scientific reports

Check for updates

OPEN

Developing a seasonal-adjusted machine-learning-based hybrid time-series model to forecast heatwave warning

Md. Mahin Uddin Qureshi[1]✉, Amrin Binte Ahmed[2]✉, Adisha Dulmini[3], Mohammad Mahboob Hussain Khan[4] & Rumana Rois[5]✉

Heatwaves pose a significant threat to environmental sustainability and public health, particularly in vulnerable regions and rapidly growing cities. They cause water shortages, stress on plants, and an overall drying out of landscapes, reducing plant growth—the basis of energy production and the food chain. Accurate heatwave forecasting is crucial for early warning systems, public health interventions, and disaster preparedness strategies, reducing heat-related mortality risk through modeling and evaluation of warnings. However, anticipating heatwave warnings requires handling the daily time series data, which is a large-scale and high-frequency time series data. High-frequency time series data forecasting presents unique challenges due to its inherent complexity and characteristics. Therefore, the study proposes two algorithms to develop Machine-Learning (ML)-based hybrid models as well as seasonal adjusted ML-based hybrid models, which can handle large datasets and reveal complex seasonal patterns. The performance of these developed ML-based hybrid models and seasonal adjusted ML-based hybrid models were compared with other traditional time series, Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing State Space (ETS), and Trigonometric Box-Cox ARMA Trend Seasonal (TBATS) and ML models, Artificial Neural Network (ANN), Support Vector Regression (SVR), Prophet, Random Forest Regression (RFR), and Long Short-Term Memory (LSTM), to forecast heatwave warnings in Rajshahi, one of Bangladesh's warmest districts, based on 42-year historical daily instances. Our findings indicate that the seasonal adjusted ML-based hybrid model, by integrating the Seasonal-Trend decomposition procedure based on LOESS (STL) approach with different time series and ML models, STL-ARIMA-LSTM, outperformed all other models with MAE (0.8974), MAPE (2.9232), RMSE (1.1794), MASE (0.3814) and ACF1 (0.0026). Hence, our suggested seasonal adjusted ML-based hybrid model, ensures a more accurate forecast and helps to determine the number and days of heatwaves, enabling people to plan ahead and take necessary safety measures before they occur.

Keywords Forecasting, Heatwave warning, High-frequency data, Hybrid model, Machine learning, Seasonal-adjusted hybrid model, STL decomposition

Severe heat poses a significant threat to human health and leads to many fatalities annually[1]. Global warming has resulted in a five-fold increase in record-breaking heat extremes over the past century, driven by growing greenhouse gasses[2–4]. Rising global temperatures lead to more extreme weather occurrences, such as heatwaves[5,6]. Globally, both historical data and future forecasts have shown that heatwaves and warm spells are growing more common, severe, and lasting longer[6–11]. Heatwaves have major societal and environmental consequences, affecting human health, agriculture, economy, natural disasters, and ecosystems[12]. The term heatwave refers to a widespread area experiencing temperatures above 36 degrees for a minimum of three consecutive days[13]. Furthermore, heatwaves occur when a high-pressure area at 10,000–25,000 feet (3000–7600 m) intensifies and persists over a region for several days or weeks[14].

[1]Department of Statistics and Data Science, Jahangirnagar University, Dhaka 1342, Bangladesh. [2]Department of Statistics and Data Science, Jahangirnagar University, Dhaka 1342, Bangladesh. [3]Department of Business and Law, University of Wollongong, Wollongong, NSW 2522, Australia. [4]Bangladesh Meteorological Department (BMD), Dhaka 1207, Bangladesh. [5]Department of Statistics and Data Science, Jahangirnagar University, Dhaka 1342, Bangladesh. ✉email: mdmahin.stu2017@juniv.edu; amrin.stu2019@juniv.edu; rois@juniv.edu

Heatwaves have become increasingly severe and frequent worldwide, significantly impacting various regions. In 2020, people worldwide watched as Australia experienced an epic heatwave[15]. However, Australia wasn't the only place dealing with the heat. In other parts of the world, Northern New England and Canada were very warm. As summer approached, Siberia, an unexpected hotspot, reached 38 °C (100 °F)—an incredible temperature for one of the world's coldest places[16]. Across the Atlantic, Death Valley in the United States made headlines by reaching a scorching 54.4 °C (129.9 °F), the hottest temperature on Earth in 100 years[17]. In British Columbia, the town of Lytton reached 49.6 °C (121.3°F), setting a new national record in 2021[18]. Japan went through its worst heatwave in 150 years at the end of June 2022[19]. China dealt with its highest temperatures and one of its lowest rainfall amounts in 61 years during a two-month summer heatwave. This started forest fires, ruined crops, and messed with power supply[20]. Countries like India, China, Thailand, and Laos felt the heat in a big way[21]. Pakistan also had to deal with a heatwave in June 2024[22]. The Edhi Foundation reported more deaths than usual in Karachi from June 20 to June 25 during this heatwave[23]. A notable climatic phenomenon, the 2024 Bangladesh heatwave was characterized by exceptionally high temperatures of up to 42 °C, which is 6 degrees warmer than the annual average[24,25]. This intense heatwave is considered one of the most severe since records began in 1948 and has led to nationwide school closures, affecting children[24,26]. At least 15 people died from heat strokes between April 22 and May 5, according to the Directorate General of Health Services (DGHS)[24].

Consequently, heatwaves are becoming a major concern in Bangladesh, as evidenced by a number of recent investigations carried out there. For instance, Solomon (2007) predicts that Bangladesh's average annual temperature would rise by 1.4 ± 0.6 °C by the year 2050[27], Karmakar and Das found that heatwaves are frequently experienced in Bangladesh before the monsoon season[28], and two other studies anticipated that Bangladesh will experience more frequent and intense heatwaves in the future[29,30]. Various climate models and statistical models have also been employed in different studies to predict heatwaves in Bangladesh. For example, Nissan et al. (2017) investigated the climate dynamics linked to heat waves based on 35 weather stations across Bangladesh and created a heat early warning system (HEWS) using a generalized additive regression model[31], Rahman et al. used the Mann–Kendall and Sen's slope techniques to determine seasonal and annual temperature trend patterns of heat wave frequencies in Bangladesh[32], and Rashid et al. (2024) investigated the frequency of heatwaves at 34 meteorological stations in Bangladesh between 1990 and 2019 using linear trend analysis and the Mann–Kendall test[13]. According to the study[13], Rajshahi had the highest frequency of heatwaves (4.2333) during the pre-monsoon season, while Jashore had the most heatwave days (30.9 days) overall. Using remote sensing and socioeconomic data, some researchers created a heatwave vulnerability index (HVI) to examine the heatwave susceptibilities in Chattogram City Corporation, Bangladesh's main corporate hub[33].

Moreover, forecasting heatwave warnings is a complex task that combines meteorological data, climate models, and various statistical and machine learning (ML) models. Several researchers investigated forecasting heatwave warnings in different regions using a variety of climate models, statistical models, and ML models. For instance, Park and Kim (2018) employed Multivariate Adaptive Regression Splines (MARS) analysis to establish heatwave thresholds for developing an effective heatwave warning system[34], Anderson et al. (2011) used Bayesian hierarchical modeling to examine the mortality risk associated with heatwaves[35], Aboubakri et al. investigated the impact of heat waves on mortality and years of life lost in Iran employing distributed lag non-linear models (DLNM) and Poisson Regression model[36], Xu et al. conducted a meta-analysis to investigate the influence of heatwaves on mortality for various heatwave definitions[37], and Zhu et al. forecasted heatwaves using Spatial–Temporal Projection Models (STPM)[38]. Recent literature[39–42] on heatwave forecasting highlights advancements in machine learning (ML) models and statistical methods to improve prediction accuracy and timeliness. ML models have gained popularity for time series forecasting due to their ability to handle large datasets and reveal intricate patterns. For instance, Iqbal et al. utilized Classification and Regression Tree (CART) models for heatwave predictions in India using daily high summer temperatures[43], and Lee et al. used explainable artificial intelligence for forecasting heatwave in South Korea[41]. However, heat wave forecasting in Bangladesh has not been adequately done using ML-based time series forecasting. As per our knowledge, Sultana et al. were the only studies to use three different types of ML-models, for example, Long-Short Term Memory (LSTM), Bi-directional LSTM (Bi-LSTM), and Convolutional Neural Network (CNN) models for forecasting heatwave in four heatwave-prone districts of Bangladesh: Dhaka, Rajshahi, Bogra and Dinajpur[44].

While ML models have been applied to heatwave prediction in Bangladesh, research suggests that further investigation and development in this area are warranted. Furthermore, several studies highlighted the advantages of hybrid models over traditional ML models for forecasting across various domains[45]. Additionally, seasonal patterns in higher frequency time series are usually more complex[61]. For instance, there might be weekly, monthly, and annual patterns in daily data, a higher frequency data. When it comes to heatwave forecasting, daily data has several advantages over lower frequency data (such as monthly or quarterly), particularly for short- to long-term predictions and capturing nuanced patterns. We were, therefore, motivated to use ML-based hybrid models to forecast daily heatwaves in a district in Bangladesh that is vulnerable to heatwaves. Rajshahi, a city in Bangladesh's northwest, is highly susceptible to summer heatwaves due to climate change and decreased water vapor[46]. Consequently, heatwave forecasting in Rajshahi, Bangladesh, is crucial due to early warning systems, public health interventions, and overall resilience to climate change. Hence, this study aims to develop ML-based hybrid models for effective daily heatwave forecasting in Rajshahi, essential not only for immediate public health and safety but also for long-term economic stability and climate resilience.

Materials and methods

Data source

The study reveals that Rajshahi experiences the highest frequency of heatwaves during the pre-monsoon season[13]. Figure 1 displays comparable information. Hence, the daily maximum temperature data for Rajshahi Station was collected for this study from the Bangladesh Meteorological Department (BMD). The data comprises 15,340

