


REVIEW

Open Access



# Recently emerging trends in big data analytic methods for modeling and combating climate change effects

Anayo Chukwu Ikegwu<sup>1\*</sup> , Henry Friday Nweke<sup>2</sup>, Emmanuel Mkpojiogu<sup>1</sup>, Chioma Virginia Anikwe<sup>3</sup>, Sylvester Agbo Igwe<sup>4</sup> and Uzoma Rita Alo<sup>3</sup>

\*Correspondence:  
anayoikegwuchukwu@gmail.com

<sup>1</sup> Software Engineering Department, Veritas University, Abuja, Nigeria

<sup>2</sup> Computer Science Department, David Umahi Federal University of Health Sciences, Uburu, Ebonyi State, Nigeria

<sup>3</sup> Computer Science Department, Alex Ekwueme Federal University, Ndufu-Alike, Ebonyi, Nigeria

<sup>4</sup> Computer Science Department, Ebonyi State University, Abakaliki, Ebonyi State, Nigeria

## Abstract

Big climate change data have become a pressing issue that organizations face with methods to analyze data generated from various data types. Moreover, storage, processing, and analysis of data generated from climate change activities are becoming very massive, and are challenging for the current algorithms to handle. Therefore, big data analytics methods are designed for significantly large amounts of data required to enhance seasonal change monitoring and understand and ascertain the health risks of climate change. In addition, analysis of climate change data would improve the allocation, and utilisation of natural resources. This paper provides an extensive discussion of big data analytic methods for climate data analysis and investigates how climate change and sustainability issues can be analyzed through these approaches. We further present the big data analytic methods, strengths, and weaknesses, and the essence of analyzing big climate change using these methods. The common datasets, implementation frameworks for climate change modeling, and future research directions were also presented to enhance the clarity of these compelling climate change analysis challenges. This big data analytics method is well-timed to solve the inherent issues of data analysis and easy realization of sustainable development goals.

**Keywords:** Big data, Big data analytics, Climate change, Methods, Data mining, Review

## Introduction

Climate change and sustainable development have become important research areas due to their application in various facets of our economic life. In this present time, climate change (CC) has become a pressing issue that governments, organizations, and researchers are seriously working on the modalities to analyze data generated from vast climate change and sustainability-related activities. Climate change is considered one of the global threats just like the coronavirus (COVID-19), a pandemic that has caused over 6.9 million deaths as of the 12th of December 2023. Hence, both phenomena constitute a big threat to global existence. The contagious disease novel coronavirus (Tao et al. 2020; Sharma and Gupta 2021) has deeply impacted the national

economic development negatively in recent times. Presently, big data analytics methods have successfully assisted in analyzing large climatic data, and contagious disease data, and have recorded impressive results. Therefore, climate change could be best described as a change in the earth's climates, at local, regional, or global scales, and attributed greatly to the increased levels of atmospheric carbon dioxide (CO<sub>2</sub>) produced by the use of fossil fuels (Sebestyén et al. 2021). It is further explained as the changes in the earth's climate driven primarily by human activity since the pre-industrial period, particularly the burning of fossil fuels, and removal of forests, resulting in a relatively rapid increase in CO<sub>2</sub> concentration in the earth's atmosphere.

Climate change as one of the sustainable development goals (SDGs) (Huan et al. 2021) has been transformed and enhanced through big data analytics (BDA). The high volumes of data generated from vast climate change environments such as satellites, fossil fuels, earth's orbit, climate simulation, etc., can be regarded collectively as big data. The process of examining and gaining knowledge of large significant climate change data to reveal hidden patterns, extract useful data, and the correlation between them to enable decision support is termed big data analytics (Abdullah et al. 2020; Ikegwu et al. 2021). BDA methods offer storage support, decision support, processing, and proper strategy against climate change effects and improve people's resilience in the face of the adverse effects of climate change through trends, patterns, prediction, and technical analysis using generated data.

However, data analysis, uncertainty in data storage, and processing related to climate change have become big threats to human and wildlife inhabitation. Quite several research studies have been carried out on big data analytics for big climate change data processing. Nonetheless, our findings clearly show, to the best of our knowledge, that none of these studies has comprehensively and specifically addressed, in detail, the peculiar BDA methods for analysing large climate change. For instance, an earlier study by Hassani and Huang (2019) reviewed big data and climate change, and the researchers pinpointed the values of big data and its applications in climate change-related studies in recent years. However, useful tools and BDA methods to tackle the diverse large climate change data were not discussed. Another related recent study (Ornella 2020) presented a survey on "why nature won't save us from climate change but technology will". This study highlighted the need to approach climate change issues with technologies where the author refers to carbon capture, storage mechanisms, etc. Nonetheless, the technological methods specifically dedicated to addressing storage, data uncertainty, and processing were left out. A more recent study by Sebestyén et al. (2021) reviewed the applications of data science, big data, and the importance of big data tools in climate change. This, however, does not cover the big data and big data analytics methods to solve the peculiar issues inherent in climate change to attain sustainable development goals (SDGs) (Huan et al. 2021; Türkel 2020).

Consequently, this review is a well-timed investigation of big data analytics methods for combating climate change for sustainable development. Based on the available studies in the literature, there are no specific reviews that provide important discussion on these all-encompassing methods-based BDA approaches. As such, the contributions of this investigation to the current body of knowledge in big data, big data

analytics, and climate change are presented in Fig. 1. The specific contributions of this paper to the body of knowledge are as follows:

- To provide extensive discussion of big data analytics methods for climate change effects, and describe how climate change issues can be analysed and resolved;
- To present the data type and source of big climatic data for climate change analysis;
- To comprehensively explore the prospective analytics methods to handle vast climate change data including their strengths and weaknesses;
- To highlight the purpose of big data analytics methods in climate change and sustainable development.

The remainder of this paper is organized as follows: “The impact of climate changes” section presents the impact of climate change. “The conventional methods” section discusses the methods for modeling and combating climate change. “The big data analytic methods” section presents the big data analytics methods. “Purpose for analysing big climate data using big data analytic methods” section discusses the purpose of analysing big climatic data using big data analytic methods. The common datasets for modeling and combating climate change effects are presented in “The common datasets for modeling and combating climate change effects” section. Implementation frameworks for climate change modeling using big data analytics are discussed in “Implementation frameworks for climate change modeling using big data analytics” section. Finally, “Open research directions” section provides the conclusions and future research directions.

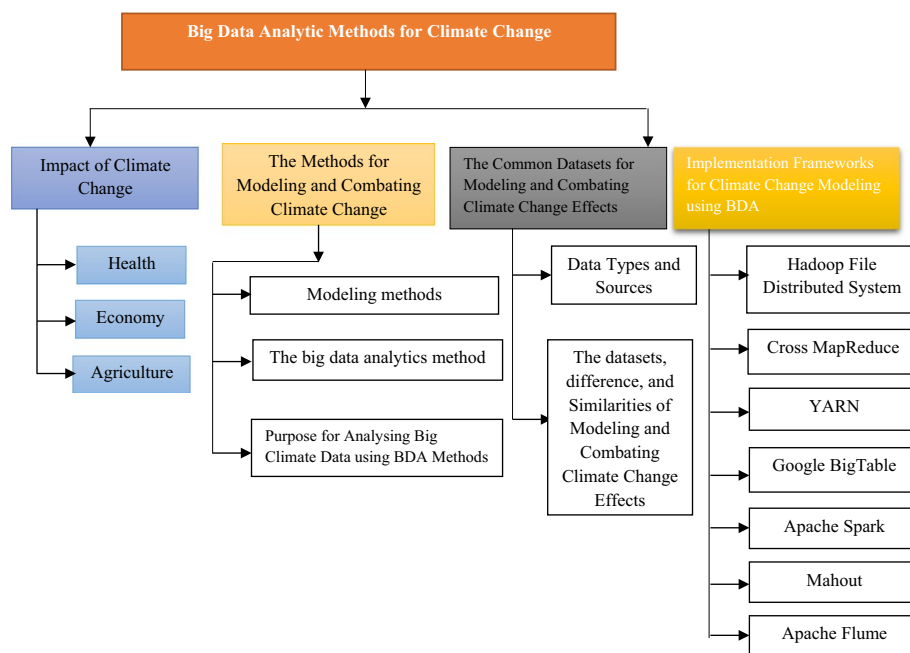


Fig. 1 Taxonomy of big data analytics methods for climate change

## The impact of climate changes

Climate change has impacted various areas of human endeavor and the ecosystem. The atmospheric Carbon Dioxide (CO<sub>2</sub>) level is increasing with speed due to the industrial revolution that has cut all edges of life. The pressing component areas that have been impacted widely include health, economy, and agriculture. The impact of climate change is represented in Fig. 2.

### Health

Climate change has revolved around human health. It is no longer a gainsaying that climate change is affecting healthy living. Due to global warming is a result of a consistent increase in levels of atmospheric carbon dioxide (CO<sub>2</sub>) generated from fossil fuels (Sebestyén et al. 2021) (CO<sub>2</sub>). These continue to cause threats and cause havoc on human life and its surroundings. For instance, the recent study by Habibullah et al. (2022) uses global data of species such as fishes, birds, amphibians, mammals, plants, reptiles, etc. from 115 countries to investigate the climate change impact on biodiversity loss. The results show three variables (temperature, precipitation, and natural disaster occurrence to increase the loss of biodiversity. Another related study by Vicedo-Cabrera et al. (2021) presented heat-related human health impacts using empirical data from 732 locations in 43 countries to estimate the rate of mortality burdens. This, however, shows how climate change has affected negatively human health. The 37.0% (range 20.5–76.3%) of heat-related deaths is as a result of anthropogenic climate change. Nevertheless, there are other factors such as government and public attitude toward climate change that were not considered. Other effects of climate change on human health include infectious diseases, air pollution, heat waves, etc. as seen in the case of the Karachi, Pakistan location, however, the weather instability and environmental policies are not addressed (Babar et al. 2021). Furthermore, psychologically, climate change causes anxiety and mental health as experienced by functional and cognitive-emotional impairment including a symptom of Major Depressive Disorder (MDD) and Generalized Anxiety Disorder (GAD) as shown from uncommon empirical research (Schwartz et al. 2023). However,



**Fig. 2** The impact of climate change

this research is based on associations between constructs and leaving directionality constructs unattained. In a related study, it was portrayed the view and obviously that climate change causes both emotional and psychological trauma even in young people between the ages of 16- and 25 years in 10 different countries surveyed (Clayton et al. 2023). The awakening clarion call on the implication of this effect of climatic change perception was deeply emphasized for collaborative action to be taken by the educators, stakeholders, international communities, etc. (Ogunbode 2025).

