Atmósfera 39, 451-467 (2025)
https://doi.org/10.20937/ATM.53438

Feasibility assessment of machine learning for predicting heatwaves in Bangladesh

Torikul ISLAM SANJID, Anock SOMADDER* and Tanvir AHMED

Department of Physics, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh.
*Corresponding author; email: anock-phy@sust.edu

Received: September 17, 2024; Accepted: April 8, 2025

RESUMEN

Como fenómenos meteorológicos extremos, las olas de calor conllevan graves riesgos para la salud humana, la sociedad y los ecosistemas. En los últimos años, Bangladesh ha experimentado olas de calor que son cada vez más frecuentes e intensas. Los sistemas de alerta temprana pueden ayudar a la minimización del daño potencial de estos sistemas, proporcionando tiempo suficiente para una preparación exhaustiva y eficaz. Tradicionalmente, la predicción numérica del tiempo (PNT) se emplea para pronosticar olas de calor, pero es costosa y requiere mucho tiempo. Este estudio explora el potencial del uso de aprendizaje automático como una alternativa más rápida y rentable a la PNT. Específicamente, nos centramos en construir una red neuronal artificial (RNA) para predecir olas de calor con tres días de antelación en Bangladesh. Nuestro modelo utiliza 28 características para predecir un valor objetivo binario (0 para ausencia de ola de calor, 1 para ola de calor). Los resultados son prometedores: el modelo logra una precisión del 91 % para distinguir los días con ola de calor y los días sin ella. Esto sugiere que el aprendizaje automático puede ser una herramienta valiosa para la predicción de olas de calor a gran escala en Bangladesh.

ABSTRACT

As extreme weather phenomena, heatwaves bring severe risks to human health, society, and ecosystems. Over the past few years, Bangladesh has experienced heatwaves that are becoming more frequent and intense. Early warning systems (EWS) can help to minimize the potential damage from these events by providing sufficient time for thorough and effective preparation. Traditionally, numerical weather prediction (NWP) is employed for heatwave forecasting, but it is both expensive and time-consuming. This study explores the potential of using machine learning as a faster and more cost-effective alternative to NWP. Specifically, we focus on building an artificial neural network (ANN) to predict heatwaves three days in advance over Bangladesh. Our model utilizes 28 features to predict a binary target value (0 for no heatwave, 1 for heatwave). The results are promising, with the model achieving an accuracy of 91% in distinguishing heatwave and non-heatwave days. This suggests that machine learning can be a valuable tool for large-scale heatwave prediction in Bangladesh.

Keywords: heatwave, machine learning, ANN, Bangladesh, extreme event prediction.

1. Introduction

A prolonged period of excessively hot weather relative to local climatology, accompanied by high humidity, is usually defined as a heatwave. Due to their potential for significant damage to human health, agriculture, infrastructure, and ecosystems, these extreme events are highly concerning (Barriopedro et al., 2011; Perkins, 2015). In recent decades, there have been several unusually scorching summers and heatwaves with unprecedented temperatures (IPCC,

2014). Some recent major heatwaves took place in Australia (2009), Russia (2010), South Asia (2015), and Southeast Asia (2018) (Dole et al., 2011). Heatwaves led to the deaths of more than 136 000 individuals between 2001 and 2010, which was 2000% higher than in the preceding decade (WMO, 2013). In 2015, four out of the 10 deadliest disasters were caused by heat, with South Asian heatwaves ranking third and fourth in terms of fatalities (UNDRR, 2015). Furthermore, climate change will result in more intense summer temperatures (Barriopedro et al., 2011; Perkins, 2015; Shahi et al., 2021) as well as a rise in the duration, severity, and frequency of heatwaves (Perkins and Alexander, 2013; Seneviratne et al., 2014; Ford et al., 2018; Perkins-Kirkpatrick and Lewis, 2020).

Heatwaves are referred to as silent killers, considering their direct and severe effect on people's health (Patz et al., 2005; Hondula et al., 2014; Heo et al., 2019; Ray et al., 2021). They can lead to various health complications, such as heatstroke, heat cramps, heat exhaustion, and heat stress (van Oldenborgh et al., 2018). Additionally, exposure to extreme and prolonged heatwaves may result in a significant rise in death rates, especially among the elderly and children (Burkart and Endlicher, 2011; Burkart et al., 2011a, 2011b, 2014). This is particularly concerning for Bangladesh, as it is one of the most densely populated countries in the world, with a population of 165.16 million (BBS, 2022).

Another major impact of heatwaves is on agriculture. Extended periods of high temperature can diminish crop yields and provoke reproductive failure in various crops (Chaudhury et al., 2000; Attri and Rathore, 2003; Dash and Mamgain, 2011; Siebert et al., 2014; Steffen et al., 2014; Chakraborty et al., 2019). Extreme temperatures can cause water deficiency in plants, potentially resulting in their death as a consequence of the interruption of photosynthesis (Schlenker and Roberts, 2009; Steffen et al., 2014). Domestic animals may suffer from decreased appetite, lower productivity, weakened immune systems, and potentially death when they are exposed to heat stress (Lefcourt and Adams, 1996; Steffen et al., 2014). These occurrences can contribute to a decline in agricultural productivity, which is particularly concerning given that 11-12% of Bangladesh's GDP relies on agriculture (Ministry of Finance, 2023). Heatwaves also harm important infrastructure and different services such as transport systems, electricity supply, etc. Additionally, intense heatwaves also result in lower labor productivity which then impacts the economy (Steffen et al., 2014).

However, it is possible to minimize the damage caused by heat-related incidents. Preparation for heatwaves can be taken in advance to some extent using early warning systems (EWS) (Lavell et al., 2012). For example, EWS can assist in determining when the crops will require more irrigation, when cooling facilities should be set up for people to rest, and when local hospitals will need to prepare for an increased number of patients (Bassil and Cole, 2010). Such systems can help vulnerable communities and regions by providing them with sufficient time in advance for preparation (Zhu and Li, 2018). Therefore, considering the upward trend in average temperature, often accompanied by an even quicker rise in the likelihood of severe heat events (Barriopedro et al., 2011; Perkins-Kirkpatrick and Lewis, 2020), the development of EWS for such occurrences becomes a crucial element in a successful adaptation strategy for climate change (Lavell et al., 2012).

Heatwaves are typically forecasted using numerical weather prediction (NWP). In a regular NWP framework, weather forecasts are generated by running computer models based on the laws of physics and fluid dynamics where the current weather is used as initial input, and then it simulates the evolution of the atmosphere over the next few days (Lynch, 2008). Despite several decades of progress in weather prediction, primarily through advancements in computationally expensive NWP models and data assimilation techniques (Bauer et al., 2015; Alley et al., 2019), cutting-edge NWP models still struggle to reliably predict the onset and endurance of blocking events (Pelly and Hoskins, 2003; Matsueda, 2011). Machine learning (ML), known for its ability to understand complex, elusive patterns and generate accurate results, has the potential to be used for data-driven forecasting of the spatiotemporal evolution of weather systems and their extreme events (Chattopadhyay et al., 2020a).

ML techniques have been utilized in climate and weather forecasting for decades, serving various purposes such as post-processing, data assimilation, and physical analysis. More recently, ML has been

increasingly applied to direct weather forecasting, demonstrating significant potential (de Burgh-Day and Leeuwenburg, 2023). Due to their ability to capture complex, nonlinear relationships, ML algorithms have been widely used to forecast climate variables such as extreme rainfall (Saha et al., 2016; Yaseen et al., 2018), wind patterns (Lagerquist et al., 2017; Tateo et al., 2019), and evapotranspiration (Tao et al., 2018). Some studies have also explored ML-based approaches for temperature and heatwave prediction (Khan et al., 2019; Chattopadhyay et al., 2020b). Since 2022, rapid advancements in ML-based weather forecasting have emerged (Bi et al., 2022; Keisler, 2022; Pathak et al., 2022; Lam et al., 2023; Chen et al., 2023). Deep learning architectures have successfully predicted long-lasting, intense heatwaves in France up to two weeks in advance (Jacques-Dumas et al., 2022), demonstrating higher accuracy than the European Centre for Medium-Range Weather Forecasts (ECMWF) at lead times of 3-4 weeks (López-Gómez et al., 2023). Furthermore, Google DeepMind's GraphCast, a weather forecasting system based on graph neural networks (GNN), has shown superior predictive performance, outperforming traditional weather models in 90% of tested cases and improving the forecasts of extreme events such as tropical cyclones, atmospheric rivers, and heatwaves (Lam et al., 2023).