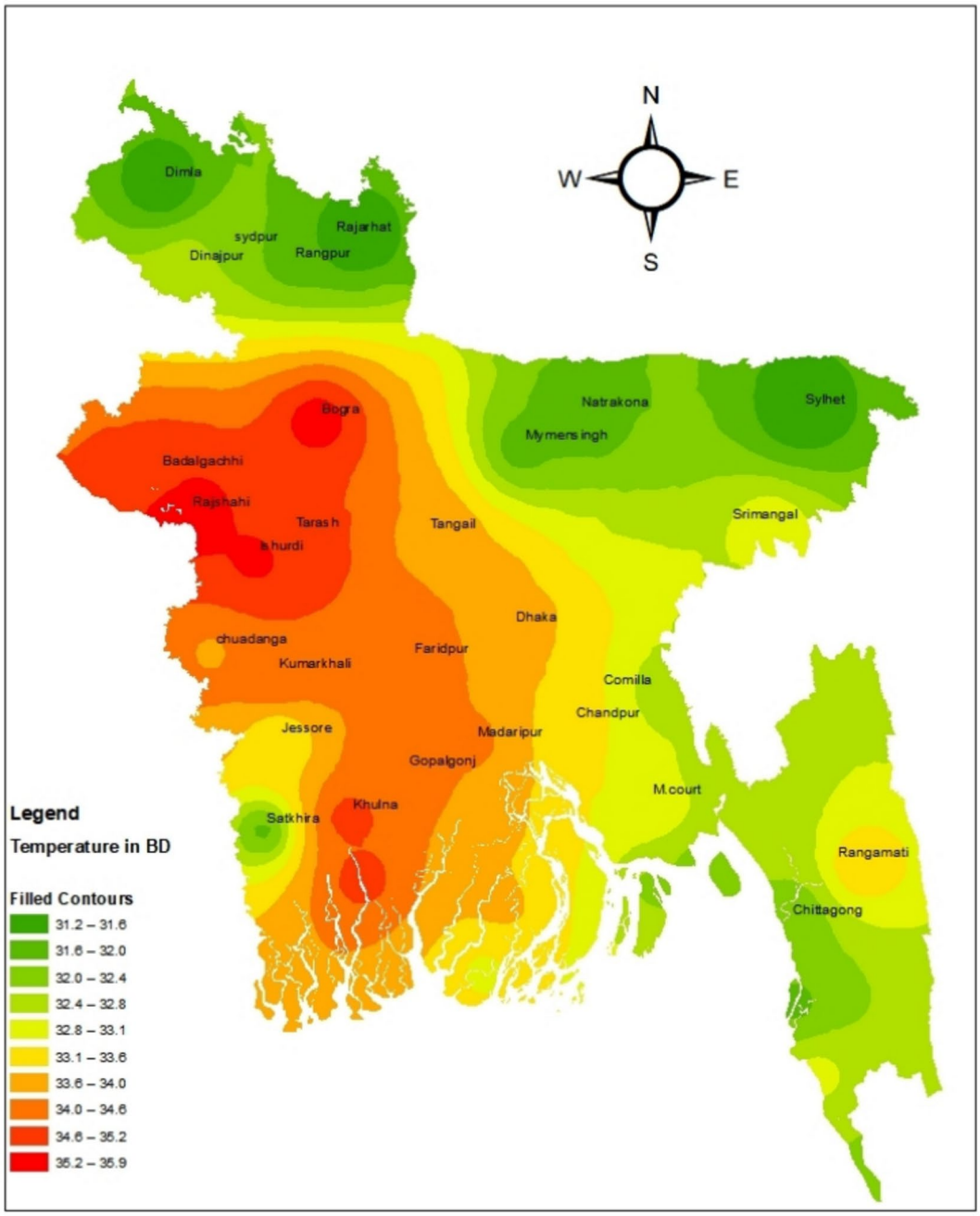

Fig. 1. Map of Bangladesh's highest temperatures by climatic zone, showing the locations of meteorological stations and study area.

observations from January 1981 to December 2022 after adjusting all missing values. A heatwave is defined in this study as a broad region that experiences temperatures above 36 degrees for at least three consecutive days[13]. We therefore intend to forecast heatwave warnings in Rajshahi based on the daily maximum temperature data from Rajshahi Station.

Accurate forecasting is essential for efficient planning and response tactics since heatwaves are a significant concern for human health and environmental sustainability. People may employ this prediction to warn of heatwaves. Therefore, to make more accurate forecasts about heatwave warning by using different time series models, i.e., AutoRegressive Integrated Moving Average (ARIMA), Seasonal AutoRegressive Integrated Moving Average (SARIMA), Exponential Smoothing State Space (ETS), and Trigonometric Box-Cox ARMA Trend Seasonal (TBATS); as well as different ML models, i.e., Artificial Neural Network (ANN), Support Vector Regression (SVR), Prophet, Random Forest Regression (RFR), and Long Short-Term Memory (LSTM); and by integrating the Seasonal-Trend decomposition procedure based on LOESS (STL) approach with different time series and ML models. All of the analysis for this investigation was conducted using R programming.

(S)ARIMA model

The Autoregressive Integrated Moving Average Model (ARIMA) is a predictive model that considers the temporal characteristics of the data and establishes a regression line based on past values, which directly impact future values[47]. It is one of the most well-known models for time series forecasting (TSF) because of its powerful capabilities[48]. In order to construct a composite model of the time series, this model integrates the Autoregressive (AR) and Moving Average (MA) processes[49].

The entire model ARIMA (p, d, q) can be specified as:[50]

$$(1-\varphi_1 B-\cdots-\varphi_p B^p)(1-B)^d y_t=(1+\theta_1 B+\cdots+\theta_q B^q)\varepsilon_t, \tag{1}$$

where $\{\varepsilon_t\}\sim WN\left(0,\sigma^2\right)$, p is the order of the autoregressive part, d is the degree of differencing, and q is the order of the moving average part. A seasonal ARIMA known as SARIMA model defined as $ARIMA(p,d,q)\times(P,D,Q)^s$ process with seasonal period *s*,

$$\phi(B)\,\phi(B^s)(1-B)^d(1-B^s)^D y_t=\theta(B)\,\Theta(B^s)\,\varepsilon_t, \tag{2}$$

where $\{\varepsilon_t\}\sim WN\left(0,\sigma^2\right)$, and the polynomial $\phi(B)$,$\Phi(B^s)$, $\theta(B)$, and $\Theta(B^s)$ are defined as $(1-\phi_1 B-\cdots-\phi_p B^p)$,$\left(1-\Phi_1 B^s-\cdots-\Phi_p B^{sP}\right)$, $(1+\theta_1 B+\cdots+\theta_p B^p)$, and $(1+\Theta_1 B^s+\cdots+\Theta_Q B^{sQ})$, respectively.

ETS model

By accounting for the seasonal and trend components, the ETS model is a method that can be employed to forecast univariate time series data[51]. The model being discussed is highly adaptable and possesses the capability to produce seasonal components for a wide range of properties and trends[52]. The model under consideration is characterized by three key factors, namely error, trend, and seasonal components. Each parameter has four values: A is additive, M is multiplicative, N is none, and Z is auto. Consequently, ETS (A, M, N) denotes the following: additive for error, multiplicative for trend, and no seasonality[53].

TBATS model

The TBATS model, also known as the Trigonometric Box-Cox ARMA Trend Seasonal model, is an enhanced version of the traditional Box-Cox ARMA model. It is specifically designed to handle complex seasonal patterns in time series data. It is particularly advantageous for series that are affected by diverse cyclical variables and have several and non-integer seasonal periods. TBATS forecasts complex seasonal trends using exponential averaging for time series data[54]. This BATS-based time series model has many adjustments. The seasonal functions are trigonometric[55,56]. This model was illustrated as follows[57]:

$$\begin{aligned} y_t &= q_{t-1}+r_{t-1}+v_t^{(1)}+v_t^{(2)}+e_t, \\ q_t &= q_{t-1}+r_{t-1}+\alpha e_t, \\ r_t &= r_{t-1}+\beta e_t, \\ v_t^{(1)} &= v_{t-m_1}^{(1)}+\gamma_1 e_t, \\ v_t^{(2)} &= v_{t-m_2}^{(2)}+\gamma_2 e_t, \end{aligned} \tag{3}$$

where $\{e_t\}\sim WN\left(0,\sigma^2\right)$, the seasonal cycles are denoted by m_1 and m_2, and the q_t and r_t are the 'level' and 'trend' components, respectively. The parameters of smoothing are represented by the coefficients α, β, γ_1 and γ_2. The ith component at the time t is $v_t^{(i)}$ and they are seasonal[52].

ANN model

ANNs are a highly successful alternative to ARIMA models for time series forecasting, with a multitude of distinguishing characteristics[58]. One important aspect is the universal approximation property, meaning an ANN can accurately estimate any nonlinear continuous function to any desired degree[59,60]. No prior assumptions regarding the model form are required during the model construction process; rather, the network model is primarily determined by the data's characteristics[61].

The model consists of three layers: an input layer, single hidden layer and an output layer. For time series forecasting, we have to obtain the relationship between the output y_t and the input $(y_{t-1},y_{t-2},\ldots,y_{t-k})$ which has the following mathematical representation[62]:

$$y_t=\alpha_0+\sum_{j=1}^{q}\alpha_j g\left(\beta_{0j}+\sum_{i=1}^{p}\beta_{ij}y_{t-1}\right)+\varepsilon_t, \tag{4}$$

where, $g(x)=\frac{1}{1+e^{-x}}$, α_j represents the weight between jth hidden neuron and output neuron, α_0 represents output neuron bias, β_{0j} represents hidden neuron bias, β_{ij} represents weight between ith input and jth hidden neuron, p is inputs, and q is hidden neurons, also the sigmoid activation function is applied to the hidden layer.

SVR model

Derived from Support Vector Machines (SVM), Support Vector Regression (SVR) is a strong class of supervised machine learning intended especially for regression applications. SVR shows exceptional mastery in efficiently managing intricate, non-linear patterns that are often seen in time series data. The algorithm employs an ε-insensitive loss function that imposes penalties on predictions that deviate more than ε units from the desired output[63]. Furthermore, it shows resistance to outliers and can produce more consistent forecasts for future time periods[64].

Prophet model

Prophet stands out as a powerful forecasting tool specifically designed for time series data, developed by Facebook's Core Data Science team[65]. It excels in accurately predicting time series data with trends, seasonality, and holidays[66]. The model uses the formula[67]:

$$y\left(t\right)=g\left(t\right)+s\left(t\right)+h\left(t\right)+\in_t \tag{5}$$

where $y\left(t\right)$ is the predicted value determined a linear or logistic equation, $g\left(t\right)$ and $s\left(t\right)$ represent seasonality or time series based on yearly, monthly, daily or another period, $h(t)$ is the holiday outliers and $\in_t$ represents unexpected error. While holiday effects don't appear relevant for the daily maximum temperature dataset, the Prophet model was chosen for this study due to its ability to capture trends and complex seasonal patterns in time series data over extended periods. The Prophet model has proven to be effective in forecasting climatological aspects due to its capacity to handle complex seasonality patterns, interannual trends, temporal variations, long-term seasonality forecasting, and anomaly detection in the required computational time.[68].

RFR model

Random forests (RF) are an ensemble learning method for classification and regression that generates many randomized decision trees during training and then averages the results to make predictions[69]. Including lagged variables or trends as predictors helps RF to address temporal dependencies in time series analysis. The RFR algorithm for time series analysis is:

1. Generate a dataset D with lagged variables for a time series y_t i.e.,

   $D=(y_{t-1},y_{t-2},\ldots,y_{t-p})_{t=p+1}^{T}$, where p is the number of lags used as predictors.

