and 47% of transnational climate initiatives are driven by SDG12[37]. The emphasis on the circular economy might promote the alignment of SDG 12 with NDC[38], influencing the degree of ambitiousness of NDC targets.

Comparing the entire list of top 50 indicators extracted from ET (Fig. 1) with the results of LR and RF, we find that the unemployment rate (n_sdg8_unemp) and mean area in marine sites protected for biodiversity (n_sdg14_cpma) are detected by all three models. Unlike the LR and RF models that select two indicators as major ones with high predictive power (top 5 and 10, respectively), the ET model places two indicators among the top 20 (Fig. 1), assigning relatively diminished significance to two indicators in predicting NDC.

Despite the common expectation of high relevance between SDG 13 and the NDC target, SDG 2 and SDG 3 most frequently occur. Indicators are oriented towards life, well-being and health, including human trophic level, prevalence of stunting in children, subjective well-being, life expectancy at birth, maternal mortality rate, and universal health coverage. This implies that public health issues or concerns have relatively intense interaction with national commitment to tackling climate change. The achievement of SDG 2, 'End of hunger,' would be hampered due to adverse climate conditions, including droughts, landslides and floods that affect food prices and food security[39,40]. As demonstrated by extreme weather events that cause physical and mental health problems[41] and renewable energy production that improves health through pollution reduction[42], SDG 3, 'Good health and well-being,' is synergistically associated with SDG 13[43]. The NDC activities are found to have strong connections to SDG 2 in that measures for sustainable and climate-smart agriculture are emphasized in the NDC report to limit temperature increase[44]. SDG 3 is also linked to the NDC by the efforts to alleviate the negative impact of climate change on health[1].

Natural language processing-based evaluation of voluntary national review

The similarity in the content of VNR documents across countries was examined using NLP (Fig. 2). With the multidimensional scaling technique, we extract x and y coordinates, the horizontal and vertical addresses of a point in a two-dimensional space, to identify the exact location of each country. The distance between dots representing countries is preserved, indicating that countries close to each other in the two-dimensional plot might have similar perspectives in their VNRs. In contrast, a vast distance illustrates a difference in their perspectives.

Unlike the geographical grouping where we do not find any distinctive differences across groups, Fig. 2 shows that a group of high-income economies creates a cluster on the upper left side of the graph (red oval shape), demonstrating a certain level of similarities. For the other income groups, no difference is distinctively observed.

The cluster in the upper left corner mainly comprises EU member countries. Having the SDGs as an intrinsic part of the political guidelines, the EU integrates the SDGs into all Commission proposals, policies and strategies[45]. Of the six priorities in the guidelines developed by European Commissions, the European Green Deal addresses 10 SDGs encompassing health, cities, circular economy, and life[46]. It is noticeable that Italy is located relatively far from the center of a group of EU countries. This might be attributable to a slightly different approach to analyzing Italy's SDG progress. While most countries evaluate progress on each SDG, focusing on each goal primarily, we find that Italy's VNR places relatively more importance on creating an enabling environment to achieve SDGs by looking at policy coherence for sustainable development, localization of the SDGs and stakeholder engagement. The integrated vision that Italy adopted is reflected in the way SDG progress is presented in VNRs. Unlike other EU countries analyzed in this study that presents the details of the national progress of each SDG, Italy appears to assess the combined impact of multiple SDGs under the National Sustainable Development Strategy, which was built upon the shared idea of sustainability across multisectoral actors aiming to strengthen the coherence of policy interventions across government ministries. Such differences in the reports' primary focus and analytic approach caused the distance between Italy and other EU countries.

Another cluster in the bottom right corner comprises the Democratic People's Republic of Korea (DPRK), Mongolia and Namibia. We find similarities across three countries by comparing the contents of three VNRs by counting pairs of two words that occur together. The word pairs, including climate change, disaster risk, and gender equality, are commonly found. Climate change and disaster risks are frequently observed in the VNRs of all three countries. Moreover, we speculate that it might be attributable to similarities in political ideology between DPRK and Mongolia, which seem to have affected EU projects for sustainable forestry to fund North Korean experts to travel to Mongolia to establish comparative frameworks and expand their capacity for forest management and technical skills.

Status of sustainable development goals, visions and nationally determined contributions

The NDC target and score of SDG 13 (climate), which is a sum of CO_2 emissions from fossil fuel combustion, CO_2 emissions embodied in imports and fossil fuel exports, are applied (Fig. 3). The level of score of SDG13 is segmented and grouped into four categories, represented as four different colors (blue, yellow, green, and orange), while the NDC emission reduction target determines the size of dots. For example, the United Kingdom has a large NDC emission reduction target compared to Japan, which has an intermediate reduction target or Vietnam, which has a small target. Countries with high emissions (in orange color) tend to have lower NDC emission reduction targets. Most of
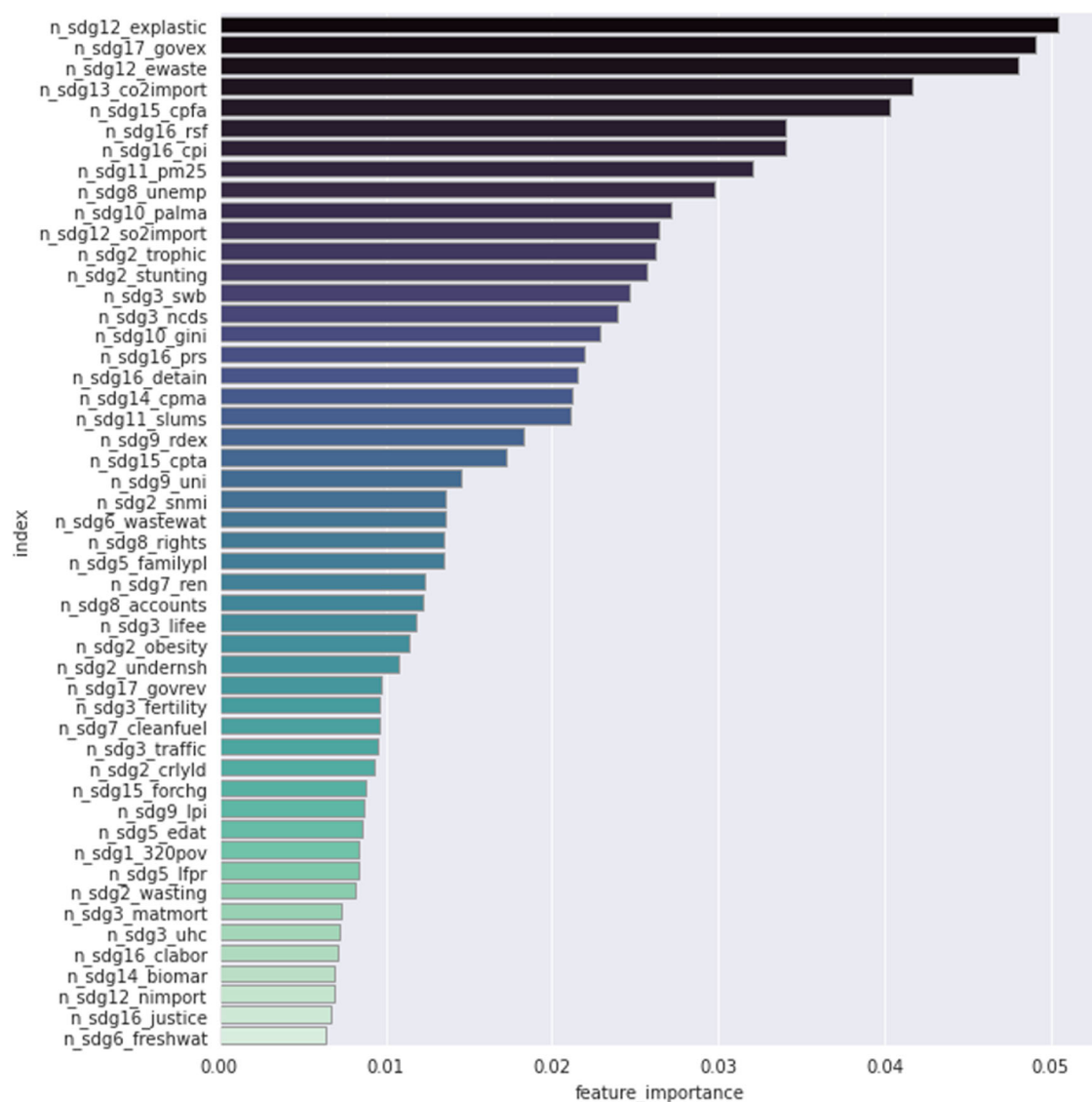
Fig. 1 | Feature importance score of sustainable development goals. The length of each bar represents the level of relative importance of each feature meaning that a longer bar means a larger effect of respective feature on the model. The explanation of individual codes is provided in Supplementary Table 1.

