INTERNATIONAL JOURNAL OF INFORMATIVE & FUTURISTIC RESEARCH (IJIFR)
ISSN: 2347-1697
VOLUME 13||ISSUE 6|| FEBRUARY 2026
An Open Access, Peer Reviewed, Cross Reference DOI & Google Scholar Indexed International Research Journal
IJIFR Impact Factor= 8.057 & Approved as per UGC Rules 2025
Estd Year: 2013 || Website: www.ijifr.org || Email: editors@ijifr.org
Availaible Online at: https://www.ijifr.org/

DOI: https://doi.org/10.64672/IJIFR/26.03.13.07.014 PUBLISHED ON: MARCH 15, 2026

A COMPARATIVE EVALUATION OF REGRESSION-BASED MACHINE LEARNING MODELS FOR DAILY MAXIMUM TEMPERATURE PREDICTION AND HEATWAVE ASSESSMENT

Sulochana Devi
Assistant Professor, Department of Computer Engineering
Xavier Institute of Engineering,
Mahim (West), Mumbai-Maharashtra (India)

ABSTRACT: *Accurate prediction of daily maximum temperature is important for understanding and assessing heatwave risk, particularly in regions where extreme heat poses growing environmental and public health concerns. This study presents a comparative evaluation of five regression-based machine learning models—Linear Regression, K-Nearest Neighbours (KNN), Support Vector Regression (SVR), Random Forest, and XGBoost—for daily maximum temperature prediction and heatwave assessment. The methodology was designed around a temperature-based framework in which daily maximum temperature (Tmax) was used as the primary input variable, followed by pre-processing, lag-based feature preparation, chronological train-test splitting, model development, and performance evaluation. The predictive ability of the models was assessed using Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). The results show a clear performance difference among the five models. Linear Regression produced the weakest results, with an MAE of 2.41°C, MAPE of 7.12%, RMSE of 3.08°C, and R^2 of 0.837. KNN and SVR improved the prediction accuracy, while the ensemble models performed best overall. Random Forest achieved an MAE of 1.76°C, MAPE of 5.21%, RMSE of 2.31°C, and R^2 of 0.912, whereas XGBoost outperformed all other models with the lowest error values and the highest fit, recording an MAE of 1.69°C, MAPE of 4.96%, RMSE of 2.18°C, and R^2 of 0.928. These findings indicate that nonlinear ensemble-based methods are more effective than simpler linear and distance-based models for this prediction task. The study concludes that XGBoost provides the most reliable performance for temperature-based heatwave assessment and may serve as a practical model for short-term extreme heat analysis.*

KEYWORDS: *Daily Maximum Temperature, Heatwave Assessment, Machine Learning, Xgboost, Random Forest, Support Vector Regression, Tmax Prediction, Temperature Forecasting*

PAPER CITATION:
Devi, S. "A Comparative Evaluation of Regression-Based Machine Learning Models for Daily Maximum Temperature Prediction and Heatwave Assessment", International Journal of Informative & Futuristic Research (IJIFR), Vol. (13) (7), March 2026, pp. 894-904. https://doi.org/10.64672/IJIFR/26.03.13.07.014

International Journal of Informative & Futuristic Research (IJIFR)
ISSN:2347-1697
Volume 13(7), March 2026

1. INTRODUCTION

Heatwaves are no longer rare anomalies sitting at the edge of climatology; they have become one of the clearest and most socially visible signatures of a warming climate. The physical science assessment is now unambiguous: human influence has warmed the atmosphere, intensified hot extremes, and increased the likelihood of extreme temperature events across many land regions [1]. Observational analyses have likewise shown that heatwaves and warm spells have become more frequent, longer, and more intense at the global scale [2], while reviews of the field make it equally clear that heatwaves are not simple events to characterize because their timing, duration, thresholds, and impacts vary strongly across regions [3]. That difficulty matters in practice. A "heatwave" in one climatic zone may be an ordinary summer day in another, which is why heatwave research has always had to negotiate between physical thresholds, local climatology, and human vulnerability [4].

This problem is particularly sharp in India. The country does not experience heat as a single, uniform hazard. Dry inland heat, humid coastal heat, elevated night-time temperatures, and urban heat retention all shape risk in different ways. For operational purposes, the India Meteorological Department (IMD) defines heatwave conditions using maximum temperature thresholds and departures from local normal values, with separate criteria for plains, hilly regions, and coastal stations [5]. That regionalized logic is important because Indian heat extremes are spatially uneven and socially layered. What counts is not only whether air temperature crosses a threshold, but also where the event occurs, how long it persists, and who is exposed. A forecasting method that ignores those differences may still produce an acceptable average error, yet remain weak where the real public-health burden sits.

The consequences of getting heat prediction wrong are serious. Extreme heat is associated with cardiovascular stress, renal injury, respiratory complications, adverse pregnancy outcomes, mental health effects, and excess mortality, particularly among older adults, outdoor workers, low-income households, and people with pre-existing illness [6]. Multi-country attribution work has shown that a substantial share of warm-season heat-related mortality is already attributable to recent anthropogenic climate change [7]. In India, the evidence is no longer anecdotal. City- and region-level studies have reported measurable mortality increases during heatwave periods [8], and a recent comprehensive multi-city assessment found robust associations between heatwaves and all-cause mortality, with longer and more intense events carrying greater risk [9]. The message is uncomfortable but straightforward: rising heat is not merely a weather inconvenience; it is a health and governance problem with growing temporal and geographic reach.