2. Create training and test sets from the dataset D accordingly D_{train} and D_{test}.
3. From the training set D_{train}, draw B bootstrap samples $\{D_b\}_{b=1}^{B}$.
4. Grow a decision tree T_b for every bootstrap sample D_b.
5. Predict the response $(\widehat{y_t})$ by averaging the predictions from all B trees, for a given input $x=(y_{t-1},y_{t-2},\ldots,y_{t-p})$ in the test set,

$$\widehat{y_t}=\frac{1}{B}\sum_{b=1}^{B}T_b\left(x\right). \tag{6}$$

LSTM model

The LSTM is a variant of the RNN that uses extra features to help it memorize temporal sequences in data[49]. It is used in time series forecasting, particularly where long-range dependencies are involved. It features a memory cell that can store its state in time, which is essential for identifying or remembering pertinent information over quite wide windows. Three gate forms in LSTM help to control flow information. Forget gate (f_t) decides from the cell state (C_t) what to forget. The input gate (i_t) helps us decide which fresh data to put into the cell state $(C_t$). The output gate (o_t) computes at a given moment the elements to be finally preserved in every dimension. The equations are illustrated below[70]:

$$f_t=\sigma\left(W_f\bullet\left[h_{t-1},X_t\right]+b_f\right), \tag{7}$$

$$i_t=\sigma\left(W_i\bullet\left[h_{t-1},X_t\right]+b_i\right), \tag{8}$$

$$\widehat{C_t}=\tanh(W_C\bullet\left[h_{t-1},X_t\right]+b_C), \tag{9}$$

$$C_t=i_t\bullet\widehat{C_t}+f_t\bullet C_{t-1}, \tag{10}$$

$$o_t=\sigma\left(W_o\bullet\left[h_{t-1},X_t\right]+b_o\right), \tag{11}$$

$$h_t=o_t\bullet\tanh(C_t) \tag{12}$$

where, h_t and h_{t-1} denote the current and previous hidden states respectively. The weight matrices W_f, W_i, W_C and W_o correspond to the forget gate (f_t), input gate (i_t), cell state (C_t) and output gate (o_t) respectively. The biases associated with these gates are represented by b_f, b_i, b_C and b_o. X_t denotes the input at time step t. The hyperbolic tangent function $(\tanh)$, which maps values to the range $[-1,1]$, is used as the activation function. $\widehat{C_t}$ represents the new memory update vector, and σ denotes the sigmoid activation function.

STL decomposition

In time series analysis, decomposition techniques are essential instruments that aid in identifying the underlying component (such as trend, seasonality, and irregular fluctuations) in observable data. Cleveland et al. (1990) developed the STL method[71]. STL can manage any kind of seasonality, meaning that it can handle daily and sub-daily series data in addition to monthly and quarterly data, unlike the classical method, SEATS(Seasonal Extraction in ARIMA Time Series), and X11[67,72]. The daily maximum temperature data, which includes complex seasonal patterns including a combination of weekly, monthly, and annual patterns of seasonality, is the basis for heatwave forecasting. Consequently, we employed the STL approach for seasonal adjustment in our ML-based hybrid algorithm.

Proposed algorithms

The study develops two algorithms, for instance, Algorithm 1 for developing ML-based hybrid models and Algorithm 2 for developing ML-based seasonal adjusted hybrid models. We initially divided the daily maximum temperature data into 70% training data and 30% test data to execute both algorithms. The detailed procedures for developing ML-based hybrid models and ML-based seasonal adjusted hybrid models are described in Algorithms 1–2. We identify the days with heatwaves and calculate the anticipated frequency of heatwaves in the future using all of the traditional time series models, ML models, and our suggested hybrid models.

1: Fit a model (TS or ML) to the daily maximum temperature data. Hence, compute the residuals (R_{t1}) and forecasted values (F_{t1}).

2: Utilize the residuals (R_{t1}) to fit another model (ML or TS). Hence, also calculate the forecasted values (F_{t2}).

3: Aggregate all forecasted values (F_{t1}) and (F_{t2}) to get forecasts of the original daily maximum temperature.

- **TS means Time Series**

Algorithm 1. For developing an ML-based hybrid model to forecast the daily maximum temperature.

1: Use the STL-based seasonal decomposition to obtain the daily maximum temperature in seasonal (S_t) and seasonal-adjusted (SA_t) components.

2: Fit a model (TS or ML) to the seasonal-adjusted daily (SA_t) data. Hence, compute the residuals (R_{t1}) and forecasted values (F_{t1}).

3: Utilize the residuals (R_{t1}) to fit another model (ML or TS). Hence, also calculate the forecasted values (F_{t2}).

4: Combine forecasts of seasonal component (S_t) with all forecasted values (F_{t1}) and (F_{t2}) to get forecasts of the original daily maximum temperature.

- **TS means Time Series**

Algorithm 2: For developing a seasonal-adjusted ML-based hybrid model to forecast thedaily maximum temperature.

Evaluation metrics

The hybrid prediction models' performances were evaluated using Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE), Mean Absolute Scaled Error (MASE), and Root Mean Squared Error (RMSE) metrics. Each measure was calculated using the following formulas:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{13}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \times 100\%, \tag{14}$$

$$\mathrm{MASE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|y_i - \hat{y}_i|}{\frac{1}{n-m}\sum_{j=m+1}^{n}|y_j - y_{j-m}|}\right), \tag{15}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{16}$$

where n indicates the number of observations, y_i denotes the original values, and $\hat{y}_i$ represents the predicted values, and m is the seasonality value. MAE, and RMSE are scale-dependent metrics based on absolute errors and squared errors, respectively. MAPE is a unit-free error measure based on percentage errors and MASE is a scale-free error metric.

Results

Data description

One of Bangladesh's warmest districts is Rajshahi, which is situated in the country's northwest. For this study, the daily maximum temperature in Rajshahi from January 1981 to December 2022 was gathered from the Bangladesh Meteorological Department (BMD). Figure 2 offers insights into the structure and underlying patterns of the trend as well as seasonal adjustments and the behavior of the maximum temperature throughout time. Rajshahi's historical data, as shown in Fig. 2, also reveals that during a 42-year period, there were 2037 days with a maximum temperature of 36 degrees or higher, leading to 224 heatwave occurrences, and no notable upward or downward trend features. The green line in Fig. 2, which represents seasonal-adjusted maximum temperature data, suggests that the seasonal-adjusted data is smoother than the original data. Therefore, Rajshahi's daily maximum temperature data fluctuated considerably due to the complex seasonality. To understand more about this complex seasonality in the daily maximum temperature data, we constructed Figs. 3 and 4 to examine an annual seasonal pattern and Figs. 4, 5 and 6 to investigate a monthly seasonal pattern.

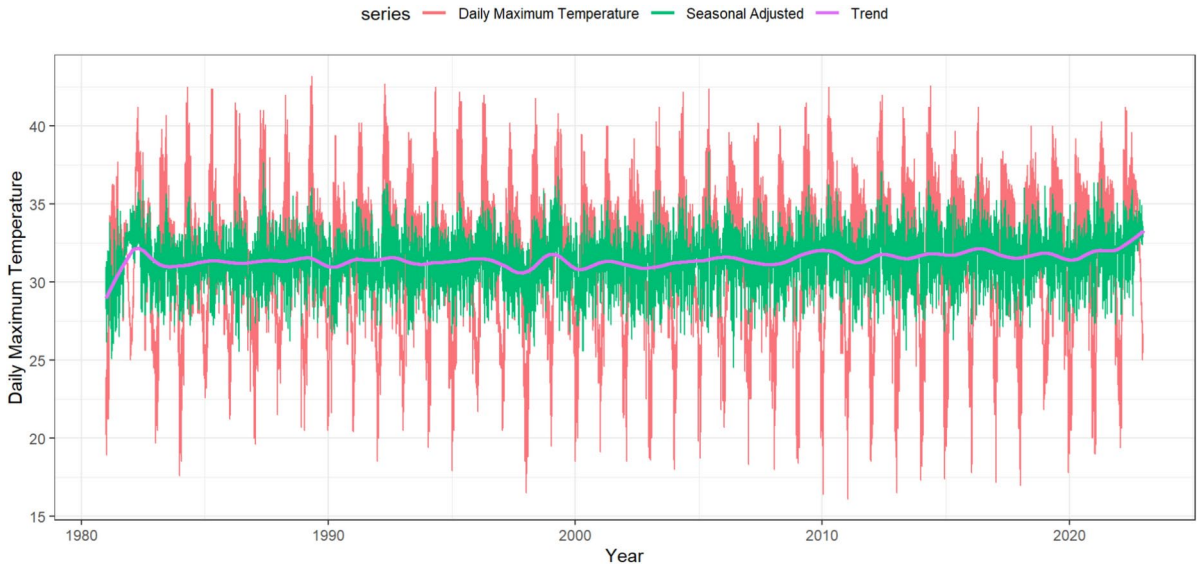

Fig. 2. Daily maximum temperature in Rajshahi with seasonal-adjusted maximum temperature (green) and trends (purple) revealed for January 1981 to December 2022.

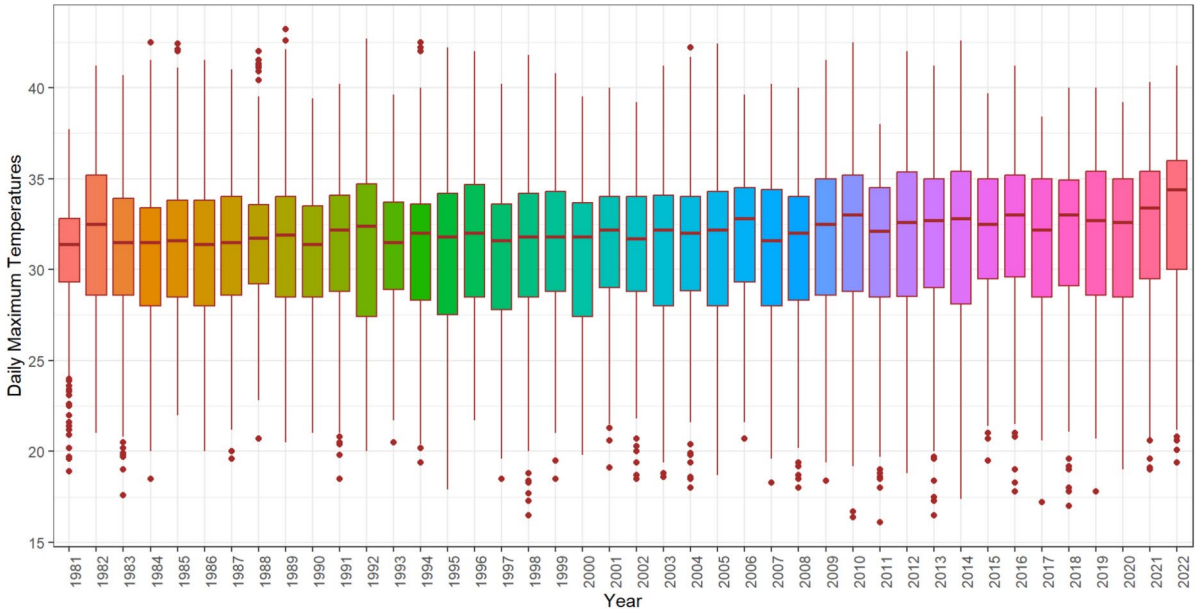

Fig. 3. Boxplot illustrates the distribution of yearly seasonal maximum temperature variation in Rajshahi from January 1981 to December 2022.

