Theoretical and Applied Climatology (2022) 149:945–963
https://doi.org/10.1007/s00704-022-04103-7

ORIGINAL PAPER

Hybrid deep learning approach for multi-step-ahead prediction for daily maximum temperature and heatwaves

Mohd Imran Khan[1] · Rajib Maity[1]

Received: 24 May 2021 / Accepted: 16 May 2022 / Published online: 24 May 2022

Abstract

Availability of increasing information and digital meteorological data leads to an opportunity for better simulations/prediction of complex hydroclimatic phenomena. However, volume and size of such data and underlying complex association pose many challenges to traditional approaches. This study focuses on the potential of a hybrid deep learning (DL) approach, a combination of one-dimensional convolutional neural network (Conv1D) and long short-term memory (LSTM) neural network (hereinafter hybrid Conv1D-LSTM), for multi-step-ahead (1-day to 10-day) daily maximum temperature prediction. The proposed approach is applied to twenty-eight major cities in India, located in different climate regimes, to explore its potential to predict the daily maximum temperature and to foresee the heatwave events. Seven meteorological precursors, closely associated with daily temperature variation along with the month index are used as input and the proposed approach is expected to efficiently learn the complex relationship between the precursors and daily maximum temperature. Apart from its alluring performance in predicting the daily maximum temperature, the results also show some promise to raise an alert for the upcoming heatwaves. The performance of the proposed hybrid model is also compared with other machine learning (ML), DL-based approaches, and three popular weather applications (weather apps) that help to portray the superiority of the proposed hybrid DL–based approach.

1 Introduction

It is observed that the global average temperatures have shown a warming trend of about 0.85 °C in the past century due to climate change and are expected to increase by upto 5.5 °C by the end of the twenty-first century (IPCC 2013; Mazdiyasni et al. 2017; Rohini et al. 2016). Considering the Indian mainland, a rise in the mean annual temperature of about 0.51 °C with an increase in warm days is observed during the period 1961–2007 (Kothawale et al. 2010). The change in mean value of maximum temperature may cause an increase in temperature extremes and thereby lead to a heatwave event, an unusual extreme temperature prevailing over for days in a region with serious consequences. Maximum temperature plays a crucial role in managing several activities ranging from ecosystem to hydrological system to social welfare, required for the prosperity of a nation, such as preparedness against heatwave, crop failure, and wild fire (Mazdiyasni et al. 2017; Murari et al. 2015). Therefore, considering the expected increase in temperature and its resulting calamitous consequences under the changing climate, there is always a need to improve the quality of daily maximum temperature prediction at a region of interest (Ma et al. 2015; Vasseur et al. 2014).

The variation in daily maximum temperature of a place is affected by numerous uncertain factors viz. local altitude, latitude, land use land cover, wind pattern, and even ocean currents and distance from the sea, in case of coastal areas. Several studies have been carried out in the recent past to forecast maximum temperature, minimum temperature, and average temperature at various spatio-temporal scale. Many of them requires an exhaustive information about the physical processes viz. laws of physics, atmospheric chemistry, and fluid motion, making the simulations computationally intensive (Luk et al. 2000). Whereas some approaches implicitly consider the inherent physical processes without any explicit requirement as it is in case of conceptual and physical models, e.g., data-driven approaches (DDAs) (Cifuentes et al. 2020; Scher 2018; Tran et al. 2021). Recent growth in data records and computational power

✉ Rajib Maity
rajib@civil.iitkgp.ac.in; rajibmaity@gmail.com

[1] Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721302, West Bengal, India

has enhanced the potential of DDAs, particularly artificial intelligence (AI)-based machine learning (ML) approaches. These approaches have proven their potential in understanding many complex phenomenon such as speech recognition, natural language processing, and image analysis and are gradually paving their way in complex hydrological processes (Khan and Maity 2020; Kratzert et al. 2019b; Krizhevsky et al. 2012; LeCun et al. 2015; Liu et al. 2017; Maity et al. 2021; Nearing et al. 2021; Pan et al. 2019). Among several ML approaches, deep learning (DL), a name coined to a new subset of ML, is the recently popularized DDA which has an ability to learn high-level abstractions from the raw data features without any human expertise and contribute information to the model by using its hierarchical architecture. DL-based models have high model efficiency, processing capability, and potential in capturing complex associations between input features and the target variables from a cluster of available data sources which is beyond the capabilities of older ML models (Khan and Maity 2020; LeCun et al. 2015; Maity et al. 2021; Matsuoka et al. 2020, 2018).

In the domain of hydroclimatic analysis and modelling, during the recent decade, a number of studies have demonstrated the effectiveness of DL algorithms over traditional/existing ML approaches in forecasting/simulating hydroclimatic variables such as temperature, streamflow, soil moisture, rainfall, and wind speed at various spatio-temporal scales. To discuss a few, Liu et al. (2016) proposed a two-dimensional convolutional neural network (CNN) model to extract the weather information from the climate dataset. The study was also successful in identifying the weather extremes. Hu et al. (2018) simulated rainfall-runoff process with the help of long short term memory neural network (LSTM) and compared the performance with other artificial neural networks (ANNs) and were found better. Wang and Li (2018) also showed the potential of LSTM in forecasting wind speed. Khan and Maity (2020) designed a DL-based hybrid architecture comprising of one-dimensional CNN (Conv1D) and multi-layer perceptron (MLP) to forecast multi-step ahead daily rainfall. The performance of the proposed hybrid model was also compared with two other popular ML approaches viz. MLP and support vector regression (SVR) and was found better. Fu et al. (2020) analyzed the potential of DL-based LSTM model in forecasting another hydroclimatic variable, i.e., streamflow, at daily scale. The accuracy of the developed LSTM model was tested with backpropagation neural network model and was found better. Fang et al. (2021) also showed the effectiveness of LSTM model in making prediction of two dynamic hydroclimatic variables viz. soil moisture and streamflow by incorporating data synergy method with DL. The proposed approach was successful in developing a more robust model by pooling a large dataset irrespective of the region homogeneity, i.e., a single model was trained on the whole dataset rather than training the model on region wise dataset. Thus, the study presents a scope of unification of the several meteorological data with DL to overcome the shortage of data in making prediction at a particular region. Similarly, several studies have been attempted to forecast hydroclimatic variables at different spatio-temporal scales in different regions of the globe (Cai et al. 2019; Chattopadhyay et al. 2020; Ham et al. 2019; Liu et al. 2018; Matsuoka et al. 2018; Oh et al. 2020; Scher 2018; Shen et al. 2019; Shi et al. 2016; Sun and Tang 2020; etc.).

Apart from the aforesaid literatures, studies specifically related to temperature forecasting have also been carried out in the recent decade, using AI-based ML/DL DDAs (Cifuentes et al. 2020; Tran et al. 2021). For instance, Kisi and Shiri (2014) assessed the potential of adaptive neuro-fuzzy inference system (ANFIS) and ANN in predicting monthly air temperature in Iran. In this study, four geographical inputs viz. station latitude, longitude, and altitude, and month number of the year (periodicity) are used as input to the model and monthly averaged air temperature was estimated. In order to perform the training of the model, 14 years of data from 20 weather stations are utilized and the performance was assessed on ten other weather stations. The testing performance of both the models are analyzed using coefficient of determination and root mean square error (RMSE) and ANN having a lower RMSE (i.e., in the range of 1.53 to 4.20 °C) across all the 10 testing stations is found better. Kisi and Sanikhani (2015) carried out another study to evaluate the performance of five different DDAs viz. gene expression programming (GEP), ANN, support vector regression (SVR), ANFIS with subtractive clustering (SC), and ANFIS with grid partition (GP), using 50 weather station data of Iran. The authors utilized the same set of geographical inputs, as mentioned in the previous study, for training. The result of the analysis shows a clear dominance of the SVR model over other models in terms of lower RMSE, ranging between 0.63 and 2.17 °C. Salcedo-Sanz et al. (2016) compared the performance of two popular regression ML algorithm namely SVR and MLP by predicting monthly mean air temperature at ten stations, eight located in Australia and two in New Zealand. In this study, time series data of past monthly average temperature, two dummy variables, and three climate indices viz. Indian ocean dipole (IOD), pacific decadal oscillation (PDO), and southern oscillation index (SOI) were used to train the model to predict next month average temperature. The result was assessed using mean absolute error (MAE), which showed that SVR yielded better performance (0.73 to 1.33 °C). Papacharalampous et al. (2018) evaluated the performance of SVR and MLP model along with four other classical algorithms by predicting monthly temperature and precipitation in Greece. The study mainly focuses on the performance of the ML models in predicting both the hydroclimatic variables at 1 month and 12 months lead. It was concluded that performance of ML and classical model depends on the criteria of interest and limitations imposed in the study and both can be equally good. The RMSE obtained by the models in case of temperature prediction at 1 month and 12-month lead ranges between 0.66 to 1 °C and 1.14 to 1.70 °C respectively. Likewise, in the case of precipitation prediction, RMSE values at 1 month and 12-month lead ranges between 39 to 72 mm and 41 to 52 mm respectively. Zhang et al. (2020) forecasted the daily average value of temperature using a hybrid of CNN and recurrent neural network (RNN) for mainland China. The model reported an RMSE of 1.67 °C during the testing period. Tran et al. (2020) compared the performance of three different neural networks viz. ANN, RNN, and LSTM, optimized using genetic algorithm (GA), in predicting daily maximum temperature in four different seasons at Cheongju station in South Korea. The study aimed to predict the maximum temperature at 15 different lead times (1-day to 15-day) using 40 years of only maximum temperature time series data. Among the three aforesaid models, GA-based LSTM was found performing better in terms of RMSE values. The lowest RMSE of 2.36 °C was achieved by the best model at 1-day lead during the summer season. Kreuzer et al. (2020) proposes a hybrid of two-dimensional CNN and LSTM to forecast hourly temperature up to 24 h. The proposed approach was applied to five different weather stations in Germany and was found better than the seasonal autoregressive integrated moving average (SARIMA), LSTM, and Naïve forecast, especially at longer leads. The author also showed the forecast at daily scale (24-h advance) by averaging the hourly performance of the model and achieved a RMSE of 2.10 °C.

Summarizing the aforesaid literatures, it can be observed that performance of DL models is better than other models. However, a reasonably higher value of RMSE and feeding of limited input variables to models creates a scope for improving the model performance. This forms the motivation of the study which aims to propose a suitable hybrid DL model to predict daily maximum temperature in major cities of India. Therefore, an analysis is carried out to test the performance with individual DL models and thereafter by combining two best performing DL models (i.e., hybrid model), which may enhance the performance by considering different aspects of modelling of different modules of DL approaches (Khan and Maity 2020). Furthermore, the potentials of DL is used to foresee the temperature-related weather hazards through the prediction of daily maximum temperature. This is in the focus of this study. Thus, the objective of this study is to develop a location-specific hybrid DL approach for predicting maximum daily temperature for 1-day to 10-day-ahead and to foresee the heatwave events, if any. The performance of hybrid model is also compared with LSTM, Conv1D, MLP, SVR, and with the performance of three popular weather apps, namely AccuWeather, real-time weather system, and weather underground to investigate the benefit of the proposed hybrid DL approach.

