Received: 26 March 2023 | Revised: 8 January 2024 | Accepted: 9 January 2024

DOI: 10.1002/joc.8366

RESEARCH ARTICLE

International Journal of Climatology RMetS

Statistical downscaling of maximum temperature under CMIP6 global climate models and evaluation of heat wave events using deep learning methods for Indo-Gangetic Plain

Manisha Chaturvedi[1] | **Rajesh Kumar Mall**[1] | **Saumya Singh**[1] | **Pawan K. Chaubey**[1] | **Ankur Pandey**[1,2]

[1]DST-Mahamana Centre of Excellence in Climate Change Research, Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi, India

[2]Department of Electrical and Electronics Engineering, Rajiv Gandhi Institute of Petroleum Technology, Amethi, India

Correspondence
Rajesh Kumar Mall, Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi, India.
Email: rkmall@bhu.ac.in

Funding information
Department of Science and Technology, Ministry of Science and Technology, India, Grant/Award Number: DST/CCP/CoE/80/2017(G)

Abstract

The rising global temperature is one of the primary concerns of the world as it impacts the economy, environment and healthcare of any country which are more pronounced a regional level. Assessment of regional impacts of climate change at a local level requires fine resolution of climate data for which a robust and fast downscaling method is needed. In this study, we use three deep learning-based methods, namely long short-term memory network (LSTM), deep neural network (DNN) and recurrent neural network (RNN), to downscale CMIP6 13 GCMS models data (1.25° × 1.25° resolution) global climate model (GCM) maximum temperature (Tmax) at a regional scale of 0.5° × 0.5° spatial resolution for the period 1991–2010 over the Indo-Gangetic Plain (IGP). In addition to the temperature prediction, heat wave events have been also analysed in the study. The study found that LSTM method performs better than DNN and RNN in downscaling of all GCM model datasets when evaluated against observed maximum temperature data from the India Meteorological Department (IMD) in terms of RMSE (0.9–3.5), average of all grid MAE value between (1.2 and 2.68), correlation (0.68–0.9) along with and spatiotemporal variability. LSTM also performed better in heat wave prediction over the region with similar temporal range (12–36 events) and spatial occurrence as compared to the observation (12–28 events). Overall, the study concludes that LSTM performs better than two methods for Indo-Gangetic Plain with best hyper parameter tuning. Hence, we propose to utilize a deep learning framework based on LSTM for downscaling GCM dataset at a finer resolution.

KEYWORDS

CMIP6, deep learning, downscaling, Indo-Gangetic Plain, temperature

1 | INTRODUCTION

The ability of general circulation models (GCMs) in reproducing weather parameters such as maximum temperature holds a critical importance in climate change studies. The purpose of GCMs is to understand the effects of climate change by simulating physical processes up to 200 years in the future. The simulation of the global climate on a large-scale necessitates a significant amount of computational resources. GCMs provide

plausible simulations of weather variables at global scales but cannot generate local climate details at a local scale needed for climate change impact assessment on different sectors such as health and agriculture. Although GCMs have been widely used as a reliable source of climate variables, they remain relatively coarse or low in resolution and are unable to resolve significant subgrid-scale features such as topography, clouds, land use, temperature and offshore wind energy (Niazkar et al., 2023; Peach et al., 2023; Zhang and Li, 2021). Therefore, there is a need to convert the GCM outputs of climate variables at a local scale. Downscaling is an efficient technique to obtain high-resolution climate information from coarse resolution GCM data for accurate regional climate projections. Downscaling can be performed using two approaches: dynamical and statistical approach. Dynamical downscaling is the method to derive smaller-scale climatic information over a bounded area using high-resolution regional climate model, driven by initial and boundary condition from GCMs. Statistical downscaling establishes a statistical relationship between large-scale climate features (predictors) and local climate data (predictands) to derive local climate information (Zhang and Li, 2021). More recently, statistical downscaling has found wide application in climate studies for scenario construction and simulation/prediction of daily mean, maximum and minimum temperature (Wang & Tian, 2022; Wilby et al., 2002; Zhang and Li, 2021). As precipitation is a fundamental component in the water cycle, a comprehensive outlook is essential for managing water systems under changing climate (Sachindra & Perera, 2016). Studies has shown that the downscale high-resolution data will significantly improve in the simulation of meteorological and hydrological variables (Ghosh and Mujumdar, 2008; Qiu et al., 2022; Singh & Mall, 2023). According to Okkan et al. (2023) downscale GCM outputs through the ANN structures will provide sufficient simulation accuracy for climate uncertainty and climate impact study. Hence, along with temperature, downscaling has been also performed for precipitation studies such as for low-frequency rainfall events (Ahmed et al., 2020; Pandey et al. 2022), potential evaporation rates (Weisse & Oestreicher, 2001) and extreme precipitation (Cavalcante et al., 2020).

Intergovernmental Panel on Climate Change (IPCC) in its sixth assessment report has stated that anthropogenic activities induced climate change has increased the frequency of extreme weather events such as heat waves, flood, drought etc., which has caused huge loss to natural and man-made ecosystem (IPCC, 2021). In previous studies, several researchers have reported increasing temperature with more extreme occurrences at different time scales over various parts of India (Mall et al., 2021; Singh et al., 2023). Studies have also reported that increase in temperature will lead to reductions wheat yield (Ishtiaque et al., 2022; Yadav et al., 2015). Extreme temperature have a detrimental impact on health leading to increase in morbidity and mortality of the population (Singh et al., 2019, 2021a, 2021b). Similarly, extreme temperature has been found to impact the water demands as well as evapotranspiration in the region (Kuttippurath et al., 2021). The regional impacts of such events are much more pronounced causing huge loss of life and property (Mall et al., 2019). In such scenario of extreme temperature causing severe impact on health, water and agriculture sector, high-resolution climate information is vital for present as well as future projection based impact assessments.

The implementation of dynamical downscaling is limited by high computational processes, a demand for enormous volumes of data, and a requirement for high levels of skill to apply and comprehend results (Trzaska & Schnarr, 2014). Following these constraints which prevents its widespread application, statistical downscaling approach has been widely used in climate change studies, as it is computationally inexpensive and convenient in implementation and interpretation of the results (Wang & Tian, 2022). Statistical downscaling has been extensively employed in studies of climate change due to the ease of implementation and interpretation of the results (Zhang and Li, 2021). The primary goal of statistical downscaling is to establish a link between predictors and predictands which has become advanced using artificial intelligence (AI) and machine learning (ML) methods which include (i) nonlinear regression models of artificial intelligence, for example, artificial neural network (ANN; Chadwick et al., 2011), support vector machine (SVM; Tripathi et al., 2006), least squares support vector machine (LSSVM) and relevance vector machines (RVM; Okkan & Inan, 2015a, 2015b); (ii) linear regression models, for example, statistical downscaling model (SDSM; Wilby et al., 2002); (iii) weather generators, for example, long ashton research station-weather generator (LARS-WG; Racsko et al., 1991). The ability of ANNs to simulate the nonlinear and time-varying properties of atmospheric variables at various scales, as well as their propensity to uncover complex patterns and linkages between predictors and predictands, have led to successful applications of ANN for downscaling purpose (Chadwick et al., 2011). Deep learning-based methods have been used for GCM bias correction and downscaling in recent years. The developed an image high-resolution architecture called YNet to downscale precipitation and monthly mean temperature, from multiple GCMs model data (Liu et al., 2020), proving that the model performed

better than a shallow plain architecture (Vandal et al., 2019) and a traditional statistical downscaling method.

All the above methods basically consider climate data to be linear and stationary. As an alternative to these methods that can address the nonlinearity in time series data, a number of ML models have been presented for climate data downscaling and multi-model ensemble forecasts (Ahmed et al., 2020; Kolluru et al., 2020). Deep neural networks (DNNs) are typically feed forward networks (FFNNs), in which data travels from the input layer to the output layer without travelling backward and the links between the layers are one way, that is, in the forward direction, never touching a node again. The recurrent neural networks (RNNs), a typical class of deep learning method which is a subset of ML with a specific internal state (memory cell), can exhibit temporally dynamic behaviour for a time sequence, and such networks have been widely used to solve time series problems involving speech recognition, time series prediction, weather events and power loads (Wang et al., 2018). Long short-term memory's (LSTMs) is an artificial neural network used in the fields of artificial intelligence and deep learning. Long short-term memory network is an advanced RNN, a sequential network that allows information to persist (Hochreiter & Schmidhuber, 1997). It is capable of handling the vanishing gradient problem faced by RNN (Tran Anh et al., 2019). These deep learning (DL) techniques feature key overcome for the training mechanisms of deep networks, and they can get over the hidden layer number and neuron number restrictions in conventional ANNs. Thus, DL methods are advantageous and used for studying the feature extraction, data mining, weather events and nonlinear function approximation, downscaling (Feng et al., 2021). DL is regarded as a high-efficiency methodology for processing and analysing the climate "big data" of the CMIP projects; especially, DL methods can assist researchers in discovering underlying physical principles in Earth's climate systems and in identifying detailed relationships (Zhang and Li, 2021).

The projection of climatic extremes from GCMs datasets show high uncertainty due to different factors like emission scenarios, regional climate variability, model parametrization schemes and internal model physics (Chaubey & Mall, 2023). To mitigate this uncertainty, previous researchers applied different corrections and deep learning methods, which have been found to be efficient at statistical downscaling (Sabarinath et al., 2023; Wang & Tian, 2022). A notable research gap in statistical downscaling lies in the effective integration of advanced machine learning techniques, such as deep learning and ensemble methods, to enhance the accuracy and robustness of downscaling models. Therefore, to address challenges associated with statistical downscaling, and improve the applicability of downscaling models to analysis the regional climatic events like heat wave at higher resolution, this study aims to use different deep learning techniques. The three novel deep learning techniques, namely long short-term memory (LSTM), deep neural network (DNN) and recurrent neural network (RNN), are used to downscale the 13 climate models (CMIP6-GCM) data and select the best suitable method to downscale the daily maximum temperature. Additionally, this study evaluates extreme temperature and heat wave events over the Indo-Gangetic Plain (IGP) of India at high resolution to facilitate an effective climate adaptation planning over the region.

2 | DATA AND METHODOLOGY

2.1 | Study area

The study was conducted for Indo-Gangetic Plain (IGP) region of India. IGP covers an area of about 700 thousand km^2 (172-million-acre) and lies between 21°35′–32°28′N latitude and 73°50′–89°49′E longitude. It is one of the primary region for producing wheat in the country (Banjara et al., 2022), which is shown in Figure 1. The Indus and Ganges rivers, which flow through this area and encircle the northern parts of the Indian subcontinent, so this region is called the IGP Plain. Figure 1 shows these agro-climatic zones that have been developed based on climate, soil and cropping patterns (Alagh, 1990). IGP plain is divided into four agro-climatic zones, namely Trans Gangetic Plain (TGP), Upper Gangetic Plain (UGP), Middle Gangetic Plain (MGP) and Lower Gangetic Plain (LGP) (Mall et al., 2021).

2.2 | Data

The observed dataset for daily maximum temperature for Jan–Dec (1991–2010) has been obtained from India Meteorological Department (IMD) for the study region. Climate models are crucial for understanding and analysing the physical variables such as temperature, rainfall, and humidity to name a few. In this regard, Coupled Model Intercomparison Project (CMIP) provides a benchmark dataset for understanding the earth and climate phenomenon (Eyring et al., 2016). In this work, we have used CMIP6 GCM 13 models data (Table 1). The CMIP6 GCM models temperature data were obtained from Coupled Model Intercomparison Project Phase 6 (CMIP6) (https://esgf-node.llnl.gov/search/cmip6/) (Chaubey & Mall, 2023).