Several studies have specifically investigated heatwave prediction using ML techniques. For instance, Chattopadhyay et al. (2020b) employed capsule neural networks (CapsNets) to predict heatwave and cold wave occurrences over North America. Khan et al. (2019) compared the performance of support vector machines (SVM), random forest (RF), and artificial neural networks (ANNs) in predicting heatwave days in Pakistan, concluding that SVM exhibited the highest predictive accuracy. Additionally, Weirich-Benet et al. (2023) used linear models and RF with regional atmospheric and surface predictors to forecast summer temperature anomalies and heatwave probabilities in Central Europe. Most of these existing ML-based heatwave prediction studies have focused on specific regions with extensive historical datasets. However, due to data limitations, similar research has not been extensively conducted for Bangladesh. The increasing availability of high-resolution, open-source meteorological datasets such as ERA5 reanalysis data (Hersbach et al., 2023a, b) has now made it feasible to explore ML-based heatwave forecasting in data-scarce regions.

This study aims to develop a data-driven ANN model for predicting heatwave occurrences over Bangladesh, utilizing the hourly ERA5 dataset, which has been applied in previous ML-based weather forecasting studies due to its ability to learn complex, nonlinear dependencies from high-dimensional meteorological data capturing both short-term fluctuations and long-term trends (Abhishek et al., 2012). The primary objective is to assess the feasibility of deploying such a model at a larger scale, with the ultimate goal of improving heatwave prediction for Bangladesh in the future. The structure of this paper is as follows: section 2 provides an overview of the study area and dataset; section 3 discusses the 2-m temperature climatology over Bangladesh; section 4 details the methodology; section 5 presents the results and discussion; and section 6 provides concluding remarks and future research directions.

2. Study area and data

Bangladesh, a small country in South Asia, is situated between 20.57º and 26.63º N, and 88.02º and 92.68º E (Fig. 1). It has common borders with India to the north, east, and west, and Myanmar to the southeast. It has a total land area of 148 460 km^2 extending 820 km from north to south and 600 km from east to west. In the south, the country has a 720-km-long coastline. Approximately 80% of the country's land area is lush, alluvial lowland known as the Bangladesh Plain, part of the larger Plain of Bengal. The majority of elevated land is less than 10 masl, with the northern part of the plain reaching altitudes up to 105 m. Some exceptions to the lowlands include the hills of Chittagong in the southeast, the low mountains of Sylhet in the northeast, and the higher terrain in the north and northwest (Wikipedia, 2023).

This study utilizes the 5th generation atmospheric reanalysis (ERA5) by the European Centre for Medium-Range Weather Forecasts (ECMWF) to develop a dataset for ML implementation. These data are available for the last eight decades, starting from 1940 onwards. The data covers the entire globe, including both ocean and land areas (Hersbach et al., 2023b). Data is also available at various vertical levels of the

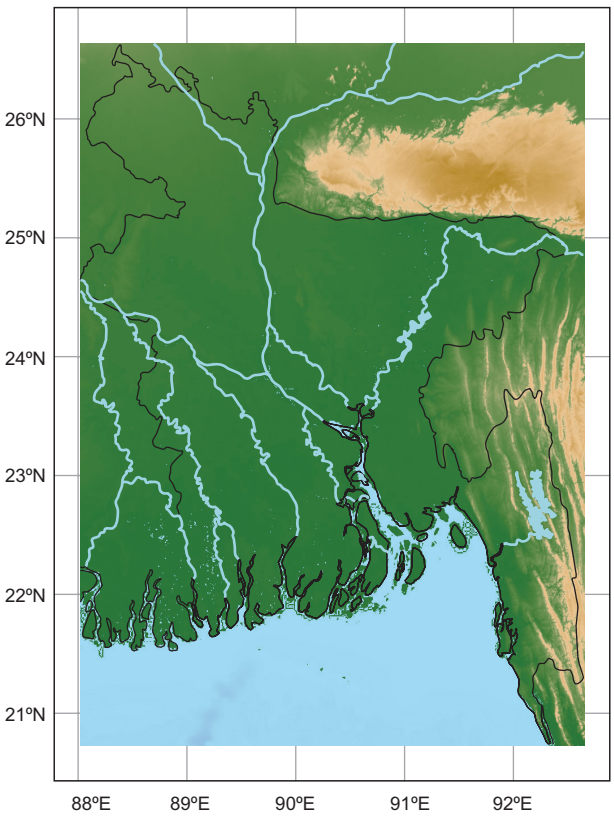

Fig. 1. Study area (Bangladesh).

atmosphere (Hersbach et al., 2023a). The temporal resolution of ERA5 data is hourly, providing detailed information on atmospheric conditions throughout the day. This high frequency is particularly useful for studying short-term weather events and variability. ERA5 has a higher spatial resolution compared to its predecessors, with grid points spaced at 0.25° in latitude and longitude. This fine spatial resolution enables a more detailed representation of atmospheric processes, especially in regions with complex terrain.

3. Two-meter temperature climatology over Bangladesh

Among the four meteorological seasons in Bangladesh, three are significantly influential: winter (DJF), pre-monsoon (MAM), and monsoon (JJAS). Post-monsoon (ON) is relatively weak and serves as a transitional period from the monsoon to winter (Khatun et al., 2016). During the pre-monsoon period, continental heating increases, developing a low-pressure monsoon trough supported by the Tibetan Plateau and the Himalayas, which extends the heating effect across the entire troposphere (Huang et al., 2023). This creates a cross-equatorial pressure gradient, which is further strengthened by convective activity, leading to the onset of monsoons over the southeast region of Bangladesh in early June. Then, at the end of September or the beginning of October, it moves northwest and withdraws from the northeast to the southwest (Khatun et al., 2016).

The typical monthly fluctuations of the mean, maximum, and minimum temperature based on ERA5 2-m temperature data from 1981-2022 are displayed in Figure 2a. The average temperature in January is at its lowest point, steadily rising from February onwards to reach a value of approximately 28 ºC in May. From May to September, the mean temperature remains relatively stable before gradually decreasing. The maximum daily temperature (daytime temperature) reaches its highest values in

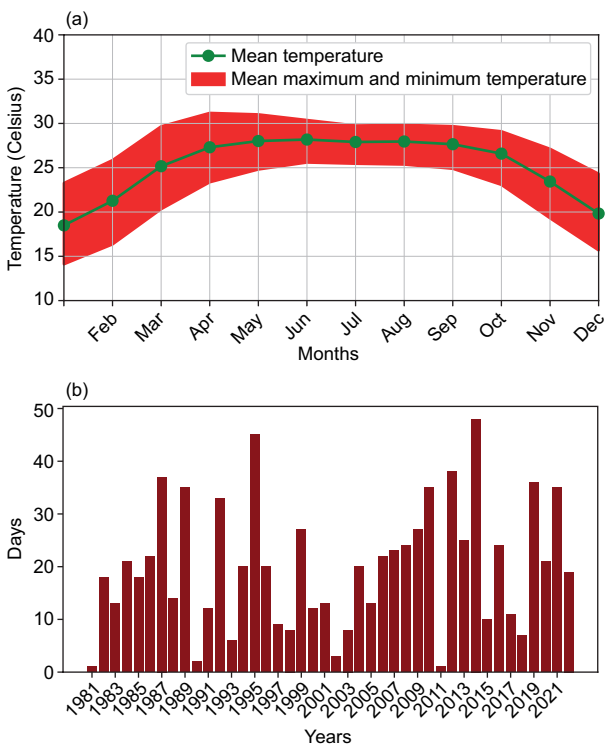

Fig. 2. (a) Monthly normal mean, maximum, and minimum 2-m temperature (1981- 2022), and (b) annual number of days that exceeded the 95th percentile of maximum 2-m temperature.

April and May, averaging around 31 ºC. However, it drops slightly due to an increase in rainfall when the monsoon season arrives early in June. The average maximum temperature then remains relatively steady until the monsoon ends in September or October. On the other hand, the minimum daily temperature (night-time temperature) does not exhibit the same crest during the pre-monsoon season as the maximum temperature does. They reach their peak value in the mid-monsoon season.

In the context of global warming, the frequency and intensity of hot days are increasing, while the number of cold days is declining (Perkins-Kirkpatrick and Lewis, 2020). This trend is evident in the annual count of days exceeding the 95th percentile of maximum temperature, as illustrated in Figure 2b. The plot demonstrates a significant upward trend in the number of such days over the study period, contributing to an increase in the occurrence of heatwave days.

Figure 3 showcases the spatial distribution of monthly mean of daily maximum 2-m temperature from 1981-2022. In January, the northwestern part of Bangladesh experiences the lowest monthly normal maximum temperature (20-24 ºC), while the rest of the country faces higher maximum temperature (24-28 ºC). In February, the normal maximum temperature is relatively consistent across the country (24-28 ºC). By March, the lower western part of the country and some areas in the Chittagong Hill Tracts witness higher temperature (32-36 ºC) compared to the rest of the country (28-32 ºC). This temperature trend intensifies, covering almost the entire country except for the northeastern part in April, and remains relatively stable during May. In June, this pattern diminishes and influences some western parts of the country. The temperature remains nearly constant (28-32 ºC) throughout the country from July to October. In November, temperature decreases (24-28 ºC) in most parts of the country, except for some lower western regions. In December, the average maximum temperature across the country falls within the 24-28 ºC range.