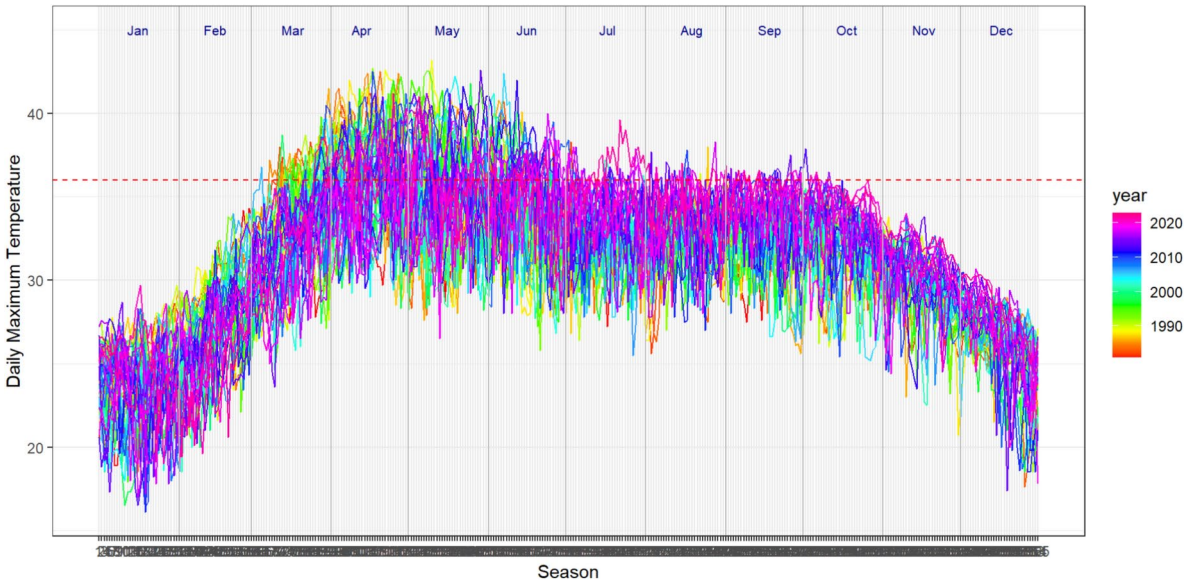

Fig. 4. Year-wise representation of daily maximum temperature with heatwave days (crossing the dotted red line) in Rajshahi from January 1981 to December 2022.

Figure 3 presents a box plot of daily maximum temperatures in Rajshahi from 1981 to 2022, revealing a subtle warming trend over a 42-year period. The median temperatures seem to be rising steadily from the previous year to the current one. Rajshahi's maximum temperatures vary greatly both within and between years, some years have more compact distributions of temperatures, while others show a greater range. Extreme temperatures are more common in earlier years, with the highest recorded temperatures reaching around 45 °C in 1989. In the most recent decade (2012–2022), there appears to be a slight increase in the median daily maximum temperatures in Rajshahi, but potentially less variability in the lower range. Hence, Fig. 3 effectively shows both the long-term warming trend and the annual seasonal pattern (i.e., year-to-year variability) in daily maximum temperatures over four decades.

Figure 4 demonstrates a monthly seasonal pattern and an annual seasonal pattern in daily maximum temperatures in Rajshahi, with the lowest in winter (December–January) and the highest in summer (April–May). The daily maximum temperature fluctuates between 20 and 40 °C throughout the year. The hottest period

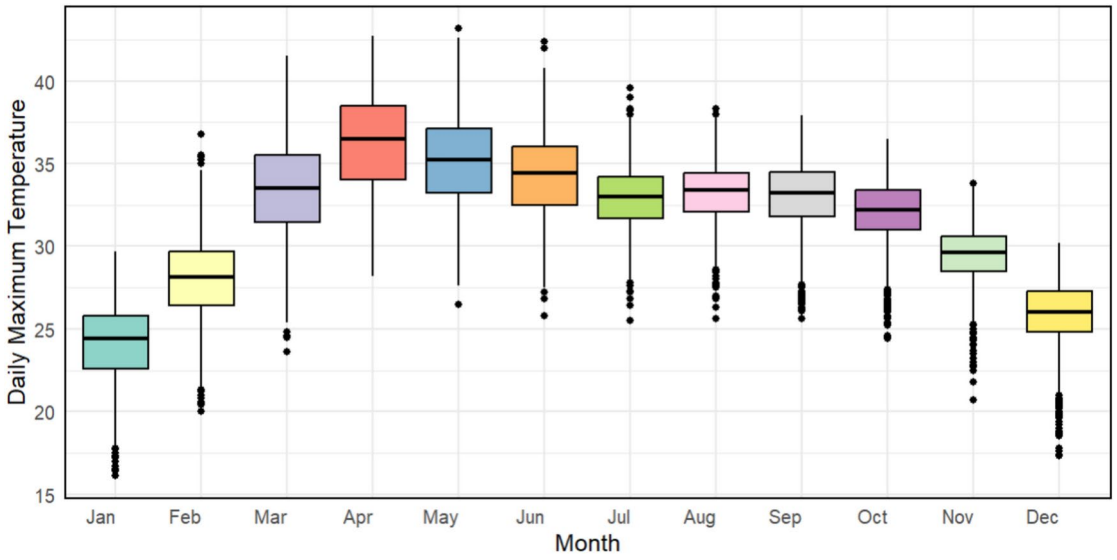

Fig. 5. Boxplot illustrates the distribution of monthly seasonal maximum temperature variation in Rajshahi from January 1981 to December 2022.

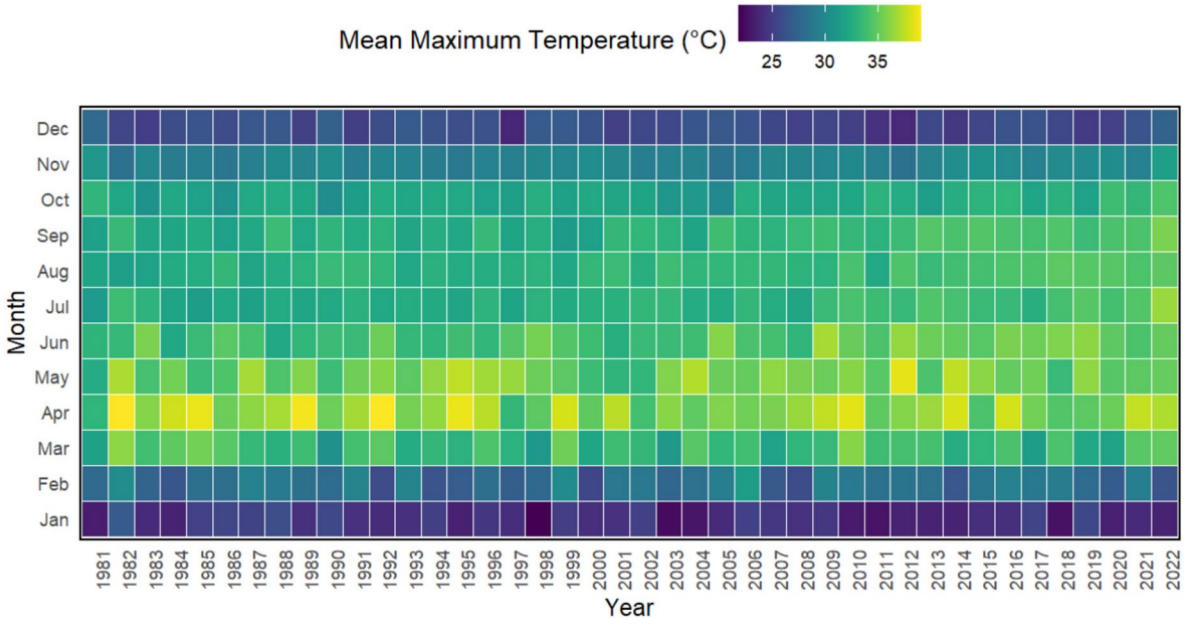

Fig. 6. Heat map illustrates the temporal analysis of monthly seasonal maximum temperature variation in Rajshahi from January 1981 to December 2022.

is from April to June, with many years reaching or exceeding 40 °C, and the coolest period is December and January, with some days dropping below 20 °C. While the long-term trends are difficult to identify in Fig. 4, year-to-year and month-by-month variabilities are significant. Moreover, Fig. 4 explores that the more recent years (e.g., 2020, shown in purple) consistently exhibit higher temperatures compared to earlier years (e.g., 1990, shown in orange/yellow). Furthermore, heatwaves were more frequent during the pre-monsoon season (March to May) and June in previous years, but in the most recent year, they have also been observed during the monsoon season (June to September). This information can be reflected more clearly in a month-wise representation (Figs. 5, 6).

The boxplot in Fig. 5 shows the seasonal variation of temperature trends, with lower temperatures in the early and late months of the year (January, February, November, December) and higher in the middle months (April, May, June). April has the highest median and overall highest temperatures. Therefore, Rajshahi's maximum temperature contains a very significant monthly seasonal pattern, which also reveals outliers with unusually high or low temperatures, providing insight into extreme weather events. In Fig. 6, each cell represents the mean maximum temperature for a specific month and year, with cooler temperatures in shades of blue/green and

warmer temperatures in shades of yellow. The annual cycle is visible, with January and December showing the lowest temperatures (around 25 °C). April and May have the highest temperatures, often reaching 36 °C. The temperatures gradually increase from January to April/May and then decrease from May/June to December (Fig. 6).

Consequently, the daily maximum temperature in Rajshahi, our study data, exhibits multiple seasonal patterns that combine monthly and annual seasonality. This complexity arises particularly in high-frequency data, such as daily or hourly observations, where both monthly and annual trends can significantly influence the outcomes. Accurate data modeling in time series analysis requires an understanding of the prevailing seasonal pattern. Hereafter, we refer to these multiple seasonal patterns as complex seasonality in this manuscript.

Heatwave forecasting without seasonal adjustment

We used the daily maximum temperatures data, a high-frequency time series data, to forecast heatwave warnings in Rajshahi, the hottest location in Bangladesh. Without seasonal adjustment, forecasting high-frequency time series is a difficult challenge[73]. To forecast heatwave warnings without seasonal adjustment in Rajshahi, we utilized two conventional time series models (ARIMA and ETS), in addition to the TBATS model, which is a time series model for high-frequency time series data; five machine learning models (ANN, Prophet, SVR, LSTM, and RFR); and thirty ML-based hybrid models using Algorithm 1 (ARIMA-ANN, ARIMA-Prophet, ARIMA-SVR, ARIMA-LSTM, ARIMA-RFR, ETS-ANN, ETS-Prophet, ETS-SVR, ETS-LSTM, ETS-RFR, TBATS-ANN, TBATS-Prophet, TBATS-SVR, TBATS-LSTM, TBATS-RFR, ANN-ARIMA, ANN-ETS, ANN-TBATS, Prophet-ARIMA, Prophet-ETS, Prophet-TBATS, SVR-ARIMA, SVR-ETS, SVR-TBATS, LSTM-ARIMA, LSTM-ETS, LSTM-TBATS, RFR-ARIMA, RFR-ETS, and RFR-TBATS). The forecasting performance of all these models based on different forecast accuracy metrics is outlined in Table 1.

Autocorrelation function (ACF) of errors at lag 1 (ACF1) for the fitted ANN, ETS, ARIMA, LSTM, TBATS, and Prophet models exhibited the highest correlation coefficients, which is a concerning observation. Conversely, the ACF1 values for the SVR (− 0.1667), ARIMA-LSTM (− 0.0012), ARIMA-RFR (− 0.0012), and TBATS-LSTM (− 0.0119) models were close to zero, suggesting minimal to no correlation. The ML-based hybrid TBATS-LSTM model's MAPE value is 4.0418, indicating that, on average, the predictions of this model deviate from the actual values by 4.0418%, which is generally considered a low error rate. Similarly, compared to all other models without seasonal adjustment based on MAE (1.2261), MAPE (4.0418), RMSE (1.6522) and MASE (0.5191), TBATS-LSTM produces more accurate results (Table 1).