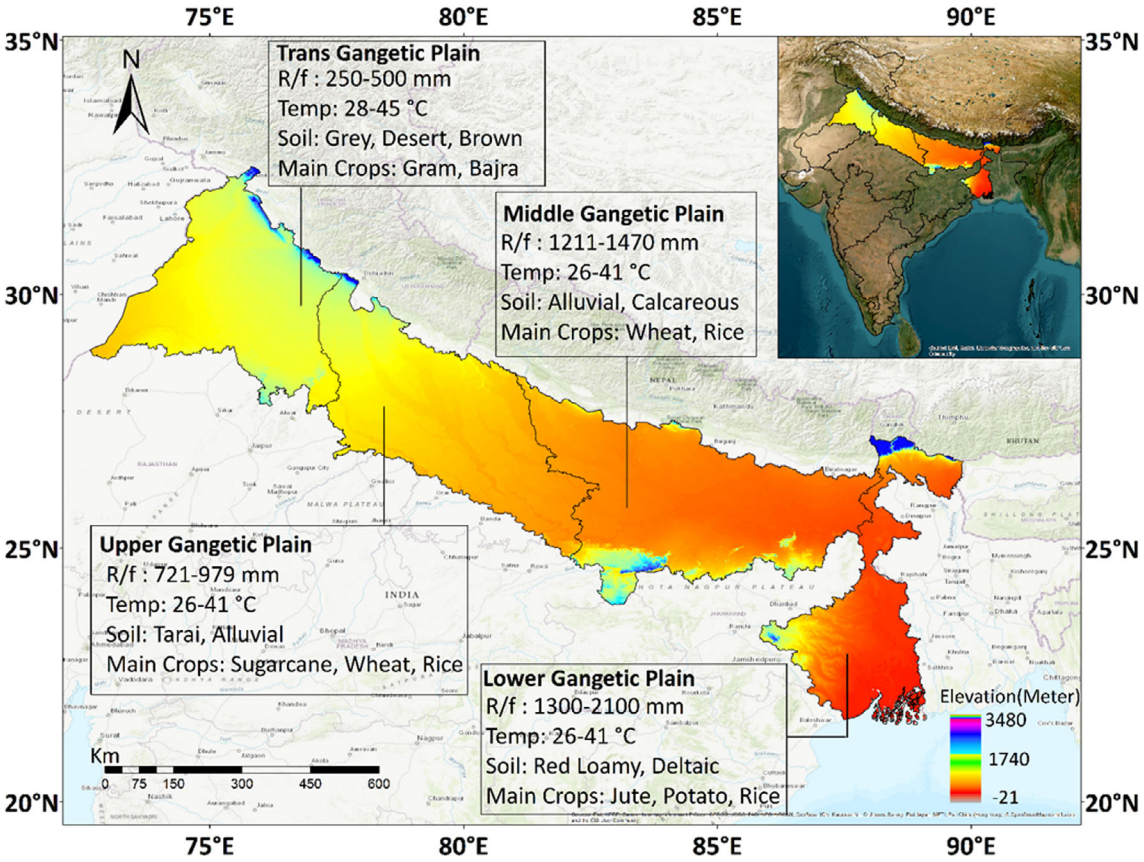

FIGURE 1 Study area of Indo-Gangetic Plain (IGP). [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 List of CMIP6 models used in this study along with resolution and country of the modelling group.

S. No.	Models	Institution/country	Resolution
1	ACCESS-CM2	Commonwealth scientific and industrial research organization/Australia	1.25° × 1.25°
2	ACCESS-ESM1-5	Commonwealth scientific and industrial research organization/Australia	1.25° × 1.25°
3	AWI-CM-1-1-MR	Alfred Wegener Institute/Germany	1.25° × 1.25°
4	AWI-ESM-1-1-LR	Alfred Wegener Institute/Germany	1.25° × 1.25°
5	BCC-CSM2-MR	Beijing Climate Center (BCC) China Meteorological Administration/China	1.25° × 1.25°
6	BCC-ESM1	Beijing Climate Center Earth System Model/China	1.25° × 1.25°
7	CMCC-ESM2	Euro-Mediterranean Center on Climate Change coupled climate model/Italy	1.25° × 1.25°
8	MPI-ESM1-2-HR	Max Plank Institute for Meteorology/Germany	1.25° × 1.25°
9	MPI-ESM1-2-LR	Max Plank Institute for Meteorology/Germany	1.25° × 1.25°
10	MRI-ESM2-0	Meteorological Research Institute (MRI)/Japan	1.25° × 1.25°
11	NorESM2-LM	Norwegian Climate Center/Norway	1.25° × 1.25°
12	NorESM2-MM	Norwegian Climate Center/Norway	1.25° × 1.25°
13	TaiESM1	Taiwan Earth System Model/Taiwan	1.25° × 1.25°

The GCM data for daily maximum temperature has been obtained at a resolution of 1.25° × 1.25°. The study extends over the period of 20 years from 1991 to 2010 and is divided into three time frames for training (1991–2002), testing (2003–2006) and validating (2007–2010) the downscaled temperature data. The goal is to learn from the training data for the period (1991–2002), predict Tmax for the calibration (2003–2006), and validate (2007–2010) results.

2.3 | Methodology

The workflow of the proposed methodology for downscaling the Tmax dataset has been shown in Figure 2. In the present study CMIP6 Models Tmax data at 1.25° resolution has been used for downscaling. First, the data is preprocessed using a Min Max Scalar that brings the data in the range of 0–1 (Smolinska et al., 2014). This normalization process helps in reducing the bias before any model fitting if the original dataset contains few outliers or values that have large deviations from the mean (Kang et al., 2023). We evaluate three AI-based techniques, namely LSTM, RNN and DNN, for downscaling CMIP6 GCM 13 models data with a spatial resolution of 1.25° to a 0.5° resolution and validated with the observed data. The dataset for the period of 1991–2010 was split into three sub data sets, with 60% of the data used for training (1991–2002), 20% used for calibration (2003–2006) and 20% used for validation (2007–2010) of daily Tmax dataset. The performance evaluation of the deep learning methods on a daily scale is done using standard performance metrics which are RMSE and MAE. The RMSE and MAE values were estimated at each grid point for both the downscaled daily Tmax, and the corresponding results are shown in Figures 8 and 9. Further, probability distribution function, cumulative density function (CDF) and Taylor diagrams are used to evaluate the performance of deep learning approach. The fidelity of the deep learning model output data is tested for the IGP by plotting the probability density function (PDF). The probability distribution is a statistical function that captures all possible values and likelihoods that a random variable can take within a given range. This range will be bounded between the maximum and minimum possible values. The probability distribution function (PDF) has been used to compare the probability between downscale models output and observations (Kumar et al., 2021).

The cumulative distribution function (CDF) determines the cumulative probability of a random variable, it is used to compare the cumulative probability between downscale models output and observed. These assessments help in selecting the best model for downscaling purpose (Singh et al., 2021a, 2021b). Taylor diagram encompasses three statistics of how well the patterns between the simulations and references are matched (Taylor, 2001). These statistics include the Pearson correlation coefficient (PCC) of the spatial patterns, which is reflected by the azimuth positions; the ratio of standard deviations (SD), which is reflected by the radial distance from the origin; and the centralized root-mean-square error (RMSE), which is reflected by the distance between the simulation results and a reference. Taylor diagram gives a concise statistical summary of correlation,

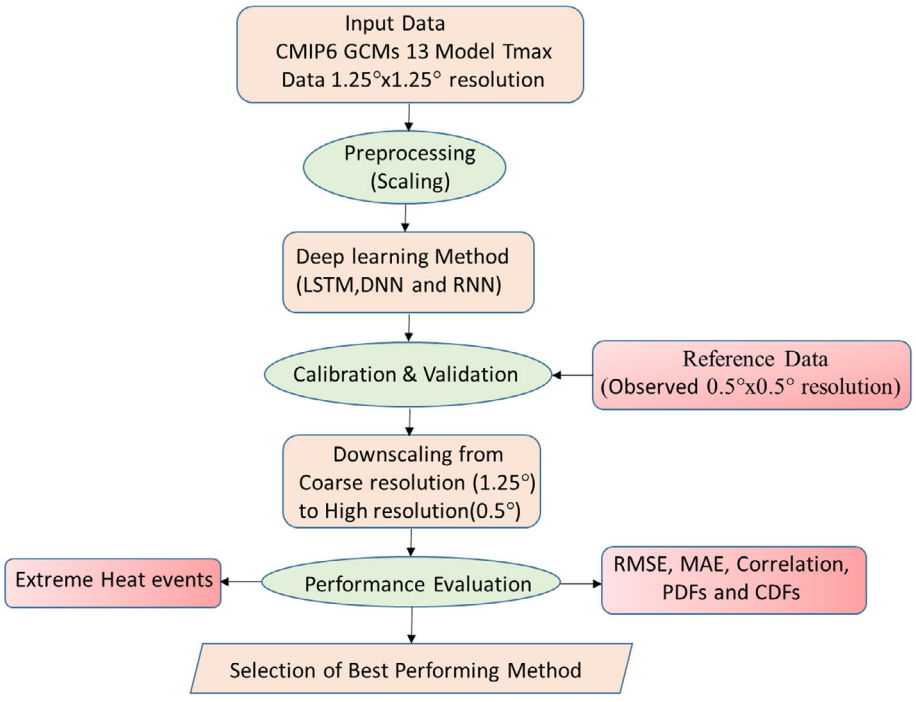

FIGURE 2 Proposed methodology for downscaling using Deep learning based method. [Colour figure can be viewed at wileyonlinelibrary.com]

root-mean square (RMS) difference and the standard deviation between observed and downscale model data. Further, few standard deep learning methods have been deployed/fitted and its performance on the validation set is assessed. The best deep learning model that provides least error as compared to the observed dataset has been selected. Apart from temperature validation the study analyses the performance of the output from deep learning methods in representing the extreme temperature for the region. For this, heat wave events have been analysed using the criteria based on exceedance of percentile threshold of daily maximum temperature over the IGP region (Perkins & Alexander, 2013). In the subsequent section, we explain the different deep learning methods utilized in this work for downscaling.

2.3.1 | Deep neural networks

A DNN is inherently an artificial neural network consisting of many hidden layers between the input layer and final output layer. DNNs have shown very good results in the field of image and video processing (LeCun et al., 2015). Further, they are not much sensitive to outliers and can be effectively applied for both univariate and multivariate datasets. The advantage with DNN over standard machine learning methods such as support vector machines (SVM), decision trees (DT) or k-nearest neighbour (k-NN) is the ability to model complex nonlinear functions and learn the hidden patterns automatically in the data. Hence, we utilize representation learning approaches (LeCun et al., 2015; Salahuddin et al., 2022) and a three layered network with one hidden layer having 8 neurons in this study. The input layer fires 32 neurons with single neuron at the output. The goal for any DNN is to minimize the cost function L(Θ) as discussed below,

$$\min_{\Theta} \frac{1}{n} \sum_{i=1}^{n} \ell(f(\mathbf{x}_i;\Theta), y_i) := L(\Theta). \tag{1}$$

Here, $\mathbf{x}_i$ denotes the input data and y_i is the target data as we are performing supervised learning. Θ denotes the learning parameters or the weights that are chosen to minimize the loss function such as mean absolute error or root mean square error in our case. The goal of all the deep learning methods used in this work is to minimize the above mentioned loss function.

2.3.2 | Recurrent neural networks

RNNs are a class of deep neural network with a short-term internal memory (LeCun et al., 2015). In contrast to DNNs, the RNNs are extensively used in time-series prediction as it considers both the current input and the learnings from the past input. DNNs only consider the current input and the information propagates only in the forward direction through the feed-forward network (Han et al., 2019). As shown in Figure 3, it is observed that due to the feedback in the RNN structure, the information is propagated in both the directions and the information of immediate past is added to the present thereby making it more suitable for time-series predictions. The RNN equations are given below: the least estimation error. However, RNNs in general encounter challenges such as vanishing gradient, and difficulties in learning long sequences due to its short memory (LeCun et al., 2015). LSTMs overcome these issues and will be discussed in the next subsection,

$$d^{<t>} = g_1\left(W_{dd} a^{<t-1>} + W_{dx} x^{<t>} + b_d\right), \tag{2}$$

$$y^{<t>} = g_2\left(W_{yd} d + b_y\right), \tag{3}$$

where W_{dx}, W_{dd}, W_{yd}, b_d, b_y are coefficients that are shared temporally and g_1, g_2 denote the activation functions.

These settings are chosen based on the experiments performed on calibration set data providing the least estimation error. However, RNNs in general encounter challenges such as vanishing gradient, and difficulties in learning long sequences due to its short memory (LeCun et al., 2015). LSTMs overcome these issues and will be discussed in the next subsection.