them are found in developing countries, which implies a need to decouple economic growth from greenhouse gas emissions and environmental degradation. EU member countries appear to be the same size regardless of the amount of emission because the EU submitted its updated NDC target of 55% emission reduction by 2030 from 1990 levels, which is applied to all EU countries.

Based on the results of classifier-based feature selection (Fig. 1), the score of plastic export under SDG12 (termed 'SDG12-explastic' below) is applied. The amount of plastic waste exported (kg per capita) is normalized between 0 and 100 as default in SDR, which is segmented into four categories. A class with a score range between 0-30 is colored blue, 30-60 in yellow, 60-90 in green, and 90 and above in orange. Compared to Fig. 3, Fig. 4 does not show clear division by color. European countries show more divisive results with SDG12 than SDG13 (Fig. 3), which is attributable to a much larger standard deviation of SDG12 than SDG13 across European countries. Finland and Italy are classified into a different group among EU countries due to a relatively higher score for SDG12-explastic.

In contrast, the remaining countries tend to be classified better per their NDC. For example, Fig. 3 presents the United Arab Emirates,

Australia and Thailand separated by colors, while the three countries set NDCs at similar levels (UAE: 23.5%, Australia: 27%, Thailand: 20%); however, SDG12 in Fig. 4 group three countries with the same color of green. The Republic of Korea (ROK) and the United Kingdom (UK) are represented in the same color group in Fig. 3, while their NDCs are 40% and 68%, respectively. This is then re-classified in Fig. 4 in a more balanced way where the ROK is in the same class as Italy (NDC 55%) while the UK is filtered in a different class colored yellow. Countries with NDC target lower than 15% tend to be found in a group with SDG12-explastic score of higher than 90 (in small orange dots in Fig. 4). The NLP analysis shows that the small orange dots representing countries with high plastic exports and low NDC targets do not tend to contain content similar to those in the VNR reports.

Although Vietnam seems different, its SDG12-explastic level is 87.5, close to 90. Bahrain, illustrated as a high SDG 12 and low NDC target, demonstrates similar visions for SDGs with Finland and the Netherlands. The VNR reports from Bahrain and Finland appear to commonly emphasize 'water' and 'society', which are relatively relevant to them rather than mitigation, which might have caused the closeness of two different colored and sized dots on the plot.
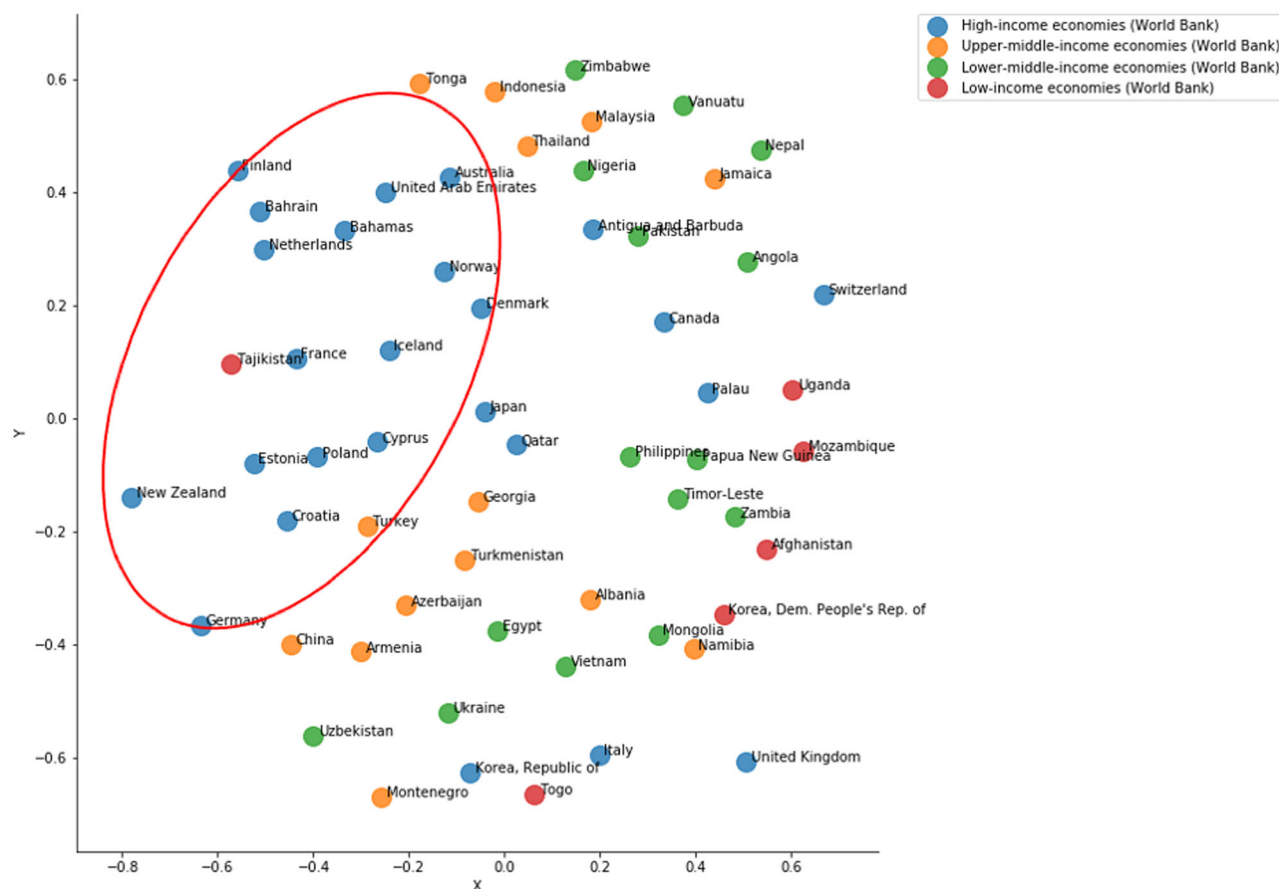
Fig. 2 | Comparative mapping of Voluntary National Reviews (VNR) across countries. The four-color categories represent the World Bank country classifications by income level. The location of each dot is determined by a multidimensional scaling technique that finds the x and y coordinates of individual dots in the reduced dimensional space (two-dimensional, in this case) by preserving distances in the original high dimensional space so that relationships between objects are graphically visualized in a perceptual dimension. Thus, X and Y axes refer to x and y coordinate, respectively, and a shorter distance between objects represent higher similarity. (World Bank country classification by income level can be found from Supplementary Table 2).