4. Methods

4.1 Heatwave definition

There is no single, universal definition of heatwave. Rather, the definition varies depending on the impact on different sectors. For example, one study defined heatwaves by examining four key variables: the heat metric (e.g., maximum/minimum/mean temperature, diurnal temperature difference), duration, threshold type, and threshold intensity (Vaidyanathan et al., 2016). One definition identifies a heatwave as a period when the temperature surpasses the long-term daily 80th percentile of the daily maximum temperature for three consecutive days (Della-Marta et al., 2007). According to another definition, a heatwave occurs when the maximum temperature in a grid exceeds the average temperature by at least 3 ºC for three or more consecutive days (Srivastava et al., 2009). Heatwaves are also sometimes defined as periods of at least six consecutive days where the daily maximum temperature exceeds the historical 99th percentile (Mishra et al., 2015). So, most definitions rely on the prolongation of air temperature at the surface surpassing a certain threshold value (Perkins, 2015).

The definition of heatwave days in this study is based on the study conducted by Nissan et al. (2017), who demonstrated a 12 and 10% rise in mortality in Bangladesh when the maximum daily temperature and maximum daily heat index, respectively, exceeded the 95th percentile for at least three days in a row. Following these criteria, a heatwave was defined as follows: when the maximum daily temperature or maximum daily heat index in any grid surpasses its respective 95th percentile threshold for at least three consecutive days, we consider it a heatwave from the first day.

4.2 Data collection and feature selection

ERA5 reanalysis data were obtained for this study from the Copernicus Climate Data Store. This data is extremely reliable and has been used in many previous studies. The ERA5 dataset provides a comprehensive range of weather parameters, including the variables required for this research, as listed in Table I, spanning the period from 1981 to 2022. Here, features or variables were selected based on their successful application in forecasting heatwaves in other studies (Fig. 4). For example, in one study, surface temperature and geopotential height were used to forecast the occurrence of intense and prolonged heatwaves (Jacques-Dumas et al., 2022). In two separate studies, geopotential height, air

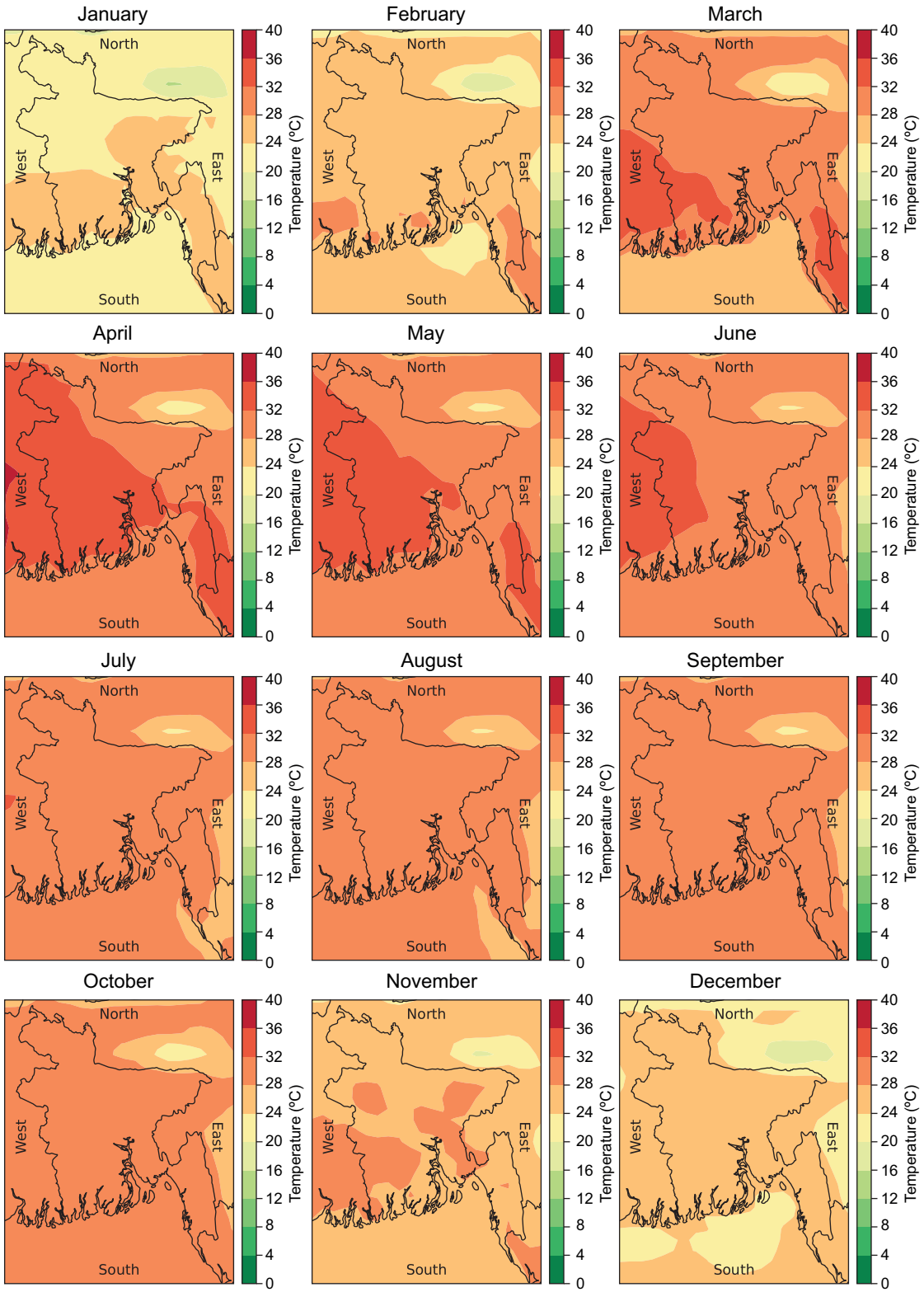

Fig. 3. Spatial distribution of monthly mean of daily maximum 2-m temperatures (1981-2022).

Table I. Atmospheric variables included before feature selection.

Atmospheric variables	Symbol	Level
2-m temperature	t2m	2 m
Air temperature	temp	1000, 850, 500 hPa
Maximum temperature	mx2t	1000, 850, 500 hPa
Minimum temperature	mn2t	1000, 850, 500 hPa
Geopotential	z	1000, 850, 500 hPa
Relative humidity	rhum	1000, 850, 500 hPa
Specific humidity	shum	1000, 850, 500 hPa
U wind	u10	10 m
V wind	v10	10 m
Surface pressure	sp	Surface
Mean sea level pressure	msl	Surface
Precipitation	tp	Surface
Soil moisture	swvl1, swvl2	0-7 cm, 7-28 cm

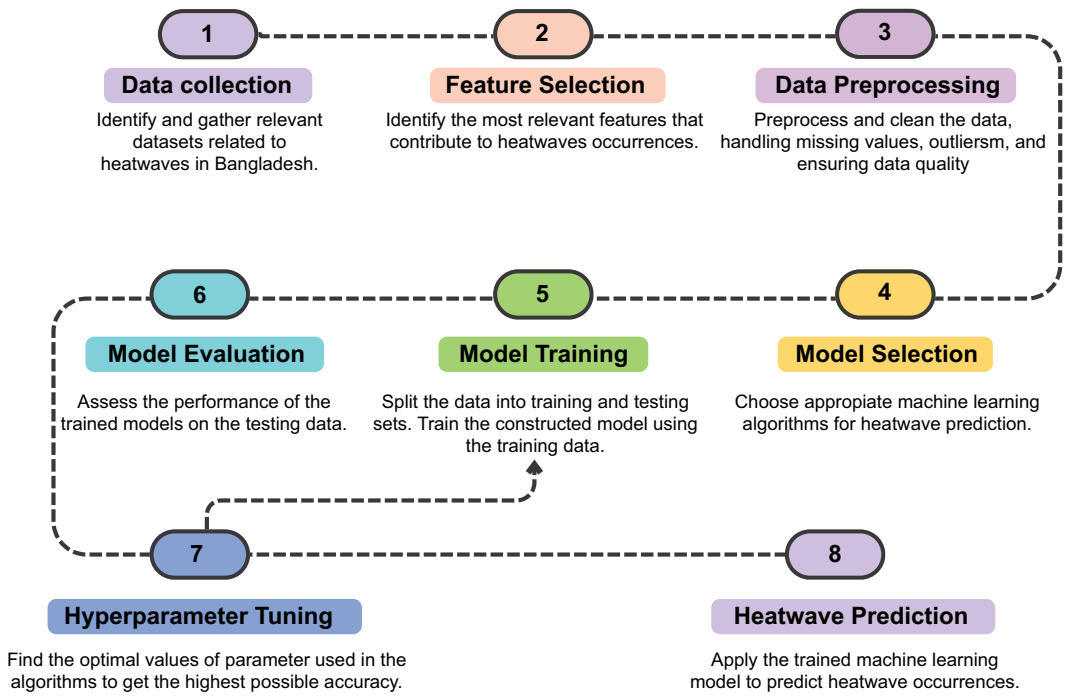

Fig. 4. General workflow for building a machine learning model.

temperature, relative humidity, specific humidity, U wind, and V wind at different pressure levels were used to predict heatwaves over Pakistan and Iran (Asadollah et al., 2021; Khan et al., 2021). In another study, researchers utilized 2-m air temperature, 500-hPa geopotential, precipitation, and soil moisture to forecast sub-seasonal summer heatwaves in Central Europe (Weirich-Benet et al., 2023).