2.3.3 | Long short-term memory networks

LSTM are basically the extension networks of RNNs that increase the memorizing ability of RNNs using the regulators known as gates. LSTM gates decide which information to store and learn, and what to discard using "forget" gates. There are three gates that control the information flow, namely an input gate, an output gate and a forget gate (Yu et al., 2019). These gates compute letting in new input (input gate), remove or forget the unimportant information, or output the current time step (output gate). The forget's gate activation function which occurs in the gradient term, and the additive structure leads to solving the vanishing gradient challenge (Hochreiter & Schmidhuber, 1997). Figure 4 shows the architecture used for downscaling the time-series data of temperature values. The LSTM update for t time-step considering the input x_t, $h_{(t-1)}$, and $c_{(t-1)}$ is given below equations,

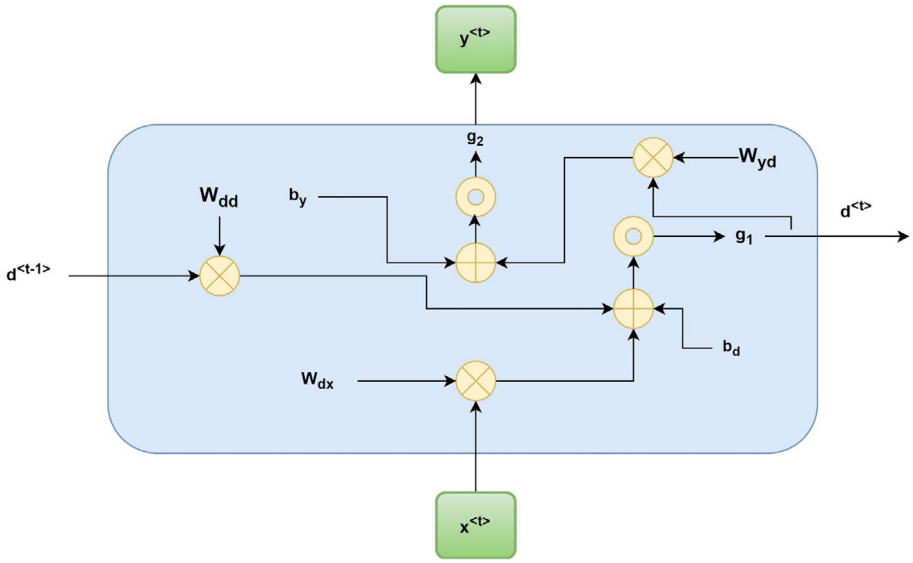

FIGURE 3 The flow diagram of RNN. [Colour figure can be viewed at wileyonlinelibrary.com]

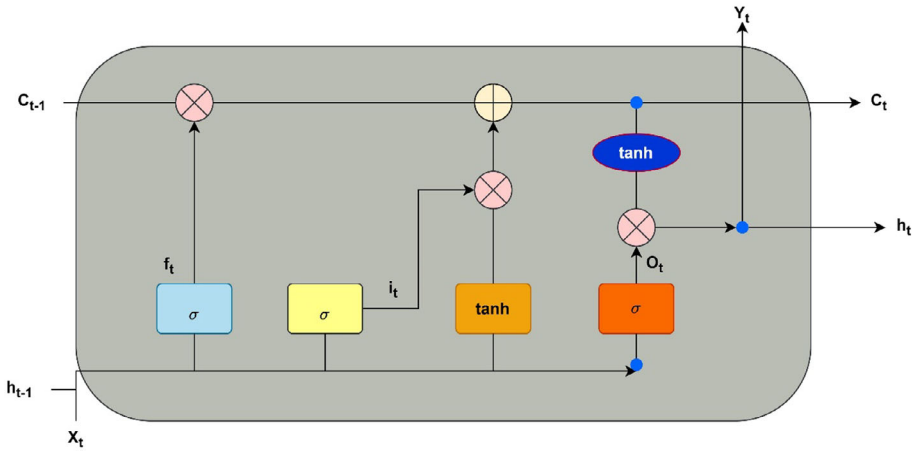

FIGURE 4 The flow diagram of LSTM. [Colour figure can be viewed at wileyonlinelibrary.com]

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \\ f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \\ o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \\ g_t &= \phi(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \end{aligned}, \tag{4}$$

$$c_t = f_t \odot c_t + i_t \odot g_t, \tag{5}$$

$$h_t = o_t \odot (c_t). \tag{6}$$

In this context, i, f, o, c and g correspond to the input gate, forget gate, output gate and cell activation, input modulation gate vectors. These vectors share the same dimensions as the h vector, which defines the hidden value. The symbols σ denote the the sigmoid (logistic) function, and x_t represents the memory cell layer during a specific time step; W_{xi}, W_{xf}, W_{xo} and W_{xc} are weight matrices; b_i, b_f, b_o and b_c signify the from-to relationships (the input–input gate matrix, the hidden-input gate matrix, etc.) are bias vectors. The symbol φ signifies the element-wise application of the hyperbolic tangent (tanh) function, and $\odot$ denotes elementwise multiplication.

The details of this architecture are mentioned in Khan et al. (2020). We tune this architecture by setting the look back to 17,730 and observe that LSTM provides better results in prediction task as compared to DNN and RNNs as it addresses the vanishing gradient problem as well as the gated architecture.

2.3.4 | Model development and selection

In this work, a LSTM based model is developed for downscaling the temperature variable from 1.25° × 1.25° resolution to 0.5° × 0.5° spatial resolution for the period 1991–2010 over the Indo-Gangetic Plain (IGP). We obtain the best estimation results by varying the hyperparameters such as

batch size, number of layers, optimizer and the activation function to name a few. It is to be noted that we performed 50 trials as suggested in (Kaveh et al., 2021) with different hyperparameters. We report that the input data dimensions for calibration stage is 181*5845, and validation stage is 181*7305. The data is first normalized using a min-max scaler as it does not change the data distribution and helps the optimizer in minimizing or maximizing the cost function. We utilize ADAM (adaptive moment estimation) optimizer as it is robust to noisy gradients and can handle large datasets (as in our case) with adaptive learning rates. Furthermore, we utilize RELU (rectified linear unit) as an activation function because it introduces nonlinearity to learn the complex pattern in the temperature data. Also, RELU helps in avoiding vanishing gradient issue that helps the deep learning models for accurate estimation. Selection of appropriate batch size is important for a LSTM model as correct size leads to accurate convergence and stability. In this work, after multiple trials we select a batch size of 30 for LSTM in contrary to the best batch size of 16 for DNN and RNN. In this work, we utilize the RNN model with two hidden layers and a look back period of 30 days. This means that we are attempting to learn from previous 30 days' temperatures to predict the next day's temperature. It can be observed from the results section that above mentioned hyperparameters provide best results for LSTM model.

2.3.5 | Heat wave analysis

In the study, heat wave events have been identified using the criteria based on exceedance of percentile threshold of daily maximum temperature over the study region. Heat wave events have been calculated for the heat wave season when heat waves are observed over most of the country during March–June (2007–2010) (Singh et al., 2021a, 2021b, 2023). For calculation of heat wave events, firstly 90th percentile of the daily maximum temperature observed over each grid of IGP over the study duration has been calculated. After the 90th percentile is estimated, a heat wave event has been declared when the daily maximum temperature exceeds the threshold temperature, that is, the 90th percentile temperature for three consecutive days over each grid during 2007–2010 (Mar–Jun) for study region (Perkins & Alexander, 2013). The total number of heat wave events obtained over each grid from the 13 model outputs has been evaluated against the observed heat wave events in the present study.

3 | RESULT AND DISCUSSION

In this study, we evaluate three AI-based techniques, namely LSTM, RNN and DNN, for downscaling CMIP6 GCMs 13 model data with a spatial resolution of 1.25° to a 0.5° resolution and validated with the observed data. The dataset for the period of 1991–2010 was split into three sub data sets, with 60% of the data used for training (1991–2002), 20% used for calibration (2003–2006) and 20% used for validation (2007–2010) of daily Tmax dataset. The performance evaluation of the deep learning methods on a daily scale is done using standard performance metrics which are RMSE, MAE. The RMSE and MAE values were estimated at each grid point for both the downscaled daily Tmax. Further, Taylor diagrams, PDFs, CDFs, line plots and spatial distribution are used to evaluate the performance of machine learning approach.

3.1 | Cumulative performance evaluation

The study evaluates the model (LSTM, RNN and DNN) performance using PDFs, CDFs and Taylor diagrams using the daily mean maximum temperature (output obtained after downscaling of 13 GCMs model) over study area. Figures 5 and 6 show the PDF and CDF of the daily maximum temperature from all downscale model outputs, GCMs 13 models output and observed maximum temperature during validation period, respectively. The distributions (PDF and CDF) aids in the better understanding of the performance of each downscaling method.

Figure 5 shows probability distribution (PDF) of the DL methods outputs against observed for the daily maximum temperature over the Indo-Gangetic Plain. Figure 5 shows an estimation of the mean of daily maximum temperature (downscale model outputs) with respect to the observed. The daily mean maximum temperature downscaled by LSTM lies within the range LSTM (17–42.5), DNN (17.2–43.5) and RNN (17.3–43.7), against the observed (17–42.28) during the validation period for all 13 GCMs CMIP6 model. From Figure 5 it is very apparent that the RNN and DNN models overestimate the observed range. For IGP plain, after downscaling through LSTM methods, data seem to be closer to observed in validation period. In this comparison, the deep learning downscaling model LSTM was found to be better among the three methods as it shows the closer too observed.

Figure 6 shows the cumulative distribution function (CDF) of the mean of daily maximum temperature for Indo-Gangetic Plain. However, all three deep learning methods performed effectively but differently to adjust the shape of distribution. The cumulative distribution function curves show a clear difference between downscale outputs and GCM 13 models outputs. For IGP plain,

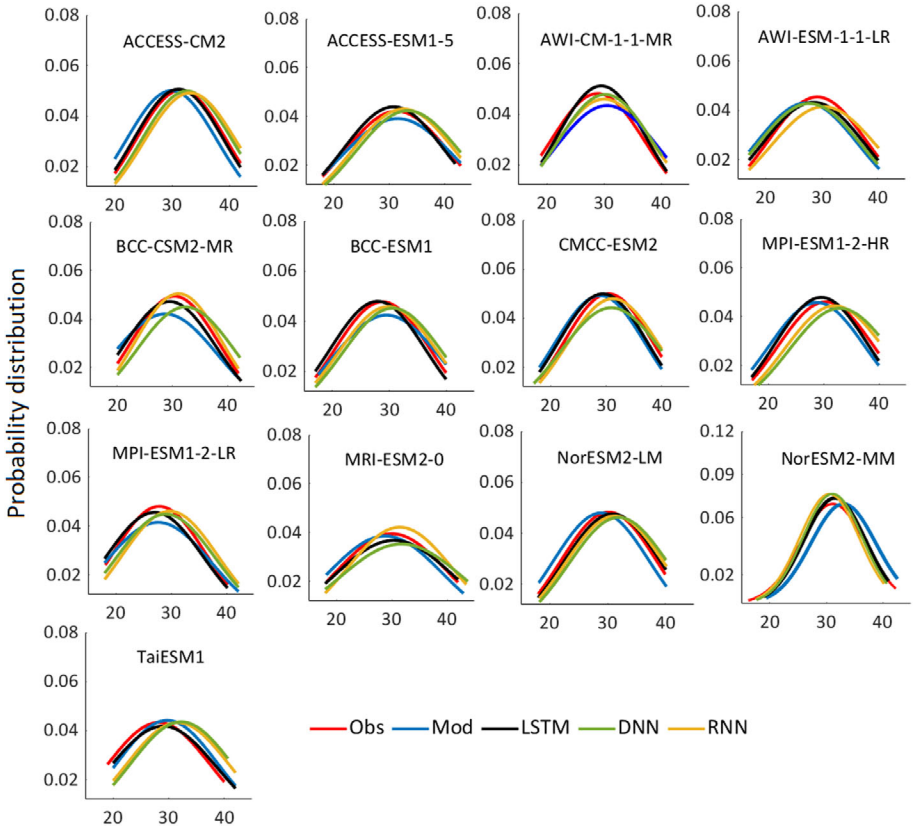

FIGURE 5 Probability distribution function (PDF) showing performance of deep learning methods (LSTM, DNN and RNN) for the validation period (2007–2010) (*x*-axis shows the temperature in °C). [Colour figure can be viewed at wileyonlinelibrary.com]