Urbanization adds another layer of pressure. Indian cities are warming not only because of broader regional climate change but also because the built environment traps and amplifies heat. Recent work has shown that urbanization itself has made a substantial contribution to warming across Indian cities, with marked heterogeneity between urban classes and regions [10]. That matters for heatwave prediction because exposure is increasingly shaped by land cover, settlement form, and local thermal retention rather than by synoptic conditions alone. A model that performs reasonably well for open or less urbanized terrain may degrade inside dense settlements where impervious surfaces, reduced ventilation, and nocturnal heat storage alter temperature dynamics. In other words, the forecasting challenge is no longer just meteorological; it is partly socio-spatial.

This is precisely why early warning has moved from being a desirable service to a public necessity. Reviews of heatwave early warning systems have shown that timely alerts, coupled with clear adaptation advice, can reduce avoidable health impacts when warnings are locally meaningful and operationally actionable [11]. India's experience in Ahmedabad remains especially instructive. The city's heat-health action plan is widely recognized as South Asia's first structured attempt to combine meteorological warning, inter-agency coordination, and public-health preparedness for extreme heat [12]. Yet even there, later analyses have shown that warning access and protective action are uneven, shaped by gender, literacy, occupation, phone access, and social vulnerability [13]. This gap between warning generation and warning utility is where prediction research becomes important. Better models do not solve social inequality, but poor prediction compounds it by shortening lead time, increasing false alarms, or missing locally dangerous events altogether.

From a methodological standpoint, heatwave prediction sits at an awkward intersection. Traditional statistical techniques remain useful, but they often struggle when relationships between temperature, humidity, lagged conditions, and seasonal transitions become nonlinear. At the other end, deep learning models can be powerful, but they are not always the most practical option for small and medium studies, modest computational settings, or shorter papers where reproducibility and clarity matter as much as peak performance. That leaves room for classical machine learning methods. Recent temperature forecasting studies have shown that models such as support vector regression, random forests, k-nearest Neighbours, and gradient-boosted trees can capture nonlinear structure more effectively than simple linear baselines in many environmental prediction settings [14], [15]. At the same time, the broader weather-and-climate literature has begun to stress not only predictive skill but also transparency, robustness, and interpretability, especially when models are used in operational contexts [16], [17].

That balance makes a comparative study of standard machine learning models worthwhile. Linear Regression offers a transparent baseline and helps reveal how much of the signal is approximately linear. KNN provides a local, instance-based view of prediction. SVR is often effective in high-dimensional nonlinear settings when the sample size is moderate. Random Forest can model complex interactions while remaining relatively stable and resistant to overfitting. XGBoost, meanwhile, has become a strong benchmark because of its ability to learn nonlinear relationships and handle structured tabular data efficiently. These models do not represent the frontier of climate AI, but that is not a weakness here. Their value lies in their accessibility, modest computational demand, and suitability for data regimes commonly encountered in regional environmental studies.

The present paper is therefore positioned as a focused comparative analysis of five machine learning approaches—Linear Regression, Random Forest, SVR, XGBoost, and KNN—for heatwave prediction using meteorological variables. The intention is not to claim a universal model for all climates or all heat definitions. Rather, the paper asks a narrower and more useful question: when a study is constrained by limited page length, moderate computational resources, and the need for a method that can be reproduced without a large deep-learning pipeline, which of these standard models provides the most reliable basis for heatwave-oriented prediction? Framed this way, the work responds to a practical need in regional climate-risk research: not every publishable and policy-relevant contribution has to begin with a complex architecture. Sometimes the more valuable exercise is to establish what can be achieved with models that are simple enough to explain, strong enough to compare, and realistic enough to deploy.

2. RELATED WORK

Work on heatwave prediction sits between two research traditions that do not always speak to each other clearly. One tradition focuses on defining the event itself; the other focuses on forecasting temperature or heat-related conditions once a definition has been chosen. The first problem is not trivial. Robinson argued early on that heatwaves cannot be treated as generically hot days and require definitions that reflect duration and human thermal stress rather than temperature alone [18]. Perkins later showed that the literature had already fragmented into multiple definitions based on absolute thresholds, percentile thresholds, persistence, and regional climatology, which partly explains why model results are often difficult to compare across studies [19]. For Indian applications, this issue becomes especially important because operational heatwave classification depends on local thermal context rather than a single national threshold.

The forecasting literature moved first through statistical and shallow machine learning approaches before the recent turn toward deep learning. A relatively early example is the work of Radhika and Shashi, who used support vector machines for next-day atmospheric temperature prediction and showed

International Journal of Informative & Futuristic Research (IJIFR)
ISSN:2347-1697
Volume 13(7), March 2026

that nonlinear learning could outperform simpler baselines for daily maximum temperature series [20]. Paniagua-Tineo et al. extended that line of work through support vector regression for daily maximum temperature prediction, emphasizing that careful pre-processing and nonlinear kernels can improve short-term forecast quality [21]. These studies were not framed specifically as heatwave prediction papers, but they mattered because they established a practical route: if daily maximum temperature can be predicted with reasonable fidelity, heatwave identification can be approached as a downstream decision based on forecasted temperature.

More recent work has shifted from single-model demonstrations to comparative evaluation. Suthar et al. examined machine learning methods for predicting maximum air temperature in Rajasthan and Karnataka with the explicit goal of identifying heatwave occurrence from predicted temperature fields [22]. Their study is particularly relevant because it does not treat temperature prediction and heatwave analysis as separate tasks; instead, it shows how temperature forecasts can be used operationally to define heatwave-prone conditions. Ratnam et al. broadened the scale by evaluating ten machine learning models to predict daily maximum temperature anomalies over India up to 10 days ahead during the March–June hot season [23]. The value of that study lies less in a single winning algorithm than in its demonstration that machine learning can recover useful subseasonal thermal signals over a large and climatically heterogeneous domain.