2 Study area and data

Twenty-eight major cities located in different states of India are considered in this study. Location of these cities are shown in Fig. 1. India is a vast country that spans over a wide range of climatological conditions. According to Koppen climate classification, the climate of India is categorized into six main subcategories. The temperature at a place in the country is classified according to the season that comprises of mainly summer, winter, and rainy season. Indian subcontinent is extremely hot and the maximum temperature at many locations in the country experiences a very hot climate where the daily maximum temperature crosses 40 °C and heatwaves are common in summer season at many places as evidenced in the recent past.

We considered daily maximum temperature from 1979 to 2020 for the analysis. Seven meteorological variables at daily scale, are considered input to the model (Table 1). These are lagged (previous 4 days) values of outgoing longwave radiation ($W/m^{2)}$, relative humidity, resultant of zonal and meridional wind speed (m/s), sea level pressure (kPa) and rainfall (mm), maximum temperature (°C), and solar radiation (J/m^2). In addition, month index, i.e., 1 for January, 2 for February, is also used as an input. Out of these variables, rainfall and maximum temperature values are observed records, converted to gridded products with a spatial resolution of 0.25° (latitude) × 0.25° (longitude) and 1° (latitude) × 1° (longitude), respectively. These are obtained from the India Meteorological Department (IMD) (URL: https://www.imdpune.gov.in/Clim_Pred_LRF_New/Grided_Data_Download.html, accessed in February 2022). Other causal variables are the reanalysis products, obtained from the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF), popularly known as ERA-5 with a spatial resolution of 0.25° (latitude) × 0.25° (longitude) (ERA5, URL: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5, accessed in February 2022).

3 Methodology

3.1 Data preparation

The preparation of dataset and its handling are entirely carried out in scientific python development environment

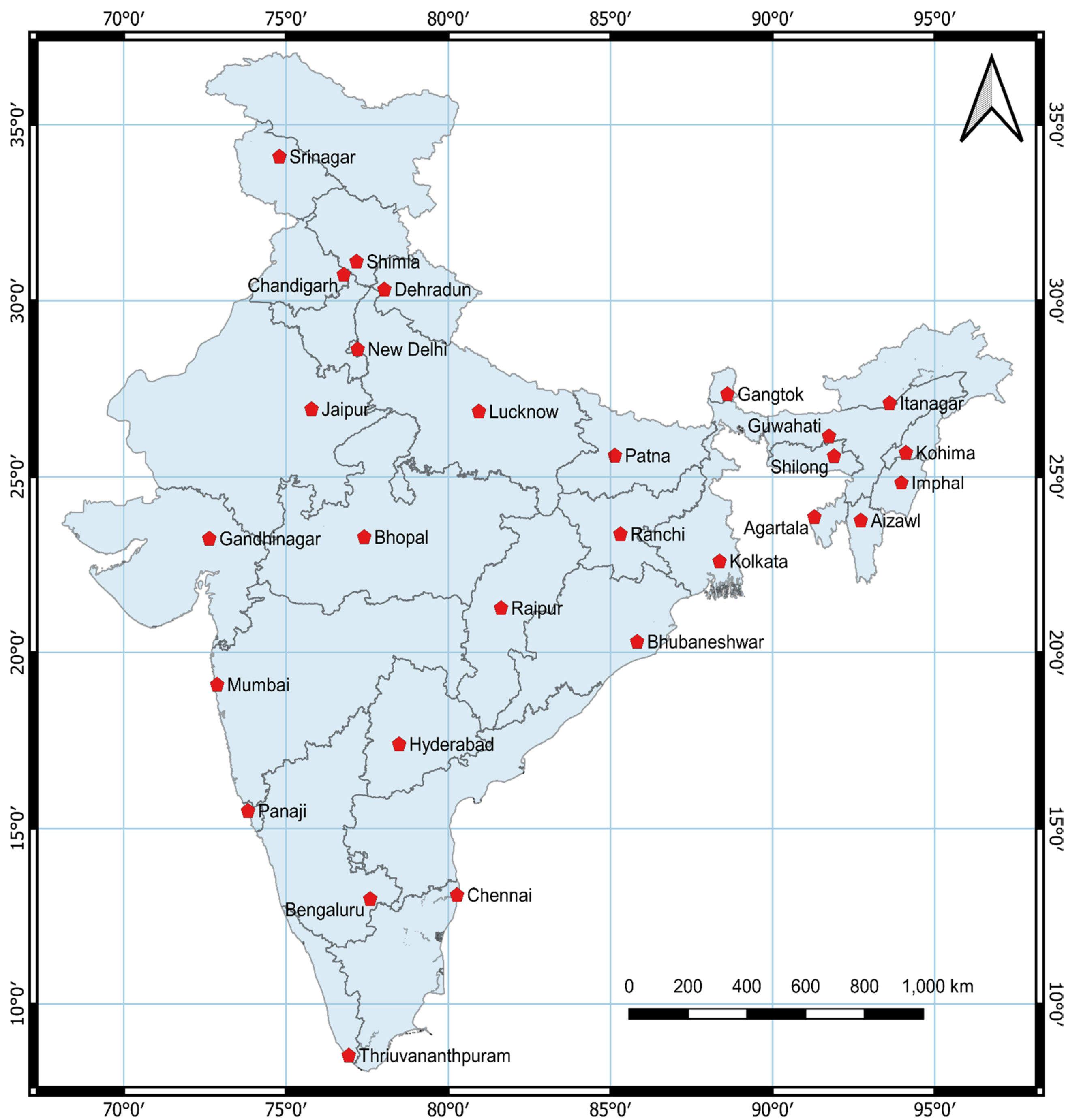

Fig. 1 Study area map showing the location of twenty-eight major cities located in different states of India

(spyder) notebook. The values of meteorological precursors, at a specific city, are computed through inverse distance weighting (IDW) method from its nearest neighbouring four grid intersections. The values obtained from ERA5 are converted to daily scale, since they are obtained at hourly scale, and are then normalized between 0 and 1 to avoid the problem of scaling between different input features. Next, the dataset is split into k ($k = 5$) parts (folds), i.e., each fold contains an approximate of 20% of the total dataset, in order to perform fivefold cross validation (CV) of the model. Thereafter, the model is repeatedly train using the aforementioned seven meteorological precursors and the month index on (k-1) fold (i.e., 80% of the data) and is tested on the kth fold (i.e., 20% of the data) to predict the daily maximum temperature with a lead time of 1-day to 10-day.

Table 1 Details of the hydrometeorological dataset used in this study

Dataset (1979–2020)	Variables	Spatial resolution	Vertical/pressure level	Units
ERA5 hourly data on single levels	Mean sea level pressure	0.25°×0.25°	Surface	Pa
	Long wave radiation flux	0.25°×0.25°	Surface	W/m.^{2}
	Solar radiation	0.25°×0.25°	Surface	J/m.^{2}
	10-m u-wind (zonal)	0.25°×0.25°	10-m above surface	m/s
	10-m v-wind (meridional)	0.25°×0.25°	10-m above surface	m/s
ERA5 hourly data on pressure levels	Relative humidity	0.25°×0.25°	1000 hPa	%
IMD gridded data	Rainfall	0.25°×0.25°	Surface	mm
	Maximum air temperature	0.1°×0.1°	Surface	°C

3.2 Proposed hybrid Conv1D-LSTM model

The proposed DL model is a hybrid composition of one-dimensional convolutional neural network (CNN), henceforth Conv1D, and long short-term memory (LSTM) neural network architecture. It is of a sequential type as shown in Fig. 2, and it is developed using a Keras library, built on the top of tensorflow used for large scale DL algorithms, in the spyder notebook.

The Conv1D comprises of three layers, i.e., an input layer, hidden layers, and an output layer. The input layer is the first layer of the model which feeds the input to the model and is of Conv1D type. It comprises of filters, kernel size, activation function, kernel initializer, and input

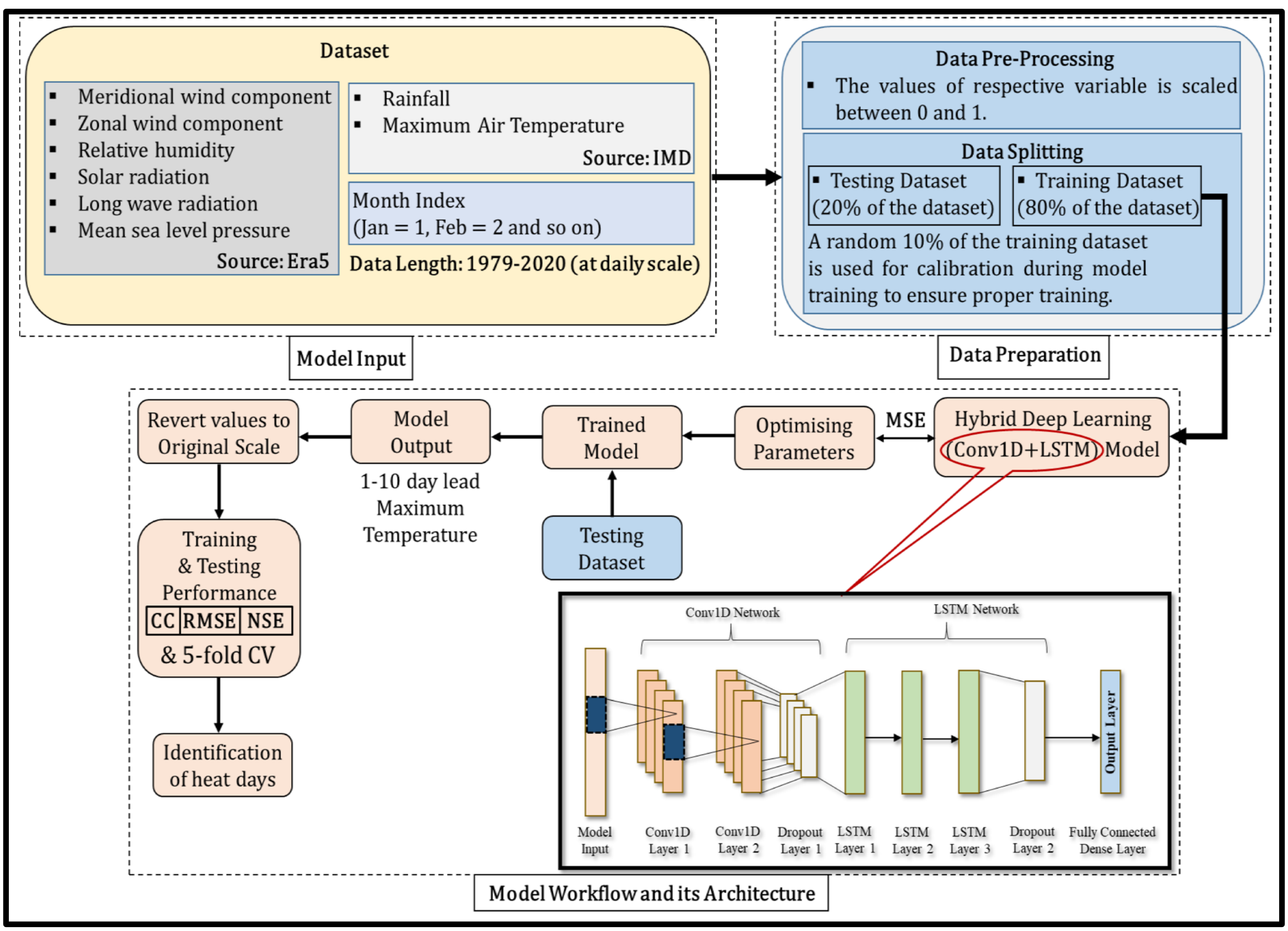

Fig. 2 Schematic representation of the proposed hybrid Conv1D-LSTM model architecture and its workflow

shape. Next, the hidden layers are added that vary in types as well as in numbers. Finally, the output layer is added.