Heatwave forecasting with seasonal adjustment

In order to anticipate heatwave warnings in Rajshahi, we seasonally adjusted the daily maximum temperatures data, a high-frequency time series data, using the STL decomposition approach. The STL decomposition method is more effective for seasonal adjustment of high-frequency time series data[74]. Our proposed Algorithm 2 utilizes the STL decomposition method to develop seasonal adjusted hybrid models. It identifies seasonal patterns in data, builds a hybrid model using seasonally adjusted data, and incorporates the actual seasonality of the test data into the final forecast. We employed the STL method with the two conventional time series models (STL-ARIMA and STL-ETS) in addition to the STL-TBATS model; five machine learning models (STL-ANN, STL-Prophet, STL-SVR, STL-LSTM, and STL-RFR); and thirty ML-based hybrid models using Algorithm 2 (STL-ARIMA-ANN, STL-ARIMA-Prophet, STL-ARIMA-SVR, STL-ARIMA-LSTM, STL-ARIMA-RFR, STL-ETS-ANN, STL-ETS-Prophet, STL-ETS-SVR, STL-ETS-LSTM, STL-ETS-RFR, STL-TBATS-ANN, STL-TBATS-Prophet, STL-TBATS-SVR, STL-TBATS-LSTM, STL-TBATS-RFR, STL-ANN-ARIMA, STL-ANN-ETS, STL-ANN-TBATS, STL-Prophet-ARIMA, STL-Prophet-ETS, STL-Prophet-TBATS, STL-SVR-ARIMA, STL-SVR-ETS, STL-SVR-TBATS, STL-LSTM-ARIMA, STL-LSTM-ETS, STL-LSTM-TBATS, STL-RFR-ARIMA, STL-RFR-ETS, and STL-RFR-TBATS) to forecast heatwave warnings with seasonal adjustment in Rajshahi. Table 2 illustrates each model's forecasting performance according to several forecast accuracy metrics.

In Table 2, our study reveals that the STL-TBATS-LSTM provides satisfactory results in terms of MAE (0.8954) and MASE (0.3806), but the STL-ARIMA-LSTM model is recommended because of its more dependable assessment metrics, such as MAE (0.8974), MAPE (2.9232), RMSE (1.1794), and MASE (0.3814). The ACF values at lag 1 for the STL-TBATS-LSTM and the STL-ARIMA-LSTM models are almost the same at 0.0018 and 0.0026, respectively. The seasonal adjusted ML-based hybrid model (the STL-ARIMA-LSTM with MAE (0.8974), MAPE (2.9232), RMSE (1.1794), MASE (0.3814) and ACF1 (0.0026) in Table 2) outperforms the ML-based hybrid model (the TBATS-LSTM with MAE (1.2261), MAPE (4.0418), RMSE (1.6522), MASE (0.5191) and ACF1 (-0.0119) in Table 1) in forecasting heatwave warnings in Rajshahi. The STL-ARIMA-LSTM model provides the most accurate forecast results, and adjusting for seasonality yields superior results when forecasting daily data. The results from using a single LSTM model were not adequate, but when integrated with a time series model, it produces the most favorable outcomes for the analysis.

The STL-ARIMA-LSTM hybrid model, based on our proposed Algorithm 2, is particularly well-suited for time series with multiple layers of complexity, such as seasonality, trends, and non-linearities. This model can handle them better than any hybrid models that rely on either traditional or machine learning models alone. Moreover, the ML-based hybrid model, TBATS-LSTM, based on our proposed Algorithm 1, skips decomposition altogether. STL's ability to isolate and focus on different components of the time series allows ARIMA and LSTM to work more effectively, leading to better predictive accuracy. The combination of STL, ARIMA, and LSTM utilizes each model's strengths in a complementary manner. STL removes noise and outliers, while ARIMA focuses on short-term linear relationships. LSTM handles non-linear aspects in residuals. Therefore, the significant outperformance of the STL-ARIMA-LSTM hybrid model is driven by its ability to combine the strengths of these three powerful techniques.

Models	MAE	MAPE	RMSE	MASE	ACF1
ARIMA	3.7371	12.6045	4.5179	1.5884	0.9174
ETS	3.8879	14.3184	5.3327	1.6525	0.9178
TBATS	1.7871	5.9380	2.3170	0.7596	0.6863
ANN	4.7405	18.6282	7.0359	2.0149	0.9628
Prophet	1.7785	6.0067	2.3449	0.7559	0.6784
SVR	1.2939	4.2495	1.7719	0.5500	− 0.1667
LSTM	3.5726	12.6656	4.5644	1.5605	0.9120
RFR	1.3142	4.2605	1.7157	0.5586	0.1156
ARIMA-ANN	3.9762	12.9619	4.6549	1.6900	0.9171
ARIMA-Prophet	3.6443	12.2805	4.4128	1.5490	0.9134
ARIMA-SVR	3.7440	13.0124	4.8056	1.5913	0.9133
ARIMA-LSTM	1.2710	4.1683	1.7053	0.5390	− 0.0012
ARIMA-RFR	1.3620	4.5076	1.7854	0.5789	− 0.0012
ETS-ANN	3.8125	14.0055	5.2174	1.6204	0.9164
ETS-Prophet	3.8930	14.3324	5.3368	1.6546	0.9180
ETS-SVR	4.6176	17.3744	6.8583	1.9626	0.9181
ETS-LSTM	1.3084	4.2991	1.7644	0.5561	0.0509
ETS-RFR	1.3872	4.5143	1.7880	0.5896	− 0.0189
TBATS-ANN	1.7889	5.9320	2.3119	0.7603	0.6862
TBATS-Prophet	1.7626	5.8684	2.2926	0.7492	0.6789
TBATS-SVR	1.7574	5.9174	2.3725	0.7470	0.6817
TBATS-LSTM	1.2261	4.0418	1.6522	0.5191	− 0.0119
TBATS-RFR	1.4028	4.6394	1.7940	0.5789	0.0922
ANN-ARIMA	4.7180	18.6088	7.0259	2.0053	0.9628
ANN-ETS	5.6003	17.9869	6.8564	2.3803	0.9614
ANN-TBATS	4.7175	18.5433	6.9995	2.0051	0.9626
Prophet-ARIMA	1.7780	6.0053	2.3446	0.7557	0.6783
Prophet-ETS	1.8823	6.0748	2.3339	0.8000	0.6784
Prophet-TBATS	1.8168	6.1839	2.4113	0.7722	0.6785
SVR-ARIMA	1.2981	4.2508	1.7719	0.5518	− 0.1667
SVR-ETS	1.2980	4.2507	1.7719	0.5517	− 0.1667
SVR-TBATS	1.2970	4.2785	1.7803	0.5513	− 0.1667
LSTM-ARIMA	3.6866	12.7048	4.5643	1.5554	0.9119
LSTM-ETS	5.9787	21.5249	7.3941	2.5411	0.9167
LSTM-TBATS	6.3904	20.8667	7.2358	2.6520	0.9120
RFR-ARIMA	1.3029	4.2386	1.7108	0.5538	0.1156
RFR-ETS	1.2900	4.2341	1.7154	0.5483	0.1197
RFR-TBATS	1.2869	4.2104	1.7056	0.5470	0.1156

Table 1. Forecast accuracy measures for the single time series and ML models, and ML-based hybrid models to forecast heatwaves in Rajshahi.

Figure 7 shows the forecasted heatwaves in Rajshahi using the STL-ARIMA-LSTM hybrid model from January 2023 to December 2026 using the developed hybrid model based on the entire dataset. Following the prediction of the daily maximum temperature from January 2023 to December 2026, we gathered the actual highest temperature data for Rajshahi from January 2023 to June 2024. This was done to compare and contrast the observed data with our previously forecasted data. Figure 8 illustrates the disparity between the observed data (which was not involved in any modelling) and the predicted data, which was trained on a 42-year historical observation.

Table 3 highlights Rajshahi's early warning of heatwaves based on the STL-ARIMA-LSTM model, a seasonal adjusted ML-based hybrid model. which indicates that Rajshahi will encounter 32 heatwaves between July 2024 and December 2026. Based on our developed Algorithm 2, the proposed hybrid model, STL-ARIMA-LSTM, ensures a more accurate forecast and helps to determine the number and days of heatwaves, allowing people to plan ahead and take necessary precautions before the onset of heatwaves.

Discussion

Global warming has led to a five-fold increase in record-breaking heat extremes over the past century, causing more extreme weather events like heatwaves[2–6]. These events have major societal and environmental

Models	MAE	MAPE	RMSE	MASE	ACF1
STL-ARIMA	1.2835	4.1158	1.6185	0.5455	0.6478
STL-ETS	1.7923	5.9974	2.2109	0.7618	0.6468
STL-TBATS	1.3314	4.4355	1.7181	0.5659	0.6468
STL-ANN	2.4774	7.6803	2.7918	1.0530	0.8889
STL-Prophet	1.3125	4.3645	1.7005	0.5578	0.6478
STL-SVR	0.9126	2.9625	1.1948	0.3879	− 0.0809
STL-LSTM	1.2261	4.0348	1.5905	0.5211	0.6433
STL-RFR	0.9164	2.9899	1.1966	0.3895	0.0932
STL-ARIMA-ANN	1.3441	4.3222	1.6922	0.5713	0.5807
STL-ARIMA-Prophet	1.2603	4.0527	1.5957	0.5357	0.6467
STL-ARIMA-SVR	1.2832	4.1209	1.6467	0.5454	0.6607
STL-ARIMA-LSTM	0.8974	2.9232	1.1794	0.3814	0.0026
STL-ARIMA-RFR	31.7401	102.0559	31.7282	13.4907	− 0.0192
STL-ETS-ANN	1.8583	6.2013	2.3004	0.7898	0.6256
STL-ETS-Prophet	1.7940	6.0028	2.2123	0.7625	0.6471
STL-ETS-SVR	1.8027	6.0464	2.2979	0.7662	0.6718
STL-ETS-LSTM	0.9604	3.1285	1.2606	0.4082	0.0688
STL-ETS-RFR	0.9528	3.0944	1.2432	0.4050	− 0.0326
STL-TBATS-ANN	1.4242	4.6452	1.7941	0.6054	0.5760
STL-TBATS-Prophet	1.3370	4.4555	1.7247	0.5683	0.6469
STL-TBATS-SVR	1.3814	4.6121	1.8102	0.5871	0.6737
STL-TBATS-LSTM	0.8954	2.9236	1.1795	0.3806	0.0018
STL-TBATS-RFR	0.9629	2.9860	1.1991	13.4907	− 0.0076
STL-ANN-ARIMA	2.4588	7.6356	2.7463	1.0451	0.8889
STL-ANN-ETS	2.3379	7.6116	2.7373	0.9937	0.8881
STL-ANN-TBATS	2.4588	7.6337	2.7465	1.0451	0.8889
STL-Prophet-ARIMA	1.3128	4.3658	1.7009	0.5580	0.6470
STL-Prophet-ETS	2.3651	7.8709	2.7549	1.0053	0.6478
STL-Prophet-TBATS	1.5890	5.3107	1.9979	0.6754	0.6476
STL-SVR-ARIMA	0.9171	2.9739	1.1981	0.3898	− 0.0809
STL-SVR-ETS	0.9104	2.9836	1.2037	0.3869	− 0.0809
STL-SVR-TBATS	0.9087	2.9765	1.2015	0.3862	− 0.0809
STL-LSTM-ARIMA	5.6591	119.5562	37.0886	2.4053	0.6417
STL-LSTM-ETS	1.3368	4.4542	1.7237	0.5682	0.6463
STL-LSTM-TBATS	4.6027	114.2280	35.4338	1.9563	0.6433
STL-RFR-ARIMA	0.9164	2.9901	1.1966	0.3895	0.0932
STL-RFR-ETS	0.9224	3.0048	1.2034	0.3920	0.1047
STL-RFR-TBATS	0.9113	2.9766	1.1927	0.3873	0.0932