Keyword network

Keyword networks are drawn to discover hidden links among keywords and investigate whether the networks could further explain what was observed in previous sections. Focusing on G8 countries that have set ambitious goals for climate actions, we select the Republic of Korea, Italy, and Canada as examples from Asia, Europe and North America, respectively. The networks are visualized in Fig. 5.

In Fig. 5, color code is determined by modularity algorithm, which detects communities by measuring connection density between words. The size indicates the most influential nodes based on betweenness centrality that measures the number of times a node lies on the shortest path between other nodes. The network from the Republic of Korea (Fig. 5a) shows a noticeable difference from the other two countries in that Economy, Environment and Society demonstrate strong centrality in Korea. In comparison, the other two countries contain none of these words as central words. While Korea shows strong centrality of the three words, Italy (Fig. 5b) and Canada (Fig. 5c) show relatively moderate levels of difference in node size. The network from Korea is spread out, while Italy and Canada show more robust interconnection among communities, which may be attributable to a higher level of attraction between nodes in both countries. The attractiveness of one word to another increases when the words are co-occurring. Moreover, as shown in the thickness of the lines, these three words connected with thin lines do not show a strong linkage. Unlike the Republic of Korea, Italy and Canada seem to place emphasis on equal access and inclusion of indigenous communities, respectively. The interconnections around the main keywords indicate

that the associated words share similar contextual features. For instance, Canadian VNR highlights challenges indigenous people face across all 17 SDGs.

Moreover, each community appearing in the same color is linked with SDG goals. The network from Korea indicates relevance with multiple SDGs. The blue-colored nodes with the economy may be relevant to SDG8 (Decent work and economic growth), and the orange-colored nodes imply two SDGs, that is, SDG16 (Inclusive society) and SDG17 (global partnership). In case of Italy, we find SDG4 (Quality education) from the blue dots, SDG5 in purple, SDG8 in light green and blue, and SDG16 in light green, dark green, and purple. In Italy and Korea, multiple communities are covered under SDG16, which may reflect countries' emphasis on SDG16.

The largest node indicates the highest interconnectedness or centrality, as it plays the most influential role in the network by linking two different nodes with the shortest path. The keywords in the most prominent nodes in Italy and Korea reflect the contents of SDG goals and demonstrate links between them, while Canada's most central words indicate beneficiaries. The word 'Indigenous' indicates its crucial role as a mediator connecting keywords found in Canada's VNR report.

Previous studies highlight connections between three fundamental pillars of sustainable development, economy, society and environment to achieve sustainability[47]. Countries studied in this chapter show various approaches to strengthen the connections. Some prioritize the progress of the three main domains, while others incorporate these three to reach final beneficiaries. Accordingly, governments implement different strategies to implement their SDGs.
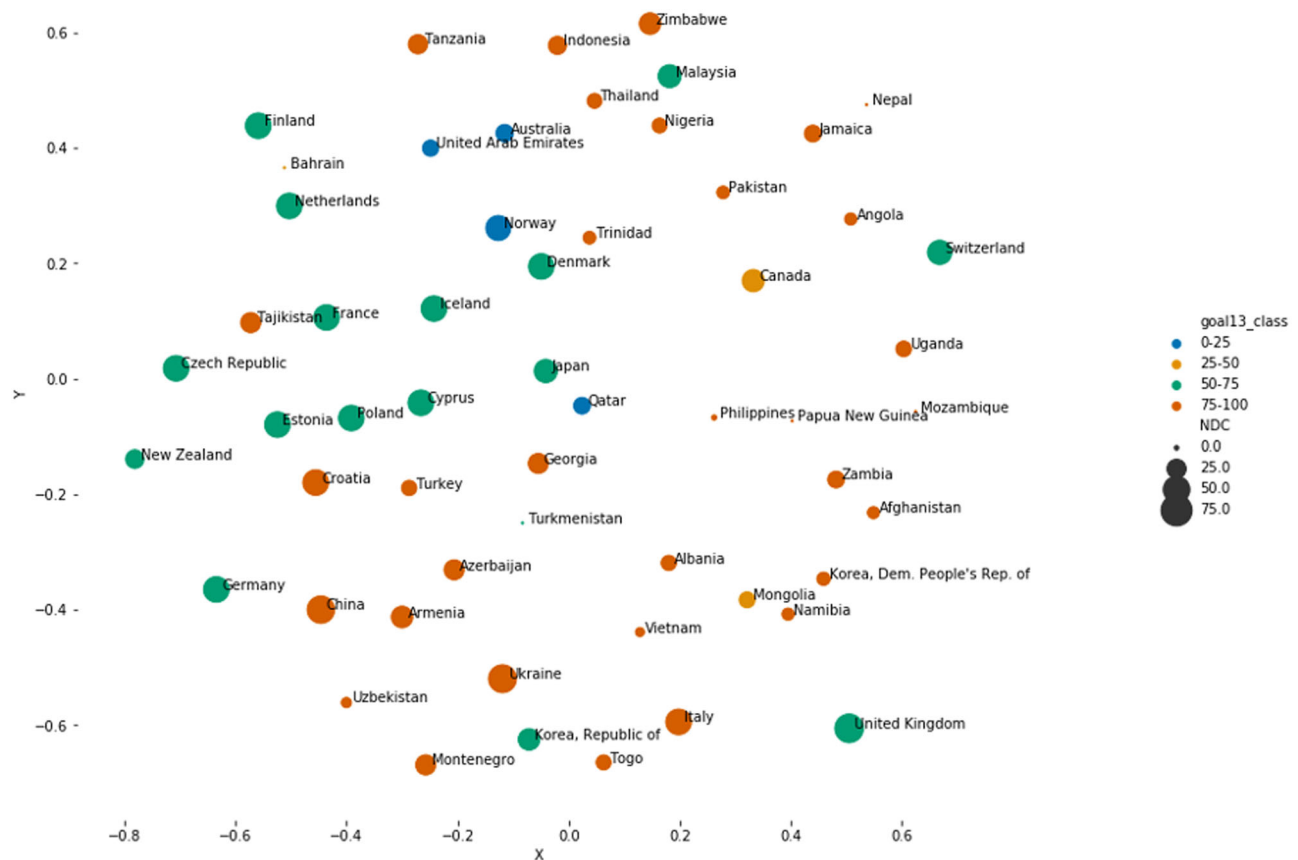
Fig. 3 | Comparative mapping of document similarity, SDG 13 scores and NDC targets across countries (SDG: Sustainable development goals, NDC: nationally determined contributions). The four colors represent the level of normalized-SDG13 score that is segmented into four different classes. The NDC target is also segmented into four different classes, which are represented by the size of dots. The location of each dot is determined by multidimensional scaling technique that finds *x* and *y* coordinates of individual dots in the reduced dimensional space (two dimensional, in this case) by preserving distances in the original high dimensional space so that relationships between objects are graphically visualized in a perceptual dimension. Thus, *X* and *Y* axes refer to *x* and *y* coordinate, respectively, and a shorter distance between objects represent higher similarity while color and size of dots provide additional information on the level of SDG score and NDC.