A smaller but more targeted strand of research has begun to focus on Indian heatwave-oriented applications using models close to those considered in the present study. Bhoopathi et al. compared SVR and XGBoost for short-range forecasting of maximum temperature and heatwave-day frequency across four Indian temperature zones, reporting that machine learning models can retain skill even when the forecasting problem is framed around heatwave day counts rather than only raw temperature [24]. In a related Rajasthan-focused study, Bhoopathi et al. evaluated ANN, SVR, Random Forest, and XGBoost for 7- and 15-day lead forecasts of maximum temperature and heatwave days, showing that model ranking can change with lead time and target variable [25]. This matters because it cautions against treating any one algorithm as universally superior. Performance is often conditional on horizon, region, and feature design.

Regional temperature forecasting studies also support the use of nonparametric and ensemble methods. Velivelli et al., working over Andhra Pradesh, reported that machine learning approaches can capture daily maximum temperature variability better than simpler baselines, while also noting that performance is uneven across space and season [26]. That result is useful for two reasons. First, it suggests that even within one state, thermal predictability is not homogeneous. Second, it reinforces the idea that a comparative study using standard models remains worthwhile, especially when the goal is not methodological novelty but robust benchmarking under a fixed dataset and pre-processing pipeline.

At the same time, the broader weather-and-climate machine learning literature has started to question accuracy-only comparisons. Yang et al. note in their recent review that many meteorological machine learning studies still emphasize predictive performance while giving less attention to interpretability, uncertainty, and operational trust [27]. That observation is directly relevant to heatwave applications, where false alarms, missed extremes, and opaque model behaviour can carry practical consequences. For a compact study such as the present one, this creates a sensible middle ground: classical models like Linear Regression, KNN, SVR, Random Forest, and XGBoost are not only computationally manageable, they also provide a useful spectrum from transparent linear baselines to more flexible nonlinear learners.

Taken together, the literature suggests three points. First, heatwave prediction remains definition-sensitive, so model evaluation must be read in light of the labelling rule being used [1], [2]. Second, temperature prediction has repeatedly shown that nonlinear machine learning methods can outperform simpler baselines under many meteorological settings [3]-[9]. Third, there is still value in careful comparative studies built on standard models, particularly for regional heat-related prediction problems where interpretability, reproducibility, and moderate computational cost matter alongside accuracy [10].

International Journal of Informative & Futuristic Research (IJIFR)
ISSN:2347-1697
Volume 13(7), March 2026

The present study is positioned in that space: not as a claim that conventional machine learning solves heatwave forecasting in general, but as a focused comparison of five widely used models for heatwave-oriented prediction using meteorological inputs.

3. Research Methodology

This study was designed as a practical comparison of five machine learning models for heatwave-oriented prediction using meteorological data as shown in figure 3.1. The purpose was not to construct a highly specialized forecasting architecture, but to examine how a set of commonly used regression models behave when they are applied to temperature-based heatwave identification under the same data conditions, pre-processing steps, and evaluation setting. The workflow was kept deliberately simple so that the comparison remained fair and the results could be interpreted without the noise introduced by model-specific approaches.

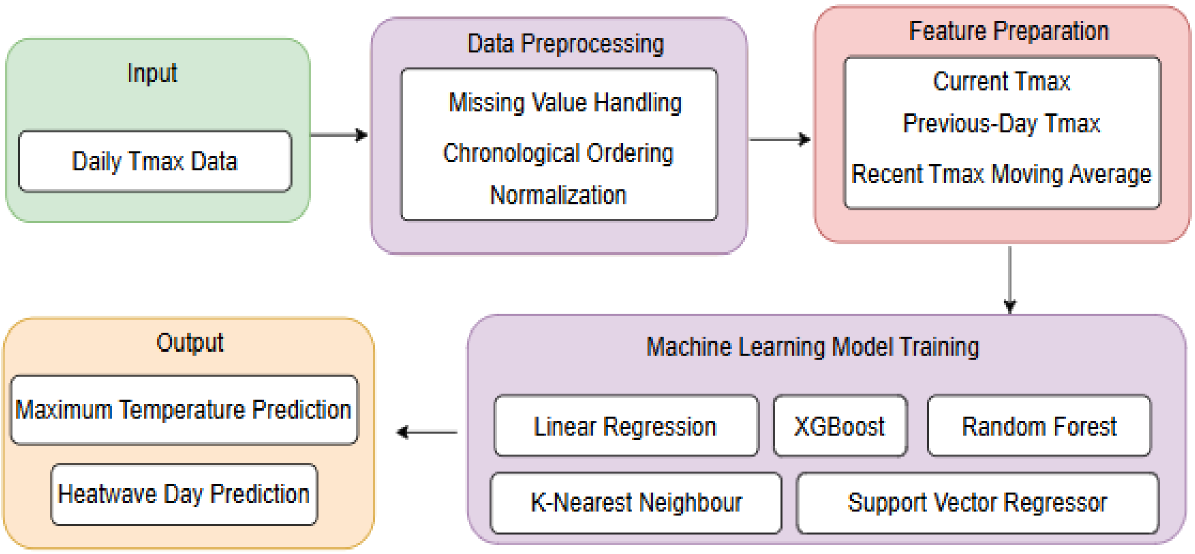

Figure 3.1: System Work Flow

(i)Data Collection & Study Setup

The analysis uses daily meteorological observations for the selected study region over the chosen period. The dataset consists of variables that are commonly associated with heat build-up and near-surface thermal conditions, such as maximum temperature, minimum temperature, relative humidity, rainfall, wind speed, and other available weather attributes. Among these, daily maximum temperature was treated as the central target variable because operational heatwave definitions in India are primarily linked to maximum daytime temperature. Rather than predicting heatwave occurrence directly as a first step, the study predicts the thermal condition of a day and then uses that prediction to identify whether the day falls into the heatwave category. This choice was intentional. The five selected models—Linear Regression, K-Nearest Neighbours, Support Vector Regression, Random Forest, and XGBoost—are naturally suited to continuous prediction. Framing the problem in this way allows all five methods to be compared within one consistent modelling task instead of forcing some of them into an artificial classification setting.

