Predicting future global temperature and greenhouse gas emissions via LSTM model

Ahmad Hamdan[1,2], Ahmed Al-Salaymeh[2], Issah M. AlHamad[3], Samuel Ikemba[4] and
Daniel Raphael Ejike Ewim[5,6*]

Abstract

This work is executed to predict the variation in global temperature and greenhouse gas (GHG) emissions resulting from climate change and global warming, taking into consideration the natural climate cycle. A mathematical model was developed using a Recurrent Neural Network (RNN) with Long–Short-Term Memory (LSTM) model. Data sets of global temperature were collected from 800,000 BC to 1950 AD from the National Oceanic and Atmospheric Administration (NOAA). Furthermore, another data set was obtained from The National Aeronautics and Space Administration (NASA) climate website. This contained records from 1880 to 2019 of global temperature and carbon dioxide levels. Curve fitting techniques, employing Sin, Exponential, and Fourier Series functions, were utilized to reconstruct both NOAA and NASA data sets, unifying them on a consistent time scale and expanding data size by representing the same information over smaller periods. The fitting quality, assessed using the R-squared measure, ensured a thorough process enhancing the model's accuracy and providing a more precise representation of historical climate data. Subsequently, the time-series data were converted into a supervised format for effective use with the LSTM model for prediction purposes. Augmented by the Mean Squared Error (MSE) as the analyzed loss function, normalization techniques, and refined data representation from curve fitting the LSTM model revealed a sharp increase in global temperature, reaching a temperature rise of 4.8 °C by 2100. Moreover, carbon dioxide concentrations will continue to boom, attaining a value of 713 ppm in 2100. In addition, the findings indicated that the RNN algorithm (LSTM model) provided higher accuracy and reliable forecasting results as the prediction outputs were closer to the international climate models and were found to be in good agreement. This study contributes valuable insights into the trajectory of global temperature and GHG emissions, emphasizing the potential of LSTM models in climate prediction.

Keywords　Climate change, Artificial intelligence, Global temperature, Carbon dioxide emissions, Weather forecasting, Recurrent Neural Network

*Correspondence:
Daniel Raphael Ejike Ewim
daniel.ewim@yahoo.com
[1] Cambridge Engineering Consultants, Amman, Jordan
[2] Mechanical Engineering Department, Faculty of Engineering and Technology, University of Jordan, Amman 11942, Jordan
[3] Mechanical and Aerospace Engineering Department, United Arab Emirates University, Al Ain, UAE
[4] Nigeria Atomic Energy Commission, Federal Capital Territory, Nigeria
[5] Department of Mechanical and Aerospace Engineering, The Ohio State University, Columbus, USA
[6] Department of Mechanical Engineering, Durban University of Technology, Durban, South Africa

Introduction

Climate change is a long-term variation in the average weather patterns that have come to define the Earth's local and international climates (Pickson et al., 2022). These fluctuations are causing a broad range of impacts on the Earth, humankind, animals, and plants. Away from the debate associated with the reasons for climate change, there is significant statistical evidence which indicates that the worldwide climate is currently changing (Dasgupta & Robinson., 2022). According to the report issued by the Intergovernmental Panel on Climate Change (IPCC, 2014), the world has witnessed a notable

climate change, which has caused the temperature values of several oceans to rise remarkably. These increases, in turn, has caused a growth in the frequency of extreme weather events, hurricanes, and tropical storms. These influences have resulted in a rise in the level of oceans, along with increased precipitation causing flash flooding in various coastal cities. It is reported that the industrial revolution contributed to a significant rise in carbon dioxide levels, which have reached 400 parts per million (ppm) compared with the 1880 levels in which carbon dioxide concentration in the atmosphere were recorded at 280 ppm (Little, 2020; Mooney, 2018). According to a study reported by the United States Environmental Protection Agency (EPA) in 2022 (United States Environmental Protection Agency, 2022), the world has witnessed a significant increase in carbon dioxide concentration, rising from 280 ppm in the late 1700s to 414 ppm nowadays due to the industrial revolution. This increase in carbon dioxide and other GHG emissions is remarkably high compared to pre-industrial levels.

The Kyoto Protocol located a group of seven GHG emissions which comprise Nitrogen Trifluoride (NF_3), CO_2, Hydrofluorocarbons (HFC), Perfluorocarbons (PFC), CH_4, N_2O, and Sulfur Hexafluoride (SF_6) which can remarkably influence the Earth's atmosphere, (NAEI-UK, 2022). The protocol sets limits for GHG emissions reduction targets in a group of industrial countries to be reduced by 5.2% in comparison with the emissions in 1990 (Najarzadeh et al., 2021). Among the seven gases, CO_2, CH_4, and N_2O have been identified by the European Commission as the most highly impacting gases on global warming (European Commission, 2018). These three gases have active lifetimes which last for decades. It strengthens the Paris Agreement, which sets an ambitious target of restricting the rise in global temperature to below 2 ℃ compared with the pre-industrial level (Chaichaloempreecha et al., 2022).

The GHG Global Warming Potential (GWP) was developed by international agencies to assess the impact of carbon dioxide. It can be described as the ratio of thermal trapping capability relative to the heat trap of carbon dioxide (AlHashmi et al., 2017). GWP factors convert GHG emissions data for non-CO_2 gases into units of carbon dioxide equivalent (CO_{2e}). By that, each $CO_2 = 1$ GWP, each $N_2O = 265$ GWP, and each $CH_4 = 28$ GWP (Pereira et al., 2022).

Studying climate change needs a long-time scale period of data (temperature and GHG concentrations) covering thousands of years. Unfortunately, temperature monitoring records only began in 1880 (Smith et al., 2008), while CO_2 began to be measured by NASA and UK-Meteorological in the 1950s. Although the Central England Temperature Data Series, which started in 1659, is the earliest continuous temperature records, the data from that series is unreliable due to the lack of precision equipment at the time. A thorough understanding of past climate change is needed to investigate present climate change and make future predictions. Researchers were able to rebuild the temperature record using tree rings to acquire temperature and GHG records for a period of 2000 years (Linderholm et al., 2018). Tree growth rings provide a rough history of the temperature, moisture, and cloudiness for each growing season.

The temperature and GHG record from 800,000 BC to the historical era might be recreated from ice cores. Measurements were taken from the center of the EPICA Dome C ice in Antarctica (Jouzel, et al., 1979). Ice cores are considered as a computer-based laboratory with extremely large memory. In this unique laboratory, scientists incorporate all of the current understanding about how the atmosphere, ocean, land, and ice function. Every layer of ice contains a tale about the state of the planet at the time that the coating of snow fell (Eyrikh, 2022). As snow deposits onto a growing glacier, the icy layers hold aerosol molecules, including sea salts, trace elements, pollen, ash, and dust, that were in the atmosphere at that interval. These molecules remain in the ice, offering physical evidence of previous worldwide events. The National Oceanic and Atmospheric Administration (NOAA) obtained climate data, including global temperature, carbon dioxide, nitrous oxide, and methane levels from ice core samples between 800,000 BC and 1950 (National Centers for Environmental Information (NCEI) 2022; National Centers for Environmental Information (NCEI) 2022). The data pattern indicated that the temperature and GHGs change in the atmosphere periodically. This can be categorized as the natural climate cycle, where no human influence existed. Model data can observe ice ages with significant temperature variations, approximately every 100,000 years. This cycle has moderately changed, and smaller changes in temperature are noted every 41,000 years. Hence, if nature had been left to itself, the Earth should be going into a cooling phase instead of getting warmer nowadays, and we should be living in an ice age for the next few tens of thousands of years.