Each Conv1D layer is the building component of model. These are also known as the computational engine of the model. The configurations (except input shape argument) and functions of all Conv1D layers are the same as that of the input layer. It consists of 1-D filter to extract the complex features from the input dataset. Though it is 1-D by name, the width of the filter by default captures the entire width of input shape at a time step and height can vary according to the provided input shape.

After the Conv1D layers, a max-pooling/dropout layer (if used) is used after the Conv1D layer. Max-pooling layer reduces the dimensionality and thereby avoids the complexity of the model output. Dropout layer helps in improving the overfitting/underfitting of the model on the testing dataset by assigning zero weights to the less contributing neurons (Srivastava et al. 2014). A more detailed background about Conv1D can be found in Kiranyaz et al. (2019) and Khan and Maity (2020).

In the proposed hybrid model architecture, a network of LSTM layers is added after the Conv1D layers. The layers of LSTM have memory units that are responsible for remembering the information from the inputs passing through the layer and deciding which information is to be memorized and kept, and which is to be forgotten. The first layer of LSTM receives the inputs from the last layer of Conv1D. Return sequence, kernel initializers, and activation functions are some of the hyper-parameters in the LSTM layers. The return sequence has Boolean value, i.e., either true or false. True value keeps the same dimension of the input data of the proceeding layer and the false value changes the dimension to 1D form to move the output to the fully connected dense/dropout layer. The LSTM network may or may not have dropout layer(s), depending on the need of the model. A more detailed background about the LSTM can be found in Hochreiter and Schmidhuber (1997). After the Conv1D and LSTM networks, a output layer is added to the model. It is a type of fully connected dense layer responsible for providing the outputs.

The aforesaid several parameters/hyper-parameters of the hybrid model need to be configured properly depending on the problem at hand. The details are provided in the next section (Section 3.3). Once the layers are configured, the input features are mapped to the target feature to learn their associations. This process involves adjustments of hyper-parameters by observing the loss function. The loss functions are defined to measure the error between observed and the modeled value of training and validation data at each time step. Once the model validation is completed, the model is ready for further application to new data (testing data).

3.3 Model parameters/configurations

In order to obtain a reliable model, several combinations of model hyper-parameters are to be examined, such as number of filters, kernel size, hidden layers, dropout rate, and LSTM units. Similarly, several hyper-parameters, such as learning rate, decay rate, momentum rate, number of epochs, batch size, kernel initializer, loss function, and activation function, are also optimized to ascertain the stability of the neural network across all the cities. Once finalized, the model configuration is kept unchanged across all the cities.

Memory blocks of LSTM layers are more wiser than the classic neurons as it memorizes the sequence of the time series, also known as serial dependency (Hochreiter and Schmidhuber 1997). It uses the sigmoid activation function and a pointwise multiplication operation for flow and change of information within the cell state. Memories perform the mapping of an input sequence to a target/output sequence using the Eqs. 1, 2, 3, 4, 5, and 6 with the help of three gates. Three different types of gates within a memory unit are as follows: (a) forget gate: conditionally take control of the unwanted information, (b) input gate: conditionally selects information from the input sequence to update the memory/cell state, and (c) output gate: responsible for giving the output after analyzing the conditions of the input and the memory cells.

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1},\ X_t\right] +\ b_i\right) \tag{1}$$

$$\tilde{C}_t = tanh\left(W_c\left[h_{t-1}, X_t\right]\ +\ b_c\right) \tag{2}$$

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1},\ X_t\right] +\ b_i\right) \tag{3}$$

$$C_t = i_t * \tilde{C}_t + f_t * C_{t-1} \tag{4}$$

$$o_t = \sigma\left(W_0\left[h_{t-1},\ X_t\right]\ +\ b_0\right) \tag{5}$$

$$h_t = o_t *\ tanh\left(C_t\right) \tag{6}$$

In these equations, i_t, f_t *and* o_t are the outputs of three sigmoid functions and their values range between 0 and 1. They control the stored information in the new cell state ($\tilde{C}_t$), forgotten information in old cell state (C_{t-1}), and the output information to the cell (h_t), respectively. X_t is the input given to the memory block at time instant t and h_{t-1} is the output of the previous cell, and $W_{i,}$ W_f, W_c, W_0, b_i, b_f, b_c, b_0 are their corresponding weights and biases.

3.4 Comparison with other models

The effectiveness of hybrid Conv1D-LSTM model is validated by comparing it with four other popular approaches widely used for the prediction of hydroclimatic variables. These are long short-term memory (LSTM) neural network, 1-dimensional convolutional neural network (Conv1D), multilayer perceptron (MLP), and support vector regression (SVR). These models are briefly described here. The designed structure of LSTM, Conv1D, and MLP, a DL-based algorithm, also contains three categories of layer, i.e., an input layer for receiving the input data, hidden layers for computation, and an output layer for receiving the predicted values.

The models are developed and configured in the same spyder environment using the Keras library with same proportion of training, validation, and testing dataset as that of proposed hybrid model. After successful configuration of all the layers of these models, the input data sets are mapped to the output data to learn the hidden associations in them (Livingstone 2008). The models are used for making predictions once training and validation are completed. More details about the working of these approaches can be found elsewhere (Haidar and Verma 2018; Kashid and Maity 2012; Khan and Maity 2020; Kiranyaz et al. 2019; Kratzert et al. 2019a; Maity et al. 2021).

The fourth model used for comparison is the ML-based SVR model, which is a regression form of support vectors in support vector machine (SVM) (Drucker et al. 1996). The modelling of SVR is performed by using the scikit-learn library, available in python. The optimization of SVR model is carried out by using cost function (C) and regularization parameter (γ) of the radial basis function (RBF) (Choy and Chan 2003). The details about its working principle may be found in the existing literature (Bhagwat and Maity 2014 and Drucker et al. 1996).

The prediction of maximum temperature by the proposed hybrid model is compared with the aforementioned models using three performance metrics, namely coefficient of correlation (CC), root mean square error (RMSE), and Nash–Sutcliffe efficiency (NSE). Apart from these, the prediction skill is also investigated in terms of identifying upcoming heatwaves, which is discussed later.

4 Results and discussion

4.1 Model configuration and calibration

In the proposed model, the finalized architecture comprises of eight layers (Fig. 2), whose configurations are discussed as follows. The prepared set of seven causal variables and month index are fed as input to the first layer. The input shape is arranged in a three-dimensional tensor form. It contains values from 4 previous time steps (days) of each causal variable and the month index. The first layer of the model, i.e., Conv1D layer, helps to identify the pattern and extract the hidden information in the input sequence. It comprises of 224 filters, kernel size/stride of 1, Glorot uniform kernel initializer (aka Xavier uniform initializer), input shape, and rectified linear unit (ReLU) activation function. Kernel/stride size signifies the height of the filter, and activation function is responsible for neuron's output. A threshold value is set by the activation function on the basis of the input and output data and is to be achieved by the neuron before it moves to the subsequent layer. The second layer added to the hybrid model is also a Conv1D layer, comprising of 192 filters and same kernel/stride size, kernel initializer, and activation function as that of the first Conv1D layer. After providing two consecutive Conv1D layers, a dropout layer (dropout rate = 0.30) is added as the third layer. Dropout layer reduces the chances of overfitting and complexity in the model as it ignores the weight of each dropout neuron during the backward pass of training (Srivastava et al. 2014). Next, three consecutive layers LSTM are added as a fourth, fifth, and sixth layer in the model architecture having same kernel initializer and activation function as that of Conv1D layers and 64 number of memory cells.

Next, a dropout layer having 10% of dropout rate (i.e., 0.10) is added in the LSTM network as a seventh layer. It is followed by a fully connected dense layer, i.e., output layer (eighth layer), consisting of ten neurons, which is added to the model. These ten neurons provide the 1-day to 10-day lead forecast of maximum temperature.

Having configured the different layers of the network architecture, the model is compiled with a batch size of 60 and 550 epochs. The mean absolute error (MAE) is defined as a loss function, and Adam optimizer function (learning rate = 0.0001, momentum rate 0.9, and decay rate 1×10^{-7}) of stochastic gradient descent is adopted for training and validation of the proposed hybrid model. A proper observation of MAE was done during the training and validation period to avoid chances (if any) of overfitting/underfitting of the hybrid Conv1D-LSTM model. After successful training of the model, the testing dataset is used to assess the model performance. The aforesaid process is repeated for each fold, i.e., the model is trained on $(k-1)$ folds and is tested on kth fold; hence, the robustness of the model is checked.

4.2 Prediction of multi-step ahead (1-day to 10-day) daily maximum temperature

The proposed DL-based hybrid Conv1D-LSTM model is applied to twenty-eight selected cities across India to predict the multi-step (1-day to 10-day) ahead daily maximum temperature. The model is trained using hydrometeorological precursors as inputs and daily maximum temperature, for the next 10 days, as output. Table 2 shows the average performance obtained across five folds for 1-day ahead prediction through *CC*, *RMSE*, and *NSE*, obtained at all twenty-eight cities. It also presents the values of these performance metrics for other four models used for comparison, i.e., LSTM, Conv1D, MLP, and SVR model. It is observed that the SVR model is the least performing model showing the highest range of *RMSE* values (0.74 to 1.64 °C), lowest range of *CC* (0.90 to 0.98) and *NSE* (0.80 to 0.95), respectively, during the training phase. Moreover, its performance on testing dataset is also poorer than the other models. The values of the performance metrics during testing period are noticed to be in the range of 0.75 to 1.70 °C (*RMSE*), 0.89 to 0.97 (*CC*), and 0.79 to 0.95 (*NSE*).

Next, the performances of MLP and Conv1D are compared. It is noticed that the performance of Conv1D model varies marginally to reasonably better than MLP across all twenty-eight cities. An investigation of quality of the performance by both the models leads to a summary as follows: the *RMSE* — 0.61 to 1.44 °C by Conv1D and 0.63 to 1.74 °C by MLP, *CC* — 0.93 to 0.99 by Conv1D and 0.93 to 0.98 by MLP, and *NSE* — 0.83 to 0.96 by Conv1D and 0.83 to 0.95 by MLP during the training period. During the testing period, *RMSE* ranges from 0.68 to 1.58 °C, *CC* ranges from 0.91 to 0.98, and NSE ranges from 0.83 to 0.96 in case of Conv1D, whereas these ranges (in same order) are 0.68 to 1.85 °C, 0.91 to 0.98, and 0.65 to 0.93, respectively, in case of MLP. Hence, from the aforementioned statistics, a lower RMSE value of the Conv1D model establishes the better performance of the model across all the cities.

Having established the better performance of the Conv1D model over MLP and SVR, its performance is compared with LSTM. It is observed that the performance of the Conv1D model is better than LSTM model in most of the cities except Bengaluru, Jaipur, New Delhi, Patna, and Ranchi. Although it is difficult to segregate the model performance of the cities based on local climatological conditions and geographical locations, an investigation of the performances metrics portrays the superiority of LSTM over Conv1D and vice versa at selected cities. The range of values of three performance metrics obtained during the evaluation of LSTM model is as follows: *RMSE* ranges from 0.63 to 1.39 °C, *CC* ranges from 0.90 to 0.98, and *NSE* 0.85 to 0.97 during the training period. Likewise, *RMSE* ranging from 0.73 to 1.52 °C, *CC* ranging from 0.90 to 0.98, and *NSE* ranging from 0.80 to 0.96 are obtained during the testing period.