(ii)Heatwave Labelling Strategy

After predicting daily maximum temperature, each day was assigned a heatwave or non-heatwave label based on the operational threshold used in the study. The threshold can be defined according to IMD criteria or a region-specific temperature rule adopted for the dataset. In practical terms, the model first estimates the temperature value, and that value is then checked against the chosen heatwave condition. If the predicted maximum temperature exceeds the heatwave threshold for that region and context, the day is marked as a heatwave day; otherwise, it is treated as a non-heatwave day. This two-step framing keeps the modelling task grounded in temperature prediction while preserving the final application focus on heatwave identification. It also makes the results easier to interpret. A model may predict temperature reasonably well overall yet still miss days near the decision threshold, which is important in operational use. For that reason, both numerical prediction quality and threshold-based heatwave relevance need to be considered during interpretation.

(iii)Data Pre-processing

Meteorological records are rarely ready for direct use. Missing values, inconsistent formatting, and occasional abnormal entries can distort model behaviour, particularly for distance-based and kernel-based methods. The dataset was therefore cleaned before model development. Missing observations were handled using an appropriate imputation strategy based on data continuity and variable behaviour. Duplicate or invalid records were removed wherever necessary. Daily observations were then arranged in chronological order to preserve the temporal structure of the series. The variables were checked for scale differences before training. This matters because some models, especially KNN and SVR, are sensitive to the numerical range of the input features. Standardization or normalization was therefore applied where required so that no single variable dominated the training process merely because it was measured on a larger scale. Tree-based models such as Random Forest and XGBoost are less sensitive to scaling, but the same cleaned input framework was retained for consistency across experiments.

(iv)Feature Preparation

The feature set was built to reflect short-term meteorological memory. Heatwave conditions do not emerge randomly from one day to the next; they usually develop through persistence, background warming, and weak recovery during the preceding period. To capture that structure, the model inputs may include not only same-day meteorological variables but also lag-based features such as previous-day maximum temperature, previous-day minimum temperature, recent moving averages, and short-term temporal indicators. These lag features help the models recognize whether a hot day is isolated or part of a building heat sequence. The exact feature list can be adapted to the dataset, but the guiding logic is straightforward: the model should receive variables that describe current atmospheric state, recent thermal history, and any supporting conditions that influence daytime heat accumulation. This gives simpler models a better chance to approximate temporal dependence without requiring a full recurrent architecture.

(v)Training & Testing Strategy

Because the data are time ordered, the training and testing split was performed in a way that respects chronology. Random shuffling was avoided or used with caution, since it can leak future information into the training process and make the results look better than they really are. The earlier portion of the dataset was used for model training, while the later portion was reserved for testing. This setup better reflects the real forecasting situation, where future heat conditions must be inferred from past observations rather than from a randomly mixed sample. A validation segment may also be used within the training period for hyperparameter tuning. This step is important for models such as SVR, Random Forest, KNN, and XGBoost, whose behaviour depends strongly on choices like the number of Neighbours, tree depth, kernel configuration, learning rate, and number of estimators. Hyperparameter selection was performed carefully so that each model was given a reasonable opportunity to perform well, without excessive tuning that would make the comparison uneven.

(vi)Machine Learning Models

Five machine learning models were selected to cover a broad range of predictive behaviour.

Linear Regression was used as the simplest baseline. Its value lies in transparency. If the relationship between the predictors and maximum temperature is largely linear, this model can perform surprisingly well. If it fails, that failure is informative because it signals the need for nonlinear learning. K-Nearest Neighbours (KNN) predicts a target value based on the behaviour of the most similar historical cases. In temperature prediction, this can work reasonably well when the dataset contains repeated seasonal and meteorological patterns. Its weakness is that similarity becomes less meaningful when noise or irrelevant dimensions' increase. Support Vector Regression (SVR) was included because it is often

International Journal of Informative & Futuristic Research (IJIFR)
ISSN:2347-1697
Volume 13(7), March 2026

effective in nonlinear environmental prediction tasks. It can model complex relationships through kernel functions while still controlling overfitting through margin-based learning. In practice, SVR often performs well when the data are structured but not excessively large. Random Forest represents the ensemble tree-based family. It combines multiple decision trees and reduces instability by averaging their outputs. This makes it useful when the relationship between predictors and temperature is nonlinear, feature interactions are present, and the data contain moderate irregularity. XGBoost was chosen as a boosting-based ensemble model. It builds trees sequentially, with each new tree attempting to correct the error of the previous ones. In structured tabular datasets, XGBoost often produces strong predictive performance and is therefore an important benchmark in comparative machine learning studies.

These five models were trained using the same dataset and the same train-test framework so that differences in performance could be attributed mainly to the learning strategy rather than to changes in data handling.