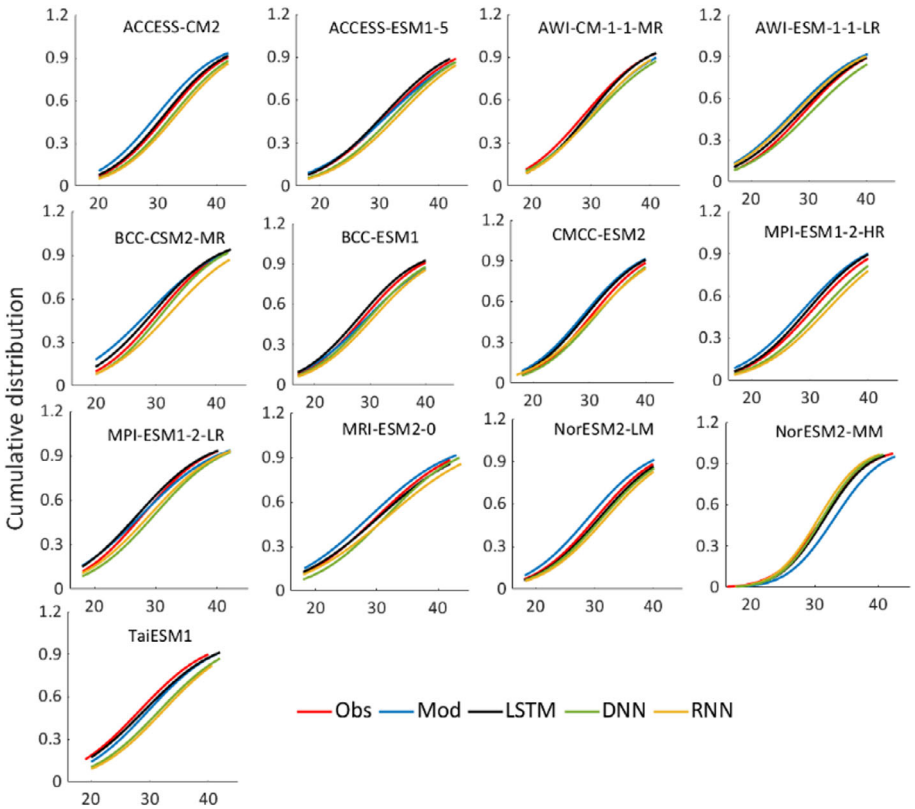

FIGURE 6 Cumulative density function (CDF) showing performance of deep learning methods (LSTM, DNN and RNN) for the validation period (2007–2010) (*x*-axis shows the temperature in °C). [Colour figure can be viewed at wileyonlinelibrary.com]

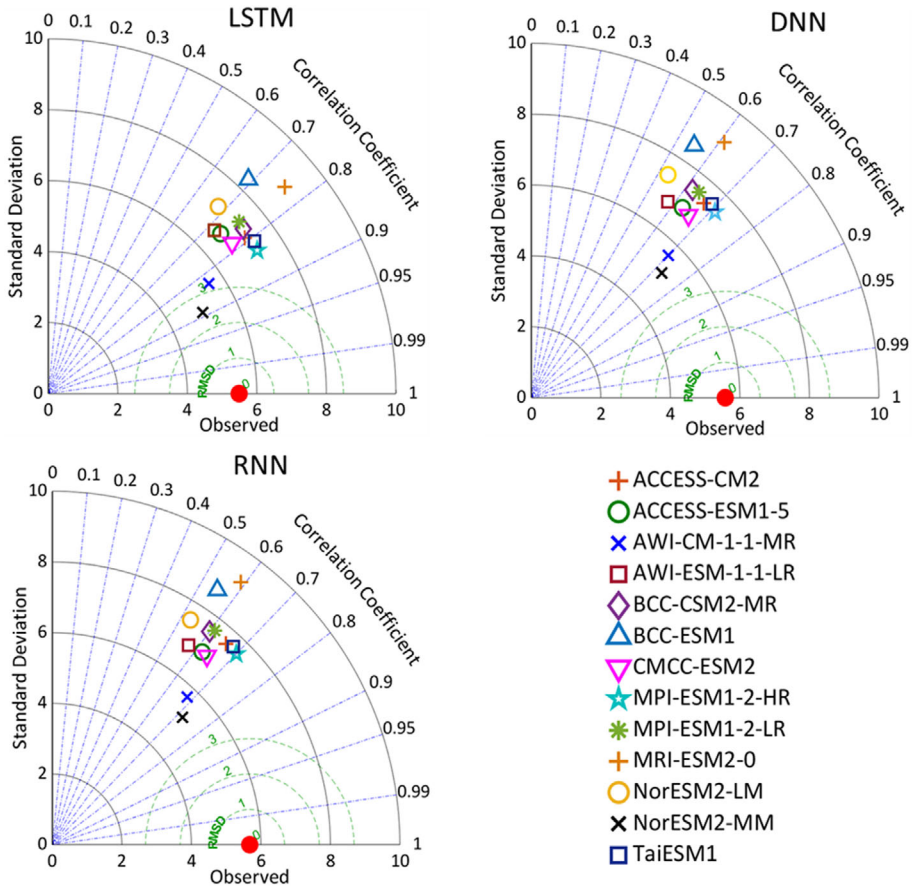

FIGURE 7 Taylor diagram showing performance of deep learning methods (LSTM, DNN and RNN) for validation period (2007–2010). [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Taylor diagram result.

	RMSE			SD			Correlation		
Model	**LSTM**	**DNN**	**RNN**	**LSTM**	**DNN**	**RNN**	**LSTM**	**DNN**	**RNN**
OBSERVED	0	0	0	5.49	5.58	5.68	1	1	1
ACCESS-CM2	4.50	4.74	4.90	7.16	7.40	7.56	0.79	0.67	0.66
ACCESS-ESM1-5	4.56	4.66	4.86	6.70	6.90	6.95	0.74	0.63	0.62
AWI-CM-1-1-MR	3.25	4.00	4.17	5.57	5.63	5.70	0.83	0.70	0.68
AWI-ESM-1-1-LR	4.43	4.83	5.02	6.63	6.78	6.87	0.72	0.58	0.57
BCC-CSM2-MR	4.73	4.90	5.23	7.28	7.48	7.55	0.77	0.62	0.60
BCC-ESM1	6.04	6.34	6.54	8.34	8.52	8.63	0.69	0.55	0.55
CMCC-ESM2	4.39	4.50	4.67	6.78	6.85	6.95	0.78	0.66	0.64
MPI-ESM1-2-HR	4.10	4.31	4.50	7.24	7.44	7.56	0.83	0.71	0.70
MPI-ESM1-2-LR	4.70	4.89	4.95	7.32	7.55	7.65	0.75	0.64	0.61
MRI-ESM2-0	5.97	6.10	6.32	8.96	9.10	9.20	0.76	0.61	0.59
NorESM2-LM	5.33	5.50	5.75	7.19	7.42	7.50	0.68	0.53	0.53
NorESM2-MM	2.52	3.98	4.08	5.00	5.15	5.20	0.89	0.73	0.72
TaiESM1	4.25	4.38	4.57	7.33	7.54	7.65	0.81	0.69	0.68

Note: RMSE, standard value (SD) and correlation values of observed and all GCMs model downscaled daily maximum temperature based on the LSTM, DNN and RNN methods of Indo-Gangetic Plain (India) for the validation period (2007–2010).

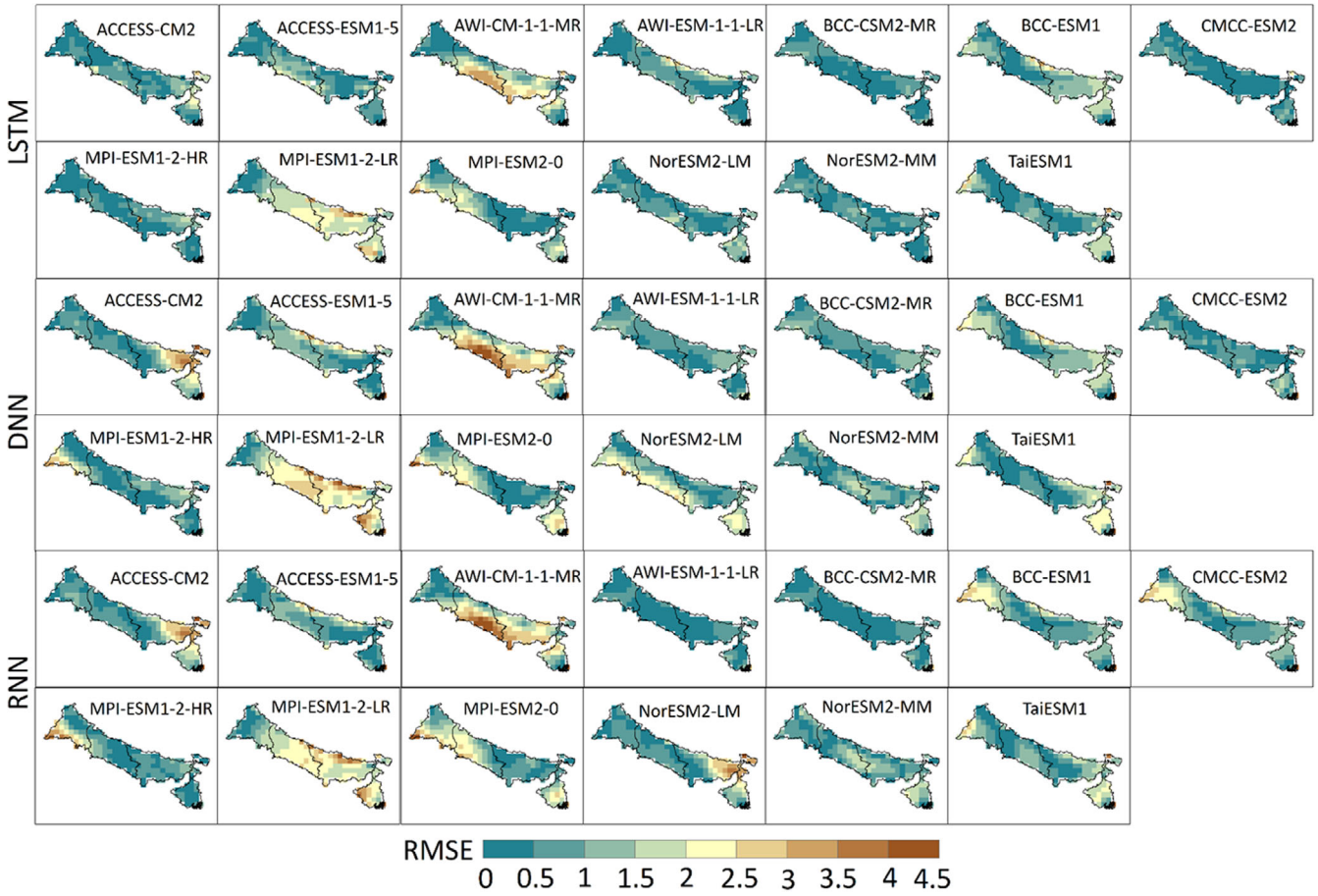

FIGURE 8 Distribution map of RMSEs (root-mean-square error) for the downscaled daily maximum temperature based on the LSTM, DNN and RNN methods of Indo-Gangetic Plain (India) for the calibration set (2003–2006). The model is trained from observed data (period 1991–2002). [Colour figure can be viewed at wileyonlinelibrary.com]

after downscaling through LSTM methods, data seem to be closer to the observed CDF curve. The RNN and DNN curves seem to slightly underestimate for all models against the observed CDF curves in validation period.

Figure 7 and Table 2 show the Taylor diagram and its summary of the observed and downscaled daily maximum temperature from three DL methods for the validation period (2007–2010). Taylor diagrams in Figure 7 show the performance of all the 13 GCMs (CMIP6) dataset downscaled by LSTM, DNN and RNN for the validation period across IGP plain. The variations were found for different deep learning approach, with PCC in the ranges of LSTM (0.68–8.96) °C, DNN (0.53–0.73) °C and RNN (0.53–0.72) °C, SD in the range of LSTM (5.0–8.96) °C, DNN (5.15–9.10) °C and RNN (5.20–9.20) °C and RMSE in the range of LSTM (2.52–6.04) °C, DNN (3.98–6.34) °C and RNN (4.08–6.54) °C against the observed during the validation period for all 13 GCMs CMIP6 model (Table 2). LSTM displayed the best performance, with PCC, SD and RMSE for all GCMs models. RNN yielded the largest RMSE and had the smallest PCC whereas DNN shows relatively better results than RNN with a RMSE, PCC and SD. Summarizing the above the results, it can be stated that LSTM Tmax shows a higher agreement with the observation as compared to other two methods for all 13 GCMs models.