Discussion

We find that the current score of SDGs may correspond to the level of NDC targets, particularly for a group of countries with low targets. However, the level of ambition of NDC and SDGs' targets seems to be influenced by national economic systems, performance and political leadership. Delivering on the Paris Agreement and SDG agenda depends largely on national resources and national capacities to enforce applicable laws, implement sustainable policies and draw economic benefits. This justifies the main principle of international treaties, underscoring the alignment between national economic growth and sustainable development.

Considering the sharp contradiction between economic growth and environmental protection that would hinder sustainable development, green growth strategies aiming to make the economy green have certainly been at the center of policy-making at the United Nations, G20, European Union, and many other countries[48]. The level of decoupling is varied across countries with different level of wealth, while a few factors such as energy intensity and the level of financial development are commonly found as drivers for decoupling energy intensity and economic development[49,50]. Future studies that identify factors that foster decoupling impact and reveal causal linkages among them are needed to assess the validity of the current policy framework for green growth, which could serve as a basis for future climate policy.

A review of the National Development Plans (NDPs) that embodies national efforts to tackle cross-cutting issues would offer insight into how climate actions are integrated into national development efforts.

A comparison of NDC reports and NDPs indicates that most of the countries are found to connect NDC actions and development plans. Moreover, strengthening the national capacity under NDP is crucial in creating the enabling environment for NDC implementation[51,52]. National capacity to mainstream climate change into national development planning would promote the implementation of climate actions by all relevant actors, enhancing coordination and collaboration across all levels of governance[53].

However, it is also found that the range of sectors included in NDC reports is narrow and does not reflect the spread of national plans[54]. This implies normative and material bias that leads policymakers to limit specific sectors as climate-relevant and reduce the scope of climate actions. Limited capacity of countries and poor coordination across institutions would cause incoherence across relevant policy interventions needed to implement NDCs' commitment[55]. Further analysis using NDCs and NDPs would allow for addressing obstacles that hinder incorporating climate actions into national development plans and developing strategies to strengthen national and institutional capacity to implement NDCs.

Moreover, it would be essential to analyze national interest equally in both mitigation and adaptation. Countries have different approaches to implementing strategies for mitigation and adaptation, which affects contextual priority setting for the national SDGs. Ambition for mitigation is highest for Europe and Central Asia, while Asia and the Pacific are the lowest. On the contrary, adaptation targets are highest for Africa and Arab States, while Europe and Central Asia record the weakest[56]. Ensuring that national targets are aligned with
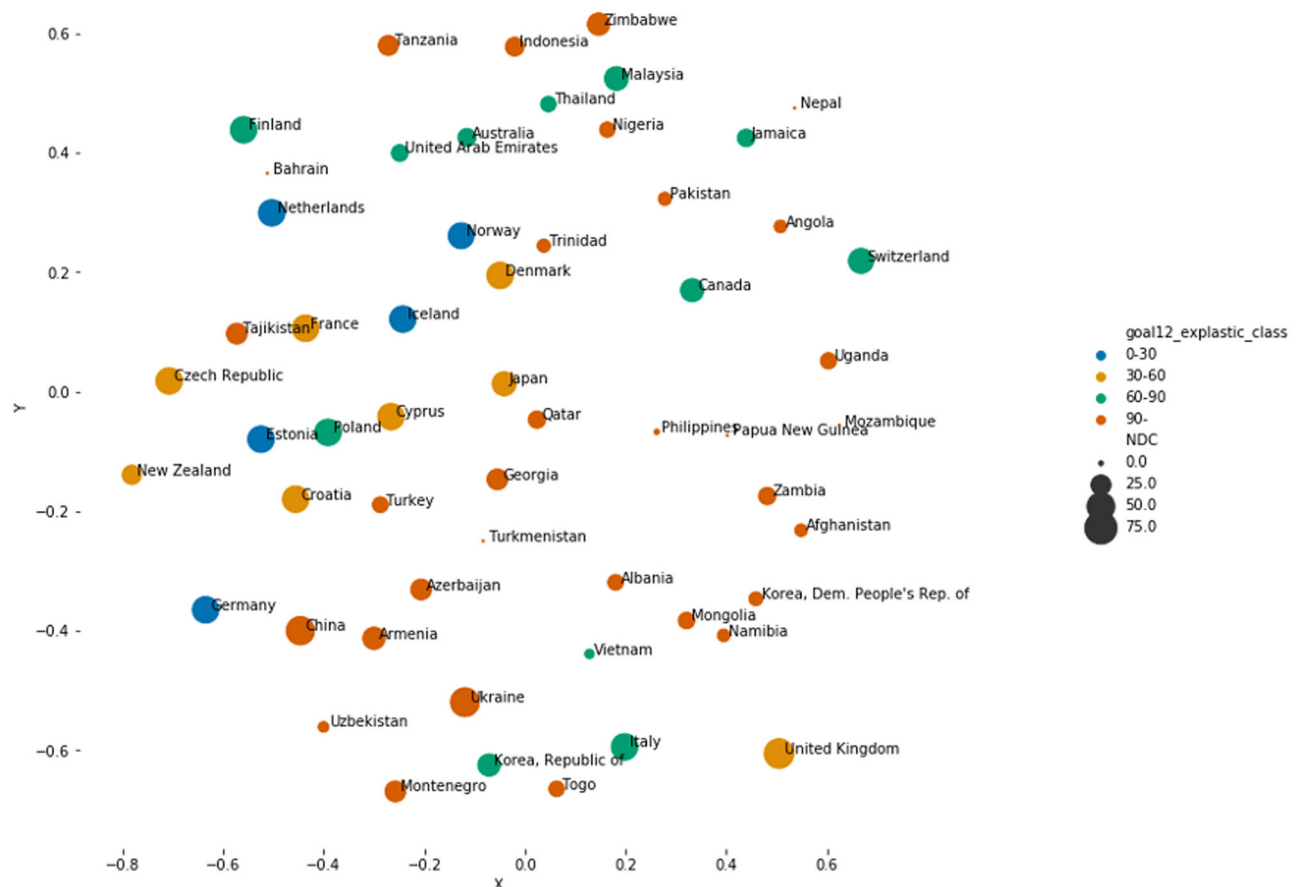
Fig. 4 | Comparative mapping of document similarity, SDG 12 scores and NDC targets across countries (SDG: Sustainable Development Goals, NDC: Nationally Determined Contributions). The four colors represent the level of normalized score for plastic export under SDG12 that is segmented into four different classes. The NDC target is also segmented into four different classes, which are represented by the size of dots. The location of each dot is determined by multidimensional scaling technique that finds x and y coordinates of individual dots in the reduced dimensional space (two dimensional, in this case) by preserving distances in the original high dimensional space so that relationships between objects are graphically visualized in a perceptual dimension. Thus, X and Y axes refer to x and y coordinate, respectively, and a shorter distance between objects represent higher similarity while color and size of dots provide additional information on the level of SDG score and NDC.