### **Economy**

The burden of climate change on the economy and general business activities is posing a serious concern. Climate change has progressively affected the small open economy as assessed using the Keynesian dynamic stochastic general equilibrium model (Economides and Xepapadeas 2019). It also affects the macroeconomy as reviewed by the central banks' monetary policy assessment of the inflation outlook (Andersson et al. 2020). Hence, this calls for technological innovation, and fiscal revenue generation as energy efficiency increases and the price of renewable energy falls. Climate change has impacted different countries such as Zambia, Pakistan, Nigeria, Bangladesh, Ethiopia, etc. are faced with high resilient intensity of atmospheric weather and anthropogenic climate change which is currently becoming a global threat. Our environment, energy usage, economic growth, and gross domestic product (GDP) are being impacted negatively. Many research efforts and global bodies are bent on providing strategies, policies, tools, etc. to ameliorate these trending impacts. For instance, the global warming of 1.5 °C presents unresolved policy issues, hence, a measure was provided by Semieniuk et al. (2021) to limit climate change by reducing energy demand, though, still practically impossible due to large energy savings in a growing global economy and industrial production. Also, the adverse effect of climate change has been causing a lot of damage to economic viability. To this concern, Bhopal et al. (2021) utilized the climate-resilient green economy (CRGE) approach in 2011 as a low-carbon development strategy in Ethiopia. Government. This is also to reduce the vulnerabilities both in health and the economy, however, the persistent nature of climate change still imposes threats. More so, a recent study by Hossain et al. (2022) saw a dynamic computable general equilibrium model to analyse the outcome of climate change on Bangladesh's economy. The model identifies a decrease of -6.17% and investment declines by -7.7% of the average gross domestic product (GDP) of Bangladesh as a result of climate change impact. Additionally, economic and biophysical uncertainties have been drastically caused by the effect of climate change. The study by Deangelo et al. (2023) attempted to utilize Seaweed growth and techno-economic models to address the costs of global seaweed production in Net-zero greenhouse gas. Farming seaweed at a relative scale to match the global carbon budget is quite challenging. However, this requires more digital models such as Artificial Intelligence (AI) and other sophisticated tools to reduce economic and biophysical problems. Similarly, the impact of climate change on economic growth is denigrating due to global warming. For instance, Tol (2024) the outcome of the research about a meta-analysis of the overall climatic change impact on economics close to the growth impact estimated as a function of weather shocks. Hence, the social cost of carbon reviews a close pattern to the overall impact estimates. The implication of social

cost beckoned on the emissions issues such as parameterization, the rate, and extent of warming, and the aggregation of impacts among the people and issues over some time. Nonetheless, more enlightenment on the impacts of moderate warming in the short and medium term should be considered. More recent studies on the climate change impact on the economy are contained in Anh et al. (2023), Javadi et al. (2023).

### **Agriculture**

Food insecurity and scarcity are seriously clocking towards the widening of climate change. The world's climate system has gradually over time changed as a result of human activities. Hence, the changes have affected natural ecosystems and the general standard of human beings. Despite strategies and planning to address this rapid occurrence, there is still an impact on agricultural food production negatively. Agriculture has played a vital role in human sustainability and the economic gross domestic product (GDP) of every nation as no nation can survive without food. The survival of agriculture, the natural ecosystem, and the nation's capacity building depends on policies, strategies, and campaigns against harmful greenhouse gas emissions. The inherent problem of meeting the continuously increasing food demand for the growing population will be further ignited by climate change across the world. Nonetheless, researchers are working around the clock to ensure finding optimum solutions. For instance, climate change has disrupted agriculture and reduced agricultural total factor productivity (TFP) by 34% as a result of high temperatures, drought, heat waves flood strikes, etc. (Fuglie 2021). These uncertainties continue to trail agriculture as farmers in Iran lamented as the changes in precipitation, temperature, and CO<sub>2</sub> fertilization increase (Karimi et al. 2018). In addition, climate change is causing agricultural food vulnerability as reported by farmers in Norte de Santander, Colombia (Núñez et al. 2018) and ASEAN agricultural sectors (Handayani and Abubakar 2020). Similarly, climate change has an impact adversely on crops (e.g. maize, rice, cotton, sugarcane, bajra, gram, etc. due to the CO<sub>2</sub> high emission observed in Pakistan, Nigeria, and at globe (Rehman et al. 2022). Additionally, Ngoma et al. (2021) analysed the climate change on household welfare and agriculture in Zambia. This was achieved using a chain of modeling analytic tools called systematic assessment of climate-resilient development framework (SACRED) to address the uncertainty trailed by the climate challenges. Hence, adaptation interventions are sorted to sustain the future of smallholder agriculture across the globe. More so, the SACRED method was utilised to provide a consistent framework for data analysis to assess temperature, precipitation, and evaporation impacts of different climate scenarios based on dynamic computable general equilibrium models in Ethiopia and Mozambique (Manuel et al. 2021; Solomon et al. 2021). Furthermore, more investigation was carried to ascertain the impact of climate change on agriculture. Anh et al. (2023) investigated climate change based on short and long-term effects on Vietnam's agriculture, at the macro stage with reference to production and values. This study utilized Autoregressive Distributed Lag (ARDL) model. The results shows that it does unfavorably affects the global food security in Vietnam and beyond, thereby, increasing poverty alleviation and sustainable development. Also, assessment carried out in Central India, particularly in Vidarbha region shows that 70 farmers from the study areas who engaged in cotton cultivation were marginalized due to climate change perception. Following perception from the designed

questionnaire, climate change has adverse impact on crop sowing, growth, post-harvest, livestock, etc. Therefore, in all cases, instances, and scenarios reviewed, it is quite understandable that climate change has greater negative effective on agricultural practice and sustainable food security development. This was a rejoinder as it is echoed and predicted at different degrees and magnitudes over the Sri Lanka and nations at large to be global warming and the resultant climatic change by the year 2025 (Nanda and Perera 2019).

### The methods for modeling and combating climate change

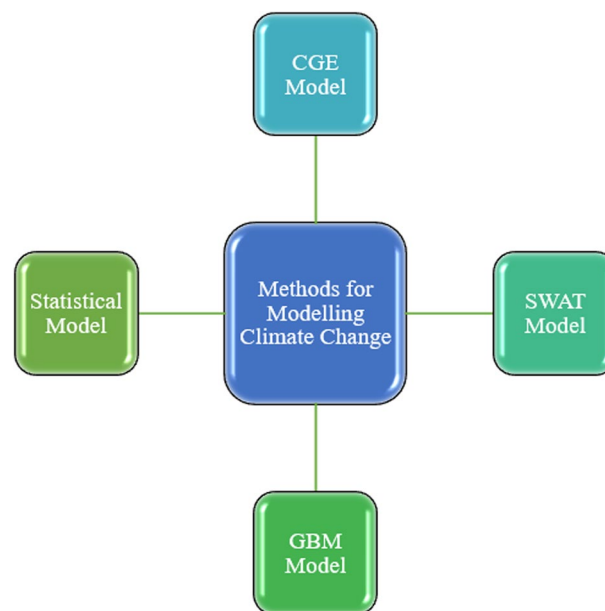
In this section, the conventional methods and big data analytic methods are discussed. Figure 3 shows the pictorial view of the methods for modeling and combating climate change. Also, the strength and weaknesses of this method are presented in Table 1.

#### The conventional methods

There are different conventions or traditional methods for modeling and combating climate change. The most effective modeling approaches include the computable general equilibrium (CGE) model, statistical model, gateway belief model (GBM), soil and water assessment tool model (SWAT), etc. These are succinctly discussed in the subsection.

#### Computable general equilibrium (CGE)

The computable general equilibrium model (Hossain et al. 2022) is a standard tool for empirical analysis and is dedicated to economic analysis world-wide. It captures the economy's supply and demand side and allows for adjustments (e.g. quantities and price) based on policy shock. The CGE could be a static or dynamic computable general equilibrium (DCGE) model. The static model is best utilized when the specific economic period is known. The DCGE model allows definite features of short and long-term



**Fig. 3** The methods for modelling and combating climate change

**Table 1** Strengths and weaknesses of the methods for modeling climate change

Methods	Description	Strength	Weakness	References
Machine learning	AI method that analyzes, informs meaning, and predicts the outcome of large climate change data	Offer fast data analysis and prediction, and improve clean energy access. It offers decision-support	Requires a lot of training data, time-consuming in learning	Roinick et al. (2019, 2022), Saba et al. (2020), Abbot and Marohasy (2017)
Deep learning	Is an ML method involving a nonlinear approach for simulation and prediction in both mining and diagnosis of large climate data	Easy discovery of climate patterns analysis, well-performed for simulation and prediction of climate change	It involves overfitting in neural networks, and hyper-parameter optimization, and is expensive to train	Huan et al. (2021), Zhang and Li (2020), Zhu et al. (2017), Kurth et al. (2019)
Artificial intelligence	Analytics method responsible for supporting simulation and decision-making infused into earth observation data and simulation climate data	Support simulation and quick decision-making. Further helps in climate change discovery	Has a high cost of implementation. Also, have data acquisition issues	Sebestyén et al. (2021), Kadow et al. (2020), Huntingford et al. (2019)
CGE	Technique for empirical and economic analysis worldwide. It captures the economy's supply and demand side and allows for adjustments based on policy shock	It allows definite features of short and long-term policy shocks to be determined. It detects any exogenous variable that changes any quantity. Also, flexible	It dwells only on economic policies and shocks and lacks computational libraries for dynamic variable data analysis	Hossain et al. (2022), Manuel et al. (2021), Solomon et al. (2021), Shahpari et al. (2021), Jensen et al. (2021)
Statistical	A key model deployed in analysing and predicting climatic conditions with regard to simple regression up to non-parametric spatiotemporal Bayesian models	It has structural simplicity and effective data modeling and analysis to foster past climate data for prediction	Limited in modeling and analyzing big climate data	Sweeney et al. (2018), Yariyan et al. (2020), Siddiqua et al. (2021), Worku et al. (2020), Jackson et al. (2018)
GBM	A dual process theory of attitude change supporting mixed perception of people oppose the scientific judgment that climate change is caused by humans	Offers theory of attitude change based on the political and personal perception that is good data modeling and implementation	Only based on theory and not analytics process of data modeling and analysis	Van Der Linden (2021), Kerr and Wilson (2018)
SWAT	A hydrological modeling tool mounted to hydrologic and environmental simulations	Good for hydrologic and environmental simulations of sensing data	It doesn't fit in the advanced analysis and prediction of climate change	Akoko et al. (2021), Li and Fang (2021), Aznarez et al. (2014), Alltane et al. (2022), Kalb et al. (2021)



policy shocks to be determined. CGE is essential and globally used to capture direct, and indirect, represent the economy, and explicitly maintain consistency at the macro and micro levels (Shahpari et al. 2021). The CGE also detects any exogenous variable that changes any quantity of the economy and gives rise to consequences throughout the system. More so, it helps the data to be numerically calibrated to their algebraic framework and mathematical programming that captures the outcome of the basic features of the dynamic model. In addition, the essence of this CGE mode is to explain the function of prices in an economy. The strength bestows in their flexibility since they can adapt to a wide range of policies and shocks. Furthermore, CGE model has three major components; data, theory, and shocks. These are combined to bring effective results. The results depend on numerical predictions of the changes in the economic system. For instance, a recent study by Hossain et al. (2022) utilised a dynamic computable general equilibrium model to analyse the impact of climate change on Bangladesh's economy. The result shows that Bangladesh's GDP is affected by  $-6.17\%$  and even declines by  $-7.76\%$  in investment. Further predictions show a reduction of  $-3.08$  and  $-3.7\%$  collectively in 2030 and 2050, especially in the rice sectors. Also, Jensen et al. (2021) applied the CGE model to analyse seven current and future climate change challenges for Myanmar, state and region-specific paddy yield change of 30-year intervals between the 2020s, 2050s, and 2080s. This application produced effective results to point out strong differences in climatic change effects in the nation's states and regions on agricultural production. Furthermore, DCGE was deployed in the analysis of Mozambique's agricultural sector and household welfare (Manuel et al. 2021). This shows acute negative impacts are experienced in the agricultural sector, especially for maize and cassava as crop yields decrease. More so, DGCE identifies the decline in production of teff, maize, and sorghum by 25.4, 21.8, and 25.2%, respectively by 2050 compared to the base period of 2025 in Ethiopia's agricultural production (Solomon et al. 2021). Further to this, Ethiopia nation will lose its agricultural GDP to 31.1% at factor cost by 2050 as a result of climate change. However, GCE is not generally utilised in all-purpose climate change modeling because it dwells more on economic policies and shocks. It lacks computational libraries for dynamic variable data analysis.