Milankovitch's theory was developed in the 1920s (Buis, 2020). It describes the collective impacts of variations in the Earth's motion on the global climate over thousands of years. Fluctuations in Earth's eccentricity, axial tilts, and initiation caused a cyclic difference in solar radiation accessing the planet. The variation in the degree of Earth rotation and tilt angles range between 22.1 and 24.5° perpendicular to the Earth's orbital plane.

Milankovitch's calculations indicated that ice ages take place approximately every 41,000 years (Buis, 2020).

Hamdan *et al. Sustainable Energy Research*   (2023) 10:21

Page 3 of 13

According to NASA (Carbon dioxide concentration, 2022), the tilt degree of the axis reached recently around 23.4°. This angle can continuously and slowly decrease in a cycle that extends about 41,000 years. It attained a maximum incline about 10,700 years ago, and it will reach its lowest level roughly 9800 years from now. These changes can cause a lower deflection, making winters warmer and summers cooler, enabling time for higher snow and ice to accumulate and promoting more cooling.

Against this backdrop, this research aims to understand the intricate relationship between climate change and the anthropogenic activities that intensify it. Leveraging extensive historical data sets from reputable sources, such as NOAA and NASA, the study endeavors to simplify the multifaceted nature of climate change for educational purposes, thereby aiming to foster better comprehension and insight. The intention is to utilize artificial intelligence (AI) methodologies to analyze vast amounts of climate data, enabling a clearer perspective on past, present, and potential future climatic trends. This would not only offer an overview of the natural and current climate cycle but also draw correlations between pivotal variables, such as temperature shifts and greenhouse gas concentrations.

Climate change models

Climate change modeling is a mathematical representation of the components of the climate system, including equations and formulas representing physical and chemical processes and their components, such as the temperature, wind, ocean variables, melting of ice sheets, sea rise levels, soil moisture drying, and many other climate system variables (Neelin, 2011). These equations are usually solved using numerical methods due to their complexity and the enormous number of variables involved. The roots of climate change modeling can be traced back to Jule Gregory Charney (Mathez & Smerdon, 2018), being the first to use computers for predicting weather conditions numerically in 1946, which paved the ground for more complex and modern climate modeling. Climate models are used for multiple purposes that range from studying the dynamics of the climate system to future climate predictions. Energy balance models (EBMs) are the simplest implementation among all the available climate models. EBMs are based on the assumption that the energy received by the Earth from the Sun is balanced by the energy radiated from the Earth back into space. The whole Earth is assumed to be at a single averaged temperature value and to behave as it is one body being heated up by the Sun and releasing its energy back into space.

General Circulation Models (GCMs) are highly complex and transverse models based on the Navier–Stokes equations, which are a group of differential equations that allow modelling the atmosphere as a continuous, compressible liquid. By converting the equations to a rotating reference frame in spherical Earth's coordinates, the primary motion equations for a "portion" of air in each direction can be obtained. Navier–Stokes equations determine the thermodynamic properties of the atmosphere of the Earth. The Geophysical Fluid Dynamics Laboratory model (GFDL) is an early GCM model and was the first climate change model by NOAA (Delworth et al., 2006). The model combined both atmospheric and oceanic processes and interaction and their influence on future climate.

Table 1 presents a comprehensive overview of predictions for global temperature rise and carbon dioxide (CO_2) levels at the year 2100 from some of the most

Table 1 Most common climate change models and their predictions at the year 2011

Model	Prediction at 2100		Methods/Software	Reference
	Temp rise (ºC)	CO_2 level (ppm)		
Geophysical Fluid Dynamics Laboratory (GFDL)	3.3	671	Python	Delworth et al., 2006)
Max Planck Institute for Meteorology(MPI-M)	3	704	OASIS CYLC (Python-based)	Baehr et al., 2015)
Hadley Centre	3.7	690	Hadley Centre special software	Pope et al., 2007)
National Center for Atmospheric Research Parallel Climate Model (NCAR PCM)	2.3	708	EOL public software packages	Washington et al., 2000)
National Center for Atmospheric Research Climate System Mode (NCAR CSM)	2.2	711	EOL public software packages	Boville et al., 2001)
Canadian Centre for Climate Modelling and Analysis (CCCma)	4	710	ArcGIS 10.5 FClimDex (R-based)	Salzen et al., 2013)
Australia's Commonwealth Scientific and Industrial Research Organization (CSIRO)	3.8	712	CSIRO special software	Gordon, et al., 2002)
Center for Climate Research Studies (CCSR) National Institute for Environmental Studies (NIES) (CCSR/NIES)	4.7	713	Statgraphics 19	Dairaku et al., 2003)

common climate change models. Each model employs distinct methods and software for their predictions, contributing to the diversity of projections. For example, the Geophysical Fluid Dynamics Laboratory (GFDL) model (Delworth et al., 2006) anticipates a temperature rise of 3.3 °C and a CO_2 level of 671 ppm at year 2100. Moreover, within the models presented, the National Center for Atmospheric Research Climate System Model (NCAR CSM) (Boville et al., 2001) stands out with a comparatively optimistic projection, estimating a temperature rise of 2.2 °C and a CO_2 level of 711 ppm at year 2100. Conversely, the Canadian Centre for Climate Modelling and Analysis (CCCma) model (Salzen et al., 2013) portrays a scenario characterized by a more conservative degree of warming, projecting a temperature increase of 4 °C with a CO_2 concentration of 710 ppm. These diverse models contribute to the understanding of potential future climate scenarios and highlights the complexity of climate modeling methodologies.

Materials and methods

This work depends on a numerical analysis through which some weather parameters are forecast until 2100 AD (NASA data set) and until 50000 AD (NOAA data set). A recurrent Neural Network (RNN) is employed to achieve the study goal. RNN is a branch of Deep Learning (DL) models that can handle sequential data of the form $(x_1, x_2, x_3...., x_n)$, such as DNA sequences, textual data, and time-series data (Goodfellow & Courville, 2016). The methodology of this work is shown in Fig. 1.