In contrast to the performance of the aforementioned four comparative models, the performance of the proposed hybrid Conv1D-LSTM model is observed to be the best in terms of its accuracy in capturing the magnitude of daily maximum temperature at all the cities. The comparison indicates that the improvement achieved by the hybrid Conv1D-LSTM model varies from "marginally" to "reasonably" across different cities. It is observed from Table 2 that *CC* values range from 0.92 to 0.98, NSE values range from 0.84 to 0.97, and RMSE values range from 0.65 to 1.40 °C during the training period, whereas these values (in the same order) range from 0.91 to 0.98, 0.83 to 0.96, and 0.68 to 1.49 °C with the testing dataset. Evidently, the performance of the hybrid Conv1D-LSTM is the best in all the cities.

For a visual impression of the performance, scatter plots between the observed and the modelled values are plotted for each pair of training and testing folds for all twenty-eight cities. However, for the brevity of presentation, a traditionally hot weather city (Jaipur) is considered for illustration. The model performance is shown in Fig. 3 that comprises of 3 sub parts. The part (i) shows the scatter plots between observed and predicted values of daily maximum temperature (1-day ahead) by all five models during the training period. The training period shown in the figure comprises of dataset from fold1 to fold4. Next, part (ii) shows a scatter plot, similar to the part (i) but for the testing period, i.e., performance of the model on unseen dataset, fold1. In addition to the above two scatter plots, another visualization, summarizing the performance of the testing dataset across all folds, is shown in Fig. 3 part (iii). It shows a scatter plot between the observed and the predicted maximum temperature (1-day lead) obtained by combining the performance of all the testing folds (fold 1 to fold 5). Apart from the general correspondence between observed and predicted values, it is further noticed that the range of daily maximum temperature extremes is better captured by the proposed hybrid model, whereas LSTM, Conv1D, SVR, and MLP models are not as efficient as the proposed one in capturing the same. This observation is more or less true for other cities also and it motivates to an investigation on the performance towards the assessment of the models in foreseeing the heatwaves.

However, so far, the discussion pertains to the 1-day ahead performance. The benefit of the proposed hybrid Conv1D-LSTM model is better realized in case of longer lead times (2- to 10- day ahead predictions. Thus, the results for all the lead times (1-day to 10-day ahead) prediction are investigated using the same performance metrics. The performance metrics values during training and testing period for all 10 lead times at all the selected cities are computed. As a typical case, a summary of the performance in terms of the statistical metrics viz. CC, RMSE, and NSE is graphically presented in case of the Jaipur city for all the lead times and for all the models (Fig. 4). The error bar in the figure indicates the variation in the performance of the models across different folds. In general, it is noticed that the model performances are gradually decreasing with the increase in the lead time of prediction, i.e., from 1-day to 10-day. However, a faster decrease is noticed in case of LSTM, Conv1D, MLP, and SVR as compared to proposed

Table 2 Performance statistics obtained at 1-day lead in predicting maximum temperature by the proposed model along with the other models used for comparison, during training (Tr) and testing (Ts) period. Each cell shows the averaged value (fold1 to fold 5) of CC, RMSE, and NSE from top to bottom

City	hybrid Conv1D-LSTM		LSTM		Conv1D		MLP		SVR	
	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts
Agartala	0.96	0.95	0.96	0.95	0.96	0.95	0.96	0.95	0.94	0.94
	0.91	0.97	0.90	1.00	0.90	0.98	0.93	1.01	1.09	1.10
	0.92	0.91	0.92	0.90	0.92	0.90	0.91	0.90	0.88	0.88
Aizawl	0.96	0.95	0.96	0.95	0.96	0.95	0.96	0.95	0.94	0.94
	0.97	1.03	0.98	1.08	0.96	1.03	0.97	1.07	1.15	1.16
	0.91	0.90	0.91	0.89	0.92	0.90	0.91	0.89	0.88	0.88
Bengaluru	0.96	0.96	0.96	0.95	0.96	0.96	0.97	0.96	0.95	0.95
	0.78	0.85	0.79	0.89	0.76	0.84	0.77	0.86	0.90	0.92
	0.92	0.91	0.92	0.90	0.93	0.91	0.93	0.91	0.90	0.90
Bhopal	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.97
	1.05	1.14	1.07	1.20	1.07	1.20	1.14	1.26	1.32	1.36
	0.96	0.96	0.96	0.95	0.96	0.95	0.96	0.95	0.94	0.94
Bhubaneswar	0.96	0.95	0.96	0.94	0.96	0.95	0.96	0.95	0.93	0.93
	0.95	1.01	0.98	1.08	0.94	1.02	0.91	1.01	1.17	1.19
	0.91	0.90	0.91	0.89	0.92	0.90	0.92	0.90	0.87	0.86
Chandigarh	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.97
	1.24	1.34	1.27	1.42	1.27	1.41	1.42	1.54	1.52	1.56
	0.96	0.96	0.96	0.95	0.96	0.95	0.95	0.94	0.95	0.94
Chennai	0.96	0.96	0.97	0.95	0.96	0.96	0.97	0.96	0.94	0.94
	0.90	0.96	0.90	1.02	0.90	0.98	0.89	0.98	1.11	1.13
	0.93	0.92	0.93	0.91	0.93	0.91	0.93	0.91	0.89	0.89
Dehradun	0.98	0.97	0.98	0.97	0.98	0.97	0.98	0.97	0.97	0.97
	1.23	1.31	1.23	1.35	1.22	1.34	1.38	1.50	1.46	1.49
	0.95	0.95	0.95	0.94	0.95	0.94	0.94	0.93	0.93	0.93
Gandhinagar	0.98	0.97	0.98	0.97	0.98	0.97	0.98	0.97	0.96	0.96
	0.93	1.02	0.95	1.07	0.94	1.05	0.97	1.09	1.19	1.22
	0.96	0.95	0.95	0.94	0.96	0.94	0.95	0.94	0.93	0.92
Gangtok	0.96	0.95	0.96	0.95	0.96	0.95	0.96	0.95	0.94	0.94
	1.08	1.13	1.04	1.19	1.05	1.14	1.06	1.17	1.23	1.24
	0.91	0.90	0.92	0.89	0.92	0.90	0.91	0.90	0.88	0.88
Guwahati	0.96	0.96	0.96	0.95	0.96	0.95	0.97	0.96	0.95	0.94
	0.99	1.04	0.98	1.09	0.99	1.07	1.00	1.09	1.18	1.19
	0.92	0.91	0.93	0.91	0.92	0.91	0.92	0.91	0.89	0.89
Hyderabad	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.97	0.96	0.96
	0.95	1.01	0.96	1.09	0.95	1.04	0.99	1.08	1.17	1.19
	0.95	0.94	0.94	0.93	0.95	0.93	0.94	0.93	0.92	0.91
Imphal	0.94	0.94	0.95	0.93	0.94	0.94	0.95	0.93	0.92	0.92
	1.12	1.19	1.11	1.26	1.10	1.18	1.08	1.20	1.31	1.32
	0.89	0.87	0.89	0.85	0.89	0.87	0.89	0.87	0.85	0.84
Itanagar	0.95	0.94	0.95	0.94	0.95	0.94	0.96	0.94	0.93	0.93
	1.14	1.23	1.13	1.27	1.13	1.23	1.14	1.26	1.33	1.34
	0.90	0.88	0.91	0.88	0.91	0.88	0.90	0.88	0.87	0.86
Jaipur	0.98	0.98	0.99	0.98	0.98	0.98	0.99	0.98	0.97	0.97
	1.07	1.18	1.08	1.24	1.12	1.26	1.22	1.35	1.39	1.43
	0.97	0.96	0.97	0.96	0.97	0.96	0.96	0.95	0.95	0.95
Kohima	0.94	0.94	0.95	0.93	0.95	0.94	0.95	0.94	0.92	0.92
	1.16	1.22	1.16	1.29	1.14	1.23	1.15	1.26	1.35	1.37
	0.89	0.88	0.89	0.86	0.89	0.87	0.89	0.87	0.85	0.84

Table 2 (continued)

City	hybrid Conv1D-LSTM		LSTM		Conv1D		MLP		SVR	
	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts
Kolkata	0.96	0.95	0.96	0.95	0.96	0.95	0.96	0.95	0.94	0.94
	0.96	1.02	0.98	1.10	0.95	1.04	0.94	1.05	1.19	1.21
	0.92	0.91	0.92	0.89	0.92	0.90	0.92	0.90	0.88	0.87
Lucknow	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.97
	1.19	1.27	1.23	1.38	1.23	1.37	1.36	1.50	1.45	1.49
	0.96	0.96	0.96	0.95	0.96	0.95	0.95	0.94	0.94	0.94
Mumbai	0.96	0.96	0.96	0.95	0.96	0.96	0.97	0.96	0.95	0.94
	0.77	0.81	0.78	0.85	0.73	0.80	0.74	0.81	0.90	0.91
	0.92	0.91	0.92	0.90	0.93	0.91	0.93	0.91	0.89	0.89
New Delhi	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.98	0.98	0.97
	1.21	1.32	1.22	1.37	1.29	1.43	1.46	1.58	1.51	1.55
	0.97	0.96	0.97	0.96	0.96	0.96	0.95	0.95	0.95	0.95
Panaji	0.96	0.95	0.96	0.94	0.96	0.95	0.96	0.95	0.93	0.93
	0.67	0.72	0.67	0.76	0.67	0.72	0.65	0.71	0.81	0.81
	0.91	0.90	0.91	0.89	0.91	0.90	0.92	0.90	0.87	0.87
Patna	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.98	0.97	0.97
	1.06	1.13	1.05	1.17	1.09	1.20	1.20	1.31	1.28	1.31
	0.96	0.95	0.96	0.95	0.96	0.95	0.95	0.94	0.94	0.94
Raipur	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.98	0.97	0.97
	1.01	1.07	1.03	1.16	1.02	1.11	1.09	1.19	1.23	1.26
	0.96	0.95	0.96	0.95	0.96	0.95	0.95	0.94	0.94	0.94
Ranchi	0.98	0.97	0.98	0.97	0.98	0.97	0.98	0.97	0.96	0.96
	1.01	1.08	1.01	1.13	1.03	1.14	1.07	1.18	1.25	1.28
	0.95	0.95	0.95	0.94	0.95	0.94	0.95	0.94	0.93	0.93
Shillong	0.96	0.95	0.96	0.95	0.96	0.95	0.96	0.95	0.94	0.94
	0.99	1.06	0.97	1.11	0.98	1.06	0.97	1.06	1.19	1.20
	0.92	0.90	0.92	0.89	0.92	0.90	0.92	0.90	0.88	0.87
Shimla	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.97	0.97	0.97
	1.23	1.31	1.23	1.37	1.24	1.37	1.40	1.51	1.47	1.50
	0.96	0.95	0.96	0.95	0.96	0.95	0.95	0.94	0.94	0.94
Srinagar	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97
	1.40	1.49	1.39	1.52	1.44	1.58	1.74	1.85	1.64	1.70
	0.97	0.96	0.97	0.96	0.96	0.96	0.95	0.94	0.95	0.95
Thiruvananthapuram	0.92	0.91	0.93	0.90	0.93	0.91	0.93	0.91	0.90	0.89
	0.65	0.68	0.63	0.73	0.63	0.68	0.61	0.68	0.74	0.75
	0.84	0.83	0.85	0.80	0.86	0.83	0.86	0.83	0.80	0.79

hybrid Conv1D-LSTM model. Thus, the benefit of proposed hybrid Conv1D-LSTM model is established for longer lead times. Apart from the aforesaid illustration of model performance at all 10-day lead of a particular city, the average performance metrics (averaged across fold1 to fold5) of all twenty-eight cities are presented in Table 3. The metrics values portrayed in the table shows the dominance of the hybrid model even at longer lead time. Hence, better performance of hybrid Conv1D-LSTM model is established for all the cities, at all lead times.