### **Statistical model**

The statistical model is one of the key models used in analysing and predicting climatic conditions. The statistical modeling methods encompass simple regression up to non-parametric spatiotemporal Bayesian models deployed for effective data modeling and analysis to foster past climate change (Sweeney et al. 2018). The statistical model is known for its structural simplicity. The kind of statistical models deployed for the assessment and evaluation of climate change data include Regression, Bivariate, Frequency Ratio (FR), Evidential Belief Function (EBF), and Ordered Weight Average (OWA) that has been widely utilized (Yariyan et al. 2020; Siddiqua et al. 2021). For instance, Yariyan et al. (2020) combined different statistical models (e.g. FR, EBF, and OWA) for flood susceptibility in Kurdistan Province, Iran to reduce the harmful effects of flooding that were highly contributed due to the impact of climate change on rainfall. The hybrid of the models achieves 95.1% efficiency. Also, the computation of the Flood Potential Index (FPI) was achieved with 2 bivariate statistical models (Costache 2019). This used

10 flood conditioning factors together with 158 flood pixels and 158 non-flood pixels that were validated through the ROC curve model. This shows that 25% of the upper and middle basin of the Prehova river is due to climatic change impact. A recent study by Yang et al. (2022) utilised a geographically and temporally weighted regression model (GTWR) to show the climate change effect on corn yield in U.S. Corn Belt making it possible to retain the original data from spatiotemporal heterogeneity data. These, however, showed over 40 years, from 1981 to 2020 the positive effect of climate change on corn yield, with temperature having a major effect compared to precipitation. More so, statistical models show off a great evaluation to assess and make future projections on the best approach, simulations, strategy, measures, and policies in different hemispheres and states such as Indiana (Hamlet et al. 2020), Jemma sub-basin, upper Blue Nile Basin of Ethiopia (Worku et al. 2020), Scotland's Atlantic salmon rivers (Jackson et al. 2018), and regions of Pakistan (Siddiqua et al. 2021). However, the current data generation is generic and spontaneous and requires advanced data analysis libraries or techniques to handle.

#### ***Gateway belief model (GBM)***

The gateway belief model (Van Der Linden 2021) is a dual process theory of attitude change, that portrays what political strategists have perceived for decades: observed scientific agreement plays a key role in people's attitudes about contested scientific issues. This model is mixed in the sense that people oppose the scientific judgment that climate change is caused by humans. So, it borders on the degree of belief of these people's perception of climate change. It tells how real people think about climate change and how much people worry about the issue. This perception continues, which depends on the level of consensus and the extent of their endorsement of those issues. However, it empirically evaluated the messages about the scientific consensus on the reality of anthropogenic climate change and the safety of genetically modified food shifting perceptions of scientific consensus (Kerr and Wilson 2018). Therefore, the consensus was perceived in informing personal beliefs about climate change, nonetheless, results indicated limitations in the impact of single and one-off messages. This perception of people's view on climate change causes arguments in the sense that some authors argue that environmentalism is not the main cause of thoughts or behaviors about climate change. Relatively, the evolved social needs for belongingness, understanding, control, self-enhancement, and trust are more practical intervention targets than the attempt to create environmentalist beliefs or identities (Brick et al. 2021). A study by Van Der Linden et al. (2019) presented a large-scale confirmatory replication of GBM on a national quota sample of the US population (N=6301). The result shows from the hypotheses of the GBM that change in perceived scientific consensus causes subsequent changes in cognitive (belief) and affective (worry) judgments about climate change, which in turn are associated with changes in support for public action.

#### ***Soil and water assessment tool model (SWAT)***

The soil and water assessment tool model (Akoko et al. 2021) is a hydrological modeling tool dedicated to hydrologic and environmental simulations. It is a physically based, semi-distributed, and continuous time hydrological model. Typically, the SWAT

model was developed to assess water resources and predict the impacts of land use/cover changes including land management practices on soil erosion, sedimentation, and non-point source pollution on large river basins/watersheds. The SWAT model has been widely applied in the assessment of climate change impact globally. For example, Li and Fang (2021) applied the SWAT model to assess the impact of climate change on the stream flow in the Mekong Basin of the Mun River, Southeast Asia. The result shows that climate change has a great impact on the river stream monthly flow changes were negatively related to temperature ( $p < 0.05$ ) in the dry season and positively linked to precipitation ( $p < 0.01$ ) in the wet season. It also projected an increase of streamflow by 10.5%, 20.1%, and 23.2% during 2020–2093 under three climate scenarios. This provides a scientific basis for adaptive management, although, this model is only good in hydrologic and environmental simulations related practices. Also, Aznarez et al. (2014) employs SWAT mode and remote sensing data to analyse the climate change impact on hydrological ecosystem services in Laguna del Sauce, Uruguay. The result shows that water resources were negatively affected in the Laguna del Sauce catchment, particularly in the representative concentration pathways (RCP) by 8.5 scenarios. However, a comparative analysis of other locations is not covered. Furthermore, water Erosion monitoring and prediction were analysed to ascertain the effect of climate change in the R' Dom watershed in Morocco using the revised universal soil loss equation (RUSLE) and SWAT Equations (Alitane et al. 2022). SWAT model was employed in the modeling of surface water availability in a Semi-Arid Basin, El Kalb River, Lebanon (Kalb et al. 2021) due to climate change, and assessment of the future impact of climate change on the Hydrology of the Mangoky River, Madagascar (Finaritra et al. 2021). However, all these studies are geared towards providing adaptation strategies, projection, and evaluation of tackling climate change issues, managing water resources, and water engineering.

### **The big data analytic methods**

This subsection discusses the big data analytic (BDA) methods or approaches employed to analyse climatic big data (CBD).

#### ***Machine learning***

Machine learning (Rolnick et al. 2019) is the analytics method that can analyze, inform meaning, and predict the outcome of large climate change data. The impact of machine learning (ML) in tackling climate change issues includes: offering fast data analysis and prediction helping to reduce greenhouse gas emissions and assisting society in adapting to a changing climate. It further accelerates the prevention of the leakage of methane from natural gas pipelines and compressor stations, modeling emissions, and improving clean energy access. More so, ML methods such as support vector machine (SVM), Decision Tree (DT), Random Forest (RF), k-Nearest Neighbors (k-NN), Naïve Bayes, etc., are utilized to access the health risk across subpopulations due to climate change effects. For example, Machine learning methods such as DT, SVM, and k-NN were utilized to detect the daily number of COVID-19 infected and death cases during the pandemic lockdown (Saba et al. 2020). Also, the ML analytic method was applied for the evaluation of anthropogenic and natural climate change, which optimize the spectral features of the component sine waves (Abbot and Marohasy 2017). Many recent other

works utilised ML methods in solving climate change effects. Few of the recent studies include (Zia 2021; Manley and Egho 2022; Davenport and Diffenbaugh 2021; Ikegwu et al. 2023).

### ***Deep learning***

Deep learning (DL) is one of the analytics methods that has played a crucial role in big data analysis. Deep learning (Zhang and Li 2020) is a nonlinear method for simulation and prediction in both mining and diagnosis of large climate data. It discovers climate patterns and predicts climate change needs. Another reason is that DL is capable of discovering unknown hidden information in big climate data. The known DL methods are neural networks (NNs) such as artificial neural networks (ANNs), convolutional neural networks (CNNs), deep neural networks (DNNs), etc., which are used for climate pattern analysis. The NNs are the interconnection of sample computing cells known as processing units or neurons. In addition, the algorithms associated with these methods include Metropolis, Gibbs sampling, simulated annealing, variation approach, etc., which are deployed for the simulation of large climate data. DL is an important aspect of analytics methods applied in significant climate data analysis to achieve sustainable development goals (Huan et al. 2021) vision on climate change. For instance, ANNs were utilized for the simulation and forecasting of climate and meteorological variables such as temperature, rainfall, solar radiation, and wind speed (Abbot and Marohasy 2017). Also, DL was applied in the analysis of remote sensing data on climate change and urbanization (Zhu et al. 2017). Recently, Kurth et al. (2019) analyzed extracted pixel-level masks of hash weather patterns using deep learning methods. This, however, achieved a parallel efficiency of 79.0%.

### ***Artificial intelligence***

Artificial intelligence (Sebestyén et al. 2021) is an analytics method for big data analysis saddled with the responsibility of supporting simulation and decision-making infused into earth observation data and simulation climate data. The methods produce a better result when combined with numerical climate model data. For example, Kadow et al. (2020) reconstructs missing climate data from global climate datasets (HadCRUT4) using the artificial intelligence (AI) image inpainting approach. Therefore, uncertainties and biases in climate records were reduced. Furthermore, AI methods discovered climate connections to enhance earth system model (ESM) simulation and weather features (Huntingford et al. 2019). Some other works that have utilised the AI approach in combating climate change effects include (Benzidia et al. 2021; Lozo and Onishchenko 2021; Narayan 2021; Cheong et al. 2022; Avand et al. 2021; Hwang et al. 2021; Cowls et al. 2021).

### **Purpose for analysing big climate data using big data analytic methods**

Climate change (CC) and sustainability development (SD) have been transformed and enhanced through big data analytics (BDA) methods. The purpose of BDA methods in CC&SD as one of 13 out of the 17 sustainable development goals (SDGs) (Huan et al. 2021) include:

- (i) Climate change detection: Big data analytics methods such as Hadoop MapReduce framework spatial cumulative sum algorithms (SCUMSUM) are used to reduce large climate data to monitor the seasonal changes exhibited (Manogaran and Lopez 2018). The spatial cumulative sum algorithms such as cumulative sum (CUSUM), bootstrap analysis, and spatial autocorrelation methods were utilized to determine the slow and drastic changes in the mean value of vast climate data. For example, (Manogaran and Lopez 2018) applied the SCUMSUM algorithms and the BDA methods to monitor the changes in rainfall, precipitation, maximum and minimum temperature, humidity, wind speed, and solar. This, however, achieved high result performance up to 81.48%.
- (ii) Disease identification: Disease identification (Lopez and Sekaran 2016) involves identifying and predicting diseases that occur due to climate change, in which analytics methods have played a vital role. These diseases could be illnesses and abnormalities in the human body system. For example, the analytics methods such as the CUSUM algorithm and Bootstrap analysis method combined with big data analytics help to identify and predict malaria, coronavirus, Ebola virus, dementia, and Parkinson's (Tao et al. 2020; Sharma and Gupta 2021; Manogaran and Lopez 2018). These methods further accelerate the monitoring of physiological and psychological changes such as high temperature, dengue fever, and emotional stress, of the human body that occur as a result of climate change. In a recent development, big data analytics methods have successfully assisted in analysing contagious disease data and have recorded impressive results as reported in the current studies. For instance, a recent study by Saba et al. (2020) utilized analytic techniques such as decision trees, support vector machines, and k-Nearest Neighbors to detect and forecast the daily overall number of COVID-19 infected and death cases during the pandemic lockdown. A more recent study by Rasheed et al. (2021) deployed linear regression and convolution neural network techniques in post-COVID-19 diagnosis from data obtained from chest X-ray images. The study achieved the accuracy of LR and CNN, 95.2–97.6% without principal component analysis (PCA) and 97.6–100% with PCA respectively.
- (iii) Emission reduction: A combination of technological, natural mechanism, and economic approaches as essential means of combating climate change will help reduce carbon emissions. However, nature alone cannot protect us from the climate change effects but technology paradigm or a hybrid approach (Ornella 2020). BDA tools such as Apache Hadoop, MongoDB, Lambdoop, etc., are utilized to reduce carbon emissions concerning oil exploitation, smart buildings, and smart city development (Zhang and Li 2020; Gomede et al. 2018). In addition, the implementation of a low fossil-carbon energy system and energy use improvement also reduce carbon emissions. More so, remote sensing data generated through earth observation and analytics emanating from the basis of satellites, aircraft, and ground-based structures are utilized to form a decision-support, prediction, and forecast for further global, regional, and field scales towards carbon emission reduction.
- (iv) Monitoring and evaluation: BDA methods facilitate monitoring, evaluation, and adaptation through the utilization of digital devices such as smartphones, sensors, social media, earth observation data, and climate simulation data. Therefore, it

examines how an adaptation program designed to enable storm warnings affects the standard of living before, during, and even after a storm event, as well as monitors if and how these changes could occur over time (Ford et al. 2016).