Two databases were employed for the research purposes which are available from the National Oceanic and Atmospheric Administration agency (NOAA) (National Centers for Environmental Information (NCEI) 2022, National Centers for Environmental Information (NCEI) 2022) and National Aeronautics and Space Administration Agency (NASA) (Carbon dioxide concentration, 2022) in a form of time-series data. With respect to NOAA data, it is represented in five variables, including temperature measured in °C, CO_2 measured in parts per millions (ppm), CH_4 and N_2O measured in parts per billion per volume (ppbv), and Earth's angle measured in (degrees). The time span at which the data represents is 240,000 years from 238,050 B.C to 1950 A.D. On the other hand, NASA's data are represented in three variables, including temperature (°C), CO_2 (ppm) and Irradiation (W/m^2). The time periods which NASA's data span is (1880–2019) for the temperature, (1958–2019) for CO_2 concentration, and (1950–2019) for irradiation in a form of one measurement per year (Carbon dioxide concentration, 2022).
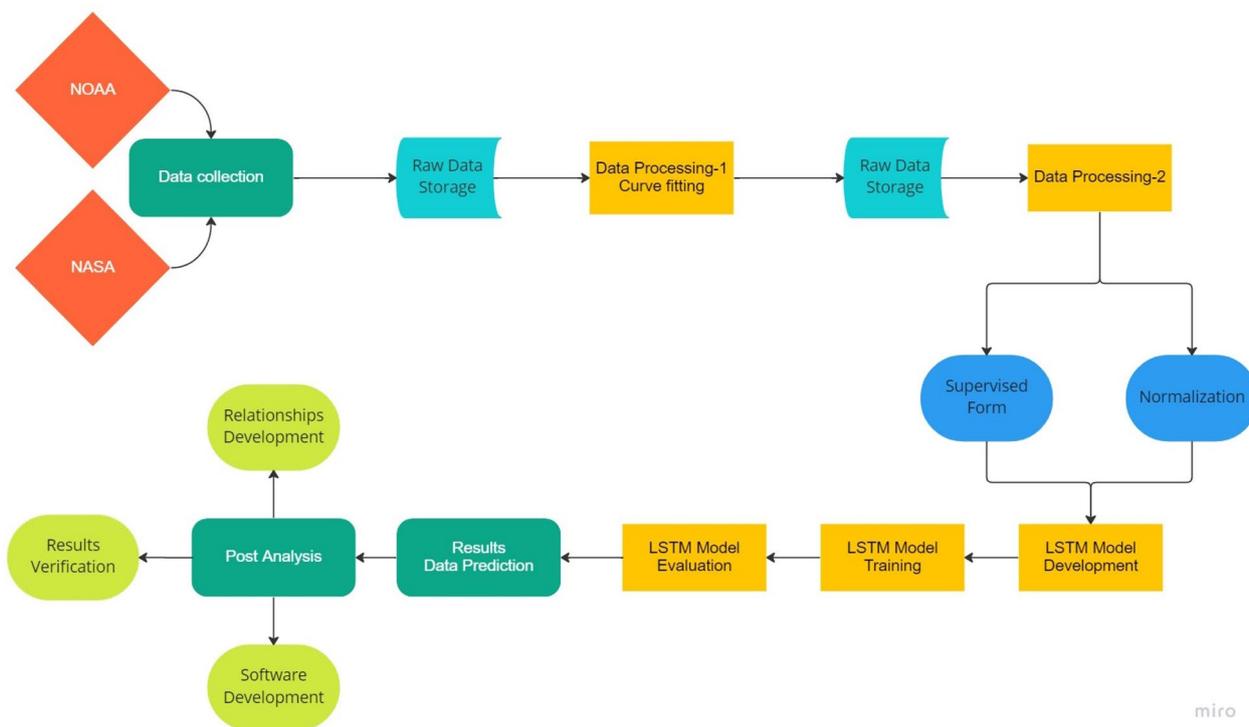

Fig. 1 Flow chart for this work methodology

Both data sources are provided at different time scales and steps. Thus, a curve fitting has been used to reconstruct both NOAA and NASA data sets for two main purposes; the first is to bring the data into a unified time scale, and the second is to increase the data size by representing the same data over a smaller period of time (on a yearly basis or daily basis according to the size of the original data). Three curve fitting functions were employed for this purpose, namely, Sin function, exponential function, and Fourier Series function. In addition, R-squared measure was employed as

an evaluation metric to judge the fitting quality. Figures 2, 3 show NOAA and NASA data sets curve fitting output for both temperature and CO_2 concentration, respectively.

The curve fitting for NOAA data sets (Fig. 2) demonstrates excellent conformance, showcasing a close alignment between the actual temperature and CO_2 levels and the fitted values. However, for NASA data sets (Fig. 3), the curve fitting results are noteworthy, with the actual temperature closely mirroring the fitted temperature, resulting in an almost perfect match. A similar pattern
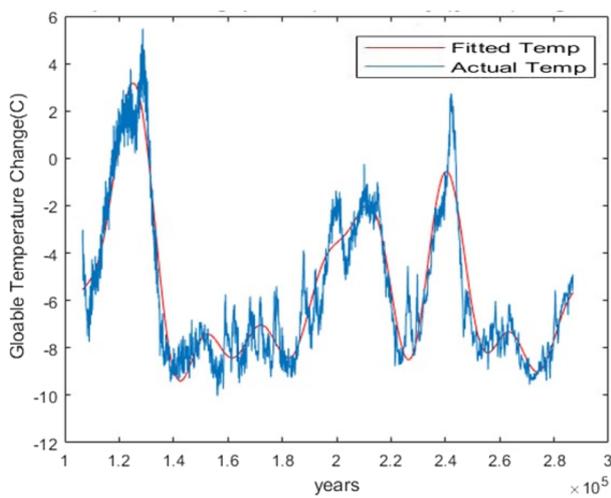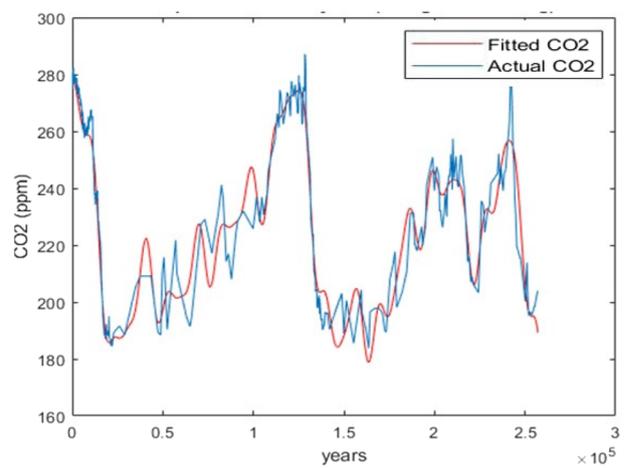

(a)   (b)

Fig. 2 NOAA curve fitting output using sin function for (**a**) Temperature and (**b**) CO_2 concentration
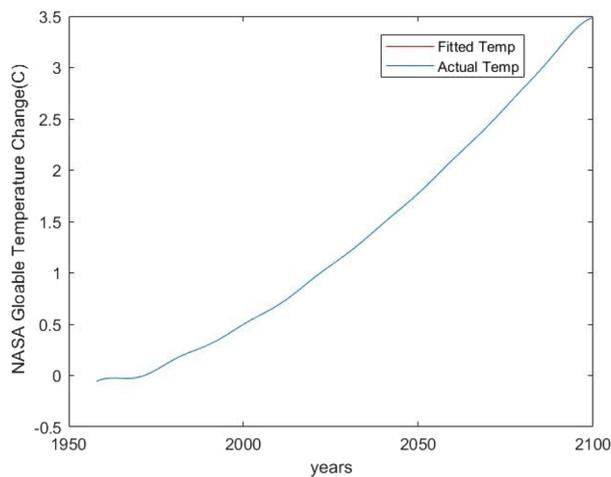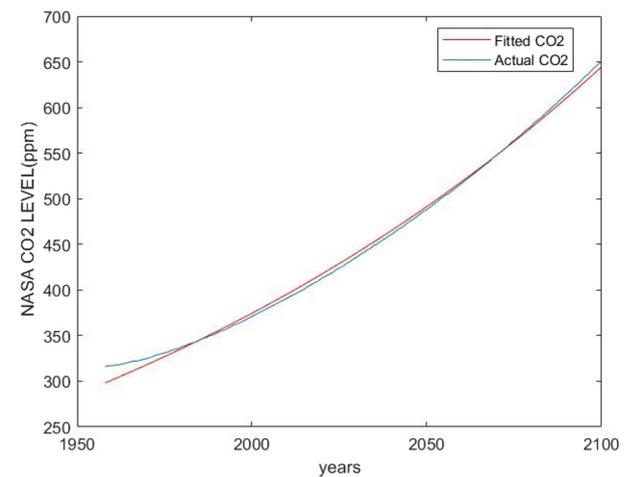
(a)   (b)

Fig. 3 NASA curve fitting output using exponential function for (**a**) Temperature and (**b**) CO_2 concentration

is observed for CO_2 levels, where the conformity is strikingly close, albeit with a minor degree of variation.