Table 2. Forecast accuracy measures for the seasonal-adjusted single time series and ML models, and ML-based hybrid models to forecast heatwaves in Rajshahi.

consequences, affecting human health, agriculture, economy, natural disasters, and ecosystems[12]. Heatwaves have become a significant part of our world's changes, with extreme temperatures affecting various regions. For instance, Australia experienced an epic heatwave in 2020[15], while Northern New England and Canada experienced extreme temperatures[16]. Siberia reached 38 °C[16], Death Valley in the US reached 54.4 °C[17], and Lytton in British Columbia set a new national record[18]. Japan experienced its worst heatwave in 150 years[19], while China experienced its highest temperatures and lowest rainfall in 61 years[20]. Countries like India, China, Thailand, and Laos also experienced heatwaves[21], with Pakistan reporting more deaths than usual in Karachi[22,23]. Bangladesh also faces climate change concerns due to increasing heatwaves[29,30], particularly in Rajshahi, a city in northwest Bangladesh. The city is particularly vulnerable to summer heatwaves due to decreased water vapor and climate change susceptibility[46]. Therefore, global organizations and climate experts are deeply concerned about the growing effects of global warming and recent heatwaves.

Moreover, heatwaves are an important issue of environmental sustainability and public health[12], making accurate forecasting crucial for effective planning and response strategies. However, anticipating heatwave warnings requires handling the daily time series data, which is a high-frequency time series data. High-frequency time series data forecasting presents unique challenges due to its inherent complexity and characteristics[73]. Nevertheless, ML models have become popular for time series forecasting because of their ability to handle large datasets and reveal complex patterns. Several studies on heatwave forecasting emphasize advances in ML models

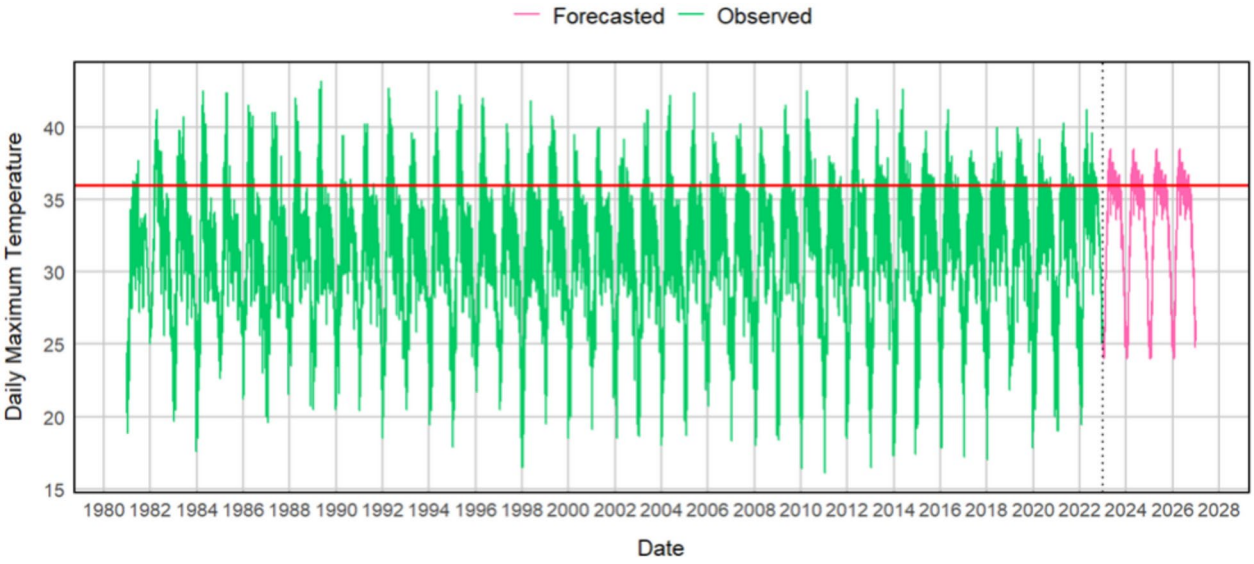

Fig. 7. Forecasted heatwave warnings from January 2023 to December 2026 in Rajshahi using the STL-ARIMA-LSTM model based on the entire dataset.

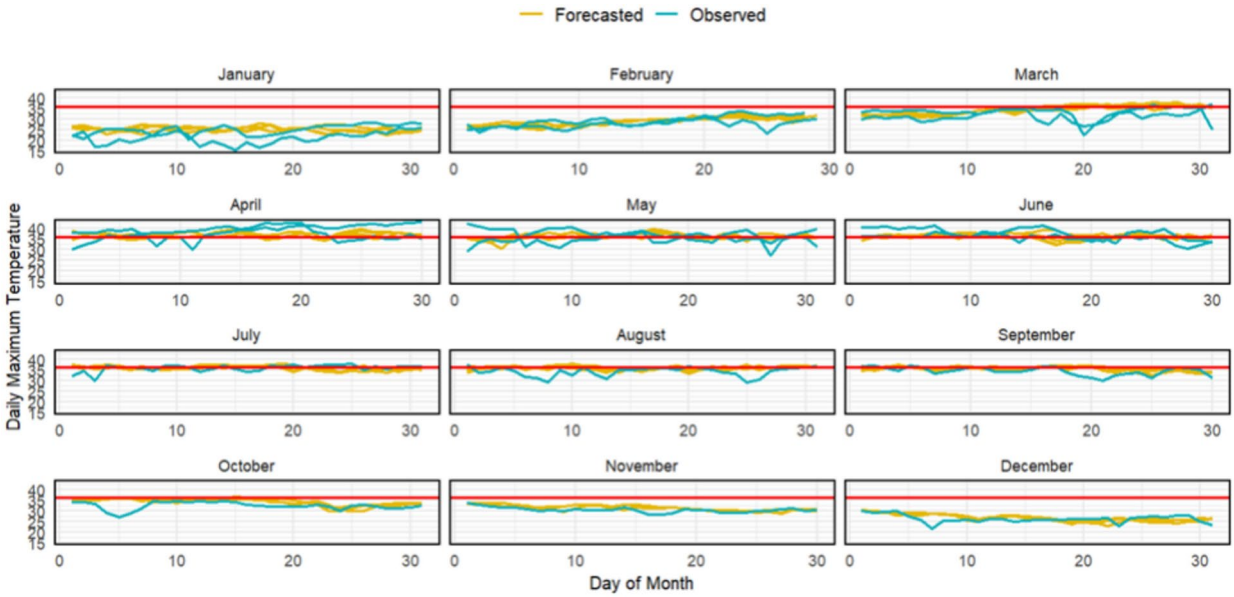

Fig. 8. Month-wise plot between actual and forecasted daily maximum temperature trend of STL-ARIMA-LSTM model from January 2023 to December 2026.

to improve prediction accuracy and timeliness. For instance, extreme heatwaves were successfully forecasted using Convolutional Neural Networks (CNNs)[42], the LSTM-based model was used on enhancing short-term temperature prediction in Korea[75], and the RFR model was employed to forecast Central European summer heat waves[76]. However, some research show that hybrid models outperform traditional ML models in a range of prediction domains[45,77]. As a result, we were inspired to design the ML-based hybrid models using Algorithm 1. Conversely, high-frequency time series forecasting is challenging in the absence of seasonal adjustment[73]. For seasonal adjustment in high-frequency time series data, the STL decomposition method works better[74]. Hence, we were also motivated to design Algorithm 2 to build seasonal adjusted ML-based hybrid models to forecast high-frequency TS observations.

Despite many advancements made in ML models, the study of ML models for weather prediction in Bangladesh is still limited. Hence, we developed thirty ML-based hybrid models to forecast heatwaves in Rajshahi using our proposed Algorithm 1, including ARIMA-ANN, ARIMA-Prophet, ARIMA-SVR, ARIMA-LSTM, ARIMA-RFR, ETS-ANN, ETS-Prophet, ETS-SVR, ETS-LSTM, ETS-RFR, TBATS-ANN, TBATS-Prophet, TBATS-SVR, TBATS-LSTM, TBATS-RFR, ANN-ARIMA, ANN-ETS, ANN-TBATS, Prophet-ARIMA,

Heatwave warning	Start date	End date	Heatwave warning	Start date	End date
1	07/11/2024	07/18/2024	17	08/26/2025	08/30/2025
2	08/06/2024	08/08/2024	18	09/14/2025	09/16/2025
3	08/26/2024	08/30/2024	19	03/19/2026	03/23/2026
4	09/14/2024	09/16/2024	20	03/25/2026	03/29/2026
5	03/19/2025	03/23/2025	21	04/05/2026	04/19/2026
6	03/25/2025	03/29/2025	22	04/22/2026	04/29/2026
7	04/05/2025	04/19/2025	23	05/04/2026	05/06/2026
8	04/22/2025	04/29/2025	24	05/08/2026	05/10/2026
9	05/04/2025	05/06/2025	25	05/13/2026	05/26/2026
10	05/08/2025	05/10/2025	26	05/31/2026	06/09/2026
11	05/13/2025	05/26/2025	27	06/11/2026	06/15/2026
12	05/31/2025	06/09/2025	28	06/22/2026	06/24/2026
13	06/11/2025	06/15/2025	29	07/11/2026	07/18/2026
14	06/22/2025	06/24/2025	30	08/06/2026	08/08/2026
15	07/11/2025	07/18/2025	31	08/26/2026	08/30/2026
16	08/06/2025	08/08/2025	32	09/14/2026	09/16/2026

Table 3. Forecasted heatwave warnings in Rajshahi from July, 2024 to December, 2026 using the STL-ARIMA-LSTM model.

Prophet-ETS, Prophet-TBATS, SVR-ARIMA, SVR-ETS, SVR-TBATS, LSTM-ARIMA, LSTM-ETS, LSTM-ETS, LSTM-TBATS, RFR-ARIMA, RFR-ETS, and RFR-TBATS. The performance of these developed ML-based hybrid models was compared with that of all other traditional time series and ML models, including ARIMA, ETS, TBATS, ANN, Prophet, SVR, LSTM, and RFR, to forecast heatwave warnings without seasonal adjustment in Rajshahi. Our study findings reveal that the TBATS-LSTM model performed better than all other models with MAE (1.2261), MAPE (4.0418), RMSE (1.6522), MASE (0.5191) and ACF1 (– 0.0119) in Table 1.

