J. Ind. Geophys. Union, 30(4) (2026), 263-275
https://doi.org/10.71122/JIGU.30(4)2026.0026

Prediction of pre-monsoon temperature of Varanasi using machine learning and deep learning techniques

R. Bhatla[1*], Mohini Dangi[2], Aashna Verma[1] and Manjari Gupta[3]
[1] Department of Geophysics, Institute of Science, BHU, Varanasi-221005, India
[2] DST- CIMS, Institute of Science, BHU, Varanasi-221005 , India
[3] Department of Computer Science, Institute of Science, BHU, Varanasi-221005, India
*Corresponding author: rbhatla@bhu.ac.in

ABSTRACT

Accurate short-term temperature forecasting is essential for understanding climate changes and supporting agricultural planning during the pre-monsoon period in northern India, when temperatures rise sharply. Although machine learning approaches have been applied in several regions, their performance for short-lead prediction, using long historical observations in densely populated urban areas such as Varanasi, has received very limited attention. In this study, we evaluate four widely used machine learning and deep learning models, Support Vector Regression (SVR), Random Forest (RF), Multi-Layer Perceptron (MLP), and Long Short-Term Memory (LSTM), to predict daily maximum temperature using a univariate approach over Varanasi, based on 71 years of IMD observations. The purpose of adopting a univariate framework is to keep forecasting operationally simple, while capturing the short-term evolution of temperature patterns without additional climatic variables. Our results show that SVR achieved the highest R^2 score (0.801) and correlation (0.897), while LSTM produced the lowest RMSE (1.673) and MAE (1.213). Overall, LSTM is considered the best performing model because the study prioritizes minimizing short-term forecasting errors, even though SVR achieves the highest explained variance. The good performance of SVR and LSTM suggests that short-term temperature forecasts can be generated reliably with limited input information, which is useful for early warnings, heatwave preparedness, and agricultural decision making during the vulnerable pre-monsoon season in the region.

Keywords: Univariate time series; Deep learning; Long Short-Term Memory (LSTM); Temperature forecasting; Support Vector Regression.

INTRODUCTION

Temperature plays a central role in shaping human health, agriculture and local environmental conditions. Rising temperatures increase the frequency and duration of heat waves, and prolonged exposure can lead to heat stress, heatstroke and cardiovascular risks, particularly among vulnerable populations such as the elderly and those with pre-existing medical conditions. Accurate temperature forecasting is essential for anticipating extreme events and improving preparedness during high-risk periods. In addition to its direct impact, temperature fluctuations are intricately linked to broader environmental and societal challenges such as urbanization, which has become a major contributor to global warming and environmental degradation (Grimmond et al., 2007, Bhatla et al., 2016a, Gupta et al., 2019). Human activities, such as deforestation for agricultural expansion and industrialization, are driven by the growing population's demands. The consequences of these activities result in disrupting the ecosystem, increasing flood and drought frequency (Verma et al., 2022; Verma et al., 2023; Maurya et al., 2023, Bhatla et al., 2025), intensifying typhoons, reducing freshwater availability, and leading to erratic and random temperature rises (Nikam and Meshram, 2013, Kumar et al., 2024). Climate models project that temperature variability in India will increase by up to 10% ºC by the end of this decade, potentially causing soil drying and shifts in atmospheric structure (Bathiany et al., 2018). Throughout the 21st century, India is expected to experience greater seasonal variation in temperature and warming beyond the global average. However, rising temperatures could lead to a 40% decline in India's agricultural productivity by the 2080s. The threats of climate change are also affecting the quality and quantity parameters of water resources and crop productivity (Kumar et al., 2014, Bhatla et al., 2016b).

Predicting these impacts requires robust forecasting methods, as effective forecasting can help mitigate risks and adapt to changing conditions. Traditional statistical approaches such as the autoregressive integrated moving average (ARIMA) model have been widely used for weather and climate time-series forecasting because they are mathematically interpretable and effective for short memory processes. However, ARIMA assumes linearity and stationarity, which limits its ability to capture rapidly changing and non-linear temperature behaviour, especially during pre-monsoon months when short-term variability and extremes are common. More importantly, statistical models depend heavily on parameter specification and often deteriorate when long historical records contain abrupt regime shifts. These limitations have motivated the increasing use of machine learning and deep learning approaches (Shrivastava et al., 2023), which are capable of modelling non-linear relationships and learning temporal patterns directly from data without strong assumptions about underlying processes. Artificial intelligent systems have revolutionized human lifestyle in various fields like healthcare, agriculture and commerce (Gosala et al., 2023; Kural et al., 2024; Pathak et al., 2024). Precisely, Artificial Neural Networks (ANN) offer an alternative due to their high accuracy and ability to efficiently learn hidden nonlinear relationships between variables. Neural networks are increasingly becoming popular in the field of forecasting because of their practicality (Chen et al., 2018). Neural Networks have the potential to