After the curve fitting process, data are now clean and ready for the second phase of data processing. For the GHG levels, there is a clearly noticed variation in the ranges of each variable, where CO_2 data values come in hundreds, while the rest of the variables have relatively small data values. For example, N_2O and CH_4 levels are of the order of ppbv, while Co_2 concentration is of the order of ppm. Having such data may affect the prediction model's performance. As a solution, all variables in both data sets were converted from their original values into a normal scale with mean ($\mu=0$) and standard deviation ($\sigma=1$), hence, all data values have been brought into the same scale with homogeneous representation.

Ultimately, time-series data in its current format is not suitable to be used directly with machine learning models for prediction purposes. Therefore, it must be converted from its time indexed format into a supervised format to overcome this issue. By supervised form of the data, we mean representing the data set as an input pattern (X) that has a certain output or target pattern (y). As a result, an algorithm can be fitted on the supervised data to learn how to map the input pattern to the target pattern (Cord & Padraig., 2008). As a result; both NOAA and NASA data sets were converted into supervised data patterns with one past data at timestep (t-1) point is used to predict the current data point (t) as follows:

$$(< x_1^{(t-1)}, x_2^{(t-1)}, ...x_n^{(t-1)} >: < x_1^t, x_2^t, ...x_n^t >) \quad (1)$$

where n is the number of input variables and t is the time step.

Consequently, having such data representation enables curve fitting for the learning algorithms and making future predictions. This representation resulted in 24,000 and 22,265 data points for NOAA and NASA data sets, respectively.

Recent advancements in the area of recurrent neural networks has contributed to the revolutionization of several disciplines in various applications, including weather forecasting, time-series anomaly detection, natural language processing, and healthcare. In its simplest form, RNN can be recognized by its potential to recall previous data to be employed in forecasting future scenarios of some variables. Nonetheless, RNN faces some issues mirrored by the longer term dependency, where the network can place larger weights for the latest inputs (xt) and lower weights for the farther past inputs ($xt-s$). Therefore, the network begins to forget the initial information defined. To address this issue, another classification related to RNNs was developed by Hochreiter et al. in 1997 (Hochreiter & Schmidhuber., 1997) known as the

Long–Short-Term Memory (LSTM) model. LSTM can be described as a practical algorithm that can effectively overcome the challenge of longer term dependency. It relies on two primary operations: (1) maintaining training data that is greatly possible to employ for future prediction and (2) forgetting the least essential information (Salehinejad et al., 2018). A group of three parameterized gates is utilized to monitor and control this operation. These gates include (a) forget gate, (b) input gate, and (c) output gate. The cell state vector can serve as model memory. At the same time, the hidden state vector (ht) generates the model outputs. Figure 4 represents a principle building block associated with the LSTM network.

It worth noting that in the domain of time series forecasting, the Auto Regressive Integrated Moving Average (ARIMA) model stands as a well-established method known for its effectiveness in capturing temporal dependencies. ARIMA model was developed by Box and Jenkins in the 1970s (Wong et al., 2005). While ARIMA has been widely employed in climate-related studies, the current work opted for an alternative approach utilizing RNN–LSTM model. The decision to choose LSTM was motivated by its inherent ability to handle long-term dependencies in sequential data, a crucial characteristic for modeling climate variables.

According to the exact functions related to LSTM gates, the critical role of 'forget' gate (ft) is to eliminate the least essential data from the cell state vector (Ct). Mathematically, this operation can be executed by calculating the sigmoid function of the weighted summation of the current input (xt) and the past hidden state (ht−1) to produce a binary mask which can be multiplied by the previous cell states (Ct−1) to update the current cell state (Ct−1) amounts. LSTM architecture was fitted on both data sets to generate future forecasting. The model comprises two main layers: (1) the first layer is the LSTM layer with 512 hidden units for NOAA data and 1024 hidden units related to the NASA database, and (2) the second layer is a dense layer with three units for NASA data
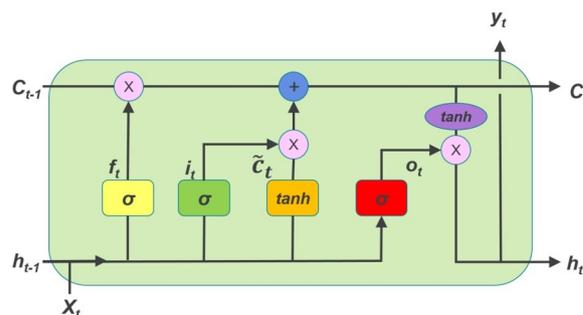

Fig. 4 LSTM building block (Aya Abdelsalam Ismail et al. 2018)

Hamdan *et al. Sustainable Energy Research*   (2023) 10:21

Page 7 of 13

and five hidden units for NOAA data. The activation of the LSTM layer is set to the default values. Meantime, the dense layer activation is set to linear activation to attain numerical predictions.

The choice of the appropriate loss function plays a crucial role in training machine learning models. In developing the current LSTM model for predicting future global temperature and greenhouse gas concentrations, different loss functions were considered, including Mean Absolute Error (MAE) and Mean Squared Error (MSE). While MAE is a viable option for regression tasks, MSE assigns greater emphasis to larger errors, aligning with the goal of achieving precise numerical predictions for climate parameters. In the context of climate science, where small variations can have substantial implications, the sensitivity of MSE to deviations proved advantageous. Hence, the decision to employ MSE was driven by the specific requirements of our forecasting objectives. The Mean Squared Error (MSE) was analyzed as a loss function along with an adaptive momentum optimization method with a learning rate value of (0.01) (Rehman & Nawi, 2011). Furthermore, both data sets were split as (80%) for training and (20%) for verification with a batch size of 32 training examples for the NASA data set and 64 training examples for the NOAA data set. Every model was fitted to the training data for 25 epochs and evaluated in terms of root MSE with a threshold (≤ 5). Table 2 summarizes the experimental settings related to each model.

Climate parameters prediction according to NOAA

Figure 5 shows the prediction performance related to the LSTM model according to the NOAA test set. It can be inferred from Fig. 5 that the model succeeded in correctly predicting the test set with minor differences between the actual and predicted values. The ultimate root indicates the squared error that has been attained on the NOAA data set is 2.018, which is lower than the

prespecified Root Mean Squared Error (RMSE) threshold of 5. Therefore, LSTM can provide higher reliability and effectiveness due to the achievement of such results, making it practical to generate future predictions and forecast weather.