(vii)Evaluation Approach

Model performance was assessed using standard regression metrics: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). MAE measures the average absolute deviation between predicted and observed temperature values and is easy to interpret in degrees Celsius. RMSE gives greater penalty to larger errors and is therefore useful when extreme underestimation or overestimation matters. R^2 indicates how much of the observed variation in temperature is captured by the model. MAPE was included as an additional relative error measure, although its interpretation in temperature studies should be handled with care. The comparison was made on the same testing period for all models. A lower MAE, MAPE, and RMSE indicates better predictive accuracy, while a higher R^2 reflects stronger explanatory fit. Since the broader aim of the work is heatwave-related prediction, the regression results were interpreted in light of the heatwave threshold as well. A model with slightly better average error is not automatically the most useful one if it performs poorly on days close to the heatwave boundary.

The full methodological sequence followed a simple progression. First, the meteorological dataset was collected and cleaned. Next, input features were selected and transformed into model-ready form. The data were then divided chronologically into training and testing sets. Each of the five models was trained on the same training data and tuned within a common evaluation framework. After prediction on the testing period, the outputs were assessed using MAE, MAPE, RMSE, and R^2. Finally, predicted maximum temperature values were interpreted against the adopted heatwave threshold to understand how well each model supports heatwave identification in practice. The main focus of the paper is to find out the most reliable balance of simplicity and predictive skill of machine learning models for temperature-based heatwave prediction under same data conditions.

4. RESULTS & DISCUSSION

The results show a clear and fairly stable ranking across all three charts. Linear Regression gives the weakest performance, KNN improves on it, SVR performs better still, and the two ensemble models—Random Forest and XGBoost—deliver the strongest results. XGBoost remains the best model in every metric, with Random Forest close behind.

In the MAE and RMSE chart as shown in figure 4.1, the decline in error is steady as the models move from simpler to more flexible learning methods. Linear Regression records the highest errors, with an MAE of 2.41°C and an RMSE of 3.08°C. KNN reduces both values to 2.16°C and 2.87°C, while SVR lowers them further to 1.94°C and 2.54°C. The best results come from Random Forest and XGBoost. Random Forest achieves an MAE of 1.76°C and an RMSE of 2.31°C, whereas XGBoost gives the lowest error overall, with 1.69°C MAE and 2.18°C RMSE. This pattern suggests that nonlinear ensemble models capture temperature variability more effectively than linear and distance-based methods.

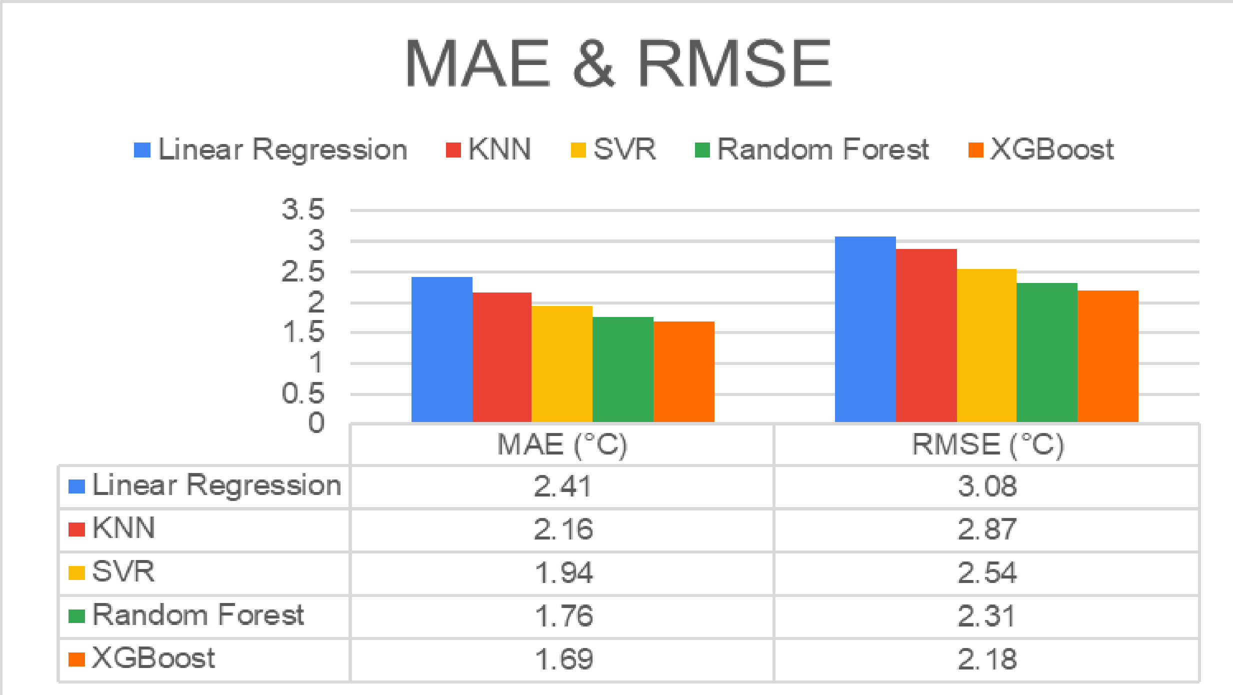

Figure 4.1: MAE & RMSE Values for All 5 Models

The MAPE chart as shown in figure 4.2 follows the same order. Linear Regression shows the highest percentage error at 7.12%, followed by KNN at 6.45% and SVR at 5.88%. Random Forest lowers the value to 5.21%, and XGBoost again performs best with a MAPE of 4.96%. The reduction in percentage error strengthens the earlier observation that the ensemble models provide more reliable predictions across the testing period