4.3 Prediction of heatwave events

How good is the potential of the proposed hybrid Conv1D-LSTM model in foreseeing the heatwaves in the coming days? To investigate this, the prediction skill of the proposed model is analyzed to identify the heat days of a heatwave event occurring during the year 2012–2020 of the dataset. Before proceeding, a brief discussion on the heatwaves is presented. Heatwaves are commonly defined as unusual extreme temperatures prevailing over for days in a region

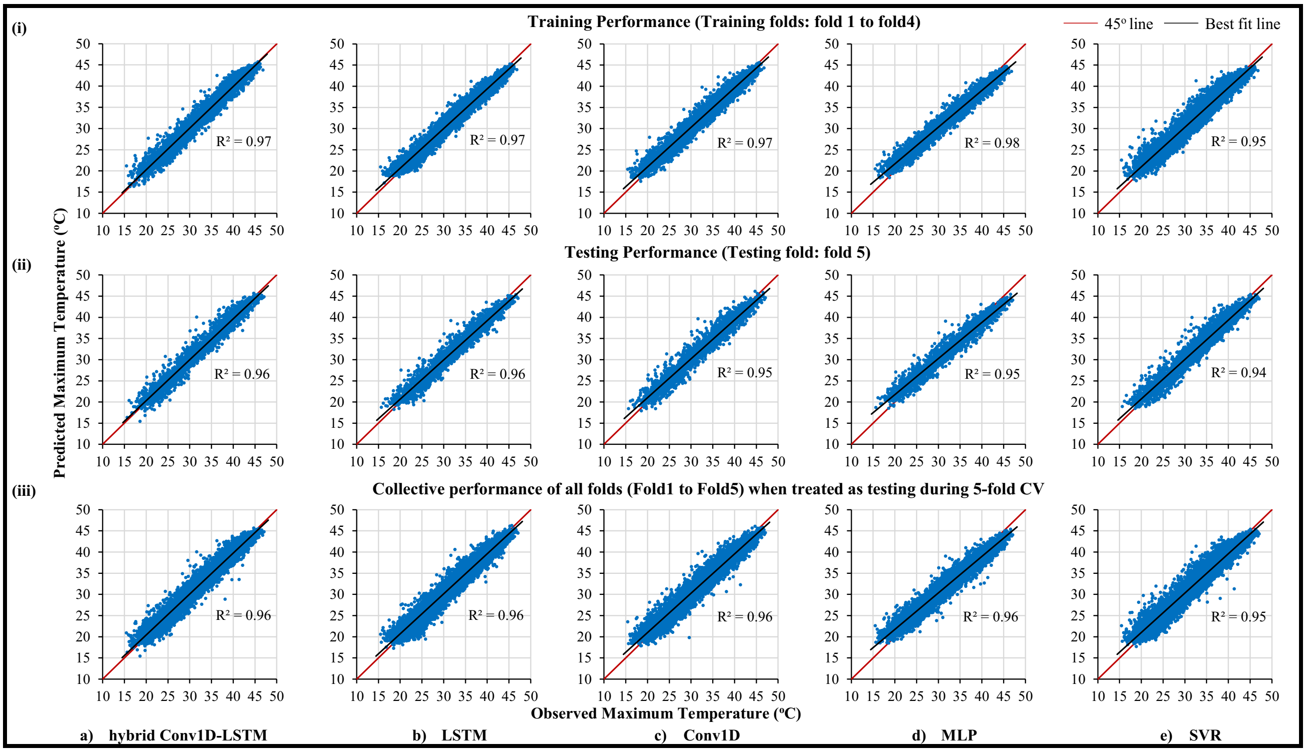

Fig. 3 Comparative scatter plots between the observed and 1-day ahead predicted maximum temperature obtained during the (i) training period (i.e., by considering fold1 to fold 4 as training dataset), (ii) testing period (i.e., by considering fold5 as testing dataset) and (iii) testing period of all 5 folds (i.e., fold1 + fold2 + fold3 + fold4 + fold 5, when each fold is treated as a testing dataset during fivefold CV), for a traditionally hot weather city (Jaipur), of (**a**) hybrid Conv1D-LSTM, (**b**) LSTM, (**c**) Conv1D, (**d**) MLP, and (**e**) SVR model run respectively

with serious consequences. The India Meteorological Department (IMD) uses the following criteria to define a heatwave: (i) on the basis of departure from normal temperature: If the departure of actual maximum temperature from the normal is 4.5 to 6.4 °C, it will be called as heatwave and when the departure is greater than 6.4 °C, it will be called as severe heatwave; (ii) on the basis of actual maximum temperature: If the actual maximum temperature is ≥ 45 °C, it is called as heatwave, and if the actual maximum temperature is ≥ 47 °C, it will be called as a severe heat wave; (iii) heatwave should be declared if actual maximum temperature in a region remains 45 °C or more, irrespective of the normal maximum temperature; (iv) heatwave should not to be considered, if the observed maximum temperature in plains and hilly regions are less than 40 °C and 30 °C respectively.

The criteria (i), (ii), and (iii) should be met at least at two meteorological station for at least 2 consecutive days, and on the second day, it is declared as a heatwave. An attempt to capture the heatwave events is carried out in the light of the aforementioned criteria. It is noticed that among the twenty-eight selected cities, twenty-four cities have faced several *heatwave* and *severe heatwave* events. The number of heat days trapped in the event of heat wave varies between 3 and 70 days across the affected cities. Table 4 shows the detailed figure of heat days occurred under several heat wave/severe heat wave events in different cities along with the predicted heat days at 1-day lead to 7-day lead, during the period 2012–2020. It may be noted that the estimation of heat days was carried out considering the aforesaid criteria. The total number of heat days occurred across the twenty-four cities as shown in the table counts to a total of 615 days, out of which the proposed model was able to capture 594 days at 1-day lead with an error of 5%, i.e., a total 31 false heat days were predicted. In other words, it can be said that accuracy of 92% was achieved by the proposed model in predicting heat days successfully at 1-day lead. Likewise, the performance of the proposed model was analyzed up to 7-day lead time in foreseeing the heat days. A gradual reduction in the accuracy was noticed as the lead time was increased. For instance, accuracy at 2-, 3-, 4-, 5-, 6-, and 7-day lead was observed to be 76% (counts to 469 heat days), 66% (counts to 407 heat days), 58% (counts to 358 heat days), 52% (counts to 318 heat days), 45% (counts to 281 heat days), and 39% (counts to 242 heat days), respectively. It may be noted that at longer leads, no false alarm was made by the model; however, performance of the model went below 50% accuracy after 5-day

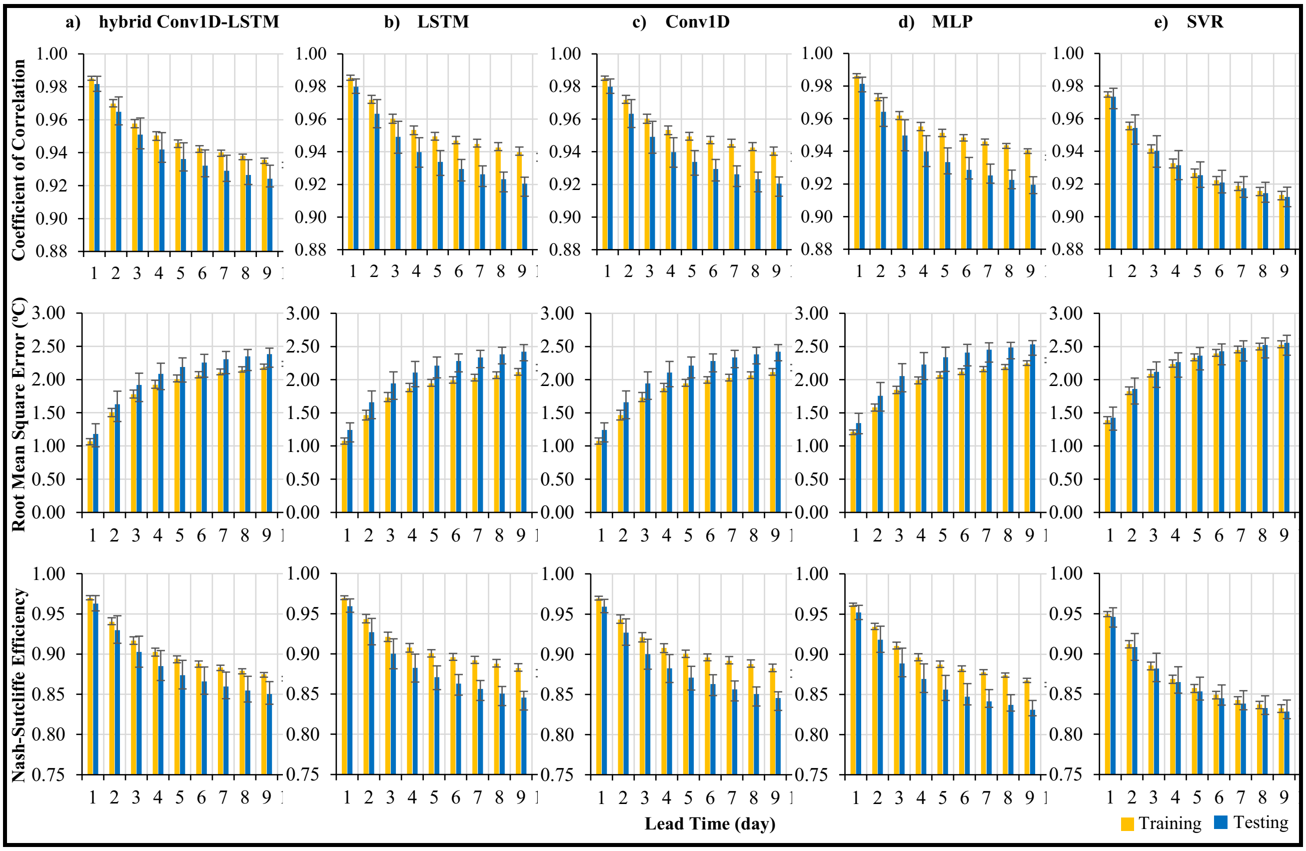

Fig. 4 Average values (fold1 to fold5) of the performance metrics obtained for multi-step-ahead (1-day to 10-day lead) daily maximum temperature prediction during training and testing period for Jaipur city using (**a**) hybrid Conv1D-LSTM, (**b**) LSTM, (**c**) Conv1D, (**d**) MLP and (**e**) SVR models. The error bar shows the range of metric values obtained across different folds (5 folds)

lead and even below 40% accuracy at 7-day lead. Hence, the prediction of foreseeing the heat wave event was stopped at 7-day lead.