SDGs, countries identify SDG priorities and set out a pathway to achieve the identified SDGs. Thus, expanding the coverage, looking at both mitigation and adaptation, would deepen our understanding of the interlinkage of climate actions and sustainable development.

As used in this study, the updated version of NDCs, unlike the first NDCs developed before the Paris Agreement, certainly benefit from a new certainty and additional clarity that the Paris Agreement brought, which forged new stability in the international climate policy-making process[56]. For example, the global stocktake every five years to assess countries' collective progress toward the Paris Agreement's long-term goals would help indicate areas and countries where increased ambitions efforts can be concentrated and supported.
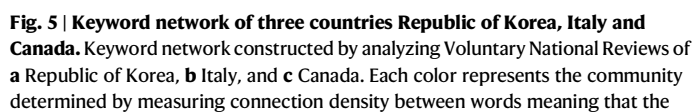
Broadening the scope of a study would provide decision-makers with tangible recommendations on how to concentrate efforts for enhanced action and support for the next round of NDCs. Similarly, applying AI to national development plans and climate actions will provide empirical knowledge to inform decision-makers on how to strengthen the NDC-SDGs interlinkages.

Many countries adopt an integrated approach to aligning NDC and SDG action plans. UNDP identified three categories as entry points critical to fostering NDC and SDG linkages: establishing an enabling environment, planning and delivering on commitments, tracking progress and reporting[57]. SEI developed NDC-SDG interaction tool that lists climate actions mentioned in NDCs related to 17 SDG goals, respectively[58]. Previous studies have applied expert interviews or case studies to identify opportunities and challenges for achieving synergies[27,28,59]. In this work, we propose to apply AI techniques to identify SDG targets closely related to NDC emissions reduction targets, measure similarities of information contained in NDC reports across countries and illustrate relatedness between SDG and NDC.

As a frontier technology, AI has the potential to be used as an enabler to improve the policy-making process for climate change and sustainable development. Detecting patterns, interactions and causalities of policy measures allows policymakers to prioritize particular action plans and integrate them into national development plans. The classification and regression tree (CART) applied in this study facilitates the decision-making process by reducing the dimensionality of an extensive database, classifying critical consequences and prioritizing risk factors to be considered. CART would select the most important climate and sustainable development actions that harness synergistic interactions between NDC and SDGs.

Correlation in data does not necessarily describe causal structure. Machine learning methods have been proven to be efficient in determining correlations in data. However, they have limitations in finding causal relationships[60]. The Granger Causality (GC) test has been widely used to investigate whether a particular period in the past helps forecast future time series. The traditional approach to GC assumes linear dynamics, which may lead to errors in describing real-world situations. New approaches to overcome the inherent challenges in GC tests have been developed, including multilayer perceptron and
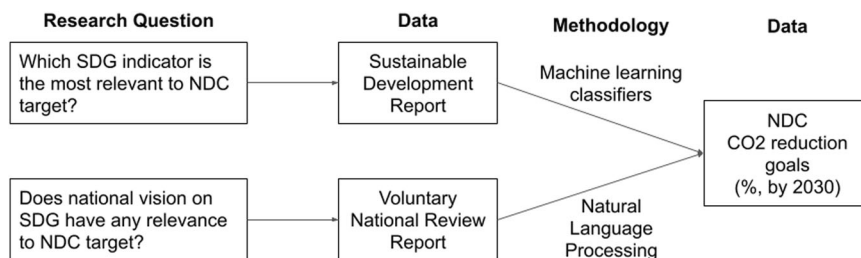
Fig. 5 | Keyword network of three countries Republic of Korea, Italy and Canada. Keyword network constructed by analyzing Voluntary National Reviews of **a** Republic of Korea, **b** Italy, and **c** Canada. Each color represents the community determined by measuring connection density between words meaning that the ones sharing substantial internal connections would create a community. The size of a word indicates the level of centrality and the strength of connection between words is described as the thickness of lines.

Fig. 6 | Overall scheme of methodology (SDG: Sustainable Development Goals, NDC: Nationally Determined Contributions). An integrated methodological framework is designed to answer two research questions that require investigations of various relationships among different types of datasets. The interlink between SDG and NDC is examined by analyzing numeric data (e.g., SDG score, NDC) using Machine learning classifiers, and by quantitatively evaluating relationships between numeric and text data using Natural Language Processing.

recurrent neural networks[61]. Neural networks that consist of several layers of artificial neurons have been applied to improve the accuracy of the prediction model further. This is particularly effective for analyzing images. With an increasing amount of image data that provides geospatial characteristics of climate, energy, and environmental conditions, neural network analysis has the potential to be used in the NDC-SDG policy-making process. Detecting the risk level at local, national, and international levels would provide optimal precautionary measures that minimize global risks and maximize global resilience.

For instance, the image of the river or glacier of specific areas or a map of areas with frequent droughts and floodings across regions over time due to climate change impact would help to predict river geometry, location or the amount of flooding. Applying Convolutional Neural Networks (CNN) for image recognition and analysis of sequential data[62], Wunsch et al. reveal groundwater levels declining until 2100 across regions in Germany due to climate change impact. An advanced version of CNN, U-Net, is deployed[63], which estimates river flooding depth and identifies river shape and wetted areas using the image of ground elevation and flooding discharge.

Despite extensive and expanded use of text and image data that ensure better prediction accuracy, AI's black box nature has been criticized for its lack of interpretability[64,65]. Interpretability is achieved through various techniques, including feature selection used in this study and newly developed techniques, such as eXplainable AI (XAI), that aim to provide human-interpretable models[66]. However, there are unresolved ambiguities over the variety of XAI methods, which would create risks to the public perception[67]. Therefore, the EU proposed the AI Act to regulate the degree of transparency that AI must present before providing conclusions by which the EU aims to bring clarity and transparency regarding the use of AI under human oversight.

Coordinated actions and concrete rules would be required for the concerted use of diverse data and access to AI. At the country level, coordination across departments for data management, including collection, verification, and data collection, is demanded. At the international level, focusing on digital cooperation between countries with AI applications would be invaluable. The High-level Panel on Digital Cooperation established by the United Nations (UN) Secretary General could, for example, be useful in providing the framework for Digital Cooperation in the application of AI to address SDGs and NDCs.