4.3 Data preprocessing

In this study, the data obtained went through the following preprocessing steps:

1. Hourly ERA5 data were converted into daily data.
2. Given that heatwaves commonly occur during the pre-monsoon, monsoon, and post-monsoon periods, we focused on data from March to October.

3. The ERA5 data was re-gridded from a high resolution (0.25º × 0.25º) to a lower resolution (1.25º × 1.25º) to enhance computational efficiency. While reducing resolution can sometimes impact accuracy, particularly in regions with complex topography, this effect is expected to be minimal for Bangladesh, as the region has relatively low topographic variation (as discussed in section 2). Therefore, the chosen resolution provides a reasonable balance between computational efficiency and predictive performance.
4. A data frame for the heat index was created. The heat index, representing perceived temperature, combines actual temperature and relative humidity to indicate human comfort levels and potential health impacts. Since ERA5 lacks surface relative humidity data, we calculated relative humidity using dew point temperature from ERA5, following the formula below (Eq. 1):

$$\textit{Relative Humidity, } rh = \frac{e}{e_s} \times 100 \tag{1}$$

where $e = 6.11 \times 10^{\frac{7.5 \times T_d}{237.7 + T_d}}$ is the actual vapor pressure, $e_s = 6.11 \times 10^{\frac{7.5 \times T}{237.7 + T}}$ is the saturated vapor pressure, T is the air temperature, and T_d is the dewpoint temperature.

After getting the relative humidity, the heat index was calculated using the formula provided by the National Oceanographic and Atmospheric Administration (NOAA), which was initially given by Steadman (1979) and further modified by Rothfusz (1990). It can be written as (Eq. 2):

$$\begin{aligned} HI = & -42.379 + 2.04901523T + 10.14333127RH \\ & - 0.22475541TRH - 0.00683783T^2 \\ & - 0.05481717RH^2 + 0.00122874T^2RH \\ & + 0.00085282TRH^2 - 0.00000199T^2RH^2 \end{aligned} \tag{2}$$

5. All data frames resulting from the previous steps were combined into a single data frame. An algorithm was then developed to identify the heatwave days in the merged data frame using the maximum temperature and maximum heat index columns, based on the heatwave definition mentioned in section 4.1. Heatwave days were indicated by 1, and non-heatwave days were indicated by 0 in a new column named "Heatwave". Finally, the target column "HW_in_3_Days" was generated using the "Heatwave" column.

4.4 Model selection

In this study, an ANN is implemented to predict heatwaves three days in advance. ANNs are a category of machine learning models that mimic the structure and functionality of the human brain. The final ANN model is structured as a sequential neural network using the Keras Sequential API, which comprises multiple densely connected (fully connected) layers (Fig. 5). The first layer is designed with an input shape to accommodate our 28 features. Following this, there were six hidden layers with 128, 64, 48, 32, 16, and eight neurons, respectively, all employing the Rectified Linear Unit (ReLU) activation function. The ReLU function, defined as $f(x) = \max(0, x)$, introduces nonlinearity by allowing positive values to pass through while setting negative values to zero. This property helps prevent the vanishing gradient problem and accelerates model convergence during training. For the binary classification problem addressed in this study, a single neuron was implemented in the output layer with a sigmoid activation function. The sigmoid function is defined as (Eq. 3):

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

This function maps the output to a probability between 0 and 1, making it suitable for our binary classification task. The model is optimized using the Adaptive Moment Estimation (Adam) algorithm with a learning rate of 0.001, ensuring efficient and stable convergence. Additionally, binary cross-entropy is used as the loss function to quantify the difference between predicted and actual labels. Binary cross-entropy measures the dissimilarity between predicted probabilities and actual class labels, ensuring that the model assigns high confidence to correct classifications while penalizing incorrect predictions more severely.

4.5 Model training

In this study, the data were split by year, with 35 years allocated for training and seven years for testing purposes. To ensure that all features contribute equally to the model, Min-Max scaling was applied to all numerical features except year, month, day, latitude, longitude, and the target variable (HW_in_3_Days). This preprocessing step was performed using

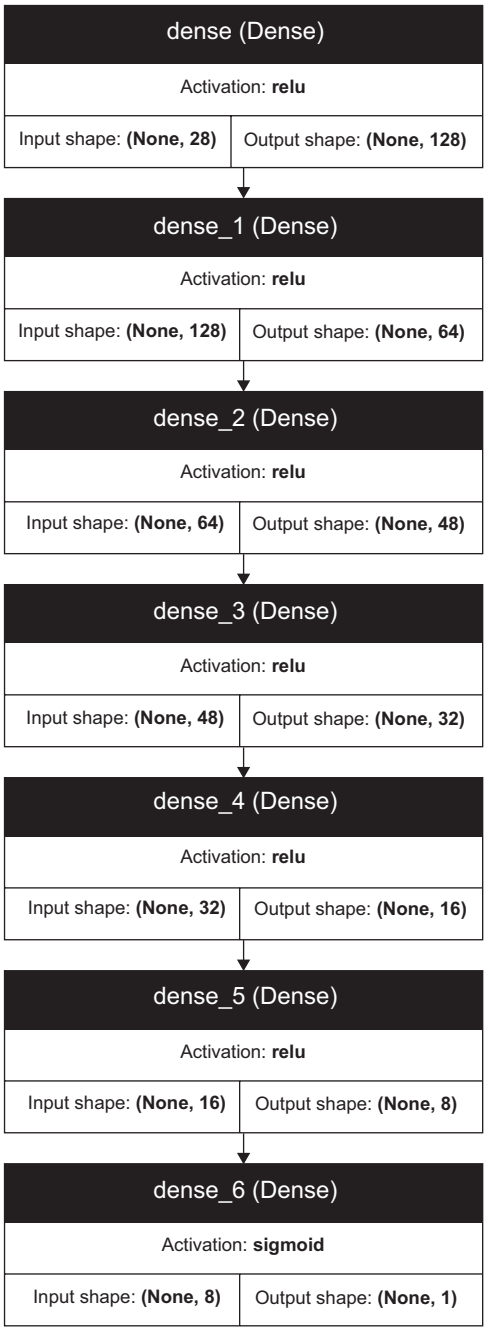

Fig. 5. Architecture diagram of the final artificial neural network.

scikit-learn's "MinMaxScaler", which normalizes each feature to a fixed range of [0,1]. The Min-Max scaling formula is given by (Eq. 4):

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4)$$

This scaling technique helps improve model convergence by ensuring that all features remain within a comparable range and preventing certain features from disproportionately influencing the learning process. To assess model performance during training, 20% of the training data was randomly set aside as a validation set using TensorFlow. This validation set helps monitor model performance at each epoch and plays a crucial role in detecting overfitting by comparing training and validation loss. Since the validation data is selected randomly from the training set, it ensures that the model's generalization ability is assessed across different atmospheric conditions rather than being biased by time or location. Finally, multiple machine learning models were trained on the preprocessed training dataset, and their performance was subsequently evaluated using the test dataset.

4.6 Model evaluation and hyperparameter tuning

Several ANN architectures were developed and systematically evaluated using key performance metrics, including the training loss curve, the area under the receiver operating characteristic (ROC) curve (AUC-ROC), and the confusion matrix. To optimize the model's performance, an iterative hyperparameter tuning process was employed, in which various architectural and training parameters were adjusted and re-evaluated.

Initially, models with three to eight hidden layers were tested, and the number of neurons in each layer was varied between eight and 256. Although the number of neurons was primarily selected in powers of two (e.g., 8, 16, 32, 64, etc.) for their computational efficiency, other values were also explored when deemed appropriate. The learning rate, a critical hyperparameter for training neural networks, was tested across multiple values (0.0001, 0.001, 0.005, and 0.01) using the Adam optimizer. These values were chosen to cover a broad range, from smaller learning rates that promote precise convergence to larger rates that can accelerate training. In addition to the learning rate, a variety of batch sizes (16, 32, 64, and 128) were tested to determine the best trade-off between model performance and training efficiency. The number of training epochs was also varied, ranging from 50 to 150 in increments of 10, to identify the optimal number of iterations required for model convergence without overfitting. To find

the best configuration, all possible combinations of these hyperparameters were systematically tested. The goal was to maximize the AUC-ROC score while minimizing both training and validation loss.