Figure 6 illustrates the actual natural climate cycle with the forecasted values for global temperature, CO_2, N_2O, and CH_4 concentrations from 1950 to 5000 AD. From Fig. 6, it can be obviously concluded that the forecasted quantities for every parameter related to the NOAA data set followed a similar behavior to the actual data. These results can indicate good validation and significant accuracy of the forecasted data. Furthermore, it can be observed that for the next 50000 years, the world will be in an Ice Age, reaching temperature values of roughly − 8 ℃ in the year 35000 AD. Then, the temperature is going to increase in subsequent years. For carbon dioxide, the prediction indicates that the CO_2 concentration will decline by approximately (35 ppm) in 35000 AD, reaching 205 ppm before starting to increase in the subsequent years. In addition, both methane and nitrous oxide possess roughly the same behavior over the prediction period in which both variables exhibit a reduction trend, reaching 0.24 and 0.45 ppm for N_2O and CH_4, respectively.

Climate parameters prediction according to NASA

Figure 7 shows the prediction performance related to the LSTM model based on the NASA test. From Fig. 7, it can be inferred that the model succeeded in generalizing the test set with minimal differences between the actual and predicted values. Notwithstanding, the maximum RMSE, according to the NASA test, is (0.814), which is lower than the prespecified RMSE threshold of five. Therefore, the LSTM model can be reliable and trustable to forecast different weather parameters with higher accuracy.

Future climate prediction results

Climate parameters are usually predicted for the year 2100 as seen in Table 1. Climate models' predictions of the global temperature for year 2100 range between a temperature rise of 2.2 ℃ (Boville et al., 2001) to 4.7 ℃ (Dairaku et al., 2003), while CO_2 concentration predictions for the same year range between 671 ppm (Delworth et al., 2006) to 713 ppm (Dairaku et al., 2003). Figure 8 presents a profile of the global temperature rise between 1958 and 2100. The figures are composed of two parts. The first part is indicated by the period from 1958 to 2020, which is the actual data set received from NASA, the second part is the future prediction from 2020 to 2100, which is generated from the data future projections based on NASA data set. Moreover, Figure 9 shows the actual CO_2 assessment between

Table 2 Overview of the experimental setting

Setting	NOAA model	NASA Model
LSTM units	512	1024
Dense units	5	3
LSTM activation	tanh	tanh
Dens activation	linear	Linear
Loss function	MSE	MSE
Optimizer	Adam	Adam
Learning rate	0.01	0.01
Batch size	64	32
Epochs number	25	25

Hamdan *et al. Sustainable Energy Research*   (2023) 10:21

Page 8 of 13
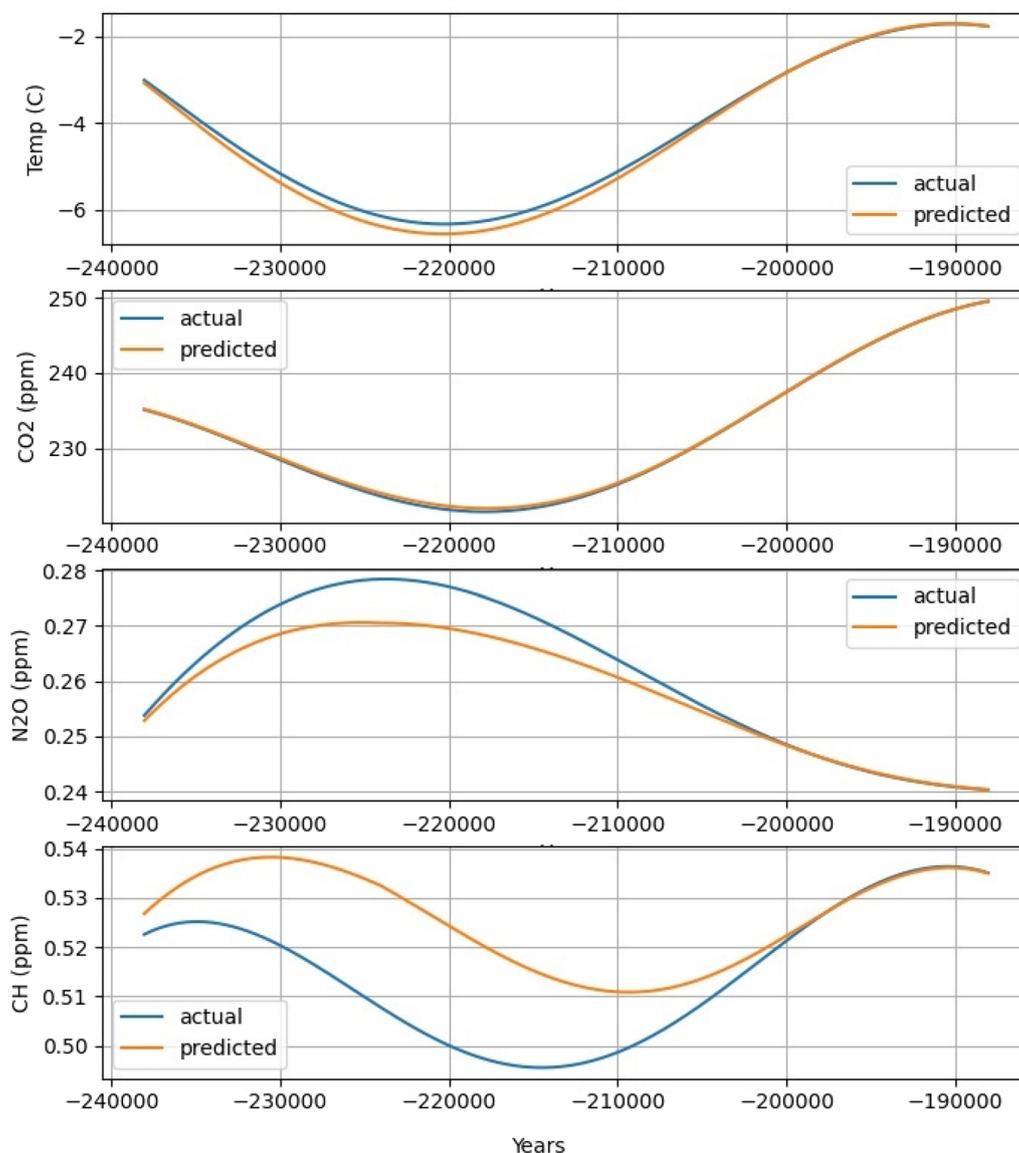
Fig. 5 Results of the performance verification related to the LSTM model according to the NOAA test (The negative numbers in the *x*-axis indicates years in the past)

1958 and 2019, followed by the predicted CO_2 profile from 2020 to 2100. Both figures show that the increasing trend of global temperature rise and carbon dioxide concentrations will continue to years 2100 and beyond.

Zooming into the period from 2020 to 2100. Figure 10 shows future predictions of both the global temperature rise and CO_2 concentration for the same period. It can be clearly seen that the global temperature rise is expected to increase from 1 ℃ in 2020 to 4.8 ℃ in 2100. Moreover, the CO_2 concentration is also expected to keep rising from 430 ppm in 2020 to 710 ppm in 2100.