- (v) Cloud-based ecosystem: BDA methods help to create a cloud-based ecosystem from vast data sources combined with the appropriate techniques and software-as-a-service (SAAS) to revolutionize the agricultural sector, marine ecosystem, and green industries; hence, leveraging the cost-effectiveness of storage of voluminous amounts of data on the cloud (Schnase et al. 2017; Jiao et al. 2015).
- (vi) Decision support: It offers a decision support system and proper strategy against climate change effects and improve people's resilience in the face of the adverse effect of climate change through trends, pattern, prediction, and technical analysis using machine learning and artificial intelligence (Knüsel et al. 2019; Heckman et al. 2018). It further offers early warning surveillance thereby enhancing the capacity to respond to climate change (Ford et al. 2016).
- (vii) Assessment of disaster damage: Humanitarian operation and crisis management are yielding excellent results using the application of big data analytics techniques (Akter and Wamba 2017). The BDA helps to visualize, analyse and predict disasters, thereby making them easy to manage. Quick disaster response is most compelling because it imposes some threats. Some of the threats include lives, meeting humanitarian needs (food, shelter, clothing, public health, and safety), clean-up, damage assessment, task assignment, resource allocation, etc. The application of big data such as satellite imageries, Global Positioning System (GPS) traces, mobile Call Detail Records (CDRs), social media posts, etc., in conjunction with advances in data analytic techniques (e.g., data mining and big data processing, machine learning, and deep learning) can facilitate the extraction of geospatial information that is highly needed for rapid and effective disaster response (Cumbane 2019). For instance, Guo et al. (2020) utilised a data mining approach, econometric regression model and input–output model are implemented in the system, based on multi-source data including hourly rainfall, geographical conditions, historical and real-time disaster information, socioeconomic data, and defense countermeasure.
- (viii) Rainfall-runoff modeling: Rainfall-runoff (Xiang et al. 2020) is viewed as a complex nonlinear time series model. It has been widely used for water resources management, hydro-power development, urban planning, irrigation, and other agro-hydrological/meteorological activities planning. Over time, it has attracted much attention from researchers to utilise the model for effective time series predictions in hydrology. Nonetheless, the application of big data analytics methods has made it more efficient for prediction. For instance, Xiang et al. (2020) utilised machine learning and deep learning (e.g. long short-term memory, sequence-to-sequence, etc.) models to predict runoff using rainfall data sets. However, the spatial inequality of rainfall is still ignored inside each sub-watershed. Also, machine learning methods were used to investigate rainfall-runoff modeling at an hourly timescale, which achieved better accuracy (Muhammad et al. 2020).
- (ix) Crop recognition: Big data from remote sensors that it possible for the identification and classification of crop species automatically, fast, and cost-effective to avoid the use of human experts (Tantalaki et al. 2019). Remote sensing helps crop map

building by pixel classification as it is important for the development of precision agriculture. Crop recognition involves two methods; a joint likelihood decision fusion multi-temporal classifier and Markov model-based technique (Zhang and Li 2020). The first method which the class with the highest likelihood of producing the observed single-date/multidate classifications for a given remote sensing image. The second method, the Markov model-based technique relates the varying spectral response along the crop cycle with plant phenology for different crop classes and recognizes different crops by analysing their spectral temporal profiles over a sequence of remote sensing images. For example, Shelestov et al. (2017) classification for multi-temporal satellite imagery for crop mapping was effectively carried through big data processing tools such as SVM, decision tree, random forest classifier, etc., and Google Earth Engine (GEE) platform.

- (x) Rainfall estimation: Seasonal (e.g. monthly, yearly, etc.) rainfall estimation and prediction are made possible with the help of BDA. This influence directly or indirectly the kind of rainfall estimation variables. This variable is dependable on data input. Examples of these input data variables include maximum, minimum, and average temperature (°C), vapor pressure (hPa), wind speed (km/h), humidity (%), and cloud cover (%) (Pundru et al. 2022). In addition, there is a variety of radar data processing algorithms involved in rainfall product data generation requests. These include data quality control, rain rate estimation, rainfall accumulation, and conversion of spherical (polar) to geographic coordinates (Seo et al. 2019; Khan and Bhuiyan 2021). These data are analysed for effective rainfall estimation and decision-making with BDA methods. For example, (Pundru et al. 2022) utilised machine learning models with singular-spectrum analysis otherwise known as least-squares support vector regression (LS-SVR), and random forest (RF) for rainfall prediction. Root Mean Square Error (RMSE) and Nash–Sutcliffe Efficiency (NSE) were used to assess the performance of the models that achieved 71.6% and 90.2% respectively. The model productively predict the rainfall.

### **The common datasets for modeling and combating climate change effects**

The extensive review showed that there are various methods or models developed for analysing a large volume of climate change data (Zhang and Li 2020). Government agencies and other philanthropic bodies are adopting the methods to study the trends, and patterns, and extract useful data in the climatic change environments by identifying new opportunities to combat climate change issues. Therefore, in this section, we focus on data types and sources, and common datasets utilised by various authors in modeling climate change. Moreover, the datasets at terabyte, petabyte, and even Exabyte scales from diverse sources are utilized to form a basis for the increase in global warming research.

#### **Data types and sources**

Recently, the epoch of big data has created a variety of datasets from vast sources in climate change domains. These datasets include several modalities, each of which has a diverse representation, distribution, scale, and density. Hence, data mining and big data

analytics methods are launched to discover hidden knowledge from the numerous datasets formed to aid quick decision-making (Zhang and Li 2020; Al-shiakhli 2019). This, in essence, helps humans better understand climate change and its analysis. Hence, large climate change data types and sources are categorically classified into two, namely; earth observation and climate simulation big data types. These are briefly discussed below:

(i) Earth observation big data (EOBD) type

Earth observation big data type involves the remote sensory observation utilized for monitoring large-scale variability of global climate and environmental changes (Zhang and Li 2020). The EOBD data source includes satellite, weather, atmosphere, hydrosphere, biosphere, lithosphere, station, and gridded data. The data generated from this source helps to obtain current climate conditions and predict future climate changes. This kind of dataset is made available in some domains or agencies such as the National Aeronautics and Space Administration (NASA), US Government Open Data Initiative (USGODI), SeaDataNet (SDN), National Centre for Environmental Information (NCEI), Climate Research Unit (CRU), and Data Distribution Centre (DDC), etc. (Grotjahn and Huynh 2018; Hartter et al. 2018; Tariq et al. 2019). For instance, Blume et al. (2023) utilized EOBD such as Sentinel-2 mosaic Bottom of Atmosphere imagery having 18,881 single 100 × 100-km tiles. The study's results support spatially-explicit seagrass and ocean ecosystem accounting, and further assist policy-making, blue carbon crediting, and all necessary financial investments. The study used random forest machine learning algorithms to assess and analyze the desired data. Additional, Béjar et al. (2023) implemented discrete global grid systems (DGGS) using earth observation data cubes having rHEALPix to enable the efficient integration of diverse spatial data. However, a lot more is anticipated from the design framework, including fully functional operations and the integration of rHEALPix-safe and rHEALPix-aware features in the Python application programming interface for the open data cube.

(ii) Climate simulation big data (CSBD) type

The data generated from these categories are generally used to predict future climate change trends and assess their impacts. The CSDB data sources are validated by foundational elements of climate science known as coupled model inter-comparison project (CMIP) standards (Colorado-Ruiz et al. 2018). The objective of CMIP is to elucidate past, present, and future climate change arising from natural and unforced variability. The example of such data generated from CSBD includes higher resolution complex physical, chemical, and biological processes. More data sources include DECK [diagnostics, evaluation, characteristics of Klima (climate)], coordination, climate projection data, infrastructure, and documentation data (Zhang and Li 2020). Hence, MCIP and DECK historical simulation formed the major source of CSBD. The datasets are useful in climate change predictions using analytics methods. These datasets of this nature can be found in National Oceanic and Atmosphere Administration (NOAA) reports, NASA Earth Observatory (NASAEO), Global Carbon Project (GCP), and Data—Climate Change (World Bank) (Tariq et al. 2019; Pinkerton and Rom 2014). Take for example, (Nikolaev et al. 2020) utilized general circulation models simulated data of an 800 year time series from CanESM2 using deep learning methods to train on them and testing



with the historical data. Nevertheless, depending on the available historical data, the simulated data needs to be fine-tuned further to lessen overfitting. Additionally, the recent study by Wang et al. (2023) implemented machine learning techniques to model snowmelt runoff in a high-altitude mountainous area in the Xiyang River Basin without the need for observation data. These techniques used meteorological data at the watershed level. In order to improve the machine learning model's prediction of snowmelt runoff in alpine mountains, snow remote sensing data was used. This was also supported by the numerical simulation analysis of surface air temperature between 1980 and 2019 as carried out Krasnodar city, Russia (Volvach et al. 2023) and related comparison with hemispheric and regional sea ice extent using NOAA and NASA passive microwave-derived (Meier et al. 2022). Following research and comparison, it was determined that the space-based observation had more local circumstances than the ground-based observation based on the data from the World Data Center (ground based) and the POWER project (space based). More recent study by Hayasaka (2024) consistently examine the fire weather conditions in Mexico by utilizing hourly weather data, simulated climate data, and 20 years' worth of satellite hotspot and rainfall data. This, however, attempted to improve local fire forecasts and trends in future fire across the globe.

#### **The datasets, differences, and similarities for modeling and combating climate change effects**

Based on the critical review, we identify 42 works that have used different or similar datasets for the implementation of climate change-related problems to identify opportunities, predict, and project effective measures, policies, strategies, and techno-drive to combat the effects of the global warming of climate change. Various data or datasets have been generated by different authors to implement climate change impacts. These datasets are structured and analysed using climate change methods (e.g. CGE, statistical, GBM, SWAT, ML, DL, AI, etc.). The climate change methods were discussed in “[The methods for modeling and combating climate change](#)” section. Therefore, some of these data are climate change earth observational big data, climate change simulated big data, and social media climate change big data. From the review, seventeen (17) out of 42 primary papers reviewed used earth observational climate change big data e.g. (Habibullah et al. 2022; Davenport and Diffenbaugh 2021; Kadow et al. 2020; Avand et al. 2021), twenty-one (21) utilised climate simulation big data, e.g. (Hossain et al. 2022; Handayani and Abubakar 2020; Yang et al. 2022; Zia 2021; Seo et al. 2019), and four (4) authors make use of social media data e.g. (Manley and Ego 2022; Hwang et al. 2021; Liu 2021).