Thus, overall, it can be said that the proposed model is able to capture the range of daily maximum temperature efficiently. However, the performance of the hybrid model also reduces with the increasing lead time.

5 Comparison between proposed DL-based model and a few existing weather applications

With the increase in advancement in the technology since twentieth century, the access to weather forecast has been drastically increased. Weather forecasts are available from different weather applications (weather apps) through the smartphones. For example, Zabini (2016) reported the functioning of 39 popular smartphone-based weather apps to communicate weather forecasts to the general public in the USA, the UK, and Italy. The study concluded that advances in mobile communication technologies could theoretically improve weather communication effectiveness. Moreover, the expectations that have been built up around weather forecasts appear to be vastly out of step with existing forecasting capabilities, especially given the inherent uncertainties in location and time, as well as the nature of the forecasted weather occurrences. It may be further noted that past values of forecasts, background methodologies/models of these weather applications (weather apps) are not openly accessible. Therefore, authors were not able to directly compare the performance with weather apps. However, an attempt is made for the comparison with the help of the information available in the existing literature on weather apps performances.

Thomas et al. (2016) has tried to assess the accuracy of weather forecast, available to the public in India by three popular weather apps, namely, AccuWeather (AccuWth), real-time weather system (RTWS), and weather underground (WUnd). The evaluation comprised of the assessment of forecasting skills of aforesaid apps in measuring maximum temperature, minimum temperature, pressure, wind speed, wind direction, and rainfall with those of observed record of IMD, made through synoptic and

Table 3 Same as Table 2 but metrics obtained during prediction of maximum temperature at lead time of 10 day

City	hybrid Conv1D-LSTM		LSTM		Conv1D		MLP		SVR	
	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts
Agartala	0.88	0.86	0.89	0.85	0.89	0.85	0.89	0.85	0.84	0.84
	1.52	1.64	1.44	1.69	1.47	1.69	1.51	1.70	1.72	1.74
	0.77	0.73	0.79	0.71	0.79	0.71	0.77	0.71	0.71	0.70
Aizawl	0.88	0.86	0.89	0.85	0.89	0.86	0.89	0.86	0.85	0.85
	1.57	1.70	1.50	1.75	1.52	1.74	1.55	1.75	1.77	1.79
	0.78	0.73	0.80	0.72	0.79	0.72	0.78	0.72	0.71	0.70
Bengaluru	0.89	0.87	0.90	0.86	0.90	0.86	0.9	0.86	0.84	0.84
	1.33	1.44	1.24	1.49	1.26	1.45	1.28	1.47	1.56	1.57
	0.78	0.74	0.81	0.73	0.81	0.74	0.80	0.73	0.70	0.70
Bhopal	0.92	0.91	0.93	0.90	0.93	0.90	0.93	0.90	0.89	0.89
	2.12	2.32	2.04	2.37	2.05	2.36	2.14	2.43	2.49	2.52
	0.85	0.82	0.86	0.81	0.86	0.81	0.85	0.80	0.79	0.78
Bhubaneswar	0.86	0.84	0.87	0.83	0.87	0.83	0.88	0.84	0.81	0.81
	1.63	1.73	1.57	1.79	1.58	1.78	1.57	1.78	1.87	1.89
	0.74	0.71	0.76	0.69	0.76	0.69	0.76	0.69	0.66	0.65
Chandigarh	0.94	0.93	0.94	0.93	0.95	0.93	0.95	0.93	0.92	0.92
	2.22	2.41	2.18	2.46	2.16	2.45	2.30	2.56	2.57	2.60
	0.89	0.86	0.89	0.86	0.89	0.86	0.88	0.85	0.85	0.84
Chennai	0.90	0.88	0.91	0.87	0.91	0.88	0.91	0.87	0.86	0.86
	1.49	1.60	1.39	1.66	1.43	1.62	1.46	1.66	1.73	1.75
	0.80	0.77	0.83	0.75	0.82	0.77	0.81	0.76	0.73	0.73
Dehradun	0.93	0.92	0.93	0.91	0.93	0.91	0.93	0.91	0.91	0.91
	2.12	2.30	2.09	2.35	2.06	2.35	2.19	2.43	2.44	2.46
	0.86	0.84	0.87	0.83	0.87	0.83	0.85	0.82	0.82	0.82
Gandhinagar	0.91	0.89	0.91	0.88	0.91	0.89	0.91	0.88	0.87	0.87
	1.90	2.06	1.85	2.10	1.83	2.08	1.88	2.12	2.19	2.21
	0.82	0.78	0.83	0.78	0.83	0.78	0.82	0.77	0.76	0.75
Gangtok	0.88	0.86	0.90	0.85	0.89	0.86	0.89	0.86	0.85	0.85
	1.71	1.84	1.61	1.92	1.65	1.88	1.68	1.90	1.94	1.96
	0.78	0.74	0.80	0.72	0.79	0.73	0.79	0.72	0.71	0.71
Guwahati	0.88	0.86	0.89	0.85	0.89	0.85	0.89	0.85	0.84	0.85
	1.71	1.84	1.62	1.92	1.67	1.90	1.71	1.91	1.92	1.94
	0.77	0.73	0.79	0.71	0.78	0.72	0.77	0.71	0.71	0.70
Hyderabad	0.90	0.89	0.91	0.88	0.91	0.88	0.91	0.88	0.85	0.85
	1.75	1.88	1.65	1.94	1.66	1.93	1.72	1.95	2.18	2.19
	0.81	0.78	0.83	0.77	0.83	0.77	0.82	0.77	0.71	0.71
Imphal	0.85	0.82	0.87	0.80	0.86	0.82	0.86	0.82	0.80	0.80
	1.77	1.92	1.67	2.02	1.72	1.96	1.74	1.96	1.99	2.01
	0.72	0.66	0.75	0.63	0.73	0.65	0.73	0.65	0.65	0.63
Itanagar	0.86	0.83	0.87	0.82	0.87	0.83	0.87	0.83	0.83	0.83
	1.89	2.04	1.81	2.12	1.84	2.09	1.88	2.11	2.07	2.09
	0.73	0.68	0.75	0.66	0.75	0.67	0.74	0.66	0.68	0.67
Jaipur	0.93	0.92	0.94	0.92	0.94	0.92	0.94	0.92	0.91	0.91
	2.22	2.41	2.18	2.45	2.18	2.46	2.29	2.55	2.56	2.59
	0.87	0.85	0.88	0.84	0.88	0.84	0.86	0.83	0.83	0.82
Kohima	0.85	0.83	0.86	0.81	0.86	0.82	0.86	0.82	0.81	0.81
	1.85	2.01	1.77	2.08	1.80	2.04	1.82	2.06	2.05	2.07
	0.72	0.67	0.75	0.64	0.74	0.65	0.73	0.65	0.66	0.64

Table 3 (continued)

City	hybrid Conv1D-LSTM		LSTM		Conv1D		MLP		SVR	
	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts	Tr	Ts
Kolkata	0.87	0.86	0.88	0.85	0.89	0.85	0.89	0.85	0.83	0.83
	1.65	1.74	1.60	1.79	1.58	1.78	1.60	1.80	1.88	1.90
	0.76	0.73	0.78	0.72	0.78	0.72	0.78	0.72	0.69	0.68
Lucknow	0.93	0.92	0.94	0.92	0.94	0.92	0.94	0.92	0.91	0.91
	2.17	2.32	2.14	2.40	2.12	2.40	2.24	2.49	2.54	2.57
	0.87	0.85	0.88	0.84	0.88	0.84	0.86	0.83	0.82	0.82
Mumbai	0.87	0.84	0.88	0.83	0.88	0.84	0.88	0.83	0.80	0.80
	1.35	1.47	1.31	1.52	1.30	1.51	1.31	1.52	1.66	1.66
	0.75	0.71	0.77	0.69	0.77	0.69	0.77	0.69	0.63	0.63
New Delhi	0.94	0.93	0.95	0.93	0.95	0.93	0.95	0.93	0.92	0.92
	2.27	2.43	2.22	2.48	2.21	2.51	2.40	2.64	2.63	2.67
	0.89	0.87	0.89	0.87	0.89	0.86	0.88	0.85	0.85	0.85
Panaji	0.87	0.83	0.88	0.82	0.88	0.82	0.88	0.82	0.75	0.75
	1.12	1.27	1.06	1.32	1.10	1.29	1.10	1.31	1.49	1.50
	0.75	0.68	0.78	0.65	0.76	0.67	0.76	0.66	0.56	0.55
Patna	0.93	0.92	0.93	0.91	0.93	0.91	0.93	0.91	0.90	0.90
	1.93	2.06	1.87	2.13	1.88	2.13	1.97	2.20	2.29	2.31
	0.86	0.84	0.87	0.83	0.87	0.83	0.85	0.82	0.80	0.80
Raipur	0.92	0.91	0.93	0.90	0.93	0.90	0.93	0.90	0.88	0.88
	1.97	2.12	1.88	2.18	1.88	2.16	1.94	2.21	2.38	2.40
	0.84	0.82	0.86	0.81	0.86	0.81	0.85	0.80	0.77	0.77
Ranchi	0.91	0.90	0.92	0.89	0.92	0.89	0.92	0.89	0.87	0.87
	1.95	2.08	1.86	2.15	1.88	2.16	1.91	2.20	2.29	2.32
	0.83	0.80	0.84	0.79	0.84	0.79	0.84	0.78	0.76	0.76
Shillong	0.87	0.84	0.89	0.83	0.87	0.84	0.88	0.84	0.83	0.83
	1.71	1.85	1.59	1.94	1.67	1.90	1.69	1.9	1.92	1.95
	0.75	0.70	0.78	0.67	0.76	0.69	0.76	0.69	0.68	0.67
Shimla	0.93	0.92	0.94	0.92	0.94	0.92	0.94	0.92	0.91	0.91
	2.17	2.36	2.09	2.41	2.10	2.40	2.24	2.50	2.47	2.50
	0.87	0.84	0.88	0.84	0.88	0.84	0.86	0.83	0.83	0.83
Srinagar	0.95	0.94	0.95	0.94	0.95	0.94	0.95	0.94	0.94	0.94
	2.36	2.57	2.33	2.60	2.31	2.63	2.59	2.84	2.63	2.66
	0.90	0.88	0.90	0.88	0.91	0.88	0.88	0.86	0.88	0.87
Thiruvananthapuram	0.81	0.78	0.84	0.76	0.83	0.77	0.83	0.77	0.75	0.74
	0.98	1.04	0.89	1.08	0.93	1.05	0.93	1.06	1.11	1.12
	0.64	0.60	0.71	0.57	0.68	0.59	0.69	0.58	0.55	0.54

automatic weather station (AWS) observations, at 1-day lead. Hence, for a comparison, with the proposed hybrid Conv1D-LSTM, we reviewed and borrowed the performance of the weather apps in forecasting maximum temperature from the aforesaid literature. The analysis in Thomas et al. (2016) was carried out into two parts: (i) Pan India Average analysis and (ii) Regional (zonal) analysis. A summary of the performance of the weather apps for each of these cases is as follows:

(i) Pan-India average analysis: Table 5 shows the result of an analysis, carried out by the authors for a period of 120 days (June to September 2012) between the synoptic observations and the three weather apps viz. RTWS, AccuWth, and WUnd, in terms of two statistical metrics (CC and RMSE). It is observed that AccuWth attains the highest CC and lowest RMSE, i.e., 0.8 °C and 2.81 °C respectively, followed by WUnd (0.78 and 2.85 °C) and RTWS (0.76 and

Table 4 Proposed model potential to foresee heat days of heatwave events occurred during the period 2012–2020, across cities at different lead time (1-day to 7 day)

City	Total no. of heat days occurred during several heat wave events (2012–2020)	Number of heat days captured by the proposed hybrid model (Conv1D-LSTM)						
		1-day lead	2-day lead	3-day lead	4-day lead	5-day lead	6-day lead	7-day lead
Agartala	9	5	2	3	1	0	0	0
Aizawl	12	8	6	4	6	4	3	1
Bhopal	27	30	21	23	20	19	15	16
Bhubaneshwar	8	7	6	5	2	2	2	1
Chandigarh	45	48	42	35	27	30	26	19
Dehradun	44	50	46	40	38	28	33	27
Gandhinagar	19	25	14	11	13	10	6	8
Gangtok	6	6	2	1	1	1	0	0
Guwahati	8	6	6	5	2	2	0	0
Hyderabad	22	20	18	15	13	12	11	11
Imphal	51	42	33	31	29	30	28	27
Itanagar	41	46	41	36	37	32	27	25
Jaipur	70	77	57	36	28	29	26	27
Kohima	60	44	32	27	24	22	17	14
Kolkata	7	9	3	2	2	3	2	2
Lucknow	38	36	27	23	22	23	21	21
Mumbai	3	2	3	2	0	0	0	0
New Delhi	28	27	23	21	16	14	17	10
Patna	17	14	13	10	9	6	7	5
Raipur	12	12	11	21	8	7	7	2
Ranchi	24	16	14	13	14	11	6	7
Shillong	15	11	7	12	9	7	3	4
Shimla	34	36	29	20	22	17	17	13
Srinagar	15	17	13	11	15	9	7	2
Total	615	594	469	407	358	318	281	242

Table 5 All India performance measure of forecast of daily maximum temperature between the weather apps and observed records (Thomas et al. 2016)

Data sources		Statistical measure	
Observed records	Forecast by weather apps	Coefficient of correlation	Root mean square error
Synoptic observations	RTWS	0.76	2.89
	AccuWth	0.80	2.81
	WUnd	0.78	2.85
AWS observations	RTWS	0.71	4.88
	AccuWth	0.37	8.67
	WUnd	0.57	4.42

2.89 °C). Apart from the comparison with synoptic observations, pan India analysis of the aforementioned three weather apps was also carried out with AWS observations (Table 5). In case of comparison with AWS, the RTWS attains the highest CC, i.e., 0.71 followed by WUnd (0.57) and AccuWth (0.37), whereas the lowest RMSE, i.e., 4.42 °C is achieved by WUnd followed by RTWS (4.88 °C) and AccuWth (8.67 °C).

(ii) Regional analysis: In case of the regional analysis, the AWS data was not available in sufficient quantity, as reported in the literature. Therefore, the regional analysis was performed only using synoptic observations. Table 6 shows the zone wise (central, east, south, northwest, west, and northeast) efficacy of the three weather apps with synoptic observation records in terms of the two statistical metrics viz. CC and RMSE. It is observed that the AccuWth is the best-performing app for the central and northwest region (CC: 0.83 and 0.81, RMSE: 2.33 °C and 2.98 °C, respectively), followed by the RTWS (CC: 0.62 and 0.70, RMSE: 3.56 °C and 4.21 °C) and WUnd (CC: 0.61 and 0.34, RMSE: 4.44 °C and 4.92 °C). Likewise, for the east region, the WUnd is found to have the highest CC (0.67) followed by the AccuWth (0.61) and RTWS (0.41), and with respect to RMSE, AccuWth is having the lowest value (2.29 °C) fol-

Table 6 Regional performance measure of forecast of daily maximum temperature between the weather apps and synoptic observed records (Thomas et al. 2016)

Region	RTWS		AccuWth		WUnd	
	CC	RMSE	CC	RMSE	CC	RMSE
Central	0.62	3.56	0.83	2.33	0.61	4.44
East	0.41	4.01	0.61	2.29	0.67	3.02
South	0.29	2.58	0.35	3.01	0.55	3.00
Northwest	0.70	4.21	0.81	2.98	0.34	4.92
West	0.35	2.96	0.51	2.57	0.50	2.67
Northeast	0.41	3.24	NA	NA	0.50	3.10

lowed by WUnd (3.02 °C) and RTWS (4.01 °C). In case of the south and the west region, the best performing apps are WUnd (CC: 0.55, RMSE: 3.0 °C) and AccuWth (CC: 0.51, RMSE: 2.57 °C), respectively, followed by the (in same order) AccuWth (CC: 0.35, RMSE: 3.01 °C) and RTWS (CC: 0.29, RMSE: 2.58 °C) and WUnd (CC: 0.50, RMSE: 2.67 °C) and RTWS (CC: 0.35, RMSE: 2.96 °C). Finally, the comparison is made for the northeast Indian region. However, in case of northeast region, the forecast from the AccuWth was not available. So, comparison is made with RTWS and WUnd apps only. It is observed that WUnd is having better performance as compared to the RTWS (CC: 0.50 and 0.41, RMSE: 3.10 °C and 3.24 °C, respectively).

Summarizing the aforesaid discussion, it may be noted that the performance of the three weather apps viz. AccuWth, RTWS, and WUnd, varies from region to region. However, it is to be observed that the best performance of these weather apps (i.e., RMSE = 2.29 °C and CC = 0.83) is far less as compared to the best and even with the worst performance of the proposed hybrid DL model. The range of RMSE and CC values are 0.68 to 1.43 °C and 0.98 to 0.91 as obtained from the proposed hybrid DL–based model (Table 2). Thus, for 1-day lead, the performance of the proposed hybrid Conv1D-LSTM model is remarkably superior as compared to all three popular weather apps in forecasting maximum daily temperature.

We could not compare the performance of the proposed hybrid DL–based model at higher lead times, i.e., 2 to 10 days in advance, as the forecasted values are not available for the aforementioned weather apps. The performance metrics are also not reported in any literature to our best knowledge. However, from Table 3, it can be noticed that the average testing performance (averaged across five folds) of the proposed model, at all twenty-eight cities at 10-day lead (maximum lead time), in terms of CC and RMSE, is in the range of 0.78 to 0.95 and 1.04 to 2.57 °C, respectively. Moreover, the model performance in foreseeing heat days is also reasonably good (i.e., efficiencies at 1-, 2-, 3-, 4-, 5-, 6-, and 7- day lead are approximately 92%, 76%, 66%, 58%, 51%, 46%, and 40%, respectively) as shown in Table 4. Thus, overall, it can be concluded that the performance of the proposed hybrid Conv1D-LSTM model is better even at higher lead times as compared to the weather apps. However, it is subjected to be proved if either the forecast results or performance metrics are available from any source.

6 Conclusions

This study presents the potential of a DL-based hybrid Conv1D-LSTM model, for multi-step-ahead (1-day to 10-day) prediction of daily maximum temperature and thereafter exploring its potential to foresee the upcoming heatwave events. It is found that the proposed DL-based hybrid model has the potential to learn the hidden complex non-linear relationship efficiently between different variables within a hydroclimatic system. Therefore, it can be successfully used in hydroclimatic modelling for prediction, a couple of days in advance. The performance of hybrid Conv1D-LSTM model for prediction of multi-step-ahead maximum temperature is better than other DL and ML-based models, such as LSTM, Conv1D, MLP, and SVR. Among the existing models, the performance of the Conv1D and LSTM model is observed to be better than the MLP and SVR model at most of the cities. The proposed DL-based hybrid Conv1D-LSTM model along with LSTM, Conv1D, and MLP is able to provide the prediction all the lead times (1-day to 10-day in advance) simultaneously, which is not possible with SVR. In general, the performance of all the models, including hybrid Conv1D-LSTM, gradually reduces as the prediction lead time increases from 1-day to 10-day in advance. However, the benefit of the hybrid Conv1D-LSTM model was better realized for the higher lead times as compared to other models. The proposed model was also able to predict the heat days with 92% accuracy at 1-day lead with only 5% of error. Although, the accuracy of the model was reduced to 50% at 5-day lead time. Furthermore, the efficacy obtained from the proposed model was also compared with three popular weather apps forecasting result which was published in Thomas et al. 2016 in terms of statistical metrics and was found much better.

Thus, results obtained from this study can be helpful in making some promise to foresee the heatwave events. Thus, timely warning will be very useful to the community to avert heat wave–related ill effects. The precise prediction of maximum temperature is also expected to be helpful in agriculture and irrigation scheduling, running various agro-based models to monitor agricultural activities and climate change study.

Acknowledgements The authors acknowledge the National Supercomputing Mission (NSM) for providing computing resources of "PARAM Shakti" at IIT Kharagpur, which is implemented by C-DAC and supported by the Ministry of Electronics and Information Technology (MeitY) and Department of Science and Technology (DST), Government of India.

Author contribution Conceptualization: Rajib Maity; methodology: Mohd Imran Khan, Rajib Maity; formal analysis: Mohd Imran Khan, investigation: Mohd Imran Khan, Rajib Maity; writing — original draft preparation: Mohd Imran Khan; writing — review and editing: Rajib Maity; funding acquisition: Rajib Maity; resources: Rajib Maity; supervision: Rajib Maity.

Funding This work is partially supported by Department of Science and Technology, Climate Change Programme (SPLICE), Government of India (Ref No. DST/CCP/CoE/79/2017(G)) through a sponsored project.

Data availability The data that support the findings of this study are available from: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5 and https://www.imdpune.gov.in/Clim_Pred_LRF_New/Grided_Data_Download.html. It is freely available and was accessed by the authors in February 2022.

Code availability The codes required for the analysis are written in scientific python development environment (spyder) notebook. The codes may be available on request from the authors.

Declarations

Ethics approval Not applicable.

Consent to participants Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

Abdel-Aal RE (2004) Hourly temperature forecasting using abductive networks. Eng Appl Artif Intell 17:543–556. https://doi.org/10.1016/j.engappai.2004.04.002

Bechini L, Bocchi S, Maggiore T, Confalonieri R (2006) Parameterization of a crop growth and development simulation model at sub-model components level. An example for winter wheat (Triticum aestivum L.). Environ Model Softw 21:1042–1054. https://doi.org/10.1016/j.envsoft.2005.05.006

Bhagwat PP, Maity R (2014) Discussion of "hydroclimatic stream flow prediction using least square-support vector regression." ISH J Hydraul Eng 20:274–275. https://doi.org/10.1080/09715010.2014.881082

Cai Y, Zheng W, Zhang X, Zhangzhong L, Xue X (2019) Research on soil moisture prediction model based on deep learning. PLoS ONE 14:1–19. https://doi.org/10.1371/journal.pone.0214508

Chattopadhyay A, Nabizadeh E, Hassanzadeh P (2020) Analog forecasting of extreme-causing weather patterns using deep learning. J Adv Model Earth Syst 12:e2019MS001958. https://doi.org/10.1029/2019MS001958

Choy KY, Chan CW (2003) Modelling of river discharges and rainfall using radial basis function networks based on support vector regression. Int J Syst Sci 34:763–773. https://doi.org/10.1080/00207720310001640241

Cifuentes J, Marulanda G, Bello A, Reneses J (2020) Air temperature forecasting using machine learning techniques: a review. Energies 13:1–28. https://doi.org/10.3390/en13164215

Drucker H, Surges CJC, Kaufman L, Smola A, Vapnik V (1996) Support vector regression machines. Adv Neural Inf Process Syst 9:155–161

Fang K, Kifer D, Lawson K, Feng D, Shen C (2021) The data synergy effects of time-series deep learning models in hydrology. arXiv Prepr. arXiv2101.01876.