4.7 Heatwave prediction

The primary objective of employing machine learning in this study is to utilize insights derived from historical data to generate accurate and meaningful predictions for new, unseen data. Following iterative evaluation and hyperparameter tuning, the final optimal ANN model was obtained. This model underwent testing using a separate, unseen testing dataset, which was preprocessed using the same normalization techniques applied to the training data. Finally, predictions were made using the model on this prepared test data. Figures 6 and 7 present a sample of the model's predictions alongside the true values for the testing data.

5. Results and discussion

The training and validation loss curve for the final ANN model is presented in Figure 8a. The loss function measures how much the model's predictions differ from the actual targets. The training loss curve in Figure 8a decreased over epochs, indicating a positive sign that the model was learning from the data. The validation loss curve, which shows how well the model performs on a separate dataset not used for training, initially fluctuated but generally went down. This overall decrease suggests that the model is getting better at making accurate predictions on new, unseen data.

Figure 8b presents the receiver operating characteristic (ROC) curve for our chosen ANN model compared to a random classifier. The ROC curve illustrates the trade-off between the true positive rate (sensitivity) and the false positive rate. The steep initial rise, followed by a gradual leveling off, indicates that the model achieves high sensitivity while maintaining a relatively low false positive rate. The dashed diagonal line represents the performance of a random classifier, serving as a baseline. The convex shape, bending toward the top-left corner, demonstrates that the model performs significantly better than a random classifier. The area under the ROC curve (AUC) for the ANN model is 0.90, suggesting strong discriminative ability in distinguishing between heatwave and non-heatwave events, making it a reliable tool for prediction.

Finally, the performance of the final ANN model was assessed using a confusion matrix, which provides a detailed breakdown of the model's classification results by analyzing true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions. The confusion matrix for the model is presented in Figure 9.

To quantitatively evaluate the model's predictive performance, several key metrics were derived from the confusion matrix, including precision, recall, F1-score, and accuracy (Fig. 10). Precision, also known as positive predictive value, measures the proportion of correctly predicted positive cases out of all predicted positives and is defined as (Eq. 5):

$$Precisoin = \frac{TP}{TP + FP} \tag{5}$$

The precision for class 0 was 0.95, indicating that 95% of instances predicted as class 0 were correctly classified. Similarly, the precision for class 1 was 0.56, meaning that 56% of instances labeled as class 1 were correctly identified. Recall quantifies the proportion of actual positive cases that the model correctly identifies and is given by (Eq. 6):

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

The recall for class 0 was 0.94, meaning that 94% of actual class 0 instances were correctly classified. For class 1, the recall was 0.61, indicating that 61% of true class 1 instances were accurately identified. To provide a balanced measure of both precision and recall, we computed the F1-score, which is the harmonic mean of the two metrics and is defined as (Eq. 7):

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

The F1-score for class 0 was 0.95, indicating strong performance in identifying non-heatwave instances. However, the F1-score for class 1, which represents heatwave instances, was lower at 0.58. This suggests that the model struggled more with correctly identifying heatwave days. The underlying reason for this discrepancy is likely the class imbalance,

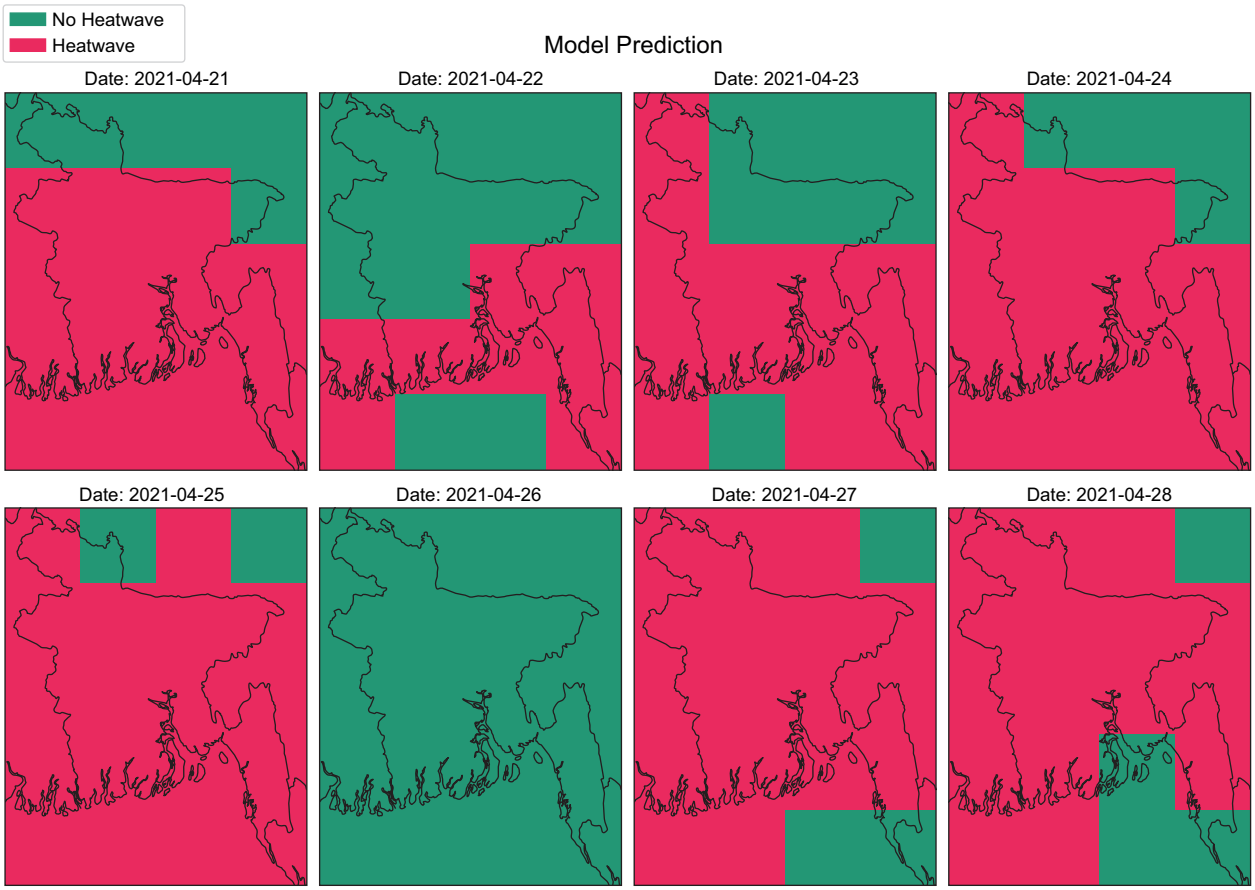

Fig. 6. Heatwave-affected regions predicted by the final artificial neural network model.