For example, some of the climate change earth observational data that have been utilised by various authors include HadCRUT4, Red list by IUCN, multi-country multi-city (MCC) that contain about 732 locations in 43 countries, social accounting matrix (SAM), DDEM, Oxford station data, 923 PRISM station data, a household that contains about 395 from 8 villages, 6301 of US Pop from Qualtrics LLC, etc. (Habibullah et al. 2022; Vicedo-Cabrera et al. 2021; Davenport and Diffenbaugh 2021; Kadow et al. 2020; Xiang et al. 2020; Atube et al. 2021). Then, some examples of climate simulation datasets include 14 Gradient Circulation Models (GCMs) from CMIPs, 20 GCMs from CMIP6, National Centre for Hydrological Meteorological

Forecasting (NCHMF), Resnet 50 dataset, NEXRAD from Amazon Web Service (AWS), Coordination of Information on Environment (CORINE), Regional Climate Model (RCM), General Inspectorate for Emergency Situation (GIES) dataset, etc. (Costache 2019; Yang et al. 2022; Van Der Linden et al. 2019; Alitane et al. 2022; Zia 2021; Narayan 2021; Muhammad et al. 2020; Seo et al. 2019; Colorado-Ruiz et al. 2018; Wang and Tian 2022).

Similarly, some of the authors who utilised datasets from Coupled Model Inter-comparison Project Phase 5 and 6 (CMIP5 and CMIP6) include (Hossain et al. 2022; Hamlet et al. 2020; Colorado-Ruiz et al. 2018; Wang and Tian 2022), social media datasets include (Manley and Egoh 2022; Hwang et al. 2021; Liu 2021). Also, National or global datasets such as GIES, CORINE, NCHMF, HadGEM, SRTMN, RCM, NASS, USDA, APHRODITE, etc., include (Costache 2019; Jackson et al. 2018; Li and Fang 2021; Kalb et al. 2021; Zia 2021; Muhammad et al. 2020; Yang et al. 2020; Nowack et al. 2018). Table 2 presents the datasets, differences, similarities, and methods for modeling and combating climate change effects.

**Table 2** The datasets, differences, similarities, and methods for modeling and combating climate change effects

Dataset/source	Methods/models	Authors
Gradient circulation model from CMIP5 and CMIP6	Statistical, ML, DL, CGE, SWAT	Hossain et al. (2022), Hamlet et al. (2020), Li and Fang (2021), Finaritra et al. (2021), Colorado-Ruiz et al. (2018), Wang and Tian (2022), Hosseini et al. (2021), O’Gorman and Dwyer (2018)
Flicker, Twitter API (7632 tweets and 7634 news articles), ZigBee, from the social media site	ML, statistical	Manley and Egoh (2022), Hwang et al. (2021), Liu et al. (2021)
Resnet50	ML, AI	Narayan (2021)
4 km daily precipitation of 923 station data from PRISM Climate Group	ML	Davenport and Diffenbaugh (2021)
HadCRUTH4, HadGEM3	AI, statistical	Kadow et al. (2020), Nowack et al. (2018)
Precipitation data from local rain gauge network, proxy temperature records, rainfall data	ML	Abbot and Marohasy (2017), Khan and Bhuiyan (2021), Hua-dong et al. (2015), Amato et al. (2020), Avand and Moradi (2021)
NCHMF dataset	ML	Zia (2021)
Panel data of land, labour, fertilizer, and temperature from the World Bank database and FAO statistics	Statistical	Handayani and Abubakar (2020)
Empirical data from 732 locations in 43 countries, and cross-national of 115 countries, Red List by IUCN, SDG index from 162 countries of 15,400 datasets	Statistical, ML	Türkeli (2020), Habibullah et al. (2022), Vicedo-Cabrera et al. (2021)
Landset8 images sensor	DL	Shelestov et al. (2017)
Global Digital elevator model (DEM), land cover, soil map	SWAT	Alitane et al. (2022), Kalb et al. (2021)
Social accounting matrices (SAM) such as temperature, precipitation, and evaporation	CGE	Manuel et al. (2021)

### Implementation frameworks for climate change modeling using big data analytics

The data processing and analysis of high-voluminous climate change data generated or sourced from public sites or databases require efficient frameworks to be handled. The big data analytics modeling platforms or tools include HDFS, cross—MapReduce, YARN, Google BigTable, Spark, Mahout, and Flume. These frameworks are briefly explained. The strength and weaknesses of these frameworks are presented in Table 3.

(i) Hadoop file distributed system (HDFS)

Hadoop file distributed system (Shih 2018) is an analytic storage framework created from the Apache family and it is responsible for storing large data. The essence of HDFS is to store large data generated from earth observation and climate simulation datasets. HDFS handles large data and maintains the simplicity of usage. It can read data into distributed arrays without introducing a single point of data conversion. For instance, (More et al. 2019) utilized HDFS to load historical weather datasets crawled from National Climatic Data Centres. This dataset was analysed to detect climate change with the help of MapReduce which was applied to remove scalability issues. Conversely, HDFS and MapReduce work in common, HDFS stores the data, and MapReduce is deployed to process the large datasets and store them back in HDFS. For example, a recent study by Greca et al. (2023) deployed big data framework such as Hive and Hadoop to store, manage, and process climate change earth surface temperature data from NOAA's MLOST, NASA'S GIS-TEMP and UK's HADCRUT. Following the research result, it shows that in the city of Durres, the temperature increased by 1.1 °C since the inception of the pre-industrial era.

(ii) Cross MapReduce

Cross MapReduce is a technological framework enabled by the Apache Hadoop ecosystem that provides data splitting into the distributed format, data mapping, shuffling, and classification to reduce document search (Mirpour et al. 2021). This is essential in the processing of large climate change data because it is capable of processing all geo-distributed data. Cross MapReduce (CMR) merges the records that have the same keys in the cluster using reduce function. CMR sequentially contains three components; MapCombine, Gshuffle, and GlobalReducer. This component is jointly utilized to minimize transferring of data volumes globally. For example, Mirpour et al. (2021) proposed a Cross MapReduce framework to minimize transferred data volume and determine the number as well as the locations of global reducers. This, however, achieved a 40% reduction in the amount of data transfer over the Internet. In addition, MapReduce helps to detect climate change from a large volume of weather data (More et al. 2019). After climate change detection, there is a need for prediction to be infused in the data to form decision support, hence; machine learning.

(iii) Yet another resource negotiator (YARN)

YARN (Cumbane 2019) is a batch-processing framework, that coordinates the components of the Hadoop framework. YARN is implemented on top of HDFS which helps the execution of multiple and distributed climate big data in parallel across applications. YARN is an important framework because it dynamically handles

**Table 3** Strengths and weaknesses of the analytics methods

Methods	Description	Strength	Weakness	References
Hadoop distributed file system	Responsible for storing large data generated from earth observation and climate simulation datasets	Maintain simplicity, reliability, cost-effectiveness, and easy storage of large climate data. Also, ensure data replication for fault-tolerance	Doesn't constitute a multi-purpose file system	Shih (2018), More et al. (2019)
Cross MapReduce	Framework that provides data splitting into the distributed format, data mapping, shuffling, and classification to reduce document search	Helps to reduce document searching for a data process, and split data into a more distributed format, data mapping, and shuffling	Has data optimization especially with small data and real-time processing issues	More et al. (2019), Mirpour et al. (2021)
Google BigTable	Handle distributed, column-oriented data storage and a large amount of climate-unstructured data	It makes mapping of spatial data easy, distribution across multiple clusters, multi-dimensionality, and persistence	High cost in reading/write operation as data are split into several columns during streaming	Ikegwu et al. (2022)
YARN	Batch-processing framework, that coordinates the components of the Hadoop framework	Cluster/resource management, running non-MapReduce applications, and failure in cascading	Cascading in cluster/resource management and running non-MapReduce applications and failure	Cumbane (2019), Ikegwu et al. (2022)
Apache Spark	Runs programs up to 100x faster compared to some other big data frameworks	Enable complexities via parallelism and fault-tolerance and support implementation	In-memory (RAM) capability and consumption are very high. Also, difficult to provide real-time processing	More et al. (2019), Ikegwu et al. (2022), Xu et al. (2020)
Mahout	Frameworks that provide extensibility libraries for big data analytics	It is scalable, easy to use, and extensibility libraries	Difficulties with configuration and compatibility with other frameworks	Zhang and Li (2020)
Apache Fume	Streaming framework platform that enables distributed, aggregate, and move large climate data to a centralized defined datastore	It is scalable, reliable, fault-tolerant, and customizable for different sources and sinks	It has a complex topology, low scalability, weak ordering guarantee, and duplication and reliability issues	Ikegwu et al. (2022), Doraswamy and Manjunatha (2017)

multiple processing and real-time interactive processing of climate big data. Furthermore, it improves multi-tenancy, cluster utilization, scalability, and compatibility in climate big data processing and management. For example, Kanwar (2018) implemented YARN-based method using climate weather datasets. This comparatively improves Hadoop execution, tightens up confines, and fixes the scheduling and resource carriage issues.

(iv) Google BigTable

Climate big data storage is important and requires an efficient implementation framework that can handle distributed, column-oriented data storage and a large amount of climate-unstructured data (Ikegwu et al. 2022). In addition, handling Google's internet search and web service operations and its related services of climate change data storage is efficiently carried out with the help of Google BigTable. For example, the database choice was taken into consideration while implementing geospatial-temporal data processing and storing; Google BigTable, Cansandra, etc. were thought to have a significant impact (Shykhmat and Verses 2023). Predictive maintenance for agricultural vehicles is included in the analysis. Meteorological stations, weather satellites, and environmental organizations can provide information about the weather, climate, and air quality. Vehicle performance is significantly impacted by weather conditions. For instance, excessive heat can cause an engine to overheat; rain and snow can muddy fields, making traction more difficult; and dust and sand are quite harsh on-air filters and radiators. However, managing geospatial-temporal data in a distributed fashion involves specific challenges and concerns.

(v) Apache Spark

Apache Spark (More et al. 2019; Ikegwu et al. 2022) is a fast and large-scale climate data processing. Major libraries (e.g., graphX, machine learning libraries (MLlib), spark streaming API, and spark SQL) are supported by a spark. Hence, runs programs up to 100× faster compared to some other big data frameworks (e.g. Hadoop, MapReduce, etc.) especially in memory or on disk up to 10×. Spark technology works well with Hadoop, YARN cluster manager, Java VM, and other architecture. Furthermore, Apache Spark can process a high volume of climate big datasets generated in memory with a high-speed response time and in addition, provides an alternative platform for stream data processing and analysis. For instance, Xu et al. (2020) accurately utilizes Spark to forecast wind speed big data in multi-step. It shows that the Spark distributed computing framework has a faster computation speed when processing climate big data, compared to other baseline processing frameworks.

(vi) Mahout

The Mahout (Zhang and Li 2020) is one of the machine learning libraries and frameworks that provides scalable, easy-to-use, and extensibility libraries for big data analytics. Mahout-Samara is a newer version, the project helps users build their own distributed algorithms, instead of using the existing library. However, it is not friendly as the configuration is not compatible with the existing Hadoop cluster. Nonetheless, some companies have utilised Mahout in big data analytics such as LinkedIn, Mendeley, etc. For example, Bhavanandam (2022) utilized Mahout to expedite unsupervised machine learning algorithms for crop yield prediction based

on weather. This paradigm facilitates the identification and forecasting of various agricultural production conditions. In addition, Mahout was used to determine the climate parameters that affect the study, evaluation, and forecast of acer mono sap liquid water using machine learning models and data from the Internet of Things (Lee 2020). Despite the study's 98.25% accuracy rate, everyday yields have persistent issues since the specific data provided by Korea Forest Service and Meteorological office are introduced.