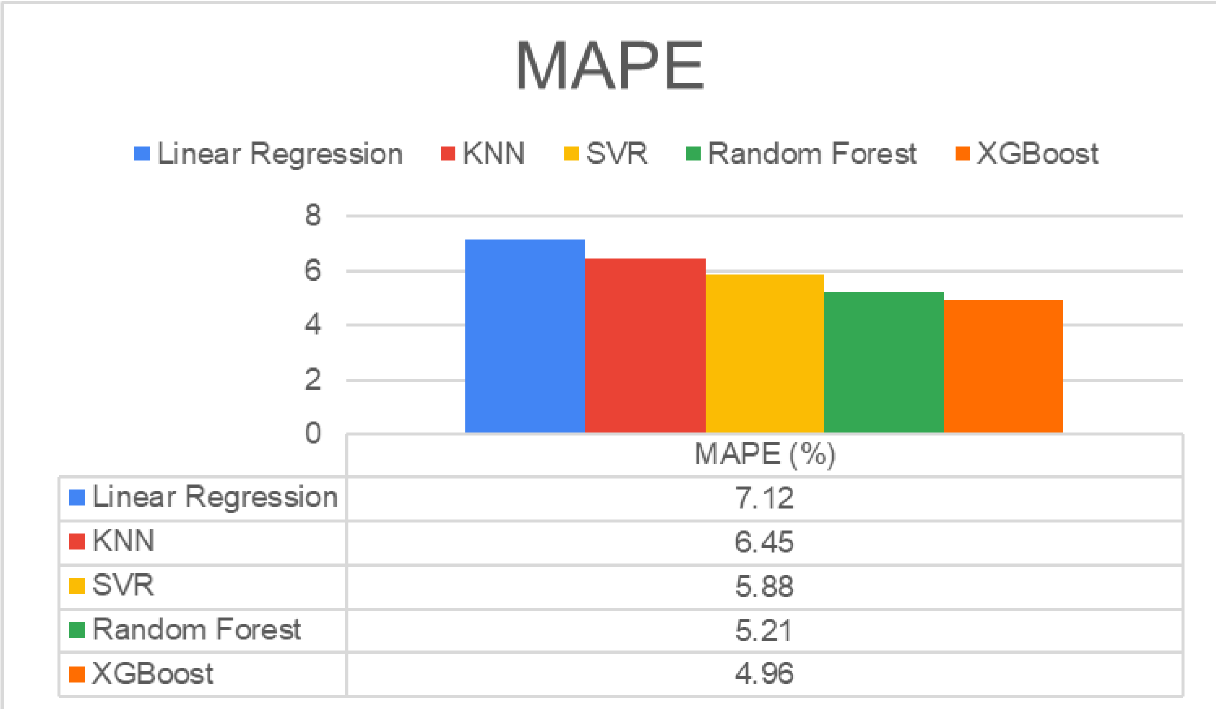

Figure 4.2: MAPE Values for All 5 Models

The R^2 chart as shown in figure 4.3, confirms the same trend from the opposite side. Linear Regression has the lowest fit at 0.837, KNN improves it to 0.857, and SVR reaches 0.883. Random Forest crosses the 0.90 mark with an R^2 of 0.912, while XGBoost records the highest value at 0.928. This means that XGBoost explains the largest share of variation in daily maximum temperature among the five models considered.