3.2 | Spatial heterogeneity of spatial performance

The performance of DL methods in downscaling of GCMs 13 model maximum temperature is shown through the spatial distribution of RMSEs for calibration (2003–2006) and validation (2007–2010) period (Figures 8 and 9) against the observed daily Tmax at each grid level in IGP. The RMSE values of all GCMs model for LSTM ranges between 0.9 and 3.0°C for both calibration and validation period. Only Model AWI-CM-1-1-MR and MPI-ESM1-2-LR for calibration (Figure 8) and ACCESS-ESM1-5, AWI-CM-1-1-MR and AWI-ESM-1-1-LR models for validation (Figure 9) period ranges between the 0.9

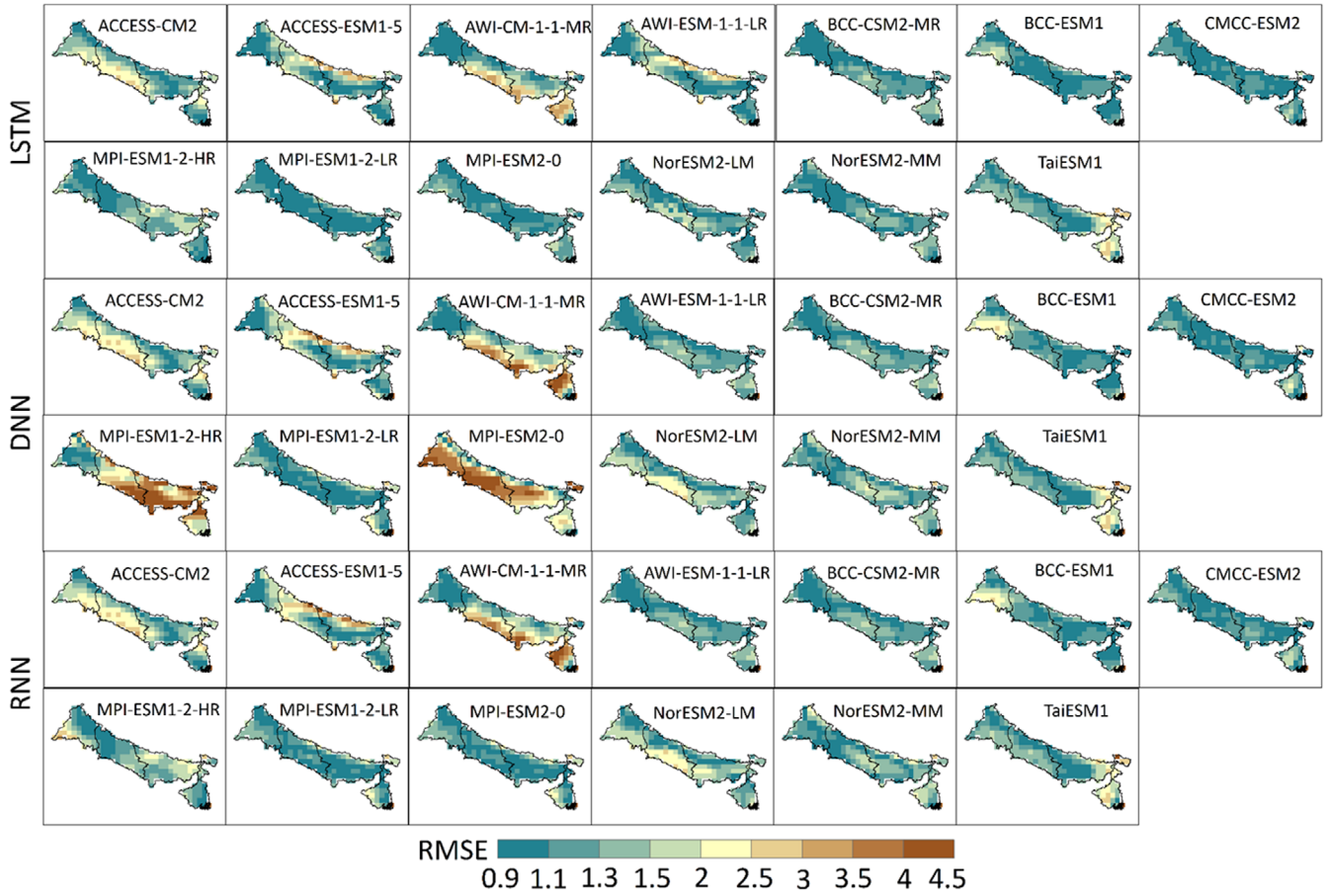

FIGURE 9 Distribution map of RMSEs (root-mean-square error) for the downscaled daily maximum temperature based on the LSTM, DNN and RNN methods of Indo-Gangetic Plain (India) for the validation set (2007–2010). The model is trained from observed data (period 1991–2002). [Colour figure can be viewed at wileyonlinelibrary.com]

and 3.5°C. In this process, climate models AWI-CM-1-1-MR and MPI-ESM1-2-LR exhibit considerable uncertainty in DNN and RNN processes; however, LSTM shows less uncertainty in these models. Furthermore, earlier studies have suggested that uncertainty is reduced after applying ANN and statistical downscaling in climate models (Qiu et al., 2022; Okkan et al., 2023).

For the calibration and validation period, the RMSE for LSTM has been observed to be as low as 0.9 in some parts of the upper, lower and some part of middle Gangetic plain for all the models except MPIESM1-2-LR in calibration period. For DNN the RMSE values, ranging from 0.9 to 4.5°C while for RNN 0.9 to 4.5°C for the both calibration and validation period, respectively. These value show the LSTM RSME value is less than DNN and RNN across the IGP plain. The MAE average values of all grids for IGP plain for all models ranging between 1.0 and 2.6°C and 1.2 and 2.68 for calibration and validation period, respectively, in the case of LSTM (Table 3). While for DNN 1.22–2.91°C and 1.26–2.78°C and for RNN 1.2–2.79°C and 1.29–2.72°C, MAE ranges for calibration and validation period, respectively, DNN and RNN MAE value is comparatively higher than the LSTM models (Table 3). Similarly, the RMSE average values of all grids for IGP plain for all model ranging between 0.5 to 1.87°C and 0.55 to 1.49 for and validation period, respectively, DNN and RNN RMSE and MAE value is comparatively higher than the calibration and validation period, respectively, in case of LSTM (Table 3). While for DNN 0.58–2.07°C and 0.67–1.77°C and for RNN 0.60–1.89°C and 0.66–1.69°C, RMSE ranges for calibration the LSTM models.

Based on above statistics it can be stated that LSTM can predict daily temperature with relatively higher accuracy than DNN and RNN for all GCMs model.

3.3 | Spatiotemporal variability of downscaled daily maximum temperature

The observed temporal temperature variability lies between the ranges of 21–38°C (Figure 10). LSTM

TABLE 3 RMSE and MAE mean values of all GCMs model downscaled daily maximum temperature based on the LSTM, DNN and RNN methods of Indo-Gangetic Plain (India) for the calibration (2003–2006) and validation period (2007–2010).

	RMSE_2003–2006			RMSE_2007–2010			MAE_2003–2006			MAE_2007–2010		
Model	**LSTM**	**DNN**	**RNN**	**LSTM**	**DNN**	**RNN**	**LSTM**	**DNN**	**RNN**	**LSTM**	**DNN**	**RNN**
ACCESS-CM2	1.06	1.1	1.11	1.41	1.43	1.44	1.47	1.95	1.92	1.47	1.5	1.49
ACCESS-ESM1-5	0.93	0.98	0.95	1.15	1.22	1.27	1.47	2.02	1.94	1.87	1.89	1.92
AWI-CM-1-1-MR	1.64	1.89	1.88	1.49	1.77	1.69	1.78	1.96	1.95	1.8	1.87	1.84
AWI-ESM-1-1-LR	0.77	0.95	0.93	1.12	1.2	1.22	2	2.03	2.02	1.82	1.85	1.87
BCC-CSM2-MR	0.55	0.69	0.6	0.75	0.76	0.73	2.1	2.3	2.15	1.66	1.67	1.68
BCC-ESM1	1.04	1.14	1.08	0.62	0.81	0.82	2.5	2.68	2.67	2.68	2.78	2.72
CMCC-ESM2	0.52	0.58	0.74	0.6	0.65	0.69	1.1	1. 23	1.3	1.3	1.5	1.4
MPI-ESM1-2-HR	0.63	0.74	0.82	0.97	2.81	1.1	1.5	1.52	1.53	1.3	2.66	1.4
MPI-ESM1-2-LR	1.87	2.07	1.89	0.58	0.67	0.66	2.6	2.91	2.79	1.78	1.81	1.79
MRI-ESM2-0	1.08	1.11	1.28	0.74	2.85	0.66	1.21	1.24	1.3	1.23	2.42	1.78
NorESM2-LM	0.82	1.26	1.02	0.94	1.17	1.22	1.82	1.87	1.93	1.9	1.95	1.97
NorESM2-MM	0.5	0.86	0.88	0.55	0.95	0.91	1	1.22	1.2	1.2	1.26	1.29
TaiESM1	0.9	0.92	0.93	1.04	1.05	1.06	1.11	1.3	1.16	1.23	1.3	1.27

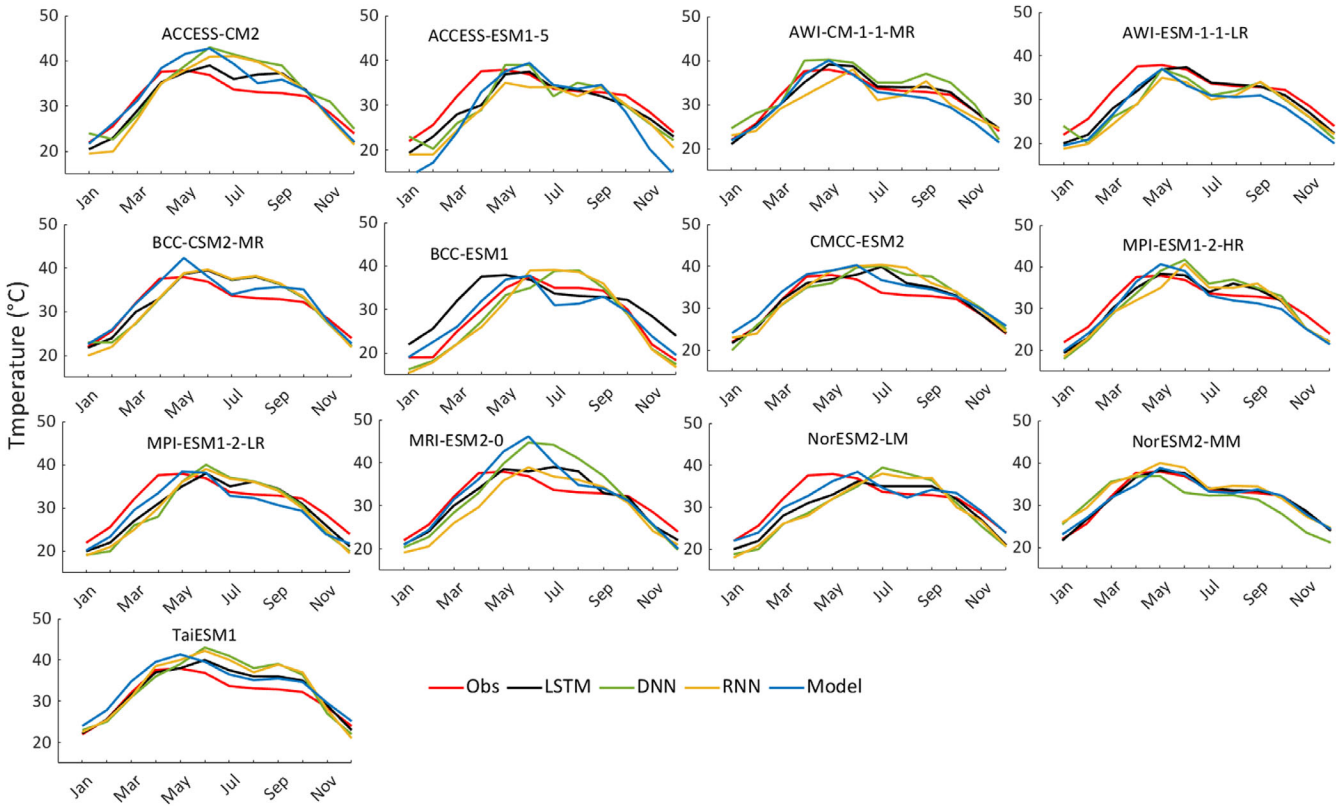

FIGURE 10 Time series diagram of the monthly downscaled maximum temperature using deep learning based models (LSTM, DNN and RNN), observed and all 13 GCMs model data outputs for the validation (2007–2010) period (*x*-axis shows the time in months). [Colour figure can be viewed at wileyonlinelibrary.com]

reproduces the temporal variability relatively closer to the observation in the range of 19–39.8°C for all GCMs model outputs in the validation period (Figure 10) compared to two other deep learning methods. DNN and RNN temporal temperature variability give the mixed (overestimate and underestimate) results behave for

different GCMS model during the validation (2007–2010) period (Figure 10).