The proposed model prediction results, based on NOAA from year 2020 to year 2100, is shown in Fig. 11. The figure shows that in 2020, the temperature fall based on the natural climate cycle is about − 2 ℃ and CO_2 concentration is around 241.686 ppm. By 2100, the temperature and CO_2 are expected to further fall down to almost − 2.02 ℃ and the 241.678 ppm. This is very close, or almost the same value as 2020, as 80 years are not long enough to witness any change in the natural climatic cycle. This proves that if nature was left to itself, without any interference from humans, Earth should be in an ice age during this period and for the
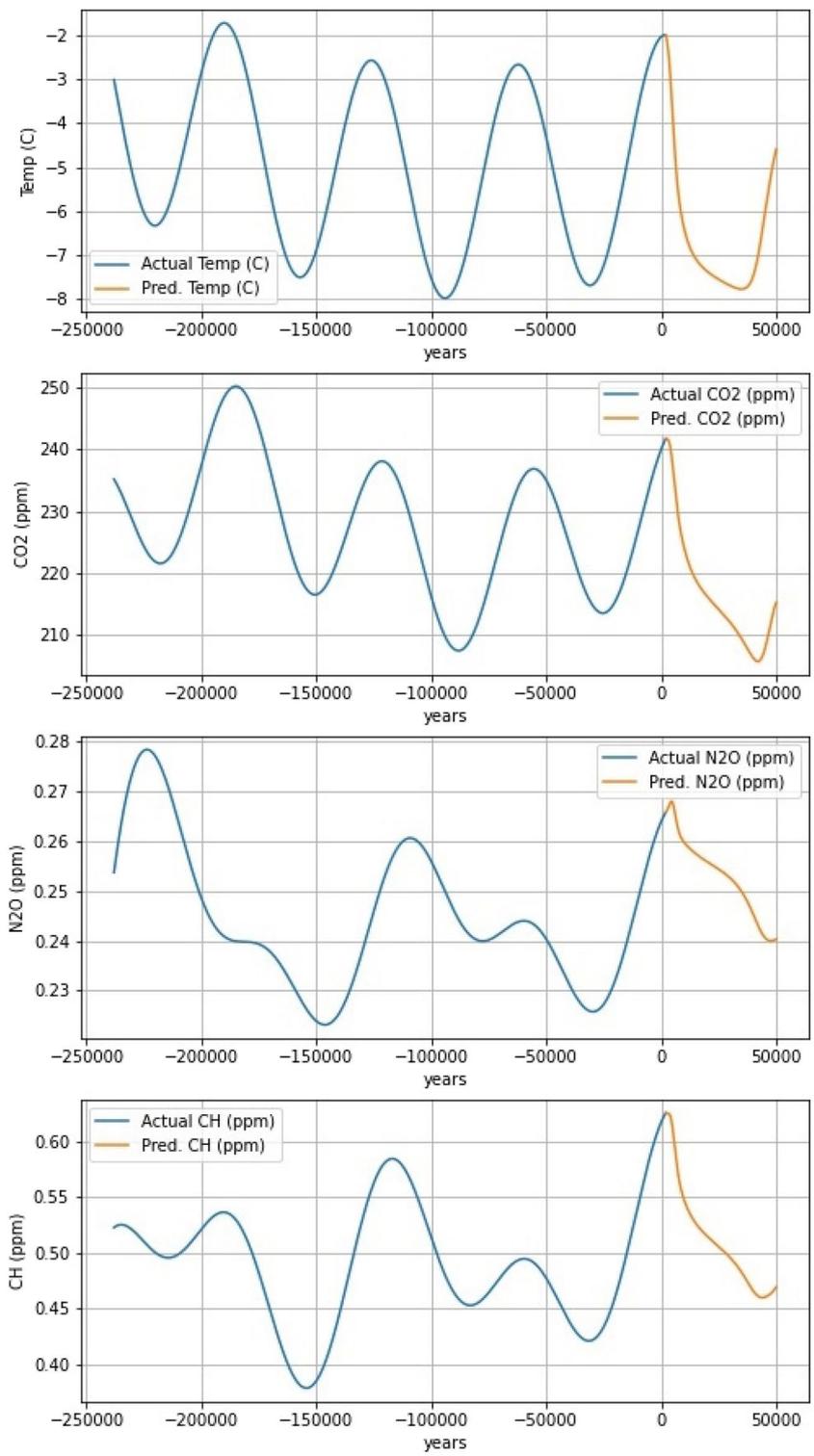
Fig. 6 NOAA actual data with prediction outputs. (The negative numbers in the *x*-axis indicates years in the past)

Hamdan *et al. Sustainable Energy Research*   (2023) 10:21

Page 10 of 13
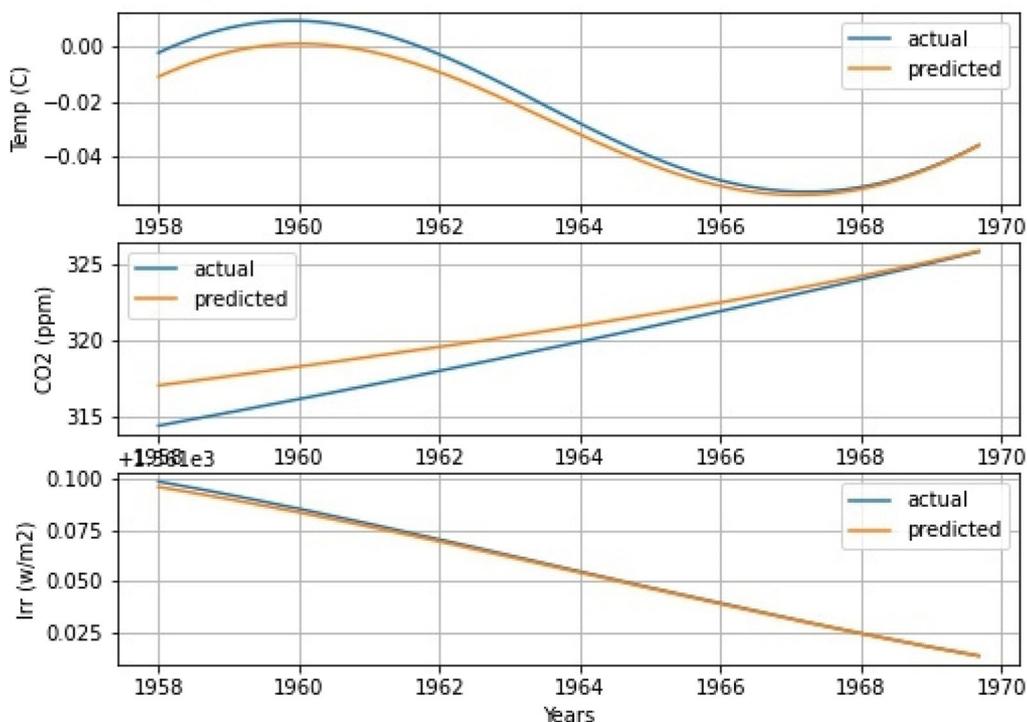
Fig. 7 The evaluation of the LSTM Model performance according to the NASA test
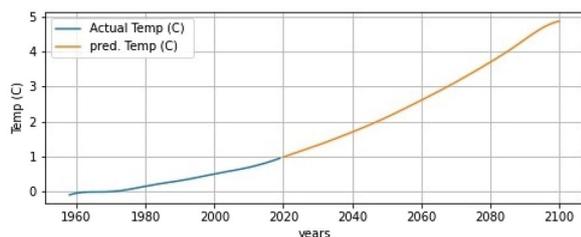
Fig. 8 Actual and predicted values of temperature trend
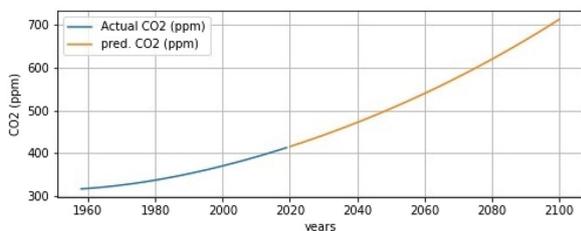
**Fig. 9** Actual and predicted values of CO_2 trend
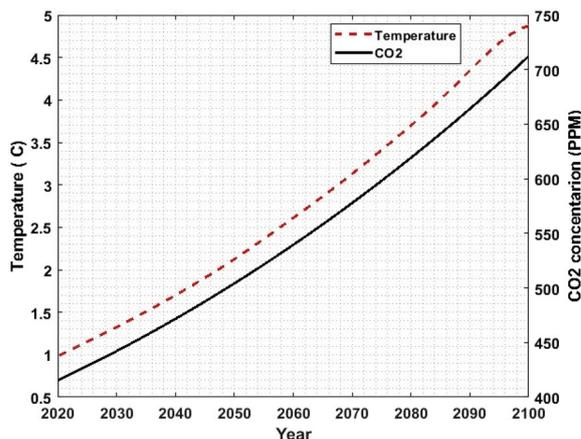
Fig. 10 Temperature rise (°C) and CO_2 concentration (ppm) prediction from 2020 to 2100 based on NASA data