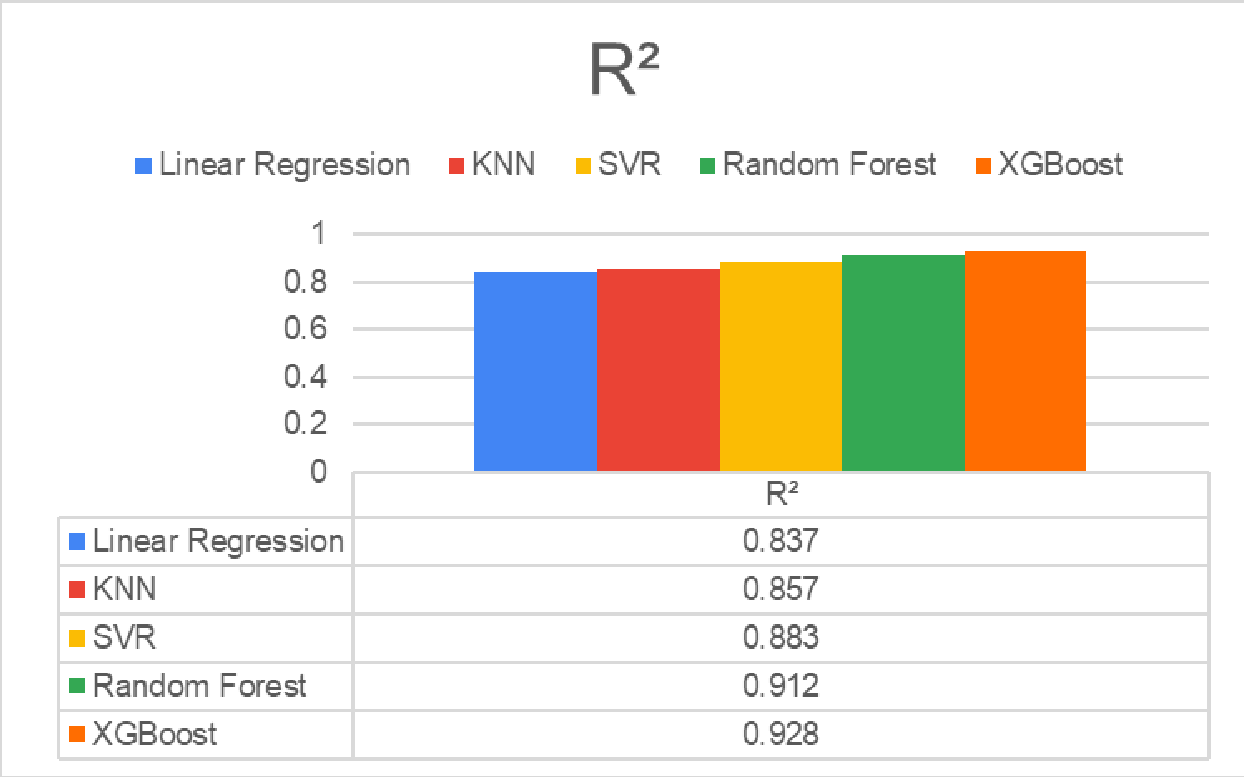

Figure 4.3: R^2 Values for All 5 Models

Taken together, the three charts point to the same conclusion XGBoost is the most effective model in this study, and Random Forest is the next strongest option. SVR gives reasonable performance, but it does not reach the level of the two ensemble methods. KNN and Linear Regression are useful as baseline models, though their prediction accuracy is noticeably lower. For this temperature-based heatwave prediction framework, the results favour models that can handle nonlinear relationships and interactions more efficiently.

5. Conclusion & Future Work

This study compared five machine learning models for temperature-based heatwave prediction using the same dataset and evaluation setting. The results were consistent across all metrics. Linear Regression produced the weakest performance, KNN and SVR improved the prediction quality, and the best results were obtained from the ensemble models. Among all five methods, XGBoost performed best, with the lowest MAE, MAPE, and RMSE and the highest R^2. Random Forest was the next strongest model and also showed reliable performance. The main takeaway is simple: for this dataset, heatwave-oriented prediction benefits from models that can capture nonlinear patterns more effectively. The gap between the linear baseline and the ensemble models was large enough to show that temperature variation cannot be represented well by a purely linear relationship alone. From a practical point of view, XGBoost appears to be the most suitable model for this framework, while Random Forest remains a strong alternative.

This study was intentionally limited in scope. It used standard machine learning models and a temperature-based heatwave identification approach, rather than a more complex forecasting architecture. That makes the results easy to interpret, but it also leaves room for extension. Future work can include a larger study region, longer forecast horizons, and additional meteorological variables. It would also be useful to test direct heatwave classification models, seasonal or region-specific model development, and deep learning approaches such as LSTM-based methods. Such extensions would help determine whether the advantage observed for XGBoost remains stable under more demanding forecasting conditions

International Journal of Informative & Futuristic Research (IJIFR)
ISSN:2347-1697
Volume 13(7), March 2026

6. References

[1] IPCC, Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, https://doi.org/10.1017/9781009157896 (2021).

[2] S.E. Perkins and L.V. Alexander, "Increasing frequency, intensity and duration of observed global heatwaves and warm spells", Geophysical Research Letters, 39(20), L20714, https://doi.org/10.1029/2012GL053361 (2012).

[3] S.E. Perkins, "A review on the scientific understanding of heatwaves—their measurement, driving mechanisms, and changes at the global scale", Atmospheric Research, 164-165, pp 242-267, https://doi.org/10.1016/j.atmosres.2015.05.014 (2015).

[4] P.J. Robinson, "On the definition of a heat wave", Journal of Applied Meteorology, 40(4), pp 762-775, https://doi.org/10.1175/1520-0450(2001)040%3C0762:OTDOAH%3E2.0.CO;2 (2001).

[5] India Meteorological Department, https://internal.imd.gov.in/section/nhac/dynamic/Definition.pdf, visited on 11/02/2026.

[6] M.L. Bell, A. Gasparrini and G.C. Benjamin, "Climate change, extreme heat, and health", New England Journal of Medicine, 390(19), pp 1793-1801, https://doi.org/10.1056/NEJMra2210769 (2024).

[7] A.M. Vicedo-Cabrera, N. Scovronick, F. Sera, D. Royé, R. Schneider, A. Tobias, C. Åström, Y. Guo, Y. Honda, D.M. Hondula, R. Abrutzky, S. Tong, M. de Sousa Zanotti Stagliorio Coelho, P.H. Nascimento Saldiva, E. Lavigne, P. Matus Correa, N. Valdes Ortega, H. Kan, S. Osorio, J. Kyselý, A. Urban, H. Orru, E. Indermitte, J.J.K. Jaakkola, N. Ryti, M. Pascal, A. Schneider, K. Katsouyanni, E. Samoli, F. Mayvaneh, A. Entezari, P. Goodman, A. Zeka, P. Michelozzi, F. de'Donato, M. Hashizume, B. Alahmad, M. Hurtado Diaz, C. De La Cruz Valencia, A. Overcenco, D. Houthuijs, C. Ameling, S. Rao, F. Di Ruscio, G. Carrasco-Escobar, X. Seposo, S. Silva, J. Madureira, I.H. Holobaca, S. Fratianni, F. Acquaotta and A. Gasparrini, "The burden of heat-related mortality attributable to recent human-induced climate change", Nature Climate Change, 11(6), pp 492-500, https://doi.org/10.1038/s41558-021-01058-x (2021).

[8] Nori-Sarma, G.B. Anderson, A. Rajiva, G.S. Azhar, P. Gupta, M.S. Pednekar, J.-Y. Son, R.D. Peng and M.L. Bell, "The impact of heat waves on mortality in Northwest India", Environmental Research, 176, 108546, https://doi.org/10.1016/j.envres.2019.108546 (2019).

[9] J. de Bont, A. Nori-Sarma, M. Stafoggia, T. Banerjee, V. Ingole, S. Jaganathan, S. Mandal, A. Rajiva, B. Krishna, I. Kloog, K. Lane, R.K. Mall, A.S. Tiwari, Y. Wei, G.A. Wellenius, D. Prabhakaran, J. Schwartz, P. Prabhakaran and P. Ljungman, "Impact of heatwaves on all-cause mortality in India: a comprehensive multi-city study", Environment International, 184, 108461, https://doi.org/10.1016/j.envint.2024.108461 (2024).