The ability of the DL methods to reproduce the spatiotemporal variability of daily maximum temperature is assessed over the IGP for all 13 GCMs model output. (Figures 11 and 12). The spatiotemporal variability of observed maximum temperature lies between the ranges (26–34°C) during the both calibration and validation period. LSTM show the spatiotemporal variability relatively closer to the observation in the range of 18–38°C for both calibration and validation period (Figures 11 and 12). LSTM outputs of all model output variation of (18–38°C) show closer to the observed variation across the IGP plain except model ACCESS-CM2 and BCC-CSM2-MR in the calibration period. Same in validation period LSTM all model outputs spatial variation shows closer to the observed spatial variation except model CMCC-ESM2 and TaiESM1. Both DNN and RNN downscale data spatial variability behave different for different model like some model overestimate (reaching up to 34°C) and some model underestimate (reaching below to 30°C) for most of the part of the IGP plain during both calibration and validation period. The mean of the temperature predicted by LSTM method range between for all model (28.36–32.57°C) is close to the observed mean (31.2°C) value compare to other two deep learning

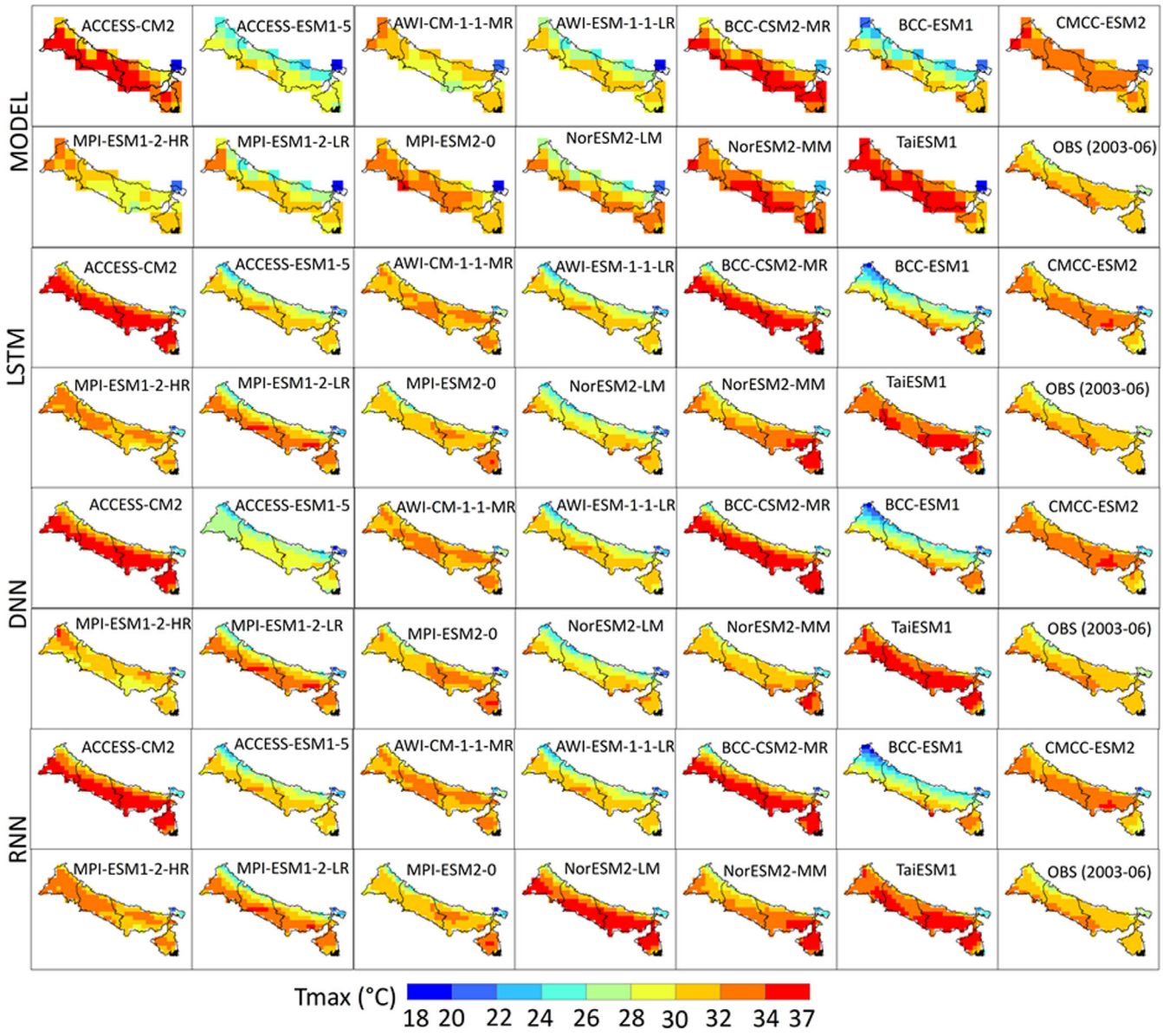

FIGURE 11 Distribution map of spatial variation for the observed, model and downscaled maximum temperature based on the LSTM, DNN and RNN methods of Indo-Gangetic Plain (India) for the calibration set (2003–2006). [Colour figure can be viewed at wileyonlinelibrary.com]

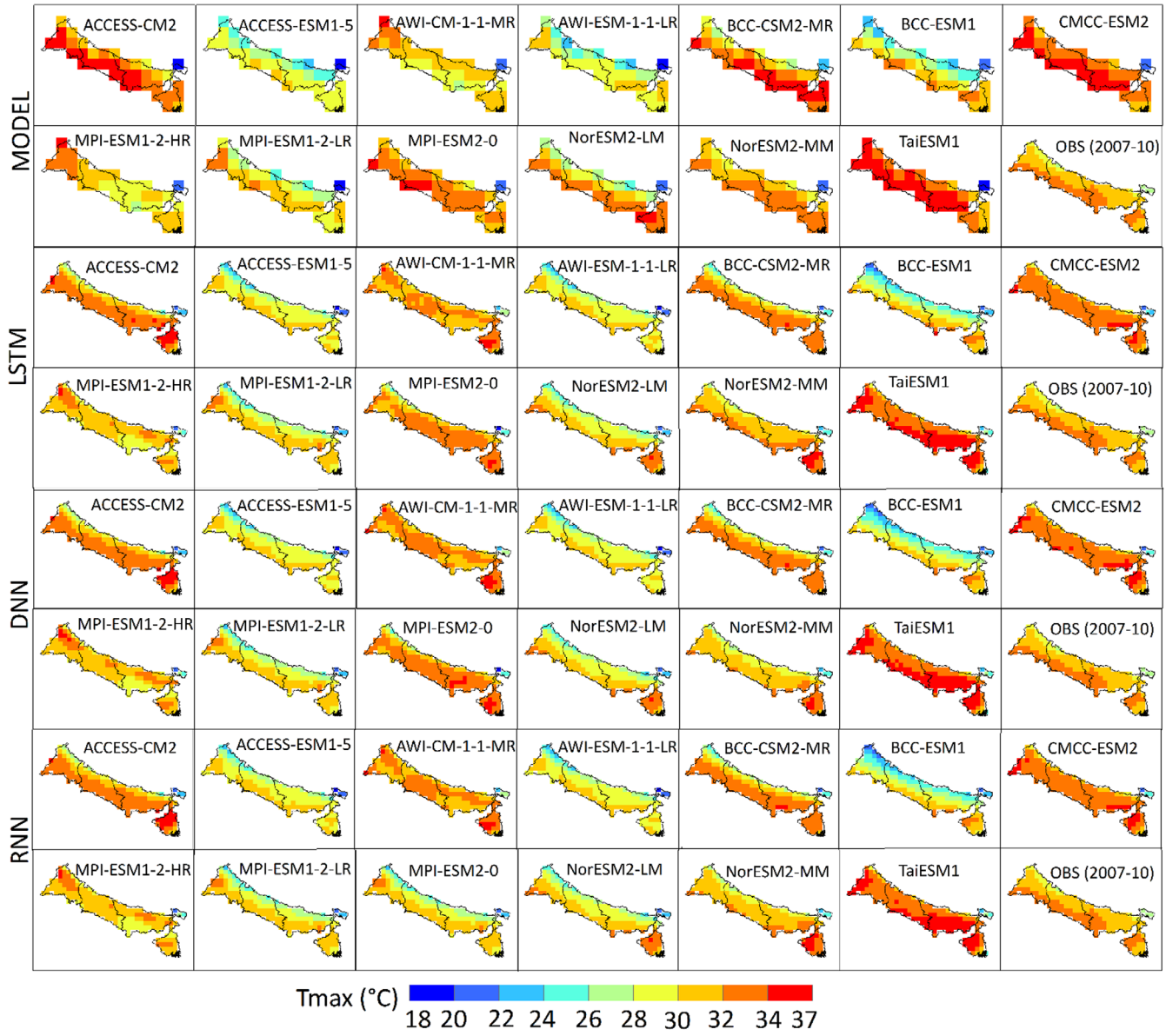

FIGURE 12 Distribution map of spatial variation for the observed, model and downscaled maximum temperature based on the LSTM, DNN and RNN methods of Indo-Gangetic Plain (India) for the validation set (2007–2010). [Colour figure can be viewed at wileyonlinelibrary.com]

methods (Table 4). The spatial distribution of downscaled maximum temperature of GCMs models shows LSTM to be best in reproducing the spatial variability of observed temperature over IGP for both calibration and validation period (Figures 11 and 12). Models simulated maximum temperature overestimated the temperature reaching up to 34°C over most of the IGP during the calibration period which improved relatively further in validation period.

Overall, the spatiotemporal variability analysis suggested that the all three DL methods output perform better than the GCMs model and out of which LSTM performs the best while RNN and DNN methods underestimate and overestimate the observed variability. The downscaled results of the LSTM method were found to have remarkably lower errors than those of the other methods. LSTMs have a "memory" which enables them to identify and capture the previously calculated useful information, and pass them along to the next iteration. Moreover, the daily temperature has a close connection with the temperature of previous days. Incorporating antecedent predictor values could improve the accuracy of temperature prediction (Coulibaly et al., 2005).

It has been observed from all the results that LSTM method outperform other two standard deep learning methods. RNNs utilize internal state memory exploiting

TABLE 4 Average mean values of all observed and GCMs model downscaled maximum temperature based on the LSTM, DNN and RNN methods of Indo-Gangetic Plain (India) for the validation period (2007–2010).

	MEAN_2007			
Model	**OBS**	**LSTM**	**DNN**	**RNN**
ACCESS-CM2	31.20	31.70	33.44	31.90
ACCESS-ESM1-5	31.20	29.95	29.59	28.13
AWI-CM-1-1-MR	31.20	31.63	33.02	30.07
AWI-ESM-1-1-LR	31.20	29.74	28.74	27.87
BCC-CSM2-MR	31.20	31.79	31.63	31.39
BCC-ESM1	31.20	28.36	27.69	27.74
CMCC-ESM2	31.20	32.21	32.66	33.01
MPI-ESM1-2-HR	31.20	30.63	31.00	30.30
MPI-ESM1-2-LR	31.20	29.73	29.31	29.28
MRI-ESM2-0	31.20	31.60	32.28	29.41
NorESM2-LM	31.20	29.59	29.29	29.16
NorESM2-MM	31.20	31.50	30.55	32.72
TaiESM1	31.20	32.57	33.37	33.46

the temporal dependencies and hence in most of the observed results perform better or closer to DNNs. However, as RNNs suffer from vanishing gradient issue; LSTMs outperform RNNs and DNNs for the time-series prediction task (Shah et al., 2018). Additionally, the time complexity of LSTMs is of O(1) (Order of one) as compared to DNNs which depends on number of hyper parameters as discussed in Pandey et al. (2021), hence, they are much faster to implement which is an advantage. Finally, the learning capabilities of both short- and long-term dependencies due to the gated LSTM architecture enables it to outperform other two deep learning architectures.