next 40,000 years (National Centers for Environmental Information (NCEI) 2022). Hence, the data from NOAA cannot be used for future prediction of the climate parameter. Yet, it can illustrate the natural climate cycle performance and show how much we deviated from it due to human activities.

Discussion

The results of this work indicate that concentrations related to major types of GHGs, such as carbon dioxide, methane, and nitrous oxide, will tend to increase sharply to year 2100. In addition, the research findings reveal that global temperature will continue to rise in the future and will reach 4.8 °C by 2100. The numerical outputs of this work are consistent with the results of several scholars, who conducted an analysis and forecasted the
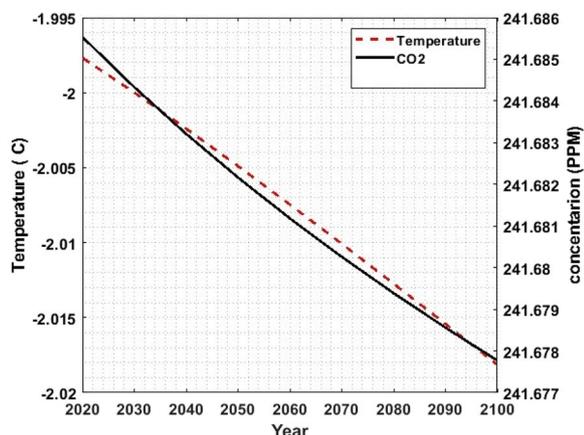
Fig. 11 Temperature rise (°C) and CO_2 concentration (ppm) prediction from 2020 to 2100 based on NOAA data

temperature and carbon dioxide emissions by 2100 and found that the levels of the worldwide temperature and CO_2 emissions will reach values greatly similar to the amounts obtained in this study. Figure 12a, b show the temperature rise and carbon concentration predictions at year 2100, respectively, for different climate models, including the proposed model.

It is important to consider the inherent limitations associated with employing LSTM models for climate prediction. Despite the promising results in forecasting future global temperature and greenhouse gas concentrations, these models exhibit sensitivity to the size and representativeness of the training data set. The complex and nonlinear nature of climate data requires diverse and extensive historical information, and the performance of LSTM models can be influenced by the availability of such data. Furthermore, the challenge of capturing

abrupt changes or extreme events poses a limitation to the model's effectiveness in certain scenarios. In addition, the interpretability of LSTM models remains a challenge, as their capacity to learn intricate patterns may come at the cost of understanding the underlying physical mechanisms that drive predictions.

Conclusions

This work is carried out by predicting the change in global temperature and concentrations of GHG emissions variation resulting from climate change and global warming, taking into account the natural climate cycle. A mathematical model was developed using the RNN and the LSTM model based on two data sets. The first raw data set was obtained from the National Oceanic and Atmospheric Administration (NOAA) for global temperature, carbon dioxide, nitrous oxide, and methane from ice core samples between 800,000 BC and 1950. The other data set was obtained from the National Aeronautics and Space Administration (NASA) climate database from 1880 to 2019 for the global temperature and from 1950 to 2019 for the carbon dioxide levels. The RNN algorithm (LSTM model) provided higher accuracy and more reliable forecasting results as the prediction outputs were closer to the international climate models. The data from NOAA revealed that, based on the natural climate cycle, which is repeated almost every 41,000 years, we should expect a temperature drop of almost 2 °C from 2020 to 2100 and CO_2 concentrations of almost 240 ppm for the same period. These climate parameters indicate that we should be living in an ice age based on the natural climate cycle, which is indicated by human activities. However, based on the numerical analysis and forecasting using the LSTM model based on the NASA data set,
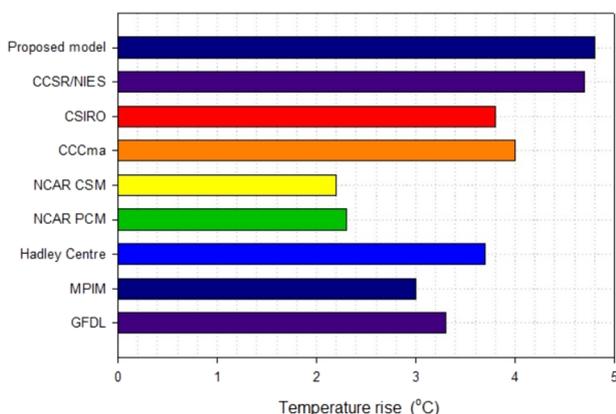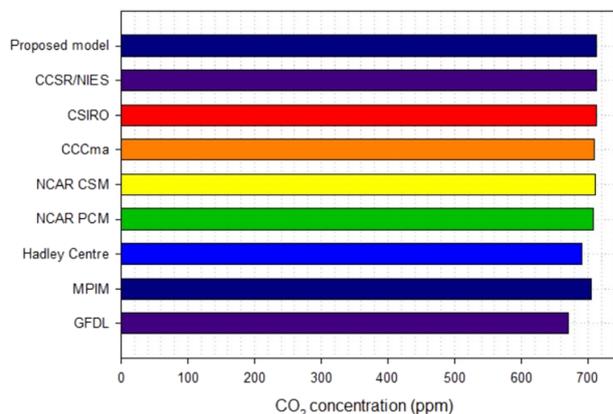

(a)   (b)

Fig. 12 Global temperature rise (**a**) and CO_2 concentrations (**b**) at year 2100 for different climate models including the proposed model

Hamdan *et al. Sustainable Energy Research*   (2023) 10:21

Page 12 of 13

it was found that the global temperature rise shows a trend of a sharp increase, and is expected to reach a value of 4.8 °C by 2100, while the carbon dioxide concentrations will continue to boom, and are expected to reach a value of 713 ppm in 2100. According to the Paris Agreement, global warming should be limited to 2 °C, preferably at 1.5 °C. However, the history of climate parameters influenced by human activities, which allowed this study to predict the future, shows that climate change consequences display a real challenge to the world's climate.

Availability of data and materials
Data are available upon request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests, financial or non-financial, that could be perceived as influencing the content or conclusions of this paper.