Methods
Data collection
This study comprises three datasets: VNRs, SDG scorecard and NDC carbon reduction targets. 247 VNRs available between 2015 and early 2021 are downloaded from the UN Sustainable Development Knowledge Platform (Supplementary Table 2), of which 67 VNRs are selected, written in English, and published most recently if multiple VNRs are found by one country. The online database for the Sustainable Development Report 2022 (SDR) is downloaded from the SDG scorecard dashboard (Supplementary Table 2). The SDR is prepared by the Sustainable Development Solutions Network (SDSN) to complement the official SDG indicators and support the country-led review process. The SDR tracks the progress of 193 UN member countries in achieving 17 SDG goals and measures the progress towards 17 SDG goals and 120 additional indicators related to SDG targets. 194 NDCs are collected from the UNFCCC website (Supplementary Table 2) and emission target value is downloaded from the website of Institute for Global Environmental Strategies (Supplementary Table 2). To maintain consistency of the analysis, this study uses only the targets expressed in percentage and chooses unconditional NDC targets in case both conditional and unconditional are available.

Methodology
Two techniques, machine learning classifier and NLP, are applied to SDR and VNR, respectively to map them to NDC targets (Fig. 6).

This study selects essential features that significantly influence the outcome to reduce dimensionality and avoid overfitting. Random Forest (RF) is used for precise selection. A random forest consists of multiple decision trees constructed during training time over different subsets of the training data. Referring to the previous studies[68–70], the random forest method is generally proven effective in analyzing the level of importance of various factors in explaining the output. Random forest works well when a large database is analyzed because it handles imperfect data effectively[70]. Compared to a neural network (e.g., ANN), a random forest better explains the mechanics behind the prediction[71]. Each time a tree is split using a randomly sampled training set, every variable is tested to find the feature that reduces the impurity of each tree. Measuring and averaging the extent to which a selected feature reduces the impurity of the split, this method computes the importance of each feature, which is indicated as 'Feature Importance'. The higher the score, the more significant its effect on the model being used to predict a particular variable.

Random Forest is a collection of multiple decision trees in which a bootstrap is created by randomly sampling the training set. Each time a tree is split, every variable is tested to find the feature that reduces the impurity of each tree. Measuring and averaging the extent to which a selected feature reduces the impurity of the split, this method computes the importance of each feature, which is called 'Feature Importance.' The higher the score, the more significant its effect on the model being used to predict a particular variable.

Known as Extremely Randomized Trees, Extra Trees selects the features and splits randomly. Unlike the Random Forest, which uses bootstrap replicas by subsampling the input data, Extra Trees uses the entire original sample. To find a split, Extra Trees randomly chooses the split points and finds the optimum split by selecting the best one among the subset of features. Compared to the Random Forest, Extra Trees is better to reduce both bias and variance. Using the whole original sample instead of a bootstrap replica reduces bias, and choosing randomly the split point reduces variance. In addition, Extra Trees' random selection rather than Random Forest's search for the

optimal split reduces execution time, which makes Extra Trees much faster.

To convert the collected text to a numeric form, this study applies the tf-idf technique (Term Frequency Inverse Document Frequency). Compared to the bag-of-words (BOW) that counts the number of appearances of each word and creates fixed-length vectors, tf-idf considers the number of times a word appears in the document while counterbalancing it by the number of documents where the word appears. Hence, words such as 'the' and 'this' that occur frequently in all documents are not given a high score, whereas a word that is present often only in a few documents is given a high score, as it might indicate the context of the document.

Term frequency is the number of times a word i appears in a document j divided by the total number of words in the document to adjust for the length of the document (Eq. 1).

$$\mathrm{tf}_{i,j} = \frac{n_{i,j}}{\sum n_{i,j}} \tag{1}$$

To measure how frequently or rarely the word appears, the number of documents that contain the term i (df_i) is calculated, which ranges between 1 and the number of all documents. The basis of the Inverse Document Frequency (idf) the probability that a randomly picked document from a corpus of N documents contains the term i (Eq. 2).

$$P_i = \frac{\mathrm{df}_i}{N} \tag{2}$$

The inverse of probability (Eq. 2) is used to calculate the idf (Eq. 3).

$$idf_i = \log\left(\frac{1}{P_i}\right) = \log\left(\frac{N}{\mathrm{df}_i}\right) \tag{3}$$

Multiplying the tf and idf (Eq. 4), the term frequency-inverse document frequency (tf-idf) measure is calculated with which we compare the level of informativeness of a word in a given document.

$$w_{i,j} = \mathrm{tf}_{i,j} \times \log\left(\frac{N}{\mathrm{df}_i}\right) \tag{4}$$

When applying tf-idf vectorization in Python sklearn, multiple steps are conducted to adjust parameters under the tf-idf vectorization. For example, we set a model to identify words (analyzyer='word'), extract unigrams and bigrams of the words (ngram_range = (1,2)), and select a word/n-gram that has a frequency of between minimum and maximum values (max_df:.95, min_df: 10).

The parameter of stop_words removes commonly found words, including 'this,' 'the,' etc. We set stop_words 'None', as tf-idf gives a low weight to those words, and simple removal of listed words might possibly affect the context of the text.

The minimum and maximum document frequency refers to either the proportion of documents (e.g., 0.95 means 95% of all documents) or the number of documents (e.g., 55 means 55 documents out of all documents). Hence, this study removes the words with frequency in the top 5% and the words that appear in less than 10 documents.

Every term in every document is given a tf-idf score, and the documents are represented as a vector of dimension equal to the number of words. The value given to each word indicates its importance in a particular document.

However, the vector values may not be directly understood if they were visualized due to the high number of dimensions. Thus, we apply multidimensional scaling (MDS) to reduce the dimensionality so that each document can be represented in 2 or 3-dimensional plots. MDS is used for non-linear dimensionality reduction through which the

results are better displayed in lower dimensional plots. The main principle is to create a distance matrix and capture the distances between all data points. This study uses Cosine similarity, which measures the cosine of the angle between two vectors in multi-dimensional space to determine similarity among documents or vectors (Eq. 5). We then construct the Cosine similarity-based distance metric and transform the high-dimensional feature space into two dimensions (Eq. 6).

$$\mathrm{similarity} = \cos(\theta) = \frac{A \cdot B}{||A||\,||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\,\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{5}$$

$$\mathrm{Cosine\ distance} = 1 - \mathrm{cosine\ similarity} \tag{6}$$

A keyword network is drawn using word pairs and weights. It shows how words relate to each other and how different groups or clusters are formed. Four essential centrality measures are applied to analyze the position of one data point in a network. First, degree centrality, which is the simplest measure of centrality is degree centrality. It counts how many edges each node has. In other words, the most degree central actor has the most ties. Second, between centrality, which captures which nodes are important in the network's flow. It uses the shortest paths in the network. A path is a series of adjacent nodes. We can find the shortest path between any two nodes, that is, the path with the least number of total steps (or edges). Third measure is modularity, which, for a given division of the network's vertices into some modules, reflects the concentration of edges within modules compared with a random distribution of links between all nodes regardless of modules. The modularity algorithm implemented in Gephi looks for the nodes that are more densely connected together than the rest of the network. The fourth one is the thickness of the edges, which indicates the weights of each word pair. The edge weight is calculated using the total number of co-occurrences of word pairs with distances. For example, the number of occurrences of a pair of word 1 and word 2 with distance 1 is counted and stored. Then, the distance 2 with the same word pair is counted.