3.4 | Heat waves events

The results show both the variations observed in capturing heat wave events due to differences in the models as well as due to the three different methods used in the study (Figure 13). It is found that the models differ in simulating heat wave events over the IGP from the

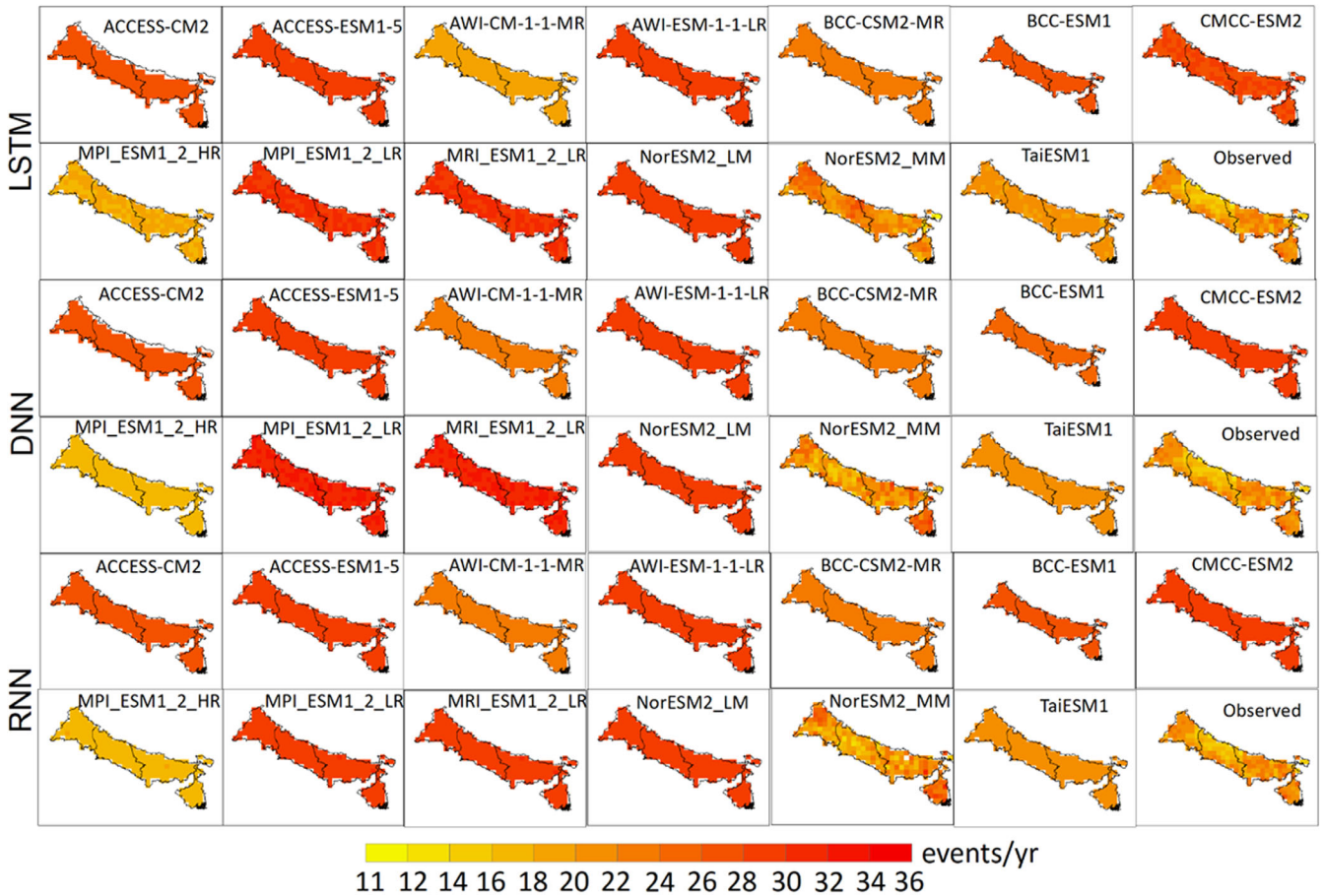

FIGURE 13 Spatial distribution of total number of heat wave events over the IGP for the validation period 2007–2010 (Mar–Jun). [Colour figure can be viewed at wileyonlinelibrary.com]

observed heat wave events which ranges from 13 to 28 in all of the three methods. The results show that downscaled model outputs observe heat waves in three classes, that is, high (28–33), intermediate (23–27) and low (17–22) ranges of heat wave events. Out of the three methods it was found that most of the models simulated heat wave in the high range of 27–30 followed by intermediate 24 and only two in the lower range ~17–19. Among the models, MPI-ESM1-2-HR (17–19) and TaiESM1 (19–22) are found to record heat wave in the lower range of heat wave events over all the grids in all of the three methods with LSTM downscaled output showing more spatial variation than DNN and RNN. ACCESS-ESM1-5 (30), AWI-ESM-1-1-LR (30), ACCESS-CM2 (27–28), CMCC-ESM2 (28–29), MPI-ESM1-2-LR (30–33) and MRI-ESM2-0 (30–33) models over predicted the heat waves as compared to the observed heat wave events throughout the IGP with spatial variability exhibited by only MPI-ESM1-2-LR, MRI-ESM2-0, and only AWI-CM-1-1-MR (21–24), BCC-CSM2-MR-LSTM (24), and BCC-ESM1 (26–27) recorded heat wave in the intermediate range (Figure 13). Among all the models it was found that NorESM2-MM performed best in simulating heat waves with similar range to the range of observation. The frequency of heat wave events ranges from 13 to 28 over the study period which is closely approximated by the model in the three methods by LSTM (12–28) followed by RNN (13–31) and DNN (15–29). The spatial distribution varies with the method and zones when compared with the observation (Figure 1). For NorESM2-MM, LSTM performs better in the Middle Gangetic Plain and Lower Gangetic Plain whereas RNN and DNN show better performance in the Trans Gangetic Plain and Upper Gangetic Plain. Further, LSTM overestimates the heat wave occurrence in Trans and Upper Gangetic Plain while DNN and RNN overestimate heat wave frequency in the Lower Gangetic Plain, and both over- and underestimation across the grids of Middle Gangetic Plain. For all the models the occurrence of heat wave events within the zones are more consistent in LSTM than DNN and RNN. It was found that the downscaled outputs of the models did not show large variation in range and spatial distribution for the three methods employed. However, among the three methods it was found that in case of DNN and RNN models could not approximate the spatial variability and simulated relatively the higher range as compared to the observed heat wave events when compared to LSTM which approximated both range and spatial variability better. It is also noteworthy that the uncertainty observed in the heat wave prediction is more due to the use of different models than due to the three methods. Out of the three methods DNN and RNN showed similar results but LSTM outperformed both the methods. Overall, all of the three deep learning methods performed well but LSTM was better and so can be used to study heat waves over India.

4 | CONCLUSION

This study investigated the performance of deep learning methods, namely LSTM, DNN and RNN, for downscaling daily Tmax for CMIP6, GCMs 13 models over IGP, India. The study found that the spatial distribution of LSTM show the best in downscaling daily temperature in validation period, while DNN and RNN mix result has been found the maximum temperature against the observation for all the 13 GCMs models. LSTM was also found to be better than DNN and RNN on the basis of other model performance metrics, spatiotemporal and intra-annual variability, as compared to the observed dataset for all the GCMs models. The improved performance of the LSTM can be explained by its ability to learn long-term dependencies in sequential datasets. Meanwhile, in case of analysing the extremes, that is, heat wave events, it was observed that LSTM method captured the range of the events similar to the observation and better than RNN and DNN. In terms of approximating the spatial distribution and variability, LSTM showed similar distribution of heat wave events in the Middle Gangetic Plain and Lower Gangetic Plain. While DNN and RNN were closer to the observation in the Trans Gangetic Plain and Upper Gangetic Plain. The intermodel variations were not much pronounced, but LSTM outperformed the two methods when both spatial and temporal characteristics were considered; therefore, it can be concluded that LSTM performed well for heat wave events. On the basis of the statistical and extreme event analysis, LSTM was found to highly resonate with the mean and extreme characteristics of the observed temperature data. The study concludes that LSTM showed better performance than DNN and RNN models in the downscaling CMIP6 GCM Tmax data and can be used for downscaling of other CMIP6-GCMs data over IGP. The use of deep learning methods is still incipient in the field of climate studies, and so our study will be instrumental in supporting and validating the use of deep learning in the field of climate science. The study provides a parallel, computationally and time efficient technique of downscaling GCM datasets for proving more accurate regional climate information which can be used in different climate impact assessment studies both for present and future period.

AUTHOR CONTRIBUTIONS

Manisha Chaturvedi: Formal analysis; investigation; validation; visualization; writing – original draft; software. **Rajesh Kumar Mall:** Conceptualization; methodology; data curation; supervision; resources; project

administration; writing – review and editing; funding acquisition. **Saumya Singh:** Writing – review and editing; visualization; data curation; software. **Pawan K. Chaubey:** Software; data curation; visualization; formal analysis. **Ankur Pandey:** Methodology; software.

ACKNOWLEDGEMENTS

Authors thank the Climate Change Programme, Department of Science and Technology, New Delhi for financial support (DST/CCP/CoE/80/2017(G)). The authors gratefully acknowledge the World Climate Research Programme's Working Groups, the former coordinating body of CMIP6 for producing and making the data available. The authors thank the Earth System Grid Federation (ESGF) infrastructure and the Climate Data Portal hosted at the Centre for Climate Change Research (CCCR), Indian Institute of Tropical Meteorology (IITM) for providing CMIP6 Asia data. The authors also thank the India Meteorology Department (IMD) for providing the observed climate dataset.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The CMIP6 GCM models temperature data available from Coupled Model Intercomparison Project Phase 6 (CMIP6) (https://esgf-node.llnl.gov/search/cmip6/). Observed data available from the India Meteorology Department (IMD) at: http://www.imdpune.gov.in/Clim_Pred_LRF_New/Grided_Data_Download.html. Authors declare that all software application and custom code support their published claims and comply with field standards. Authors declare that all data and materials support their published claims and comply with field standards. Authors promote data transparency and agree to provide in case of request from the journal.

ORCID

Rajesh Kumar Mall https://orcid.org/0000-0002-3118-096X

REFERENCES

Ahmed, K., Sachindra, D.A., Shahid, S., Iqbal, Z., Nawaz, N. & Khan, N. (2020) Multi-model ensemble predictions of precipitation and temperature using machine learning algorithms. *Atmospheric Research*, 236, 104806. Available from: https://doi.org/10.1016/j.atmosres.2019.104806

Alagh, Y.K. (1990) Agro-climatic planning and regional development. *Indian Journal of Agricultural Economics*, 45, 244–268. Available from: https://doi.org/10.1007/s12571-021-01250-z

Banjara, T.R., Bohra, J.S., Kumar, S., Singh, T., Shori, A. & Prajapat, K. (2022) Sustainable alternative crop rotations to the irrigated rice-wheat cropping system of Indo-Gangetic Plains of India. *Archives of Agronomy and Soil Science*, 68(11), 1568–1585. Available from: https://doi.org/10.1080/03650340.2021.1912324

Cavalcante, R.B.L., da Silva Ferreira, D.B., Pontes, P.R.M., Tedeschi, R.G., da Costa, C.P.W. & de Souza, E.B. (2020) Evaluation of extreme rainfall indices from CHIRPS precipitation estimates over the Brazilian Amazonia. *Atmospheric Research*, 238, 104879. Available from: https://doi.org/10.1016/j.atmosres.2020.104879

Chadwick, R., Coppola, E. & Giorgi, F. (2011) An artificial neural network technique for downscaling GCM outputs to RCM spatial scale. *Nonlinear Processes in Geophysics*, 18(6), 1013–1028. Available from: https://doi.org/10.5194/npg-18-1013-2011

Chaubey, P.K. & Mall, R.K. (2023) Intensification of extreme rainfall in Indian river basin: using bias corrected CMIP6 climate data. *Earth's Futures*, 11, e2023EF003556. Available from: https://doi.org/10.1029/2023EF003556

Coulibaly, P., Dibike, Y.B. & Anctil, F. (2005) Downscaling precipitation and temperature with temporal neural networks. *Journal of Hydrometeorology*, 6(4), 483–496.