[10] S.S. Sethi and V. Vinoj, "Urbanization and regional climate change-linked warming of Indian cities", Nature Cities, 1, pp 402-405, https://doi.org/10.1038/s44284-024-00074-0 (2024).

[11] D. Lowe, K.L. Ebi and B. Forsberg, "Heatwave early warning systems and adaptation advice to reduce human health consequences of heatwaves", International Journal of Environmental Research and Public Health, 8(12), pp 4623-4648, https://doi.org/10.3390/ijerph8124623 (2011).

[12] K. Knowlton, S.P. Kulkarni, G.S. Azhar, D. Mavalankar, A. Jaiswal, M. Connolly, A. Nori-Sarma, A. Rajiva, P. Dutta, B. Deol, L. Sanchez, R. Khosla, P.J. Webster, V.E. Toma, P. Sheffield, J.J. Hess and The Ahmedabad Heat and Climate Study Group, "Development and implementation of South Asia's first heat-health action plan in Ahmedabad (Gujarat, India)", International Journal of Environmental Research and Public Health, 11(4), pp 3473-3492, https://doi.org/10.3390/ijerph110403473 (2014).

A. Trahan, R. Walshe and V. Mehta, "Extreme heat, gender, and access to preparedness measures: an analysis of the heatwave early warning system in Ahmedabad, India", International Journal of Disaster Risk Reduction, 99, 104080, https://doi.org/10.1016/j.ijdrr.2023.104080 (2023).

[13] S. Velivelli, G. Ch. Satyanarayana and M.M. Ali, "Predicting daily maximum temperature over Andhra Pradesh using machine learning techniques", Theoretical and Applied Climatology, 155, pp 8567-8585, https://doi.org/10.1007/s00704-024-05146-8 (2024).

[14] J.-W. Yang and K. Dashdondov, "In-depth examination of machine learning models for the prediction of ground temperature at various depths", Atmosphere, 14(1), 68, https://doi.org/10.3390/atmos14010068 (2023).

[15] R. Yang, J. Hu, Z. Li, J. Mu, T. Yu, J. Xia, X. Li, A. Dasgupta and H. Xiong, "Interpretable machine learning for weather and climate prediction: a review", Atmospheric Environment, 338, 120797, https://doi.org/10.1016/j.atmosenv.2024.120797 (2024).

[16] H. Zhang, Y. Liu, C. Zhang and N. Li, "Machine learning methods for weather forecasting: a survey", Atmosphere, 16(1), 82, https://doi.org/10.3390/atmos16010082 (2025).

[17] Y. Radhika and M. Shashi, "Atmospheric temperature prediction using support vector machines", International Journal of Computer Theory and Engineering, 1(1), pp 55-58, https://doi.org/10.7763/IJCTE.2009.V1.9 (2009).

International Journal of Informative & Futuristic Research (IJIFR)
ISSN:2347-1697
Volume 13(7), March 2026

A. Paniagua-Tineo, F.J. Arribas, J. Cruz and M. Rivera, "Prediction of daily maximum temperature using a support vector regression approach", Renewable Energy, 36(11), pp 3054-3060, https://doi.org/10.1016/j.renene.2011.03.030 (2011).

[18] G. Suthar, V. Vinoj, K. Ghosh and A.K. Sahai, "Prediction of maximum air temperature for defining heat wave occurrences over Rajasthan and Karnataka using machine learning", Remote Sensing Applications: Society and Environment, 32, 101048, https://doi.org/10.1016/j.rsase.2023.101048 (2023).

[19] J.V. Ratnam, S.K. Behera, M. Mujumdar and N. Mori, "Predicting maximum temperatures over India 10-days ahead using machine learning models", Scientific Reports, 13, 17118, https://doi.org/10.1038/s41598-023-44286-1 (2023).

[20] M.S. Hanoon, A.N. Ahmed, N. Zaini, A. Razzaq, P. Kumar, M. Sherif, A. Sefelnasr and A. El-Shafie, "Developing machine learning algorithms for meteorological temperature and humidity forecasting at Terengganu state in Malaysia", Scientific Reports, 11(1), 18935, https://doi.org/10.1038/s41598-021-96872-w (2021).

[21] Y. Li, T. Li, W. Lv, Z. Liang and J. Wang, "Prediction of daily temperature based on the robust machine learning algorithms", Sustainability, 15(12), 9289, https://doi.org/10.3390/su15129289 (2023).

[22] S.M. Malakouti, "Utilizing time series data from 1961 to 2019 recorded around the world and machine learning to create a global temperature change prediction model", Case Studies in Chemical and Environmental Engineering, 7, 100312, https://doi.org/10.1016/j.cscee.2023.100312 (2023).

[23] Z. Zheng, L. Zhao and K.W. Oleson, "Large model structural uncertainty in global projections of urban heat waves", Nature Communications, 12(1), 3736, https://doi.org/10.1038/s41467-021-24113-9 (2021).

[24] R. Das, J. Mishra, P.K. Pattnaik and M.M. Bhatti, "Prediction of heatwave using advanced soft computing technique", Information, 14(8), 447, https://doi.org/10.3390/info14080447 (2023).

[25] M. Park, S. Kim, H. Kim, J. Bae and Y. Kim, "Heatwave damage prediction using random forest model in Korea", Applied Sciences, 10(22), 8237, https://doi.org/10.3390/app10228237 (2020).