(vii) Apache Flume

Apache Flume is a streaming framework platform that enables distributed, reliable, and accessible web services from different sources to efficiently collect, aggregate, and move large amounts of climate data to a centralized defined datastore (Ikegwu et al. 2022; Doreswamy and Manjunatha 2017). Apache Flume provides simple and flexible architecture based on streaming data flows with robust, fault-tolerant, reliable, failover and recovery mechanisms to ensure efficient big data processing and management. This architecture has been widely utilized in climate change-related data processing and streaming. For example, Bouziane et al. (2021) suggested using a cloud-based architecture like Apache Flume, IoT, Hadoop, etc. to speed up the intelligent management of water resources because climate change raises energy and water consumption globally.

### Open research directions

The identified areas of further research are highlighted below:

(i) Data understanding challenges

The understanding of data is essential as it forms the basic subjects utilized in climate change prediction and forecast to aid strategic planning, policy formulation, modeling, and implementation of climate change impacts. The challenge in data encompasses data observation, structure, format, pre-processing, feature extraction, modeling, etc., which is still inherent to the implementation of climate change effect (Avand et al. 2021; Dueben and Bauer 2018). Take, for instance, the voluminous geographical dataset, spatial-temporal dataset, and meteorological dataset from public climatic databases such as CMIP5, HadGEM3, NCHMF, social media site, CORINE, IFC-Cloud-NEXRADGIES, NASS, USDA, etc. (Hossain et al. 2022; Costache 2019; Yang et al. 2022; Zia 2021; Manley and Egoh 2022; Seo et al. 2019; Colorado-Ruiz et al. 2018) that are being extracted require full training using big data analytic models before knowledge can be informed with it for effective data analysis, prediction, and forecasting. Further research is also beckoned on how to obtain updated stored data automatically and when needed on public climatic database domain, government documentaries, and high precision real-time map big data (Guo et al. 2020). Furthermore, understanding and familiarization of the data storage, processing, analysis, modeling, and visualization to aid organizational policy-makings are necessary and this area requires urgent research to ensure comprehensive knowledge on the digital trends of the data collected.

(ii) Problems with climate change selection methods

Selecting climate change analytics methods (Huan et al. 2021; Hossain et al. 2022; Siddiqua et al. 2021; Van Der Linden 2021; Kalb et al. 2021; Kadow et al. 2020; Rolnick

et al. 2022) is difficult due to the nature of the model to be used. Each method requires extensive and indebt training. Each method has strengths and weaknesses and requires special skills to choose which method is best suited for climate data modeling and analysis. Further comparison is needed to determine the more efficient approach to select the methods based on the processing and analysing tasks. More so, more dynamic updates are required by the developers of General Circulation Models (GCMs) to improve climate modeling resulting in high emissions (Finaritra et al. 2021). More focus is required specifically to choose the appropriate methodologies and theoretical sensors or lenses to expedite the particularities of the selected methods.

(iii) Issues with climate big data management

The data cleaning reliability, a large volume of data aggregation of different climate big data generated, etc. requires an improvement. There is also an issue in encoding data for climate data security and privacy are most essential to the data-driven environment and for organizational success. Finding elastic measures to handle large datasets is challenging due to the progressive generation of climate big data. However, some architectures such as Spark, Hadoop libraries, and MapReduce have been considered by different authors (Mirpour et al. 2021; Xu et al. 2020; Manogaran and Lopez 2018), though much improvement on the frameworks and libraries for effective data processing and evaluation is achieved.

(iv) Technological trends challenges

As big data analytics improves the modeling and analysis of climate big data, improving the climate processes still required further studies. The emergence and convergence of different technology and embedded system day-by-day has made the implementation of climate big data storage and processing quite difficult. For instance, big data has its peculiar challenge, a lot of data generation, data integration, real-time streaming of data, network speed accessibility, data diversity and security, data cloud storage, etc. (Ikegwu et al. 2022). Although, AI and intelligent system (Kadow et al. 2020; Narayan 2021; Irrgang et al. 2021) that sense and learn from the environment helps to solve the identified complex earth processes. However, it still further requires research studies. Enhanced sensors, better satellite imagery, faster data storage, better software and hardware, and smarter intelligent systems are required to solve the inherent embedded system challenges (Preteek et al. 2020) and these requires an indebt research.

## Conclusion

Climate science and big climate data-intensive areas has been drastically affected by the emergence of big data analytics and essential technological revolutions. Big data analytic methods have unprecedentedly impacted large climate data analysis. Researchers have hitherto realized the need to increase research on climate change. However, climate change as the global warming effect has negatively impacted the economic growth and human standard of living across the globe. All these ravaging issues can be tackled with a technological approach. In this paper, we present the impact of climate change, the methods for modeling and combating climate change effects, the purpose of deploying big data analytics to analysis climate big data, and big data analytics methods. In

addition, the source of data and the type of datasets utilized for modeling climate change and implementation frameworks to combat climate change effects were also discussed. Furthermore, open research directions were highlighted to give insight for future studies. The practical implications of this study are multi-faceted. First, it provide vast array of data analytics methods for climate scientists to model climate change effects for appropriate policy formulation by governmental organizations. Second, the review unveil wide varieties of climate data recently collected for climate change effects modelling. Third, the study act as avenue to promote the need for the global community to adopt and speedy up campaign against climate change. Finally, the research revealed important areas where big data analytics based climate change modelling could be deployed. Various strategies were outlined in literature to mitigate the impacts of climate change. These include reduction in CO<sub>2</sub> emission by industries, adopting efficient implementation of renewable energy, and afforestation.

#### Author contributions

ACI coined out the title, surfed the internet for materials, and participated in manuscript preparation especially "Introduction" and "Implementation frameworks for climate change modeling using big data analytics" sections, and formatting of the manuscripts. HFN drafted out the structures, wrote "The methods for modeling and combating climate change" section and participated in the proofreading of the entire manuscript sections. ME does the grammatical check, correction, and editing. CVA screened the materials and participated in writing "The impact of climate changes" section. EM filtered the selected journals and participated in manuscript editing. SAI worked on "The common datasets for modeling and combating climate change effects" section. In addition, URA participated in the manuscript draft, proofreading, and reviewing of the entire manuscript.

#### Funding

There is no external funding received by the authors.

#### Availability of data and materials

Not applicable.

#### Code availability

Not applicable.

#### Declarations

##### Ethics approval and consent to participate

The authors certify that the study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. The full APC waiver was approved by the management of Springer Nature after written documents informed consent was obtained from all the participants/authors. The waiver approval was contained via <https://article-discounts-and-waivers.springernature.com/request-summary/ededaf80-77f2-4641-90ee-5bf2a33fb39e>.

##### Consent for publication

No applicable.

##### Competing interests

The authors declare no competing interests.

Received: 4 December 2023 Accepted: 5 February 2024

Published online: 07 February 2024

#### References

- Abbot J, Marohasy J (2017) The application of machine learning for evaluating anthropogenic versus natural climate change. *GeoResJ* 14:36–46. <https://doi.org/10.1016/j.grj.2017.08.001>
- Abdullah MF, Amin MZM, Zainol Z, Ideris MM (2020) Big data analytics as game changer in dealing impact of climate change in Malaysia: present and future research. In: *IoTBDs 2020—Proc. 5th Int. Conf. Internet Things, Big Data Secur.*, no. IoTBDs, pp 461–469. <https://doi.org/10.5220/0009794404610469>
- Akoko G, Le TH, Gomi T, Kato T (2021) A review of SWAT model application in Africa. *Water* 13:1313
- Akter S, Wamba SF (2017) Big data and disaster management: a systematic review and agenda for future research. *Ann Oper Res* 283:1–21. <https://doi.org/10.1007/s10479-017-2584-2>



- Alitane A et al (2022) Water erosion monitoring and prediction in response to the effects of climate change using RUSLE and SWAT equations: case of R' Dom watershed in Morocco. *Land* 11:93
- Al-Shiakhi S (2019) Big data analytics: a literature review perspective. Luleå University of Technology, Luleå
- Amato F, Guignard F, Robert S, Kanevski M (2020) A novel framework for spatio-temporal prediction of environmental data using deep learning. *Sci Rep* 10(1):1–11
- Andersson M, Baccianti C, Morgan J (2020) Climate change and the macro economy. *Econstor*, No. 243. <http://hdl.handle.net/10419/234484%0AStandard-Nutzungsbedingungen>
- Anh DLT, Anh NT, Chandio AA (2023) Climate change and its impacts on Vietnam agriculture: a macroeconomic perspective. *Ecol Inform* 74:101960. <https://doi.org/10.1016/j.ecoinf.2022.101960>
- Atube F, Malinga GM, Nyeko M, Okello DM, Alarakol SP, Uma IO (2021) Determinants of smallholder farmers' adaptation strategies to the effects of climate change: evidence from northern Uganda. *Agric Food Secur* 10:1–14. <https://doi.org/10.1186/s40066-020-00279-1>
- Avand M, Moradi H (2021) Using machine learning models, remote sensing, and GIS to investigate the effects of changing climates and land uses on flood probability. *J Hydrol* 595:125663
- Avand M, Moradi HR, Lasboyee MR (2021) Spatial prediction of future flood risk: an approach to the effects of climate change. *Geosciences* 11(1):1–20. <https://doi.org/10.3390/geosciences11010025>
- Aznarez C, Jimeno-s P, Pacheco JP, Senent-aparicio J (2014) Analysing the impact of climate change on hydrological ecosystem services in Laguna del Sauce (Uruguay) using the SWAT model and remote sensing data. *Remote Sens* 13:2021
- Babar MS, Tazyeen S, Khan H, Tsagkaris C, Essar MY, Ahmad S (2021) Impact of climate change on health in Karachi, Pakistan. *J Clim Change Health* 2:100013
- Béjar R, Lacasta J, Lopez-Pellicer FJ, Nogueras-Iso J (2023) Discrete global grid systems with quadrangular cells as reference frameworks for the current generation of Earth observation data cubes. *Environ Model Softw* 162:105656. <https://doi.org/10.1016/j.envsoft.2023.105656>
- Benzidia S, Makaoui N, Bentahar O (2021) The impact of big data analytics and artificial intelligence on green supply chain process integration and hospital environmental performance. *Technol Forecast Soc Change* 165:120557
- Bhavanandam S (2022) Weather-based crop yield prediction using machine learning and big data analytics. *Easy Chair Prepr* 7443:1–12
- Bhopal A, Medhin H, Bærøe K, Norheim OF (2021) Climate change and health in Ethiopia: to what extent have the health dimensions of climate change been integrated into the climate-resilient green economy? *World Med Health Policy* 13(2):293–312
- Blume A, Pertiwi AP, Lee CB, Traganos D (2023) Bahamian seagrass extent and blue carbon accounting using earth observation. *Front Mar Sci* 10:1–10. <https://doi.org/10.3389/fmars.2023.1058460>
- Bouziane S, Aghoutane B, Moumen A, Sahlaoui A, El Ouali A (2021) Proposal of a big data system for an intelligent management of water resources. *E3S Web Conf* 314:02002. <https://doi.org/10.1051/e3sconf/202131402002>
- Brick C, Bosshard A, Whitmarsh L (2021) Motivation and climate change: a review. *Curr Opin Psychol* 42:82–88. <https://doi.org/10.1016/j.copsyc.2021.04.001>
- Cheong SM, Sankaran K, Bastani H (2022) Artificial intelligence for climate change adaptation. *Wiley Interdiscip Rev Data Min Knowl Discov* 12:e1459
- Clayton SD, Pihkala P, Wray B (2023) Psychological and emotional responses to climate change among young people worldwide: differences associated with gender, age, and country. *Sustainability* 15:3540. <https://doi.org/10.3390/su15043540>
- Colorado-Ruiz G, Cavazos T, Salinas JA, De Grau P, Ayala R (2018) Climate change projections from coupled model inter-comparison project phase 5 multi-model weighted ensembles for Mexico, the North American monsoon, and the mid-summer drought region. *Int J Climatol* 38(15):5699–5716. <https://doi.org/10.1002/joc.5773>
- Costache R (2019) Flood susceptibility assessment by using bivariate statistics and machine learning models—a useful tool for flood risk management. *Water Resour Manag* 33:3239–3256
- Cowls J, Tsamados A, Taddeo M, Floridi L (2021) The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations. *Ai Soc* 38:1–25
- Cumbane SP (2019) Review of big data and processing frameworks for disaster response applications. *Int J Geo-Inf* 8:387. <https://doi.org/10.3390/ijgi8090387>
- Davenport FV, Diefenbaugh NS (2021) Using machine learning to analyze physical causes of climate change: a case study of U.S. Midwest extreme precipitation. *Geophys Res Lett* 48(15):e2021GL093787. <https://doi.org/10.1029/2021GL093787>
- Deangelo J et al (2023) Economic and biophysical limits to seaweed farming for climate change mitigation. *Nat Plants* 9:45–57. <https://doi.org/10.1038/s41477-022-01305-9>
- Doreswamy I, Gad, Manjunatha BR (2017) Hybrid data warehouse model for climate big data analysis. In *Proceedings of IEEE international conference on circuit, power and computing technologies, ICCPCT 2017*. <https://doi.org/10.1109/ICCPCT.2017.8074229>
- Dueben PD, Bauer P (2018) Challenges and design choices for global weather and climate models based on machine learning. *Geosci Model Dev* 11(10):3999–4009. <https://doi.org/10.5194/gmd-11-3999-2018>
- Economides G, Xepapadeas A (2019) The effects of climate change on a small open economy. *Econstor*, No. 7582
- Finaritra M, Tantelinaiaina R, Zhai J (2021) Assessment of the future impact of climate change on the hydrology of the Mangoky River, Madagascar using ANN and SWAT. *Water* 13:1239
- Ford JD et al (2016) Big data has big potential for applications to climate change adaptation. *Proc Natl Acad Sci USA* 113:10729–10732. <https://doi.org/10.1073/pnas.1614023113>
- Fuglie K (2021) Climate change upsets agriculture. *Nat Clim Change* 11:293–299. <https://doi.org/10.1038/s41558-021-01018-5>
- Gomede E, Gaffo FH, Briganó GU, de Barros RM, de Mendes LS (2018) Application of computational intelligence to improve education in smart cities. *Sensors* 18(1):1–26. <https://doi.org/10.3390/s18010267>