We also utilized our efforts to develop seasonal adjusted ML-based hybrid models based on our suggested Algorithm 2 to forecast heatwave warnings with seasonal adjustment in Rajshahi. We compared our developed seasonal adjusted ML-based hybrid models, including STL-ARIMA-ANN, STL-ARIMA-Prophet, STL-ARIMA-SVR, STL-ARIMA-LSTM, STL-ARIMA-RFR, STL-ETS-ANN, STL-ETS-Prophet, STL-ETS-SVR, STL-ETS-LSTM, STL-ETS-RFR, STL-TBATS-ANN, STL-TBATS-Prophet, STL-TBATS-SVR, STL-TBATS-LSTM, STL-TBATS-RFR, STL-ANN-ARIMA, STL-ANN-ETS, STL-ANN-TBATS, STL-Prophet-ARIMA, STL-Prophet-ETS, STL-Prophet-TBATS, STL-SVR-ARIMA, STL-SVR-ETS, STL-SVR-TBATS, STL-LSTM-ARIMA, STL-LSTM-ETS, STL-LSTM-TBATS, STL-RFR-ARIMA, STL-RFR-ETS, and STL-RFR-TBATS, with seasonal adjusted time series models (STL-ARIMA, STL-ETS, and STL-TBATS) and seasonal adjusted ML models (STL-ANN, STL-Prophet, STL-SVR, STL-LSTM, STL-RFR). Our research results indicate that the STL-ARIMA-LSTM model outperformed all other models with MAE (0.8974), MAPE (2.9232), RMSE (1.1794), MASE (0.3814) and ACF1 (0.0026) in Table 2.

Therefore, the seasonal adjusted ML-based hybrid model based on Algorithm 2, STL-ARIMA-LSTM, performs better than the ML-based hybrid model, TBATS-LSTM, using Algorithm 1 in forecasting heatwave warnings in Rajshahi. The LSTM model by itself was unable to generate adequate forecast results. Nevertheless, when LSTM was integrated with time series models, it became highly effective. Furthermore, by combining ML-based hybrid models with the STL decomposition method, we discovered an even better performing model that forecasts heatwave warnings more accurately. Importantly, the top-performing hybrid models consistently incorporated LSTM in our analysis. Table 3 shows Rajshahi's early warning of 32 heatwaves between July 2024 and December 2026 using the STL-ARIMA-LSTM hybrid model. The significant outperformance of the STL-ARIMA-LSTM hybrid model is driven by its ability to combine the strengths of three powerful techniques. STL decomposes the data effectively, ARIMA handles linear components, and LSTM captures non-linear dependencies and long-term relationships. This synergy allows the model to outperform other hybrids that may not integrate these elements as effectively, resulting in better predictive accuracy and robustness, especially for complex time series data. Therefore, our suggested seasonal adjusted ML-based hybrid model, using Algorithm 2, ensures a more accurate forecast and helps to determine the number and days of heatwaves, enabling people to plan ahead and take necessary safety measures before they occur.

Heatwave predictions are influenced by a variety of external factors that can complicate forecasting efforts. These factors include atmospheric and oceanic phenomena, jet stream patterns, volcanic eruptions, wildfires, climate change, technological limitations, human activity, geopolitical or economic events, and air pollution[42,78–81] Climate events like El Niño or La Niña can significantly alter global temperature patterns, leading to deviations from typical trends[42,80]. Unexpected atmospheric phenomena like temperature variations, cold fronts, or changes in cloud cover can cause immediate heatwaves that are difficult to predict accurately[42,79–81] Technological limitations, such as malfunctioning satellite observations or lack of data from weather stations, can also lead to inaccurate predictions[42,78]. Human activity, such as urbanization, deforestation, and land use changes, can also affect local climates and pollution levels[79,81]. Geopolitical or economic events, such as industrial shutdowns, can also lead to shifts in atmospheric conditions[79,80]. Each of these elements plays a significant role in shaping global

weather patterns. Although a variety of external factors or events can influence the accuracy of temperature forecasts, our study forecasts heatwave alerts using univariate time series analysis. Because of this, multivariate time series analysis, which is one of our study's future research areas, can account for any external factors or events. The study acknowledges this as a significant limitation and intends to consider this fact in future research.

The 2024 Bangladesh heatwave, which began in early April and persisted until May 5, is considered one of the most severe since records began in 1948 and has led to nationwide school closures, affecting children[42,78]. According to DGHS, at least 15 people died from heat strokes between April 22 and May 5 in 2024[42]. BMD issued heat alerts and implemented measures to mitigate the impact on vulnerable populations. The government generated a record 16,477 MW of electricity to meet the increased demand during the heatwave[25,42]. Therefore, heatwaves are increasingly becoming a significant concern in Bangladesh due to climate change, leading to rising temperatures and associated health risks. Accurate heatwave forecasting is crucial for early warning systems, public health interventions, and disaster preparedness strategies, reducing heat-related mortality risk through modeling and evaluation of warnings. Effective heatwave forecasting in Rajshahi is essential not only for immediate public health and safety but also for long-term economic stability and climate resilience.

Our study findings revealed that Rajshahi's heatwaves were more frequent during the pre-monsoon season (March to May) and June in previous years. This is consistent with Karmakar and Das's (2020) findings that heatwaves are common in Bangladesh before the monsoon season[28]. However, our research also revealed that in the most recent year, heatwaves were also recorded during the monsoon season, which runs from June to September. Two previous studies anticipated that Bangladesh will experience more frequent and intense heatwaves in the future[29,30]. Our research confirms these findings as well. However, Bangladesh has already implemented early action protocols due to heatwave forecasting, highlighting the increasing impact of climate change on the country[82,83]. The Bangladesh Red Crescent Society activated its Early Action Protocol (EAP) for Dhaka, anticipating temperatures exceeding 38 degrees Celsius and high humidity pushing the heat index over 38 for two or more consecutive days[83]. The government has ordered the closure of schools and colleges in response to the heatwave, disrupting people's lives and livelihoods, especially for low-income groups[82,83]. Our research highlighted that future vulnerable heatwaves will affect Rajshahi more frequently. This triggers actions such as awareness campaigns, water bottle distribution, public announcements, and cash assistance to vulnerable residents. Therefore, our proposed ML-based hybrid model and seasonal adjusted ML-based hybrid model can serve as a blueprint for predicting heatwaves in other regions with high susceptibility to such weather events. By prioritizing this capability, local authorities can better protect their communities from the growing threat of extreme heat.

Data availability

The datasets that support the findings of this study are available on request. All correspondence and requests for materials should be sent to MMUQ, ABA, or RR.

Received: 20 October 2024; Accepted: 5 March 2025
Published online: 13 March 2025

References

1. Smith, K. R. et al. Human health: impacts, adaptation, and co-benefits. In *Climate Change 2014: Impacts, Adaptation, and Vulnerability* 709–754 (Cambridge University Press, 2014).
2. Coumou, D., Robinson, A. & Rahmstorf, S. Global increase in record-breaking monthly-mean temperatures. *Clim. Change* **118**, 771–782 (2013).
3. Rahmstorf, S. & Coumou, D. Increase of extreme events in a warming world. *Proc. Natl. Acad. Sci.* **108**, 17905–17909 (2011).
4. Hegerl, G. C. et al. Causes of climate change over the historical record. *Environ. Res. Lett.* **14**, 123006 (2019).
5. Schär, C. et al. The role of increasing temperature variability in European summer heatwaves. *Nature* **427**, 332–336 (2004).
6. Argüeso, D., Di Luca, A., Perkins-Kirkpatrick, S. E. & Evans, J. P. Seasonal mean temperature changes control future heat waves. *Geophys. Res. Lett.* **43**, 7653–7660 (2016).
7. Perkins, S. E., Alexander, L. V. & Nairn, J. R. Increasing frequency, intensity and duration of observed global heatwaves and warm spells. *Geophys. Res. Lett.* **39**, (2012).
8. Meehl, G. A. & Tebaldi, C. More intense, more frequent, and longer lasting heat waves in the 21st century. *Science* **305**, 994–997 (2004).
9. Della-Marta, P. M., Haylock, M. R., Luterbacher, J. & Wanner, H. Doubled length of western European summer heat waves since 1880. *J. Geophys. Res. Atmos.* **112**, (2007).
10. Cowan, T. et al. More frequent, longer, and hotter heat waves for australia in the twenty-first century. *J. Clim.* **27**, 5851–5871 (2014).
11. Russo, S. *et al.* Magnitude of extreme heat waves in present climate and their projection in a warming world. *J. Geophys. Res. Atmos.* **119**, (2014).
12. Trancoso, R. et al. Heatwaves intensification in Australia: A consistent trajectory across past, present and future. *Sci. Total Environ.* **742**, 140521 (2020).
13. Rashid, G. M., Hossain, M. M. T., Akhter, M. A. E. & Mallik, M. A. K. A study on the heat wave conditions over Bangladesh during 1990–2019. *J. Eng. Sci.* **14**, 59–67 (2024).
14. Heatwave. *World Meteorological Organization* https://wmo.int/topics/heatwave (2023).
15. NASA Earth Observatory. Spring heats up down under. https://earthobservatory.nasa.gov/images/147619/spring-heats-up-down-under.
16. It just hit 100 degrees Fahrenheit in Siberia, the hottest temperature on record so far north in the Arctic. *The Weather Channel* https://weather.com/news/climate/news/2020-06-21-siberia-russia-100-degrees-heat-record-arctic (2020).
17. Wikipedia contributors. Death Valley. *Wikipedia* https://en.wikipedia.org/wiki/Death_Valley (2024).
18. Duff, R. 116-degree temperature marks city's third consecutive all-time record. *AccuWeather* (2021).
19. BBC News. Japan swelters in its worst heatwave ever recorded. https://www.bbc.com/news/world-asia-61976937 (2022).
20. Yu, V. China reports 'most severe' heatwave and third driest summer on record. *The Guardian* (2022).
21. Ratcliffe, R. & Ellis-Petersen, H. Severe heatwave engulfs Asia causing deaths and forcing schools to close. *The Guardian* (2023).
22. Carbon Brief. Pakistan: More than 500 die in six days as heatwave grips country—Carbon Brief. *Carbon Brief* https://www.carbonbrief.org/daily-brief/pakistan-more-than-500-die-in-six-days-as-heatwave-grips-country (2024).