This study applies a few additional metrics to compare the similarity of VNR documents in relation to each country's economic, geographical, and environmental features. Countries are classified according to region (Asia, Africa, America, Europe, Oceania) and the World Bank's classification (Supplementary Table 2) of countries by income level (low, lower-middle, upper-middle, and high-income countries).

Data availability
Data underlying the results is available in Figshare (https://figshare.com/s/d9a68059ef887e14c614), and can be downloaded from the web links provided in Supplementary Table 2.

Code availability
The python libraries used in this work are provided in Supplementary Figs. 1–6. Code is available upon request.

References
1. Stockholm Environment Institute. *Connections between the Paris Agreement and the 2030 Agenda*. (2019).
2. United Nations Framework Convention on Climate Change. *Key aspects of the Paris Agreement | UNFCCC*. https://unfccc.int/most-requested/key-aspects-of-the-paris-agreement.
3. UNDP. *Strengthening Resilience To Natural Disasters*. (2020).
4. Mehmood, K., Saifullah, Qiu, X. & Abrar, M. M. Unearthing research trends in emissions and sustainable development: potential implications for future directions. *Gondwana Res* **119**, 227–245 (2023).

5. Crumpler, K., Bloise, M., Meybeck, A., Salvatore, M. & Bernoux, M. Linking nationally determined contributions and the sustainable development goals through agriculture a methodological framework. *FAO* (2019).

6. Zhang, X., Singh, A., Raman, A., Deng, L.United Nations Framework Convention on Climate Change. Catalysing the Implementation of Nationally Determined Contributions in the Context of the 2030 Agenda through South-South Cooperation. *The Routledge Handbook of Global Development* https://doi.org/10.4324/9781003017653-9 (2022).

7. United Nations. *Leveraging Energy Action For Advancing The Sustainable Development Goals*. https://sustainabledevelopment.un.org/contact (2021).

8. Dzebo, A., Janetschek, H., Brandi, C. & Lacobuta, G. *The Sustainable Development Goals viewed through a climate lens*. https://www.sei.org/wp-content/uploads/2018/12/181213b-gill-dzebo-climate-actionsdgs-pb-1811m-web.pdf (2018).

9. Janetschek, H., Brandi, C., Dzebo, A. & Hackmann, B. The 2030 Agenda and the Paris Agreement: voluntary contributions towards thematic policy coherence. **20**, 430–442 (2019).

10. Berger, G. *Country Approaches to the SDGs and Well- being: Overview Survey on National Activities in Europe and Beyond First Report*. (2022).

11. World Economic Forum. Trade, climate and finance: Breaking the silos to drive action. https://www.weforum.org/agenda/2024/01/trade-climate-finance-breaking-silos-ambitious-inclusive-action/ (2024).

12. Oseland, S. E. Breaking silos: can cities break down institutional barriers in climate planning? *J. Environ. Policy Plan.* **21**, 345–357 (2019).

13. United Nations. Breaking 'Silo' approach key in toppling barriers to merging three pillars of sustainable development, speaker tells high-level political forum. https://press.un.org/en/2015/ecosoc6705.doc.htm (2015).

14. Mantlana, K. B., Maoela, M. A. & Nhamo, G. Mapping South Africa's nationally determined contributions to the targets of the sustainable development goals. *Nat. Resour. Forum* **45**, 3–17 (2021).

15. Iacobuță, G. I., Brandi, C., Dzebo, A. & Elizalde Duron, S. D. Aligning climate and sustainable development finance through an SDG lens. The role of development assistance in implementing the Paris Agreement. *Global Environmental Change*. **74**, (2022).

16. Aldoseri, A., Al-Khalifa, K. N. & Hamouda, A. M. Re-thinking data strategy and integration for artificial intelligence: concepts, opportunities, and challenges. *Appl. Sci.* **13**, 7082 (2023).

17. Organisation for Economic Co-operation and Development. Artificial Intelligence, Machine Learning and Big Data in Finance. *OECD Paris* 1–72 (2021).

18. Marginean, F. et al. An Artificial Intelligence-based Support Tool for Automation and Standardisation of Gleason Grading in Prostate Biopsies. *Eur. Urol. Focus* https://doi.org/10.1016/j.euf.2020.11.001 (2020).

19. Jha, K., Doshi, A., Patel, P. & Shah, M. A comprehensive review on automation in agriculture using artificial intelligence. *Artif. Intell. Agric.* https://doi.org/10.1016/j.aiia.2019.05.004 (2019).

20. Alowais, S. A. et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med. Educ.* **23**, 1–15 (2023).

21. Zhou, W. et al. Ensembled deep learning model outperforms human experts in diagnosing biliary atresia from sonographic gallbladder images. *Nat. Commun.* **12**, 1–14 (2021).

22. Grossmann, I. et al. AI and the transformation of social science research. *Science* **380**, 1108–1109 (2023).

23. Joshi, R., Gupte, R. & Saravanan, P. A random forest approach for predicting online buying behavior of Indian customers. *Theor. Econ. Lett.* **08**, 448–475 (2018).

24. Grajzl, P. & Murrell, P. Toward understanding 17th century English culture: a structural topic model of Francis Bacon's ideas. *J. Comp. Econ.* **47**, 111–135 (2019).

25. Utibe Monday, T. Impacts of interview as research instrument of data collection in social sciences. *J. Digit. Art. Humanit.* **1**, 15–24 (2020).

26. Hussein, H. Interview Method. *Princ. Soc. Res. Methodol.* 207–219 https://doi.org/10.1007/978-981-19-5441-2_14 (2022).

27. Vaio, A. D. & Khalid, A. M. Creating synergies among the sustainable development goals and climate action: insights from a developing economy. *Sustain* **15**, 13137 (2023).

28. Stevenson, S. et al. A hybrid approach to identifying and assessing interactions between climate action (SDG13) policies and a range of SDGs in a UK context. *Discov. Sustain.* **2**, 1–18 (2021).

29. van Soest, H. L. et al. Analysing interactions among Sustainable Development Goals with integrated assessment models. *Glob. Transit.* **1**, 210–225 (2019).

30. Sebestyén, V., Domokos, E. & Abonyi, J. Focal points for sustainable development strategies—Text mining-based comparative analysis of voluntary national reviews. *Journal of Environmental Management* (2020).

31. Hsu, A., Khoo, W., Goyal, N. & Wainstein, M. Next-generation digital ecosystem for climate data mining and knowledge discovery: a review of digital data collection technologies. *Front. Big Data* **3**, 525690 (2020).

32. Big shifts, small steps: The KPMG. Survey of Sustainability Reporting 2022. *KPMG* (2022).

33. Muttarak, R. & Lutz, W. Is education a key to reducing vulnerability to natural disasters and, hence, unavoidable climate change? *Ecol. Soc.* **19**, (2014).

34. Danish Ministry of the Environment. *5th Danish Country Report To the Convention on Biological Diversity*. (2014).

35. IISD. *Nature-Based Solutions as a Catalyst for Achieving Mutual Benefits for People, Nature, and Climate: Lessons learned from China and globally*. (2021).