- Greca S, Shehi I, Nuhi J (2023) Analyzing climate changes impacts using big data Hadoop. *CEUR Workshop Proc* 3402:21–27
- Grotjahn R, Huynh J (2018) Contiguous US summer maximum temperature and heat stress trends in CRU and NOAA Climate Division data plus comparisons to reanalyses. *Sci Rep* 8(1):1–18. <https://doi.org/10.1038/s41598-018-29286-w>
- Guo J, Wu X, Wei G (2020) A new economic loss assessment system for urban severe rainfall and flooding disasters based on big data fusion. *Environ Res* 188:109822. <https://doi.org/10.1016/j.envres.2020.109822>
- Habibullah MS, Din BH, Tan SH, Zahid H (2022) Impact of climate change on biodiversity loss: global evidence. *Environ Sci Pollut Res* 29(1):1073–1086
- Hamlet AF, Byun K, Robeson SM, Widhalm M, Baldwin M (2020) Impacts of climate change on the state of Indiana: ensemble future projections based on statistical downscaling. *Clim Change* 163(4):1881–1895. <https://doi.org/10.1007/s10584-018-2309-9>
- Handayani T, Abubakar L (2020) The impact of climate change on agriculture sector in ASEAN. *J Phys Conf Ser* 1651:012026. <https://doi.org/10.1088/1742-6596/1651/1/012026>
- Hartter J et al (2018) Does it matter if people think climate change is human caused? *Clim Serv* 10(January):53–62. <https://doi.org/10.1016/j.cliser.2017.06.014>
- Hassani H, Huang X (2019) Big data and climate change. *Big Data Cogn Comput* 3:1–17. <https://doi.org/10.3390/bdcc3010012>
- Hayasaka H (2024) Synoptic-scale wildland fire weather conditions in Mexico. *Atmosphere* 15:96
- Heckman JJ, Pinto R, Savelyev PA (2018) Federal data science: transforming government and agricultural policy using artificial intelligence. Academic Press, London
- Hossain SS, Delin H, Mingying M (2022) Aftermath of climate change on Bangladesh economy: an analysis of the dynamic computable general equilibrium model. *J Water Clim Change* 13(7):2597–2609
- Hosseini M, Bigtashi A, Lee B (2021) Generating future weather files under climate change scenarios to support building energy simulation—a machine learning approach. *Energy Build* 230:110543
- Hua-dong G, Li Z, Lan-wei Z (2015) Earth observation technology has provided highly useful information in global. *Adv Clim Change Res*. <https://doi.org/10.1016/j.accre.2015.09.007>. This
- Huan Y, Liang T, Li H, Zhang C (2021) A systematic method for assessing progress of achieving sustainable development goals: a case study of 15 countries. *Sci Total Environ* 752:141875. <https://doi.org/10.1016/j.scitotenv.2020.141875>
- Huntingford C, Jeffers ES, Bonsall MB, Christensen HM, Lees T, Yang H (2019) Machine learning and artificial intelligence to aid climate change research and preparedness. *Environ Res Lett* 14(12):124007. <https://doi.org/10.1088/1748-9326/ab4e55>
- Hwang H, An S, Lee E, Han S, Lee CH (2021) Cross-societal analysis of climate change awareness and its relation to SDG 13: a knowledge synthesis from text mining. *Sustain* 13(10):1–21. <https://doi.org/10.3390/su13105596>
- Ikegwu AC, Nweke HF, Anikwe CV, Alo UR, Okonkwo OR (2022) Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions and research directions. *Clust Comput* 25:3343–3387. <https://doi.org/10.1007/s10586-022-03568-5>
- Ikegwu AC, Nweke HF, Alo UR, Okonkwo OR (2021) HMCPAED: a new framework for students' dropout prediction. In: *ICT4NDS2021: ICT and sustainability in the 5th industrial revolution*, Ilorin, pp 131–140. [lorinuijpc.com.ng](http://lorinuijpc.com.ng)
- Ikegwu AC, Friday H, Chioma N, Anikwe V (2023) Recent trends in computational intelligence for educational big data analysis. *Iran J Comput Sci* 6(3):1–28. <https://doi.org/10.1007/s42044-023-00158-5>
- Irrgang C et al (2021) Will artificial intelligence supersede earth system and climate models? <https://doi.org/10.1038/s42256-021-00374-3>
- Jackson FL, Fryer RJ, Hannah DM, Millar CP, Malcolm IA (2018) A spatio-temporal statistical model of maximum daily river temperatures to inform the management of Scotland's Atlantic salmon rivers under climate change. *Sci Total Environ* 612:1543–1558. <https://doi.org/10.1016/j.scitotenv.2017.09.010>
- Javadi A, Ghahremanzadeh M, Sassi M, Javanbakht O, Hayati B (2023) Economic evaluation of the climate changes on food security in Iran: application of CGE model. *Theor Appl Climatol* 151(1–2):567–585. <https://doi.org/10.1007/s00704-022-04289-w>
- Jensen HT, Keogh-Brown M, Tarp F (2021) Climate change and agricultural productivity in Myanmar: application of a new computable general equilibrium (CGE) model. *Econstor*. <https://doi.org/10.35188/UNU-WIDER/2021/121-1>
- Jiao NZ et al (2015) Climate change and anthropogenic impacts on marine ecosystems and countermeasures in China. *Adv Clim Change Res* 6(2):118–125. <https://doi.org/10.1016/j.accre.2015.09.010>
- Kadow C, Hall DM, Ulbrich U (2020) Artificial intelligence reconstructs missing climate information. *Nat Geosci*. <https://doi.org/10.1038/s41561-020-0582-5>
- Kalb E, Saade J, Atieh M, Ghanimeh S (2021) Modeling impact of climate change on surface water availability using SWAT model in a semi-arid basin. *Hydrology* 8:34
- Kanwar K (2018) Improving Hadoop performance using yarn-based architecture with weather datasets. In: 2018 international conference on automation and computational engineering (ICACE), IEEE, pp. 178–186
- Karimi V, Karami E, Keshavarz M (2018) Climate change and agriculture: impacts and adaptive responses in Iran. *J Integr Agric* 17(1):1–15. [https://doi.org/10.1016/S2095-3119\(17\)61794-5](https://doi.org/10.1016/S2095-3119(17)61794-5)
- Kerr JR, Wilson MS (2018) Changes in perceived scientific consensus shift beliefs about climate change and gm food safety. *PLoS ONE* 13(7):1–17. <https://doi.org/10.1371/journal.pone.0200295>
- Khan RS, Bhuiyan AE (2021) Artificial intelligence-based techniques for rainfall estimation integrating multisource precipitation datasets. *Atmosphere* 12:1239
- Knüsel B et al (2019) Applying big data beyond small problems in climate research. *Nat Clim Change* 9(3):196–202. <https://doi.org/10.1038/s41558-019-0404-1>
- Kurth et al (2019) Exascale deep learning for climate analytics. In: *Proceedings—international conference for high performance computing, networking, storage, and analysis, SC 2018*, pp. 649–660. <https://doi.org/10.1109/SC.2018.00054>