Fu M, Fan T, Ding Z, Salih SQ, Al-Ansari N, Yaseen ZM (2020) Deep learning data-intelligence model based on adjusted forecasting window scale: application in daily streamflow simulation. IEEE Access 8:32632–32651. https://doi.org/10.1109/ACCESS.2020.2974406

Haidar A, Verma B (2018) Monthly rainfall forecasting using one-dimensional deep convolutional neural network. IEEE Access 6:69053–69063. https://doi.org/10.1109/ACCESS.2018.2880044

Ham Y, Kim J, Luo J, Southern N, Enso O (2019) Deep learning for multi-year ENSO Forecasts. Nature 573:567–572. https://doi.org/10.1038/s41586-019-1559-7

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

Hong S, Zhou Z, Zio E, Hong K (2014) Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method. Digit Signal Process A Rev J 27:159–166. https://doi.org/10.1016/j.dsp.2013.12.010

Hu C, Wu Q, Li H, Jian S, Li N, Lou Z (2018) Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. Water (switzerland) 10:1–16. https://doi.org/10.3390/w10111543

IPCC (2013) Summary for policymakers. In: Climate change 2013: the physical science basis. contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Kashid SS, Maity R (2012) Prediction of monthly rainfall on homogeneous monsoon regions of India based on large scale circulation patterns using Genetic Programming. J Hydrol 454–455:26–41. https://doi.org/10.1016/j.jhydrol.2012.05.033

Kearney M, Shine R, Porter WP (2009) The potential for behavioral thermoregulation to buffer "cold-blooded" animals against climate warming. Proc Natl Acad Sci U S A 106:3835–3840. https://doi.org/10.1073/pnas.0808913106

Khan MI, Maity R (2020) Hybrid deep learning approach for multi-step-ahead daily rainfall prediction using GCM simulations. IEEE Access 8:52774–52784. https://doi.org/10.1109/ACCESS.2020.2980977

Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman DJ (2019) 1D convolutional neural networks and applications: a survey. arXiv Prepr arXiv1905.03554:1–20

Kisi O, Sanikhani H (2015) Modelling long-term monthly temperatures by several data-driven methods using geographical inputs. Int J Climatol 35:3834–4346

Kisi O, Shiri J (2014) Prediction of long-term monthly air temperature using geographical inputs. Int J Climatol 34:179–186

Kothawale DR, Munot AA, Kumar KK (2010) Surface air temperature variability over India during 1901–2007, and its association with ENSO. Clim Res 42:89–104. https://doi.org/10.3354/cr00857

Kratzert F, Klotz D, Shalev G, Klambauer G, Hochreiter S, Nearing G (2019) Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrol Earth Syst Sci 23:5089–5110. https://doi.org/10.5194/hess-23-5089-2019

Kratzert F, Herrnegger M, Klotz D, Hochreiter S, Klambauer G (2019a) Neuralhydrology – interpreting LSTMs in hydrology, in: Samek, W., Vedaldi, A., Müller, K.-R., Montavon, G., Hansen, L.K. (Eds.), Explainable AI: interpreting, explaining and visualizing deep learning. Springer Nature Switzerland, Switzerland, pp. 347–362. https://doi.org/10.1007/978-3-030-28954-6_19

Kreuzer D, Munz M, Schlüter S (2020) Short-term temperature forecasts using a convolutional neural network —an application to different weather stations in Germany. Mach Learn with Appl 2:100007. https://doi.org/10.1016/j.mlwa.2020.100007

Krizhevsky A, Sutskever I, Hinton GE (012) Imagenet classification with deep convolutional neural networks, in: Proceedings of the 25th International Conference on Neural Information Processing Systems. pp. 1097–1105. https://doi.org/10.1145/3065386

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

Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE (2017) A survey of deep neural network architectures and their applications. Neurocomputing 234:11–26. https://doi.org/10.1016/j.neucom.2016.12.038

Liu H, Mi X, Li Y (2018) Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM Network and ELM. Energy Convers Manag 159:54–64. https://doi.org/10.1016/j.enconman.2018.01.010

Liu, Y., Racah, E., Prabhat, Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M., Collins, W., 2016. Application of deep convolutional neural networks for detecting extreme weather in climate datasets. arXiv Prepr. arXiv1605.01156. 10.475/123

Livingstone DJ (2008) Artificial neural networks: methods and applications. Humana Press, Totowa, NJ, USA

Luk KC, Ball JE, Sharma A (2000) A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. J Hydrol 227:56–65. https://doi.org/10.1016/S0022-1694(99)00165-1

Ma G, Hoffmann AA, Ma CS (2015) Daily temperature extremes play an important role in predicting thermal effects. J Exp Biol 218:2289–2296. https://doi.org/10.1242/jeb.122127

Maity R, Khan MI, Sarkar S, Dutta R, Maity SS, Pal M, Chanda K (2021) Potential of deep learning in drought assessment by extracting information from hydrometeorological precursors. J Water Clim Chang. https://doi.org/10.2166/wcc.2021.062

Matsuoka D, Nakano M, Sugiyama D, Uchida S (2018) Deep Learning approach for detecting tropical cyclones and their precursors in the simulation by a cloud-resolving global nonhydrostatic atmospheric model. Prog Earth Planet Sci 5:1–16. https://doi.org/10.1186/s40645-018-0245-y

Matsuoka D, Watanabe S, Sato K, Kawazoe S, Yu W, Easterbrook S (2020) Application of deep learning to estimate atmospheric gravity wave parameters in reanalysis data sets. Geophys Res Lett 47:e2020GL089436. https://doi.org/10.1029/2020GL089436

Mazdiyasni O, AghaKouchak A, Davis SJ, Madadgar S, Mehran A, Ragno E, Sadegh M, Sengupta A, Ghosh S, Dhanya CT, Niknejad M (2017) Increasing probability of mortality during Indian heat waves. Sci Adv 3:1–6. https://doi.org/10.1126/sciadv.1700066

Murari KK, Ghosh S, Patwardhan A, Daly E, Salvi K (2015) Intensification of future severe heat waves in India and their effect on heat stress and mortality. Reg Environ Chang 15:569–579. https://doi.org/10.1007/s10113-014-0660-6

Nearing GS, Kratzert F, Sampson AK, Pelissier CS, Klotz D, Frame JM, Prieto C, Gupta HV (2021) What role does hydrological science play in the age of machine learning? Water Resour Res 57:e2020WR028091. https://doi.org/10.1029/2020WR028091

Oh JW, Ngarambe J, Duhirwe PN, Yun GY, Santamouris M (2020) Using deep-learning to forecast the magnitude and characteristics of urban heat island in Seoul Korea. Sci Rep 10:1–13. https://doi.org/10.1038/s41598-020-60632-z

Pan B, Hsu K, AghaKouchak A, Sorooshian S (2019) Improving precipitation estimation using convolutional neural network. Water Resour Res 55:2301–2321. https://doi.org/10.1029/2018WR024090

Paniagua-Tineo A, Salcedo-Sanz S, Casanova-Mateo C, Ortiz-García EG, Cony MA, Hernández-Martín E (2011) Prediction of daily maximum temperature using a support vector regression algorithm. Renew Energy 36:3054–3060. https://doi.org/10.1016/j.renene.2011.03.030

Papacharalampous G, Tyralis H, Koutsoyiannis D (2018) Univariate time series forecasting of temperature and precipitation with a focus on machine learning algorithms: a multiple-case study from Greece. Water Resour Manag 32:5207–5239. https://doi.org/10.1007/s11269-018-2155-6

Rohini P, Rajeevan M, Srivastava AK (2016) On the variability and increasing trends of heat waves over India. Sci Rep 6:1–9. https://doi.org/10.1038/srep26153

Salcedo-Sanz S, Deo RC, Carro-Calvo L, Saavedra-Moreno B (2016) Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms. Theor Appl Climatol 125:13–25. https://doi.org/10.1007/s00704-015-1480-4

Scher S (2018) Toward data-driven weather and climate forecasting: approximating a simple general circulation model with deep learning. Geophys Res Lett 45:12616–12622. https://doi.org/10.1029/2018GL080704

Shen R, Huang A, Li B, Guo J (2019) Construction of a drought monitoring model using deep learning based on multi-source remote sensing data. Int J Appl Earth Obs Geoinf 79:48–57. https://doi.org/10.1016/j.jag.2019.03.006

Shi M, Xie F, Zi Y, Yin J (2016) Cloud detection of remote sensing images by deep learning. Int. Geosci. Remote Sens. Symp. 2016-Novem, 701–704. https://doi.org/10.1109/IGARSS.2016.7729176

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15:1929–1958. https://doi.org/10.1109/ICAEES.2016.7888100

Sun AY, Tang G (2020) Downscaling satellite and reanalysis precipitation products using attention-based deep convolutional neural nets. Front Water 2:1–22. https://doi.org/10.3389/frwa.2020.536743

Thomas A, Kashid S, Kaginalkar A, Islam S (2016) How accurate are the weather forecasts available to the public in India? Weather 71:83–88. https://doi.org/10.1002/wea.2722

Tran TTK, Lee T, Kim JS (2020) Increasing neurons or deepening layers in forecasting maximum temperature time series? Atmosphere (basel) 11:1072. https://doi.org/10.3390/atmos11101072

Tran TTK, Bateni SM, Ki SJ, Vosoughifar H (2021) A review of neural networks for air temperature forecasting. Water (Switzerland) 13.https://doi.org/10.3390/w13091294

Vasseur DA, DeLong JP, Gilbert B, Greig HS, Harley CDG, McCann KS, Savage V, Tunney TD, O'Connor MI (2014) Increased temperature variation poses a greater risk to species than climate warming. Proc. R. Soc. B Biol. Sci. 281.https://doi.org/10.1098/rspb.2013.2612

Wang J, Li Y (2018) Multi-step ahead wind speed prediction based on optimal feature extraction, long short term memory neural network and error correction strategy. Appl Energy 230:429–443. https://doi.org/10.1016/j.apenergy.2018.08.114

Ye L, Xiong W, Li Z, Yang P, Wu W, Yang G, Fu Y, Zou J, Chen Z, Van Ranst E, Tang H (2013) Climate change impact on China food security in 2050. Agron Sustain Dev 33:363–374. https://doi.org/10.1007/s13593-012-0102-0

Zabini F (2016) Mobile weather apps or the illusion of certainty. Meteorol Appl 23:663–670. https://doi.org/10.1002/met.1589

Zhang Z, Dong Y, Yuan Y (2020) Temperature forecasting via convolutional recurrent neural networks based on time-series data. Complexity 2020.https://doi.org/10.1155/2020/3536572