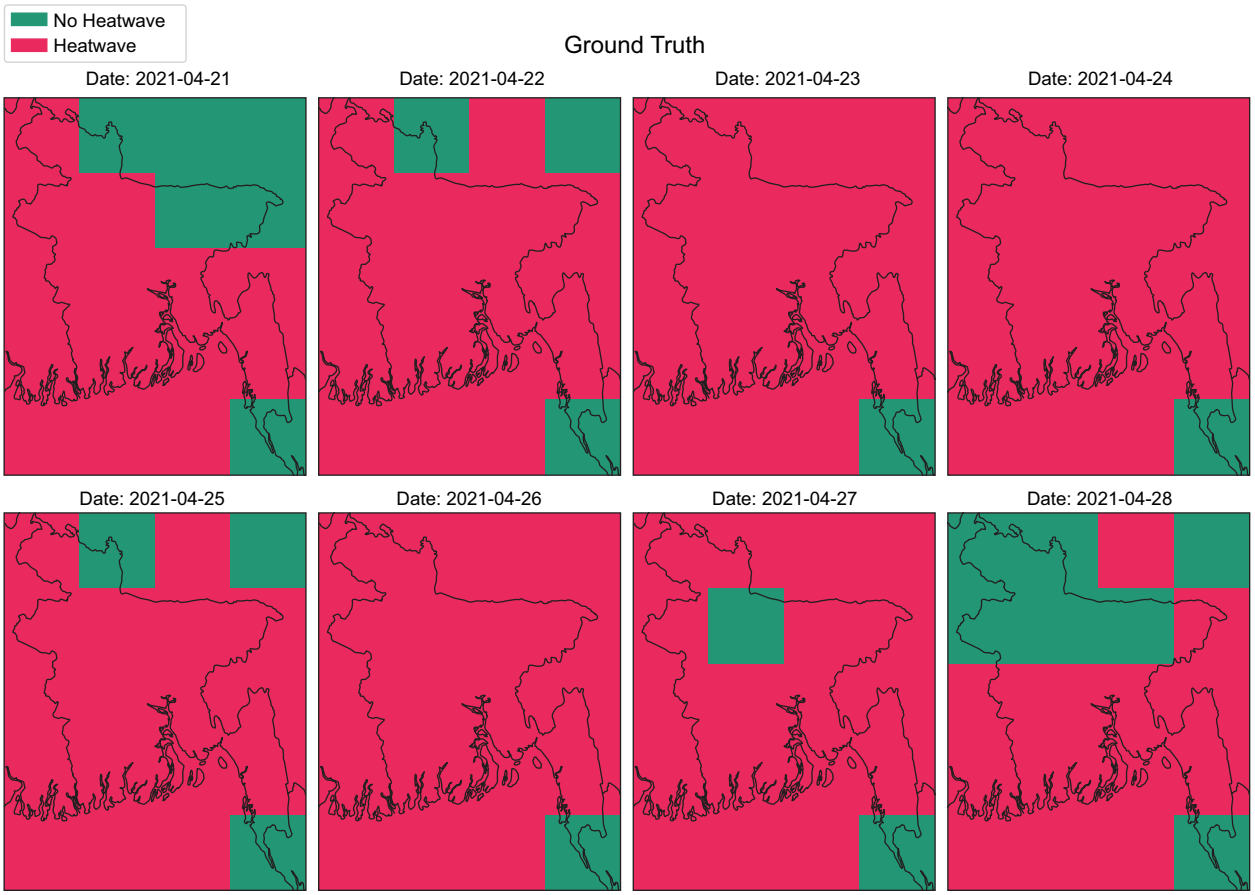

Fig. 7. Actual heatwave-affected regions.

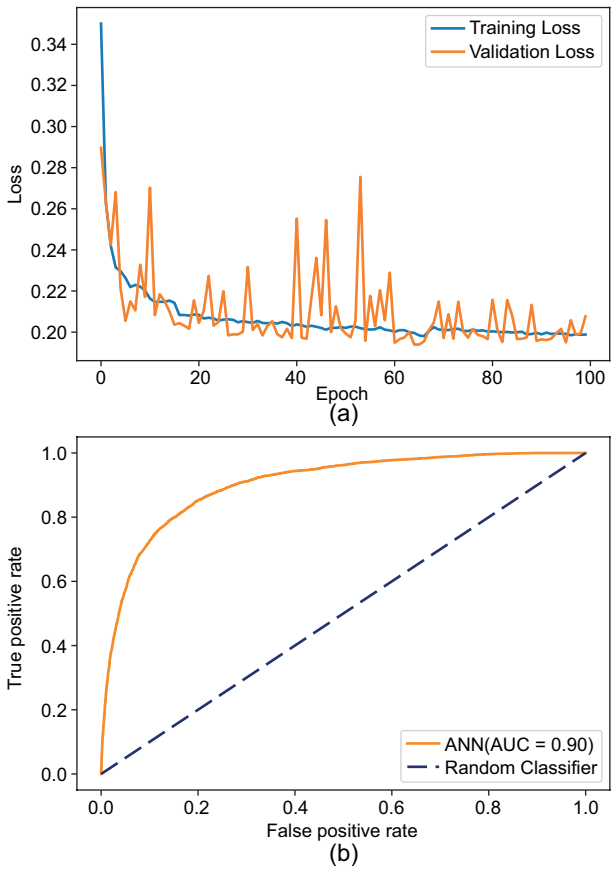

Fig. 8. (a) Training and validation loss, and (b) area under the receiver operating characteristic curve (AUC-ROC).

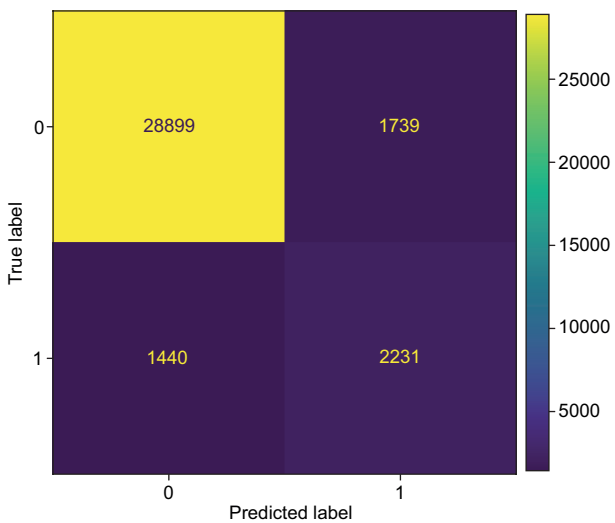

Fig. 9. Confusion matrix.

	precision	recall	f1-score	support
0.0	0.95	0.94	0.95	30629
1.0	0.56	0.61	0.58	3671
accuracy			0.91	34300
macro avg	0.76	0.78	0.77	34300
weighted avg	0.91	0.91	0.91	34300

Fig. 10. Accuracy, precision, recall, and F1 score of our artificial neural network model.

with significantly fewer heatwave days (class 1) compared to non-heatwave days (class 0). Such an imbalance is a common challenge in predictive modeling tasks, and it may have affected the model's ability to accurately classify the minority class (heatwaves). Finally, accuracy, which represents the proportion of correctly classified instances across all classes, was computed using (Eq. 8):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

The overall accuracy of the model was 0.91, meaning that 91% of the total predictions were correct. While the model demonstrates strong overall performance, there remains room for improvement, particularly in enhancing the classification of class 1 instances.

6. Conclusion

In conclusion, given the significant adverse impacts of heatwaves on human health, agriculture, livestock, infrastructure, and essential public services such as hospitals and education, the implementation of an EWS is imperative. This study has demonstrated the feasibility of utilizing machine learning, specifically employing an ANN, for predicting heatwaves using ERA5 reanalysis datasets. A comprehensive set of 28 weather variables was selected as features, and the target variable was binary, with 0 representing "No heatwave" and 1 representing "Heatwave". To train and test the model, a dataset spanning 42 years was utilized. In particular, 35 years of data were allocated for training, and the remaining seven years of data were reserved for testing. Multiple ANN architectures were trained and evaluated using various metrics, including training and validation loss curves, confusion matrices, and AUC values. The final ANN architecture

has a precision of 0.56 for class 1 and 0.95 for class 0, indicating that our model struggled more with predicting class 1. This performance issue is likely due to the class imbalance, as heatwave days (class 1) are significantly fewer than non-heatwave days (class 0). However, the final ANN model achieved an impressive 91% accuracy, showcasing the potential of machine learning for heatwave prediction in Bangladesh.

Class imbalance is a common challenge in predictive modeling tasks, and it likely contributed to the model's difficulty in accurately classifying heatwave instances. To address this issue, future work will explore several strategies such as oversampling the minority class (heatwave days) and undersampling the majority class (non-heatwave days), as well as adjusting class weights in the loss function to give more importance to the minority class, thereby improving the model's ability to detect heatwaves. In addition to addressing class imbalance, future studies should also investigate the effect of using intermediate spatial resolutions (e.g., 0.5º × 0.5º or 1.0º × 1.0º) to determine whether finer spatial detail can enhance the accuracy of heatwave predictions. These improvements could provide more accurate and region-specific predictions, further enhancing the potential of machine learning for early heatwave detection in Bangladesh.

Acknowledgments

We extend our special thanks to Tahmidul Azom Sany (George Mason University, Virginia, USA) for his valuable insights and thoughtful guidance. We also express our deepest gratitude to the Department of Physics at Shahjalal University of Science and Technology (SUST) for their unwavering support and resources throughout this research.