- Lee HS (2020) Analysis model evaluation based on IoT data and machine learning algorithm for prediction of acer mono sap liquid water. *J Korea Multimed Soc* 23(10):1286–1295
- Li C, Fang H (2021) Catena assessment of climate change impacts on the streamflow for the Mun River in the Mekong Basin, Southeast Asia: using SWAT model. *CATENA* 201:105199. <https://doi.org/10.1016/j.catena.2021.105199>
- Liu W (2021) Smart sensors, sensing mechanisms and platforms of sustainable smart agriculture realized through the big data analysis. *Cluster Comput* 26:1–15
- Liu J, Yang L, Zhou H, Wang S (2021) Impact of climate change on hiking: quantitative evidence through big data mining. *Curr Issues Tour* 24(21):3040–3056. <https://doi.org/10.1080/13683500.2020.1858037>
- Lopez D, Sekaran G (2016) Climate change and disease dynamics—a big data perspective. *Int J Infect Dis* 45:23–24. <https://doi.org/10.1016/j.ijid.2016.02.084>
- Lozo O, Onishchenko O (2021) The potential role of the artificial intelligence in combating climate change and natural resources management: political, legal and ethical challenges. *J Nat Resour* 4(3):111–131
- Manley K, Egoh BN (2022) Mapping and modeling the impact of climate change on recreational ecosystem services using machine learning and big data. *Environ Res Lett* 17(5):054025
- Manogaran G, Lopez D (2018) Spatial cumulative sum algorithm with big data analytics for climate change detection. *Comput Electr Eng* 65:207–221. <https://doi.org/10.1016/j.compeleceng.2017.04.006>
- Manuel L, Orcidia C, Gaby M, Faaiqa H (2021) Impact of climate change on the agriculture sector and household welfare in Mozambique: an analysis based on a dynamic computable general equilibrium model. *Clim Change* 2100:1–18
- Meier WN, Stewart JS, Windnagel A, Fetterer FM (2022) Comparison of hemispheric and regional sea ice extent and area trends from NOAA and NASA passive microwave-derived climate records. *Remote Sens* 14:619
- Mirpour S, Savadi A, Toosi AN, Naghibzadeh M (2021) Cross-MapReduce: data transfer reduction in geo-distributed MapReduce. *Futur Gener Comput Syst* 115:188–200
- More P, Nandgave S, Kadam M (2019) Climate change detection using Hadoop with MapReduce. *Int J Innov Res Comput Commun Eng* 7(3):1925–1932. <https://doi.org/10.15680/IJIRCCCE.2019>
- Muhammad R, Andrea A, Kisi O (2020) Short term rainfall-runoff modelling using several machine learning methods and a conceptual event-based model. *Stoch Environ Res Risk Assess*. <https://doi.org/10.1007/s00477-020-01910-0>
- Nanda S, Perera MT (2019) Global climate change: effect on agriculture food sector in Sri Lanka in the year 2025
- Narayan Y (2021) DeepWaste: instantaneous and ubiquitous waste classification using artificial intelligence for combating climate change. *CJSJ*
- Ngoma H, Lupiya P, Kabisa M, Hartley F (2021) Impacts of climate change on agriculture and household welfare in Zambia: an economy-wide analysis. *Clim Change* 167(3):1–20
- Nikolaev A, Richter I, Sadowski P (2020) Deep learning for climate models of the Atlantic ocean. In: In AAAI spring symposium: MLPS. <http://chfcs.cima.fcen.uba.ar/>
- Nowack P, Braesicke P, Haigh J, Abraham NL, Pyle J, Voulgarakis A (2018) Using machine learning to build temperature-based ozone parameterizations for climate sensitivity simulations. *Environ Res Lett* 13(10):104016. <https://doi.org/10.1088/1748-9326/aae2be>
- Núñez et al (2018) Assessing the impact of climate change on agriculture in Norte de Santander, Colombia. *J Phys Conf Ser* 1126:012045
- O’Gorman PA, Dwyer JG (2018) Using machine learning to parameterize moist convection: potential for modeling of climate, climate change, and extreme events. *J Adv Model Earth Syst* 10(10):2548–2563
- Ogunbode TO (2025) Climate change scenario in Nigeria: local perceptions and the way forward futuro scenario. *Sustain Agric Food Environ Res* (**Ahead of Print**)
- Ornella AD (2020) Why nature won’t save us from climate change but technology will: creating a New Heaven and a new earth through carbon capture technologies. *Immanente Relig Technol*, pp 1–35
- Pinkerton KE, Rom WN (2014) Global climate change and public health. Springer, New York, pp 1–406. <https://doi.org/10.1007/978-1-4614-8417-2>
- Preteek M, Anupama R, Risha M (2020) Big data in climate change research: opportunities and challenges. In: 2020 international conference on intelligent engineering and management (ICIEM)
- Pundru CSR, Yadala S, Goddumarri SN (2022) Development of rainfall forecasting model using machine learning with singular spectrum analysis. *IJUM Eng J* 23(1):172–186
- Rasheed J, Hameed AA, Djeddi C, Jamil A, Al-Turjman F (2021) A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images. *Interdiscip Sci Comput Life Sci* 13(1):103–117
- Rehman A, Ma H, Ozturk I, Ahmad MI (2022) Examining the carbon emissions and climate impacts on main agricultural crops production and land use: updated evidence from Pakistan. *Environ Sci Pollut Res* 29(1):868–882
- Rolnick D et al (2019) Tackling climate change with machine learning. *arXiv Prepr. arXiv*, no. 1906.05433. <http://arxiv.org/abs/1906.05433>
- Rolnick D et al (2022) Tackling climate change with machine learning. *ACM Comput Surv* 55(2):1–96
- Saba T, Abunadi I, Shahzad MN, Khan AR (2020) Machine learning techniques to detect and forecast the daily total COVID-19 infected and deaths cases under different lockdown types. *Microsc Res Tech* 84:1–13. <https://doi.org/10.1002/jemt.23702>
- Schnase JL et al (2017) MERRA analytic services: meeting the big data challenges of climate science through cloud-enabled climate analytics-as-a-service. *Comput Environ Urban Syst* 61:198–211. <https://doi.org/10.1016/j.compeurbysys.2013.12.003>
- Schwartz SEO, Benoit L, Clayton S, Parnes MF, Swenson L, Lowe SR (2023) Climate change anxiety and mental health: environmental activism as buffer. *Curr Psychol* 42:16708–16721. <https://doi.org/10.1007/s12144-022-02735-6>
- Sebestyén V, Czvetkó T, Abonyi J (2021) The applicability of big data in climate change research: the importance of system of systems thinking. *Front Environ Sci* 9(March):1–26. <https://doi.org/10.3389/fenvs.2021.619092>
- Semieniuk G, Taylor L, Rezai A, Foley DK (2021) Plausible energy demand patterns in a growing global economy with climate policy. *Nat Clim Change* 11(4):313–318

- Seo BC, Keem M, Hammond R, Demir I, Krajewski WF (2019) A pilot infrastructure for searching rainfall metadata and generating rainfall product using the big data of NEXRAD. *Environ Model Softw* 117:69–75. <https://doi.org/10.1016/j.envsoft.2019.03.008>
- Shahpari G, Sadeghi H, Ashena M, García D (2021) Drought effects on the Iranian economy: a computable general equilibrium approach. *Environ Dev Sustain*. <https://doi.org/10.1007/s10668-021-01607-6>
- Sharma S, Gupta YK (2021) Predictive analysis and survey of COVID-19 using machine learning and big data. *J Interdiscip Math* 24(1):175–195. <https://doi.org/10.1080/09720502.2020.1833445>
- Shelestov A, Lavreniuk M, Kussul N, Novikov A, Skakun S (2017) Exploring Google earth engine platform for big data processing: classification of multi-temporal satellite imagery for crop mapping. *Front Earth Sci* 5(February):1–10. <https://doi.org/10.3389/feart.2017.00017>
- Shih Y (2018) MASS HDFS: multi-agent spatial simulation hadoop distributed file system. In: MS capstone final report, MS in Computer Science & Software Engineering, Univ. of Washington Bothell
- Shykhmat A, Verses Z (2023) Selection of databases to store geospatial-temporal data. *Meas Comput Dev Technol Process* 4:7–12. <https://doi.org/10.31891/2219-9365-2023-76-1>
- Siddiqua A, Anwar A, Anwar MM (2021) Cotton yield and climate change adaptation in Pakistan: application of multinomial endogenous switching regression model. *J Bus Soc Rev Emerg Econ* 7(3):491–502
- Solomon R, Simane B, Zaitchik BF (2021) The impact of climate change on agriculture production in Ethiopia: application of a dynamic computable general equilibrium model. *Am J Clim Change* 10:32–50. <https://doi.org/10.4236/ajcc.2021.101003>
- Sweeney J, Salter-Townshend M, Edwards T, Buck CE, Parnell AC (2018) Statistical challenges in estimating past climate changes. *Wiley Interdiscip Rev Comput Stat* 10(5):e1437. <https://doi.org/10.1002/wics.1437>
- Tantalaki N, Souravlas S, Roumeliotis M (2019) Data-driven decision making in precision agriculture: the rise of big data in agricultural systems. *J Agric Food Inf*. <https://doi.org/10.1080/10496505.2019.1638264>
- Tao Z, Tian J, Pei Y, Yuan M, Zhang Y, Dai F (2020) A new coronavirus associated with human respiratory disease in China. *Nature* 579(7798):265–269. <https://doi.org/10.1038/s41586-020-2008-3>
- Tariq N et al (2019) The security of big data in fog-enabled iot applications including blockchain: a survey. *Sensors* 19(8):1–33. <https://doi.org/10.3390/s19081788>
- Tol RSJ (2024) A meta-analysis of the total economic impact of climate change. *Energy Policy* 185(2023):113922. <https://doi.org/10.1016/j.enpol.2023.113922>
- Türkeli S (2020) Complexity and the sustainable development goals: a computational intelligence approach to support policy mix designs. *J Sustain Res*. <https://doi.org/10.20900/jsr20200006>
- Van Der Linden S (2021) The gateway belief model (GBM): a review and research agenda for communicating the scientific consensus on climate change. *Curr Opin Psychol* 42:7–12. <https://doi.org/10.1016/j.copsyc.2021.01.005>
- Van Der Linden S, Leiserowitz A, Maibach E (2019) The gateway belief model: a large-scale replication. *J Environ Psychol*. <https://doi.org/10.1016/j.jenvp.2019.01.009>
- Vicedo-Cabrera AM et al (2021) The burden of heat-related mortality attributable to recent human-induced climate change. *Nat Clim Change* 11(6):492–500
- Volvach A, Kurbasova G, Volvach L (2023) Analysis and numerical simulation of temperature measurements made on earth and from space. *Heliyon* 9(2):e12999. <https://doi.org/10.1016/j.heliyon.2023.e12999>
- Wang F, Tian D (2022) On deep learning-based bias correction and downscaling of multiple climate models simulations. *Clim Dyn* 59:1–18
- Wang G, Hao X, Yao X, Wang J, Li H, Chen R (2023) Simulations of snowmelt runoff in a high-altitude mountainous area based on big data and machine learning models: taking the Xiyang river basin as an example. *Remote Sens* 15:1118
- Worku G, Teferi E, Bantider A, Dile YT (2020) Statistical bias correction of regional climate model simulations for climate change projection in the Jemma sub-basin, upper Blue Nile Basin of Ethiopia. *Theor Appl Climatol* 139(3–4):1569–1588. <https://doi.org/10.1007/s00704-019-03053-x>
- Xiang Z, Yan J, Demir I (2020) A rainfall-runoff model with LSTM-based sequence-to-sequence learning water resources research. *Water Resour Res*. <https://doi.org/10.1029/2019WR025326>
- Xu Y, Liu H, Long Z (2020) A distributed computing framework for wind speed big data forecasting on Apache Spark. *Sustain Energy Technol Assess* 37:100582. <https://doi.org/10.1016/j.seta.2019.100582>
- Yang R, Yu L, Zhao Y, Yu H, Xu G, Wu Y (2020) Big data analytics for financial market volatility forecast based on support vector machine. *Int J Inf Manage* 50:452–462. <https://doi.org/10.1016/j.ijinfomgt.2006.01.003>
- Yang B, Wu S, Yan Z (2022) Effects of climate change on corn yields: spatiotemporal evidence from geographically and temporally weighted regression model. *Int J Geo-Inf* 11:1–23
- Yariyan P et al (2020) Flood susceptibility mapping using an improved analytic network process with statistical models. *Geomat Nat Hazards Risk* 11(1):2282–2314. <https://doi.org/10.1080/19475705.2020.1836036>
- Zhang Z, Li J (2020) Big data mining for climate change. Elsevier, Amsterdam
- Zhu XX et al (2017) Deep learning in remote sensing: a review. *IEEE Geosci Remote Sens Mag Press* 41501462:1–60. <https://doi.org/10.1109/MGRS.2017.2762307>
- Zia S (2021) Climate change forecasting using machine learning SARIMA model. *IRASD J Comput Sci Inf Technol* 2(1):01–12

## Publisher's Note

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