36. Košir, L. G., Korpar, N., Potočnik, J. & Kocjančič, R. *Roadmap towards the Circular Economy in Slovenia*. (2018).

37. Coenen, J., Glass, L.-M. & Sanderink, L. Two degrees and the SDGs: a network analysis of the interlinkages between transnational climate actions and the Sustainable Development Goals. **17**, 1489–1510 (2022).

38. UNDP. *Circular Economy Principles for NDCs and Long-term Strategies*. (2019).

39. Hansen, J. et al. Impact pathways from climate services to SDG2 ("zero hunger"): a synthesis of evidence. *Clim. Risk Manag.* **35**, 100399 (2022).

40. Food and Agriculture Organization of the United Nations. *Climate change and food security: risks and responses*. (2015).

41. Cruz, J., White, P. C. L., Bell, A. & Coventry, P. A. Effect of extreme weather events on mental health: a narrative synthesis and meta-analysis for the UK. *Int. J. Environ. Res. Public Health* **17**, 1–17 (2020).

42. Braspenning Radu, O. et al. Exploring synergies between climate and air quality policies using long-term global and regional emission scenarios. *Atmos. Environ.* **140**, 577–591 (2016).

43. Parliament, E. *Achieving the UN Agenda 2030: Overall actions for the successful implementation of the Sustainable Development Goals before and after the 2030 deadline*. (2023).

44. Brandi, C., Dzebo, A. & Janetschek, H. *The Case for Connecting the Implementation of the Paris Climate Agreement and the 2030 Agenda for Sustainable Development*. vol. 21 (2017).

45. European Commission. EU whole-of-government' approach. https://commission.europa.eu/strategy-and-policy/sustainable-development-goals/eu-whole-government-approach_en.

46. Koundouri, P., Devves, S., Plataniotis, A. & Koundouri, P. Alignment of the European green deal, the sustainable development goals and

the European semester process: method and application keywords. *Theor. Econ. Lett.* **11**, 743–770 (2030).

47. Mangukiya, R. D. & Sklarew, D. M. Analyzing three pillars of sustainable development goals at sub-national scales within the USA. *World Dev. Sustain.* **2**, 100058 (2023).

48. Parrique, T. et al. *Decoupling debunked: Evidence and arguments against green growth as a sole strategy for sustainability.* European Environmental Bureau www.eeb.org (2019).

49. Wang, Q. & Su, M. Drivers of decoupling economic growth from carbon emission – an empirical analysis of 192 countries using decoupling model and decomposition method. *Environ. Impact Assess. Rev.* **81**, 106356 (2020).

50. Addai, K., Serener, B. & Kirikkaleli, D. Can environmental sustainability be decoupled from economic growth? empirical evidence from Eastern Europe using the common correlated effect mean group test. *Reg. Sustain.* **4**, 68–80 (2023).

51. UNDP. *Mainstreaming into National Planning Processes.* vol. 31 (2016).

52. UNFCCC. *National Capacity in the Least Developed Countries (LDCs) for Accelerated Climate Action and Launch of the Capacity Building Program.* (2023).

53. UNEP & UNDP. *NDC enhancement: lessons learned in capacity needs, challenges and opportunities. Online Event Summary Report.* https://unfccc.int/sites/default/files/resource/Summary report - NDC enhancement lessons learned in capacity needs%2C challenges and opportunities.pdf (2020).

54. Abanda, F. H. et al. Catalysing the monitoring and evaluation of Nationally Determined Contributions through North–South cooperation. *Environ. Dev. Sustain.* 1–24 https://doi.org/10.1007/S10668-023-03092-5/FIGURES/6 (2023).

55. Dombrowsky, I. et al. Effects of policy and functional (in)coherence on coordination – a comparative analysis of cross-sectoral water management problems. *Environ. Sci. Policy* **131**, 118–127 (2022).

56. UNDP. *UNDP Climate Promise Progress Report.* (2022).

57. UNDP. *Aligning Nationally Determined Contributions And Sustainable Development Goals.* https://www.wri.org/sites/default/files/WRI_INDCs_v5.pdf.

58. NDC-SDG Connections - SEI. https://www.sei.org/projects-and-tools/tools/ndc-sdg-connections/.

59. Enevoldsen, P., Baum, C. M., Low, S. & Sovacool, B. K. Examining the synergies and tradeoffs of net-zero climate protection with the Sustainable Development Goals. *Science Progress.* **105**, (2022).

60. Lecca, P. Machine learning for causal inference in biological networks: perspectives of this challenge. *Front. Bioinforma.* **1**, 746712 (2021).

61. Tank, A., Covert, I., Foti, N., Shojaie, A. & Fox, E. Neural granger causality. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 4267–4279 (2018).

62. Wunsch, A., Liesch, T. & Broda, S. Deep learning shows declining groundwater levels in Germany until 2100 due to climate change. *Nat. Commun.* **13**, 1–13 (2022).

63. Hosseiny, H. A deep learning model for predicting river flood depth and extent. *Environ. Model. Softw.* **145**, 105186 (2021).

64. Bathaee, Y. The artificial intelligence black box and the failure of intent and causation. *Harvard Journal of Law & Technology.* **31**, (2018).

65. Petch, J., Di, S. & Nelson, W. Opening the black box: the promise and limitations of explainable machine learning in cardiology. *Can. J. Cardiol.* **38**, 204–213 (2022).

66. Ali, S. et al. Explainable Artificial Intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. *Inf. Fusion* **99**, 101805 (2023).

67. Nannini, L., Balayn, A. & Smith, A. L. Explainability in AI policies: a critical review of communications, reports, regulations, and standards in the EU, US, and UK. *ACM Int. Conf. Proceeding Ser.* 1198–1212 https://doi.org/10.1145/3593013.3594074 (2023).

68. Sylvester, E. V. A. et al. Applications of random forest feature selection for fine-scale genetic population assignment. *Evol. Appl.* **11**, 153–165 (2018).

69. Lima, M. S. M. & Delen, D. Predicting and explaining corruption across countries: a machine learning approach. *Gov. Inf. Q.* **37**, 101407 (2020).

70. Jalal, N., Mehmood, A., Choi, G. S. & Ashraf, I. A novel improved random forest for text classification using feature ranking and optimal number of trees. *J. King Saud Univ. Inf. Sci.* **34**, 2733–2742 (2022).

71. Luan, J., Zhang, C., Xu, B., Xue, Y. & Ren, Y. The predictive performances of random forest models with limited sample size and different species traits. *Fish. Res.* **227**, 105534 (2020).

Acknowledgements

Author contributions

H.C. designed the research, performed all analyzes, interpreted the results, prepared the figures, and wrote and revised the paper. E.A. contributed to the writing of policy implications in the discussion section and provided expertise to revise the paper. H.C. and E.A. participated in the reviewing and editing of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53956-1.

Correspondence and requests for materials should be addressed to Haein Cho.

Peer review information *Nature Communications* thanks Anastasios Zafeiropoulos, Khanyisa B. Mantlana and Mariana Reis Maria for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

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