References

Abhishek K, Singh MP, Ghosh S, Anand A. 2012. Weather forecasting model using artificial neural network. Procedia Technology 4: 311-318. https://doi.org/10.1016/j.protcy.2012.05.047

Alley RB, Emanuel KA, Zhang F. 2019. Advances in weather prediction. Science 363: 342-344. https://doi.org/10.1126/science.aav7274

Asadollah SBHS, Khan N, Sharafati A, Shahid S, Chung ES, Wang XJ. 2022. Prediction of heat waves using meteorological variables in diverse regions of Iran with advanced machine learning models. Stochastic Environmental Research and Risk Assessment 36: 1959-1974. https://doi.org/10.1007/s00477-021-02103-z

Attri SD, Rathore LS. 2003. Simulation of impact of projected climate change on wheat in India. International Journal of Climatology 23: 693-705. https://doi.org/10.1002/joc.896

Barriopedro D, Fischer EM, Luterbacher J, Trigo RM, García-Herrera R. 2011. The hot summer of 2010: Redrawing the temperature record map of Europe. Science 332: 220-224. https://doi.org/10.1126/science.1201224

Bassil KL, Cole DC. 2010. Effectiveness of public health interventions in reducing morbidity and mortality during heat episodes: A structured review. International Journal of Environmental Research and Public Health 7: 991-1001. https://doi.org/10.3390/ijerph7030991

Bauer P, Thorpe A, Brunet G. 2015. The quiet revolution of numerical weather prediction. Nature 525: 47-55. https://doi.org/10.1038/nature14956

BBS. 2022 Population and housing census (2022): Preliminary report. Bangladesh Bureau of Statistics. Available at https://sid.portal.gov.bd/sites/default/files/files/sid.portal.gov.bd/publications/01ad1ffe_cfef_4811_af97_594b6c64d7c3/PHC_Preliminary_Report_(English)_August_2022.pdf (accessed 2023 November 15).

Bi K, Xie L, Zhang H, Chen X, Gu X, Tian Q. 2022. Pangu-Weather: A 3D high-resolution model for fast and accurate global weather forecast. arXiv preprint. https://doi.org/10.48550/arXiv.2211.02556

Burkart K, Endlicher W. 2011. Human bioclimate and thermal stress in the megacity of Dhaka, Bangladesh: Application and evaluation of thermophysiological indices. In: Health in megacities and urban areas (Krämer A, Khan MH, Kraas F, Eds.). Physica Heidelberg: 153-170. https://doi.org/10.1007/978-3-7908-2733-0_10

Burkart K, Khan MH, Krämer A, Breitner S, Schneider A, Endlicher WR. 2011a. Seasonal variations of all-cause and cause-specific mortality by age, gender, and socioeconomic condition in urban and rural areas of Bangladesh. International Journal for Equity in Health 10: 32. https://doi.org/10.1186/1475-9276-10-32

Burkart K, Schneider A, Breitner S, Khan MH, Krämer A, Endlicher W. 2011b. The effect of atmospheric thermal

conditions and urban thermal pollution on all-cause and cardiovascular mortality in Bangladesh. Environmental Pollution 159: 2035-2043. https://doi.org/10.1016/j.envpol.2011.02.005

Burkart K, Breitner S, Schneider A, Khan MMH, Krämer A, Endlicher W. 2014. An analysis of heat effects in different subpopulations of Bangladesh. International Journal of Biometeorology 58: 227-237. https://doi.org/10.1007/s00484-013-0668-5

Chakraborty D, Sehgal VK, Dhakar R, Ray M, Das DK. 2019. Spatio-temporal trend in heat waves over India and its impact assessment on wheat crop. Theoretical and Applied Climatology 138: 1925-1937. https://doi.org/10.1007/s00704-019-02939-0

Chattopadhyay A, Hassanzadeh P, Subramanian D. 2020a. Data-driven predictions of a multiscale Lorenz 96 chaotic system using machine-learning methods: Reservoir computing, artificial neural network, and long short-term memory network. Nonlinear Processes in Geophysics 27: 373-389. https://doi.org/10.5194/npg-27-373-2020

Chattopadhyay A, Nabizadeh E, Hassanzadeh P. 2020b. Analog forecasting of extreme-causing weather patterns using deep learning. Journal of Advances in Modeling Earth Systems 12: e2019MS001958. https://doi.org/10.1029/2019MS001958

Chaudhury SK, Gore JM, Sinha Ray KC. 2000. Impact of heat waves over India. Current Science 79: 153-155.

Chen K, Han T, Gong J, Bai L, Ling F, Luo JJ, Chen X, Ma L, Zhang T, Su R, Ci Y, Li B, Yang X, Ouyang W. 2023. FengWu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. arXiv preprint. https://doi.org/10.48550/arXiv.2304.02948

Dash SK, Mamgain A. 2011. Changes in the frequency of different categories of temperature extremes in India. Journal of Applied Meteorology and Climatology 50: 1842-1858. https://doi.org/10.1175/2011JAMC2687.1

De Burgh-Day CO, Leeuwenburg T. 2023. Machine learning for numerical weather and climate modelling: A review. Geoscientific Model Development 16: 6433-6477. https://doi.org/10.5194/gmd-16-6433-2023

Della-Marta PM, Luterbacher J, von Weissenfluh H, Xoplaki E, Brunet M, Wanner H. 2007. Summer heat waves over western Europe 1880-2003, their relationship to large-scale forcings and predictability. Climate Dynamics 29: 251-75. https://doi.org/10.1007/s00382-007-0233-1

Dole R, Hoerling M, Perlwitz J, Eischeid J, Pegion P, Zhang T, Quan XW, Xu T, Murray D. 2011. Was there a basis for anticipating the 2010 Russian heat wave? Geophysical Research Letters 38: L06702. https://doi.org/10.1029/2010GL046582

Ford TW, Dirmeyer PA, Benson DO. 2018. Evaluation of heat wave forecasts seamlessly across subseasonal timescales. NPJ Climate and Atmospheric Science 1: 20. https://doi.org/10.1038/s41612-018-0027-7

Heo S, Bell ML, Lee JT. 2019. Comparison of health risks by heat wave definition: Applicability of wet-bulb globe temperature for heat wave criteria. Environmental Research 168: 158-170. https://doi.org/10.1016/j.envres.2018.09.032

Hersbach H, Bell B, Berrisford P, Biavati G, Horányi A, Muñoz Sabater J, Nicolas J, Peubey C, Radu R, Rozum I, Schepers D, Simmons A, Soci C, Dee D, Thépaut J-N. 2023a. ERA5 hourly data on pressure levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.bd0915c6 (accessed 2023 June 3).

Hersbach H, Bell B, Berrisford P, Biavati G, Horányi A, Muñoz Sabater J, Nicolas J, Peubey C, Radu R, Rozum I, Schepers D, Simmons A, Soci C, Dee D, Thépaut J-N. 2023b. ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.adbb2d47 (accessed 2023 June 7).

Hondula DM, Vanos JK, Gosling SN. 2014. The SSC: A decade of climate-health research and future directions. International Journal of Biometeorology 58: 109-120. https://doi.org/10.1007/s00484-012-0619-6

Huang J, Zhou X, Wu G, Xu X, Zhao Q, Liu Y, Duan A, Xie Y, Ma Y, Zhao P, Yang S, Yang K, Yang H, Bian J, Fu Y, Ge J, Liu Y, Wu Q, Yu H, Wang B, Bao Q, Qie K. 2023. Global climate impacts of land-surface and atmospheric processes over the Tibetan Plateau. Reviews of Geophysics 61: e2022RG000771. https://doi.org/10.1029/2022RG000771

IPCC. 2014. Climate change 2013 – The physical science basis: Working Group I contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge and New York, 1535 pp. https://doi.org/10.1017/CBO9781107415324

Jacques-Dumas V, Ragone F, Borgnat P, Abry P, Bouchet F. 2022. Deep learning-based extreme heatwave

forecast. Frontiers in Climate 4: 789641. https://doi.org/10.3389/fclim.2022.789641

Keisler R. 2022. Forecasting global weather with graph neural networks. arXiv preprint. https://doi.org/10.48550/arXiv.2202.07575

Khan N, Shahid S, Juneng L, Ahmed K, Ismail T, Nawaz N. 2019. Prediction of heat waves in Pakistan using quantile regression forests. Atmospheric Research 221: 1-11. https://doi.org/10.1016/j.atmosres.2019.01.024

Khan N, Shahid S, Ismail TB, Behlil F. 2021. Prediction of heat waves over Pakistan using support vector machine algorithm in the context of climate change. Stochastic Environmental Research and Risk Assessment 35: 1335-1353. https://doi.org/10.1007/s00477-020-01963-1

Khatun MA, Rashid MB, Hygen HO. 2016. Climate of Bangladesh, MET report. Bangladesh Meteorological Department and Norwegian Meteorological Institute. https://www.researchgate.net/publication/348363003_Climate_of_Bangladesh (accessed 2023 November 16).

Lagerquist R, McGovern A, Smith T. 2017. Machine learning for real-time prediction of damaging straight-line convective wind. Weather and Forecasting 32: 2175-2193. https://doi.org/10.1175/WAF-D-17-0038.1

Lam R, Sánchez-González A, Willson M, Wirnsberger P, Fortunato M, Alet F, Ravuri S, Ewalds T, Eaton-Rosen Z, Hu W, Merose A, Hoyer S, Holland G, Vinyals O, Stott J, Pritzel A, Mohamed S, Battaglia P. 2023. Learning skillful medium-range global weather forecasting. Science 382: 1416-1421. https://doi.org/10.1126/science.adi2336

Lavell A, Oppenheimer M, Diop C, Hess J, Lempert R, Li J, Myeong S. 2012. Climate change: New dimensions in disaster risk, exposure, vulnerability, and resilience. In: Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of Working Groups I and II of the Intergovernmental Panel on Climate Change (Field CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, Mastrandrea MD, Mach KJ, Plattner G-K, Allen SK, Tignor M, Midgley PM, Eds.). Cambridge University Press, Cambridge and New York, 25-64. Available at https://www.ipcc.ch/site/assets/uploads/2018/03/SREX_Full_Report-1.pdf (accessed 2023 November 13).