References
AlHashmi, M., Haider, H., Hewage, K., & Sadiq, R. (2017). Energy efficiency and global warming potential in the residential sector: comparative evaluation of Canada and Saudi Arabia. *Journal of Architectural Engineering, 23*, 3. https://doi.org/10.1061/(ASCE)AE.1943-5568.0000253

Aya Abdelsalam Ismail, Timothy Wood, Héctor Corrada Bravo. Improving Long-Horizon Forecasts with Expectation-Biased LSTM Networks. *arXiv preprint arXiv*, pp. 1–9, 2018.

Baehr, J., et al. (2015). The prediction of surface temperature in the new seasonal prediction system based on the MPI-ESM coupled climate model. *Climate Dynamics, 44*(9–10), 2723–2735. https://doi.org/10.1007/s00382-014-2399-7

Boville, B., et al. (2001). Improvements to the NCAR CSM-1 for transient climate simulations. *Journal of Climate, 14*(2), 164–179.

Buis, A. (2020). *Milankovitch (Orbital) cycles and their role in earth's climate*. USA: NASA.

Carbon dioxide concentration (2022) NASA. NASA. https://climate.nasa.gov/vital-signs/carbon-dioxide/. Accessed: December 3, 2022.

Chaichaloempreecha, A., Chunark, P., Hanaoka, T., & Limmeechokchai, B. (2022). Thailand's mid-century greenhouse gas emission pathways to achieve the 2 degrees celsius target. *Energy Sustain Soc, 12*(1), 22. https://doi.org/10.1186/s13705-022-00349-1

Cord, M., & Cunningham, P. (2008). *Machine learning techniques for multimedia*. Berlin: Springer.

K Dairaku, S Emori, T Nozawa, N Yamazaki, M Hara, H. Kawase. Regional climate simulation over Asia under the global warming nested in the CCSR/NIES AGCM. In Proceedings of the Symposium on Water Resource and Its Variability in Asia in the 21st Century (pp. 756–764). 2003.

Dasgupta, S., & Robinson, E. J. Z. (2022). Attributing changes in food insecurity to a changing climate. *Science and Reports, 12*(1), 4709. https://doi.org/10.1038/s41598-022-08696-x

Delworth, T. L., et al. (2006). GFDL's CM2 global coupled climate models. Part I: formulation and simulation characteristics. *Journal of Climate, 19*(5), 643–674. https://doi.org/10.1175/JCLI3629.1

European Commission. Supporting policy with scientific evidence, Greenhouse gas. European Commission. 2018.

Eyrikh, S. S. (2022). Mercury in paleoarchives as a proxy of environmental and climate changes. *Limnol Freshw Biol, 3*, 1355–1358. https://doi.org/10.31951/2658-3518-2022-A-3-1355.

Goodfellow, I., Bengio, Y., Courville, A. (2016). Sequence modeling: recurrent and recursive nets. *Deep learning*, pp. 367–415.

HB Gordon et al. The CSIRO Mk3 climate system model. Report number: 60. 2002.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

IPCC. Intergovernmental Panel on Climate Change, Fifth Assessment Report. Cambridge, 2014.

Jouzel, J., et al. (2007). Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science, 317*(5839), 793–796. https://doi.org/10.1126/science.1141038

Linderholm, H. W., et al. (2018). Arctic hydroclimate variability during the last 2000 years: current understanding and research challenges. *Climate of the past, 14*(4), 473–514. https://doi.org/10.5194/cp-14-473-2018

Little, S. (2020). Message in a fossil? Lessons from the last plants on Antarctica. *Weather, 75*(1), 30–31. https://doi.org/10.1002/wea.3519

Mathez, E., & Smerdon, J. (2018). *Climate change: the science of global warming and our energy future*. Cambridge: Columbia University Press.

Mooney, C. (2018). *Earth's atmosphere just crossed another troubling climate change threshold*. Washington: The Washington Post.

NAEI-UK. (2022). *Overview of greenhouse gases*. UK: National Atmospheric Emissions Inventory.

Najarzadeh, R., Dargahi, H., Agheli, L., & Khameneh, K. B. (2021). Kyoto protocol and global value chains: trade effects of an international environmental policy. *Environment and Behaviour, 40*, 100659. https://doi.org/10.1016/j.envdev.2021.100659

National Centers for Environmental Information (NCEI) (no date) National Centers for Environmental Information (NCEI). https://www.ncdc.noaa.gov/paleo/study/6080. Accessed: December 3, 2022.

National Centers for Environmental Information (NCEI) (no date) National Centers for Environmental Information (NCEI). https://www.ncdc.noaa.gov/paleo/study/2426. Accessed: December 3, 2022.

Neelin, D. (2011). *Climate change and climate modeling*. Cambridge: Cambridge University Press.

O R Edenhofer et al. IPCC, 2014: Summary for Policymakers. In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. United Kingdom and New York, NY, USA. 2014.

Pereira, J. L. S., Perdigão, A., Tavares, A., Silva, M. E. F., Brás, I., & Wessel, D. F. (2022). Effects of the addition of different additives before mechanical separation of pig slurry on composition and gaseous emissions. *Agronomy, 12*(7), 1618. https://doi.org/10.3390/agronomy12071618

Pickson, R. B., He, G., & Boateng, E. (2022). Impacts of climate change on rice production: evidence from 30 Chinese provinces. *Environment, Development and Sustainability, 24*(3), 3907–3925. https://doi.org/10.1007/s10668-021-01594-8

Pope, V., et al. (2007). The met office hadley centre climate modelling capability: the competing requirements for improved resolution, complexity and dealing with uncertainty. *Philosophical Transactions of the Royal Society a: Mathematical, Physical and Engineering Sciences, 365*(1860), 2635–2657. https://doi.org/10.1098/rsta.2007.2087

Rehman, M. Z., & Nawi, N. M. (2011). *The effect of adaptive momentum in improving the accuracy of gradient descent back propagation algorithm on classification problems*. Berlin: Springer.

Hojjat Salehinejad, Sharan Sankar, Joseph Barfett, Errol Colak, and Shahrokh Valaee, "Recent Advances in Recurrent Neural Networks. *arXiv preprint arXiv:1801.01078*. pp. 1–21, 2018.

Smith, T. M., Reynolds, R. W., Peterson, T. C., & Lawrimore, J. (2008). Improvements to NOAA's Historical Merged Land-Ocean surface temperature analysis (1880–2006). *Journal of Climate, 21*(10), 2283–2296. https://doi.org/10.1175/2007JCLI2100.1

United States Environmental Protection Agency, "Climate Change Indicators: Atmospheric Concentrations of Greenhouse Gases. epa.org, 2022.

US Department of Commerce, N.O.A.A. (2005) Global Monitoring Laboratory - Carbon Cycle Greenhouse Gases, GML. http://www.esrl.noaa.gov/gmd/ccgg/trends/. Accessed: December 3, 2022.

von Salzen, K., et al. (2013). The Canadian fourth generation atmospheric global climate model (CanAM4). Part I: representation of physical processes. *Atmosphere-Ocean, 51*(1), 104–125. https://doi.org/10.1080/07055900.2012.755610

Washington, W. M., et al. (2000). Parallel climate model (PCM) control and transient simulations. *Climate Dynamics, 16*(10–11), 755–774. https://doi.org/10.1007/s003820000079

Wong, J. M., Chan, A. P., & Chiang, Y. H. (2005). Time series forecasts of the construction labour market in Hong Kong: the box-Jenkins approach. *Construction Management and Economics, 23*(9), 979–991.

Publisher's Note