23. Bloomberg. https://www.bloomberg.com/news/articles/2024-06-27/karachi-sees-a-surge-in-deaths-as-heat-wave-sears-pakistan (2024).
24. Rahman, M. A. Record heatwave in Bangladesh make LNG, power generation surge. Gas Outlook. https://gasoutlook.com/analysis/record-heatwave-in-bangladesh-make-lng-power-generation-surge/ (2024).
25. Haider, Q. Why is Bangladesh in the grip of a sizzling heatwave? The Daily Star. https://www.thedailystar.net/opinion/views/news/why-bangladesh-the-grip-sizzling-heatwave-3595916 (2024).
26. Yett, S. United Nations Children's Fund, Children at high risk amid countrywide heatwave in Bangladesh, UNICEF, South Asia. https://www.unicef.org/rosa/press-releases/children-high-risk-amid-countrywide-heatwave-bangladesh (2024).
27. Solomon, S. *et al. Climate Change 2007: The Physical Science Basis*. (2007).
28. Karmakar, S. & Das, M. K. On the heat waves in Bangladesh, their trends and associated large scale tropospheric conditions. *J. Eng. Sci.* **11**, 19–36 (2020).
29. Huq, S. Climate change and Bangladesh. *Science* **294**, 1617 (2001).
30. Kirtman, B., Power, S.B., Adedoyin, A.J., Boer, G.J., Bojariu, R., Camilloni, I., Doblas-Reyes, F., Fiore, A.M., Kimoto, M., Meehl, G., Prather, M., Sarr, A., Schar, C., Sutton, R., van Oldenborgh, G.J., Vecchi, G., & Wang, H.-J. Chapter 11—Near-term climate change: Projections and predictability. In *Climate Change 2013: The Physical Science Basis. IPCC Working Group I Contribution to AR5*. (eds. IPCC) (Cambridge University Press, 2013).
31. Nissan, H., Burkart, K., Coughlan de Perez, E., Van Aalst, M. & Mason, S. Defining and predicting heat waves in Bangladesh. *J. Appl. Meteorol. Climatol.* **56**(10), 2653–2670 (2017).
32. Rahman, M. M. et al. Are hotspots and frequencies of heat waves changing over time? Exploring causes of heat waves in a tropical country. *PLoS One* **19**(5), e0300070 (2024).
33. Raja, D. R., Hredoy, M. S. N., Islam, Md. K., Islam, K. M. A. & Adnan, M. S. G. Spatial distribution of heatwave vulnerability in a coastal city of Bangladesh. *Environ. Chall.* **4**, 100122 (2021).
34. Park, J. & Kim, J. Defining heatwave thresholds using an inductive machine learning approach. *PLoS One* **13**, e0206872 (2018).
35. Anderson, G. B. & Bell, M. L. Heat Waves in the United States: Mortality Risk during Heat Waves and Effect Modification by Heat Wave Characteristics in 43 U.S. Communities. *Environ. Health Perspect.* **119**, 210–218 (2011).
36. Aboubakri, O., Khanjani, N., Jahani, Y. & Bakhtiari, B. The impact of heat waves on mortality and years of life lost in a dry region of Iran (Kerman) during 2005–2017. *Int. J. Biometeorol.* **63**, 1139–1149 (2019).
37. Xu, Z., FitzGerald, G., Guo, Y., Jalaludin, B. & Tong, S. Impact of heatwave on mortality under different heatwave definitions: A systematic review and meta-analysis. *Environ. Int.* **89–90**, 193–203 (2016).
38. Zhu, Z. & Li, T. Extended-range forecasting of Chinese summer surface air temperature and heat waves. *Clim. Dyn.* **50**, 2007–2021 (2018).
39. Suthar, G., Singh, S., Kaul, N., Khandelwal, S. & Singhal, R. P. Prediction of maximum air temperature for defining heat wave in Rajasthan and Karnataka states of India using machine learning approach. *Remote Sens. Appl. Soc. Environ.* **32**, 101048 (2023).
40. Leach, N. J. et al. Heatwave attribution based on reliable operational weather forecasts. *Nat. Commun.* **15**, 4530 (2024).
41. Lee, Y. et al. Unveiling teleconnection drivers for heatwave prediction in South Korea using explainable artificial intelligence. *NPJ Clim. Atmos. Sci.* **7**, 1–12 (2024).
42. Jacques-Dumas, V., Ragone, F., Borgnat, P., Abry, P. & Bouchet, F. Deep learning-based extreme heatwave forecast. *Front. Clim.* **4**, 789641 (2022).
43. Iqbal, G. M. D. *et al.* A supervised learning tool for heatwave predictions using daily high summer temperatures. *Expert Syst.* e13656.
44. Sultana, S.S., Mumu, S.S., Fahim, R.R., Neha, N.N. and Mahbub, M.S. Regional heatwave prediction using deep learning models in Bangladesh. In *2024 2nd International Conference on Information and Communication Technology (ICICT)*. 125–129 (IEEE, 2024)
45. Wen, X., Liao, J., Niu, Q., Shen, N. & Bao, Y. Deep learning-driven hybrid model for short-term load forecasting and smart grid information management. *Sci. Rep.* **14**, 13720 (2024).
46. The Daily Star. Rajshahi faces a hotter future. *The Daily Star* (2023).
47. Garg, R., Barpanda, S., S, G. R. S. N. & S, R. Machine learning algorithms for time series analysis and forecasting. *arXiv.org* https://arxiv.org/abs/2211.14387 (2022).
48. Kotu, V. & Deshpande, B. Time series forecasting. In *Data Science* 395–445. https://doi.org/10.1016/b978-0-12-814761-0.00012-5 (2019).
49. Siami-Namini, S., Tavakoli, N. & Namin, A. S. A comparison of ARIMA and LSTM in forecasting time series. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)* 1394–1401. https://doi.org/10.1109/icmla.2018.00227 (2018).
50. Ho, S. L. & Xie, M. The use of ARIMA models for reliability forecasting and analysis. *Comput. Ind. Eng.* **35**, 213–216 (1998).
51. Jofipasi, C. A., Miftahuddin, & Hizir. Selection for the best ETS (error, trend, seasonal) model to forecast weather in the Aceh Besar District. *IOP Conf. Ser. Mater. Sci. Eng.* **352**, 012055 (2018).
52. Hyndman, R. J. & Athanasopoulos, G. *Forecasting: Principles and Practice* (OTexts, 2018).
53. Naher, S. et al. Forecasting the incidence of dengue in Bangladesh—Application of time series model. *Health Sci. Rep.* **5**, e666 (2022).
54. Zhao, D. & Zhang, H. The research on TBATS and ELM models for prediction of human brucellosis cases in mainland China: a time series study. *BMC Infect. Dis.* **22**, 934 (2022).
55. Xiao, Y. et al. Estimating the long-term epidemiological trends and seasonality of hemorrhagic fever with renal syndrome in China. *Infect. Drug Resist.* **14**, 3849–3862 (2021).
56. Yu, C. et al. Time series analysis and forecasting of the hand-foot-mouth disease morbidity in China using an advanced exponential smoothing state space TBATS model. *Infect. Drug Resist.* **14**, 2809–2821 (2021).
57. De Livera, A. M., Hyndman, R. J. & Snyder, R. D. Forecasting time series with complex seasonal patterns using exponential smoothing. *J. Am. Stat. Assoc.* **106**, 1513–1527 (2011).
58. Khandelwal, I., Adhikari, R. & Verma, G. Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. *Proc. Comput. Sci.* **48**, 173–179 (2015).
59. Zhang, G., Eddy Patuwo, B. & Hu, Y. M. Forecasting with artificial neural networks: The state of the art. *Int. J. Forecast.* **14**, 35–62 (1998).
60. Hamzaçebi, C. Improving artificial neural networks' performance in seasonal time series forecasting. *Inf. Sci.* **178**, 4550–4559 (2008).
61. Zhang, G. P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **50**, 159–175 (2003).
62. Panigrahi, S. & Behera, H. S. A hybrid ETS–ANN model for time series forecasting. *Eng. Appl. Artif. Intell.* **66**, 49–59 (2017).
63. Comito, C. & Pizzuti, C. Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review. *Artif. Intell. Med.* **128**, 102286 (2022).
64. GeeksforGeeks. Time Series Forecasting with Support Vector Regression. *GeeksforGeeks* https://www.geeksforgeeks.org/time-series-forecasting-with-support-vector-regression/ (2024).
65. Aditya Satrio, C. B., Darmawan, W., Nadia, B. U. & Hanafiah, N. Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Proc. Comput. Sci.* **179**, 524–532 (2021).
66. Hasnain, A. *et al.* Time series analysis and forecasting of air pollutants based on prophet forecasting model in Jiangsu Province, China. *Front. Environ. Sci.* **10**, (2022).

67. Hyndman, R. J., & Athanasopoulos, G. Forecasting: Principles and Practice. 3rd ed., Otexts. https://otexts.com/fpp3/ (2021).
68. Elneel, L., Zitouni, M. S., Mukhtar, H. & Al-Ahmad, H. Examining sea levels forecasting using autoregressive and prophet models. *Sci. Rep.* **14**, 14337 (2024).
69. Scornet, E., Biau, G. & Vert, J.-P. Consistency of random forests. *Ann. Stat.* **43**, (2015).
70. The ultimate guide to building your own LSTM models. *ProjectPro* https://www.projectpro.io/article/lstm-model/832 (2024).
71. Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. J. STL: A seasonal-trend decomposition procedure based on loess. *J. Off. Stat.* **6**(1), 3–33 (1990).
72. Kolassa, S., Rostami-Tabar, B. & Siemsen, E. *Demand Forecasting for Executives and Professionals* (CRC Press, 2023).
73. Ridwan, M., Sadik, K. & Afendi, F. Comparison of ARIMA and GRU models for high-frequency time series forecasting. *Sci. J. Inform.* **10**, 389–400 (2024).
74. Bandara, K., Hyndman, R. J. & Bergmeir, C. MSTL: A Seasonal-Trend Decomposition Algorithm for Time Series with Multiple Seasonal Patterns. *arXiv (Cornell University)* (2021)
75. Kim, Y. I., Kim, D. & Lee, S. O. Prediction of temperature and heat wave occurrence for summer season using machine learning. *J. Korean Soc. Disaster Secur.* **13**, 27–38 (2020).
76. Weirich-Benet, E. *et al.* Subseasonal prediction of central European summer heatwaves with linear and random forest machine learning models. *Artif. Intell. Earth Syst.* **2**, (2023).
77. Parviz, L., Rasouli, K. & Torabi Haghighi, A. Improving hybrid models for precipitation forecasting by combining nonlinear machine learning methods. *Water Resour. Manag.* **37**, 3833–3855 (2023).
78. Szagri, D., Nagy, B. & Szalay, Z. How can we predict where heatwaves will have an impact? A literature review on heat vulnerability indexes. *Urban Clim.* **52**, 101711 (2023).
79. Heat Health Risks, Global Heat Health Information Network. Understanding Heat: From Urban Heat Islands to Heatwaves. Global Heat Health Information Network. https://ghhin.org/understanding-heat/ (2024).
80. Climate change indicators: Heat waves | US EPA. US EPA. https://www.epa.gov/climate-indicators/climate-change-indicators-heat-waves (2025).
81. Domeisen, D. I. et al. Prediction and projection of heatwaves. *Nat. Rev. Earth Environ.* **4**(1), 36–50 (2023).
82. Bangladesh reels under longest heat wave in 76 years. https://www.aa.com.tr/en/asia-pacific/bangladesh-reels-under-longest-heat-wave-in-76-years/3203850 (2024).
83. The Bangladesh Red Crescent Society activates its Early Action Protocol for Heatwaves in Dhaka. Anticipation Hub. https://www.anticipation-hub.org/news/the-bangladesh-red-crescent-society-activates-its-early-action-protocol-for-heatwaves-in-dhaka (2024).

Acknowledgements
The authors express their gratitude to the editor and the anonymous reviewers for their insightful comments, which significantly enhanced the quality of the work. The Bangladesh Meteorological Department's Climate Division provided the data required for this analysis, for which the authors are very appreciative.

Author contributions
M.M.U.Q. and A.B.A. jointly contributed in processing, and analyzing data, and interpreting the results, editing, and drafting the manuscript. A.D. edited and critically examined the manuscript in addition to providing crucial comments on the study's design and result interpretation. Pre-processing the data and analyzing the findings were the main contributions made by M.M.H.K. R.R. conceptualized the research idea, design, and supervised the study. Every author has given their approval to the final version.

Funding
There is no funding for this work.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to M.M.U.Q., A.B.A. or R.R.

Reprints and permissions information is available at www.nature.com/reprints.

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

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com