Lefcourt AM, Adams WR. 1996. Radiotelemetry measurement of body temperatures of feedlot steers during summer. Journal of Animal Science 74: 2633-2640. https://doi.org/10.2527/1996.74112633x

López-Gómez I, McGovern A, Agrawal S, Hickey J. 2023. Global extreme heat forecasting using neural weather models. Artificial Intelligence for the Earth Systems 2: e220035. https://doi.org/10.1175/AIES-D-22-0035.1

Lynch P. 2008. The origins of computer weather prediction and climate modeling. Journal of Computational Physics 227: 3431-3444. https://doi.org/10.1016/j.jcp.2007.02.034

Matsueda M. 2011. Predictability of Euro-Russian blocking in summer of 2010. Geophysical Research Letters 38: L06801. https://doi.org/10.1029/2010GL046557

Ministry of Finance. 2023. Bangladesh economic review (2023). Ministry of Finance, Bangladesh. Available at: https://mof.gov.bd/site/page/44e399b3-d378-41aa-86ff-8c4277eb0990/Bangladesh-Economic-Review (accessed 2023 November 16).

Mishra V, Ganguly AR, Nijssen B, Lettenmaier DP. 2015. Changes in observed climate extremes in global urban areas. Environmental Research Letters 10: 024005. http://doi.org/10.1088/1748-9326/10/2/024005

Nissan H, Burkart K, Coughlan de Pérez E, van Aalst M, Mason S. 2017. Defining and predicting heat waves in Bangladesh. Journal of Applied Meteorology and Climatology 56: 2653-2670. https://doi.org/10.1175/JAMC-D-17-0035.1

Pathak J, Subramanian S, Harrington P, Raja S, Chattopadhyay A, Mardani M, Kurth T, Hall D, Li Z, Azizzadenesheli K, Hassanzadeh P, Kashinath K, Anandkumar A. 2022. FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv preprint. https://doi.org/10.48550/arXiv.2202.11214

Patz JA, Campbell-Lendrum D, Holloway T, Foley JA. 2005. Impact of regional climate change on human health. Nature 438: 310-317. https://doi.org/10.1038/nature04188

Pelly JL, Hoskins BJ. 2023. How well does the ECMWF ensemble prediction system predict blocking? Quarterly Journal of the Royal Meteorological Society 590: 1683-1702. https://doi.org/10.1256/qj.01.173

Perkins SE, Alexander LV. 2013. On the measurement of heat waves. Journal of Climate 26: 4500-4517. https://doi.org/10.1175/JCLI-D-12-00383.1

Perkins SE. 2015. A review on the scientific understanding of heatwaves – Their measurement, driving mechanisms, and changes at the global scale. Atmospheric Research 164-165: 242-267. https://doi.org/10.1016/j.atmosres.2015.05.014

Perkins-Kirkpatrick SE, Lewis SC. 2020. Increasing trends in regional heatwaves. Nature Communications 11: 3357. https://doi.org/10.1038/s41467-020-16970-7

Ray K, Giri RK, Ray SS, Dimri AP, Rajeevan M. 2021. An assessment of long-term changes in mortalities due to extreme weather events in India: A study of 50 years' data, 1970-2019. Weather and Climate Extremes 32: 100315. https://doi.org/10.1016/j.wace.2021.100315

Rothfusz LP. 1990. The heat index "equation" (or, more than you ever wanted to know about heat index). NWS Southern Region Headquarters technical attachment SR 90-23. Forth Worth, TX, 2 pp. Available at: https://www.weather.gov/media/ffc/ta_htindx.PDF (accessed 2023 November 16)

Saha M, Mitra P, Nanjundiah RS. 2016. Autoencoder-based identification of predictors of Indian monsoon. Meteorology and Atmospheric Physics 128: 613-628. https://doi.org/10.1007/s00703-016-0431-7

Schlenker W, Roberts MJ. 2009. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. Proceedings of the National Academy of Sciences 106: 15594-15598. https://doi.org/10.1073/pnas.0906865106

Seneviratne SI, Donat MG, Mueller B, Alexander LV. 2014. No pause in the increase of hot temperature extremes. Nature Climate Change 4: 161-163. https://doi.org/10.1038/nclimate2145

Shahi NK, Das S, Ghosh S, Maharana P, Rai S. 2021. Projected changes in the mean and intra-seasonal variability of the Indian summer monsoon in the RegCM CORDEX-CORE simulations under higher warming conditions. Climate Dynamics 57: 1489-1506. https://doi.org/10.1007/s00382-021-05771-3

Siebert S, Ewert F, Eyshi Rezaei E, Kage H, Graß R. 2014. Impact of heat stress on crop yield – On the importance of considering canopy temperature. Environmental Research Letters 9: 044012. http://doi.org/10.1088/1748-9326/9/4/044012

Srivastava AK, Rajeevan M, Kshirsagar SR. 2009. Development of a high resolution daily gridded temperature data set (1969-2005) for the Indian region. Atmospheric Science Letters 10: 249-254. https://doi.org/10.1002/asl.232

Steadman RG. 1979. The assessment of sultriness. Part I: A temperature-humidity index based on human physiology and clothing science. Journal of Applied Meteorology and Climatology 18, 861-873. https://doi.org/10.1175/1520-0450(1979)018<0861:TAOSPI>2.0.CO;2

Steffen W, Hughes L, Perkins S. 2014. Heatwaves: Hotter, longer, more often. Climate Council of Australia. Available at https://www.climatecouncil.org.au/uploads/9901f6614a2cac7b2b888f55b4dff9cc.pdf (accessed 2023 November 12).

Tao H, Diop L, Bodian A, Djaman K, Ndiaye PM, Yaseen ZM. 2018. Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: Regional case study in Burkina Faso. Agricultural Water Management 208: 140-151. https://doi.org/10.1016/j.agwat.2018.06.018

Tateo A, Miglietta MM, Fedele F, Menegotto M, Pollice A, Bellotti R. 2019. A statistical method based on the ensemble probability density function for the prediction of "Wind Days". Atmospheric Research 216: 106-116. https://doi.org/10.1016/j.atmosres.2018.10.001

UNDRR. 2015. 2015 disasters in numbers. Centre for Research on the Epidemiology of Disasters, United Nations Office for Disaster Risk Reduction. Available at https://www.preventionweb.net/files/47804_2015disastertrendsinfographic.pdf (accessed 2023 November 12).

Vaidyanathan A, Kegler SR, Saha SS, Mulholland JA. 2016. A statistical framework to evaluate extreme weather definitions from a health perspective: A demonstration based on extreme heat events. Bulletin of the American Meteorological Society 97: 1817-1830. https://doi.org/10.1175/BAMS-D-15-00181.1

Van Oldenborgh GJ, Philip S, Kew S, van Weele M, Uhe P, Otto F, Singh R, Pal I, Cullen H, AchutaRao K. 2018. Extreme heat in india and anthropogenic climate change. Natural Hazards and Earth System Sciences 18: 365-381 https://doi.org/10.5194/nhess-18-365-2018

Weirich-Benet E, Pyrina M, Jiménez-Esteve B, Fraenkel E, Cohen J, Domeisen DIV. 2023. Subseasonal prediction of Central European summer heatwaves with linear and random forest machine learning models. Artificial Intelligence for the Earth Systems 2: e220038. https://doi.org/10.1175/AIES-D-22-0038.1

Wikipedia. 2023. Geography of Bangladesh. Available at https://en.wikipedia.org/wiki/Geography_of_Bangladesh (accessed 2023 November 16).

WMO. 2013. The global climate 2001-2010: A decade of climate extremes. WMO No. 1103. World Meteorological Organization, Geneva. Available at https://library.

wmo.int/records/item/49934-the-global-climate-2001-2010-a-decade-of-climate-extremes (accessed 2023 November 16).

Yaseen ZM, Ghareb MI, Ebtehaj I, Bonakdari H, Siddique R, Heddam S, Yusif AA, Deo R. 2018. Rainfall pattern forecasting using novel hybrid intelligent model based ANFIS-FFA. Water Resources Management 32: 105-122. https://doi.org/10.1007/s11269-017-1797-0

Zhu Z, Li T. 2018. Extended-range forecasting of Chinese summer surface air temperature and heat waves. Climate Dynamics 50: 2007-2021. https://doi.org/10.1007/s00382-017-3733-7