Eyring, V., Gleckler, P.J., Heinze, C., Stouffer, R.J., Taylor, K.E., Balaji, V. et al. (2016) Towards improved and more routine Earth system model evaluation in CMIP. *Earth System Dynamics*, 7(4), 813–830. Available from: https://doi.org/10.5194/esd-7-813-2016

Feng, Z.K., Niu, W.J., Tang, Z.Y., Xu, Y. & Zhang, H.R. (2021) Evolutionary artificial intelligence model via cooperation search algorithm and extreme learning machine for multiple scales nonstationary hydrological time series prediction. *Journal of Hydrology*, 595, 126062. Available from: https://doi.org/10.1016/j.jhydrol.2021.126062

Ghosh, S. & Mujumdar, P.P. (2008) Statistical downscaling of GCM simulations to streamflow using relevance vector machine. *Advances in Water Resources*, 31(1), 132–146. Available from: https://doi.org/10.1016/j.advwatres.2007.07.005

Han, Z., Zhao, J., Leung, H., Ma, K.F. & Wang, W. (2019) A review of deep learning models for time series prediction. *IEEE Sensors Journal*, 21(6), 7833–7848. Available from: https://doi.org/10.1109/JSEN.2019.2923982

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

IPCC. (2021) New physical science behind climate change: Zhou, T. J (2021). What does IPCC AR6 tell us? *The Innovation (NY)*, 2, 100173. Available from: https://doi.org/10.1016/j.xinn.2021.100173

Ishtiaque, A., Singh, S., Lobell, D., Fishman, R. & Jain, M. (2022) Prior crop season management constrains farmer adaptation to warming temperatures: evidence from the Indo-Gangetic Plains. *Science of the Total Environment*, 807, 151671. Available from: https://doi.org/10.1016/j.scitotenv.2021.151671

Kang, Y.J., Noh, Y., Jang, M.S., Park, S. & Kim, J.T. (2023) Hierarchical level fault detection and diagnosis of ship engine systems. *Expert Systems with Applications*, 213, 118814. Available from: https://doi.org/10.1016/j.eswa.2022.118814

Kaveh, K., Kaveh, H., Bui, M.D. & Rutschmann, P. (2021) Long short-term memory for predicting daily suspended sediment concentration. *Engineering with Computers*, 37, 2013–2027.

Khan, P., Reddy, B.S.K., Pandey, A., Kumar, S. & Youssef, M. (2020) Differential channel state-information-based human

activity recognition in IoT networks. *IEEE Internet of Things Journal*, 7(11), 11290–11302. Available from: https://doi.org/10.1109/JIOT.2020.2997237

Kolluru, V., Kolluru, S., Wagle, N. & Acharya, T.D. (2020) Secondary precipitation estimate merging using machine learning: development and evaluation over Krishna river basin, India. *Remote Sensing*, 12(18), 3013. Available from: https://doi.org/10.3390/rs12183013

Kumar, B., Chattopadhyay, R., Singh, M., Chaudhari, N., Kodari, K. & Barve, A. (2021) Deep learning-based downscaling of summer monsoon rainfall data over Indian region. *Theoretical and Applied Climatology*, 143, 1145–1156. Available from: https://doi.org/10.1007/s00704-020-03489-6

Kuttippurath, J., Murasingh, S., Stott, P.A., Sarojini, B.B., Jha, M.K., Kumar, P. et al. (2021) Observed rainfall changes in the past century (1901–2019) over the wettest place on Earth. *Environmental Research Letters*, 16(2), 024018. Available from: https://doi.org/10.1088/1748-9326/abcf78

LeCun, Y., Bengio, Y. & Hinton, G. (2015) Deep learning. *Nature*, 521(7553), 436–444. Available from: https://doi.org/10.1038/nature14539

Liu, Y., Jing, W., Wang, Q. & Xia, X. (2020) Generating high-resolution daily soil moisture by using spatial downscaling techniques: a comparison of six machine learning algorithms. *Advances in Water Resources*, 141, 103601.

Mall, R.K., Chaturvedi, M., Singh, N., Bhatla, R., Singh, R.S., Gupta, A. et al. (2021) Evidence of asymmetric change in diurnal temperature range in recent decades over different agro-climatic zones of India. *International Journal of Climatology*, 41(4), 2597–2610. Available from: https://doi.org/10.1002/joc.6978

Mall, R.K., Srivastava, R.K., Banerjee, T., Mishra, O.P., Bhatt, D. & Sonkar, G. (2019) Disaster risk reduction including climate change adaptation over South Asia: challenges and ways forward. *International Journal of Disaster Risk Science*, 10, 14–27. Available from: https://doi.org/10.1007/s13753-018-0210-9

Niazkar, M., Goodarzi, M.R., Fatehifar, A. & Abedi, M.J. (2023) Machine learning-based downscaling:application of multi-gene genetic programming for downscaling daily temperature at Dogonbadan, Iran, under CMIP6 scenarios. *Theoretical and Applied Climatology*, 151(1), 153–168. Available from: https://doi.org/10.1007/s00704-022-04274-3

Okkan, U., Fistikoglu, O., Ersoy, Z.B. & Noori, A.T. (2023) Investigating adaptive hedging policies for reservoir operation under climate change impacts. *Journal of Hydrology*, 619, 129286.

Okkan, U. & Inan, G. (2015a) Bayesian learning and relevance vector machines approach for downscaling of monthly precipitation. *Journal of Hydrologic Engineering*, 20(4), 04014051.

Okkan, U. & Inan, G. (2015b) Statistical downscaling of monthly reservoir inflows for Kemer watershed in Turkey: use of machine learning methods, multiple GCMs and emission scenarios. *International Journal of Climatology*, 35(11), 3274–3295.

Pandey, A., Sequeira, R. & Kumar, S. (2021) SELE: RSS-based Siamese embedding location estimator for a dynamic IoT environment. *IEEE Internet of Things Journal*, 9(5), 3672–3683. Available from: https://doi.org/10.1109/JIOT.2021.3098356

Pandey, V., Srivastava, P.K., Mall, R.K., Munoz-Arriola, F. & Han, D. (2022) Multi-satellite precipitation products for meteorological drought assessment and forecasting in Central India. *Geocarto International*, 37(7), 1899–1918.

Peach, L., da Silva, G.V., Cartwright, N. & Strauss, D. (2023) A comparison of process-based and data-driven techniques for downscaling offshore wave forecasts to the near shore. *Ocean Modelling*, 182, 102168. Available from: https://doi.org/10.1016/j.ocemod.2023.102168

Perkins, S.E., & Alexander, L.V. (2013). On the Measurement of Heat Waves. *Journal of Climate*, 26(13), 4500–4517. Available from: https://doi.org/10.1175/jcli-d-12-00383.1

Qiu, L., Kim, J.B., Kim, S.H., Choi, Y.W., Im, E.S. & Bae, D.H. (2022) Reduction of the uncertainties in the hydrological projections in Korean river basins using dynamically downscaled climate projections. *Climate Dynamics*, 59(7–8), 2151–2167.

Racsko, P., Szeidl, L. & Semenov, M.A. (1991) A serial approach to local stochasticmodels. *Journal of Ecological Modeling*, 57, 27–41. Available from: https://doi.org/10.1016/0304-3800(91)90053-4

Sabarinath, A., Naga Rajesh, A., Gunthe, S.S. & Lakshmi Kumar, T.V. (2023) Application of deep learning algorithms to correct bias in CMIP6 simulations of surface air temperature over the Indian monsoon core region. *International Journal of Climatology*, 43, 7496–7515. Available from: https://doi.org/10.1002/joc.8276

Sachindra, D.A. & Perera, B.J.C. (2016) Statistical downscaling of general circulation model outputs to precipitation accounting for non-stationarities in predictor-predictand relationships. *PLoS One*, 11(12), e0168701. Available from: https://doi.org/10.1371/journal.pone.0168701

Salahuddin, Z., Woodruff, H.C., Chatterjee, A. & Lambin, P. (2022) Transparency of deep neural networks for medical image analysis: a review of interpretability methods. *Computers in Biology and Medicine*, 140, 105111. Available from: https://doi.org/10.1016/j.compbiomed.2021.105111

Shah, D., Campbell, W. & Zulkernine, F.H. (2018) A comparative study of LSTM and DNN for stock market forecasting. In: *2018 IEEE international conference on big data (big data)*. IEEE, Seattle, WA, USA, pp. 4148–4155. Available from: https://doi.org/10.1109/BigData.2018.8622462

Singh, N., Chaturvedi, M. & Mall, R.K. (2023) Unraveling diurnal asymmetry of surface temperature under warming scenarios in diverse agroclimate zones of India. *Theoretical and Applied Climatology*, 152, 321–335. Available from: https://doi.org/10.1007/s00704-023-04407-2

Singh, N., Mall, R.K., Banerjee, T. & Gupta, A. (2021b) Association between climate and infectious diseases among children in Varanasi city, India: a prospective cohort study. *Science of the Total Environment*, 796, 148769. Available from: https://doi.org/10.1016/j.scitotenv.2021.148769

Singh, N., Mhawish, A., Ghosh, S., Banerjee, T. & Mall, R.K. (2019) Attributing mortality from temperature extremes: a time series analysis in Varanasi, India. *Science of the Total Environment*, 665, 453–464. Available from: https://doi.org/10.1016/j.scitotenv.2019.02.074

Singh, S. & Mall, R.K. (2023) Frequency dominates intensity of future heat waves over India. *Iscience*, 26(11), 108263. Available from: https://doi.org/10.1016/j.isci.2023.108263

Singh, S., Mall, R.K. & Singh, N. (2021a) Changing spatio-temporal trends of heat wave and severe heat wave events over India: an

emerging health hazard. *International Journal of Climatology*, 41, E1831–E1845. Available from: https://doi.org/10.1002/joc.6814

Smolinska, A., Hauschild, A.C., Fijten, R.R.R., Dallinga, J.W., Baumbach, J. & Van Schooten, F.J. (2014) Current breathomics a review on data pre-processing techniques and machine learning in metabolomics breathe analysis. *Journal of Breath Research*, 8(2), 027105. Available from: https://doi.org/10.1088/1752-7155/8/2/027105

Taylor, K.E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183–7192. Portico. Available from: https://doi.org/10.1029/2000jd900719

Tran Anh, D., Van, S.P., Dang, T.D. & Hoang, L.P. (2019) Downscaling rainfall using deep learning long short-term memory and feed forward neural network. *International Journal of Climatology*, 39(10), 4170–4188. Available from: https://doi.org/10.1002/joc.6066

Tripathi, S., Srinivas, V.V. & Nanjundiah, R.S. (2006) Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology*, 330(3–4), 621–640. Available from: https://doi.org/10.1016/j.jhydrol.2006.04.030

Trzaska, S. & Schnarr, E. (2014) A review of downscaling methods for climate change projections. United States Agency for International Development by Tetra Tech ARD (September), pp. 1–42.

Vandal, T., Kodra, E. & Ganguly, A.R. (2019) Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation. *Theoretical and Applied Climatology*, 137, 557–570.

Wang, B., Zheng, L., Liu, D.L., Ji, F., Clark, A. & Yu, Q. (2018) Using multi-model ensembles of CMIP5 global climate models to reproduce observed monthly rainfall and temperature with machine learning methods in Australia. *International Journal of Climatology*, 38(13), 4891–4902. Available from: https://doi.org/10.1002/joc.5705

Wang, F. & Tian, D. (2022) On deep learning-based bias correction and downscaling of multiple climate models simulations. *Climate Dynamics*, 59(11–12), 3451–3468. Available from: https://doi.org/10.1007/s00382-022-06277-2

Weisse, R. & Oestreicher, R. (2001) Reconstruction of potential evaporation for water balance studies. *Climate Research*, 16(2), 123–131. Available from: https://doi.org/10.3354/cr016123

Wilby, R.L., Dawson, C.W. & Barrow, E.M. (2002) SDSM—a decision support tool for the assessment of regional climate change impacts. *Environmental Modelling & Software*, 17(2), 145–157. Available from: https://doi.org/10.1016/S1364-8152(01)00060-3

Yadav, M.K., Singh, R.S., Singh, K.K., Mall, R.K., Patel, C.B., Yadav, S.K. et al. (2015) Assessment of climate change impact on productivity of different cereal crops in Varanasi, India. *Journal of Agrometeorology*, 17(2), 179–184.

Yu, Y., Si, X., Hu, C. & Zhang, J. (2019) A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation*, 31(7), 1235–1270. Available from: https://doi.org/10.1162/neco_a_01199

Zhang, S. & Li, X. (2021) Future projections of offshore wind energy resources in China using CMIP6 simulations and a deep learning-based downscaling method. *Energy*, 217, 119321. Available from: https://doi.org/10.1016/j.energy.2020.119321

How to cite this article: Chaturvedi, M., Mall, R. K., Singh, S., Chaubey, P. K., & Pandey, A. (2024). Statistical downscaling of maximum temperature under CMIP6 global climate models and evaluation of heat wave events using deep learning methods for Indo-Gangetic Plain. *International Journal of Climatology*, 1–20. https://doi.org/10.1002/joc.8366