RESEARCH ARTICLE OPEN ACCESS

A Recent Trends in Drought Prediction: A Comparative Study of Machine Learning and Deep Learning Models (2020–2025)

Ajay Verma[1]*, Vinay Pandey[1]
[1]Department of Computer Science and Engineering, Chouksey Engineering College, Bilaspur, Chhattisgarh, India

Abstract
Drought has emerged as a critical environmental challenge, significantly affecting agriculture, water resources, and socio-economic stability. Between 2020 and 2025, utilizing deep learning (DL) and machine learning (ML) methods has expanded rapidly for drought prediction. Commonly applied models include Long Short-Term Memory (LSTM), XGBoost, Random Forest (RF), Support Vector Machines (SVM), CNN–LSTM hybrids, and emerging Transformer-based architectures. This review presents a comparative analysis of recent literature, focusing on datasets, drought indices (SPI, SPEI, NDVI), and performance metrics, including R^2, RMSE, and accuracy. Results indicate that RF and SVM remain effective for short-term drought forecasting, while LSTM and hybrid DL models show superior performance for long-term predictions. Looking ahead, integrating Transformer-based hybrid frameworks with satellite-derived indices offers a promising direction for more accurate and reliable drought monitoring systems.

Keywords: Drought prediction, Machine learning, Deep learning, Climate indices, Transformer models

1. Introduction

Among the most common is drought, complex and destructive natural hazards, affecting agriculture, ecosystems, water availability, and human livelihoods worldwide. Unlike other disasters, drought develops gradually and persists for long periods, making its detection and prediction particularly challenging. Globally, prolonged droughts have intensified under the influence of climate change, with several regions reporting higher frequencies of extreme events, prolonged dry spells, and shifts in monsoon dynamics. Global warming, according to the Intergovernmental Panel on Climate Change (IPCC), and altered rainfall variability are expected to exacerbate drought risks in many semi-arid and tropical regions during the twenty-first century [1]. In the Indian context, drought poses a recurring challenge to food security and rural economies, particularly in regions that are highly dependent on monsoon rainfall. Central India, encompassing areas such as Satpura and Bundelkhand, is among the most vulnerable zones. Bundelkhand, a semi-arid region, frequently experiences erratic rainfall, declining groundwater, and recurring agricultural distress. On the other hand, Satpura, despite being a relatively humid region, has shown a consistent rise in drought intensity in recent decades due to irregular monsoon patterns and rising evapotranspiration [2]. The situation highlights that both traditionally drought-prone and comparatively humid regions are becoming increasingly vulnerable under climate stress.

Earlier research on drought prediction relied primarily on traditional statistical measures such the Palmer Drought Severity Index (PDSI) and the Standardized Precipitation Index (SPI). While these indices provided useful insights, they often failed to capture the combined impact of rainfall and temperature variability. The development of the Precipitation Evapotranspiration Index Standardized (SPEI) addressed this limitation by integrating temperature-driven evapotranspiration, making it more suitable for drought monitoring under climate change scenarios [3]. Between 2010 and 2019, most studies focused on applying Artificial Neural Networks, Random Forest (RF), and Support Vector Machines (SVM) are examples of machine learning models for drought prediction in India and abroad [4]. However, since 2020, a significant methodological shift has occurred, with deep learning techniques like Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), and hybrid CNN–LSTM models becoming increasingly popular. These models are particularly effective in capturing long-term temporal dependencies and non-linear climate interactions. Despite these advances, a systematic comparative analysis of deep learning and machine learning models for drought prediction, specifically covering the period from 2020 to 2025, has yet to be conducted. Most existing reviews either focus on pre-2020 works or provide only general discussions without benchmarking performance across models.

Therefore, this paper aims to fill this gap by presenting a comparative review of drought prediction studies conducted between 2020 and 2025. The primary objectives are:

1. To analyze recent advancements in drought prediction models, focusing on both machine learning (RF, SVM, XGBoost) and deep learning approaches (LSTM, CNN-LSTM, Transformer-based models).
2. To compare their performance across different datasets, regions, and indices (SPI, SPEI, NDVI).
3. To identify emerging trends, highlight research gaps, and provide insights for future work in developing robust and region-specific drought forecasting frameworks.

By synthesizing recent literature, this paper contributes to a clearer understanding of how modern computational models can support early warning systems, agricultural planning, and climate-resilient policies in drought-prone regions like Central India [2], while also providing lessons relevant at the global scale.

2. Literature Review (2020–2025)

2.1 Climate Indices and Statistical Models

The earliest attempts at drought monitoring and prediction relied heavily on statistical indices. The most widely adopted index has been the Standardized Precipitation Index (SPI), which is based solely on rainfall data. Another The Palmer Drought Severity indicator (PDSI), a widely used indicator, provided useful insights but was found to be less effective in the context of climate change. To address these limitations, the Standardized Precipitation Evapotranspiration Index (SPEI) was introduced, incorporating temperature-driven evapotranspiration and thereby offering a more reliable measure for climate-sensitive regions [3].

After 2020, several studies combined SPI and SPEI with statistical techniques such as the Mann–Kendall (MK) trend test and the Autoregressive Integrated Moving Average (ARIMA) model. For example, Singh et al. (2023) examined rainfall trends in Central India using SPI and SPEI, reporting that SPI was effective for short-term drought monitoring, while SPEI was better suited for long-term prediction [6]. Similarly, Dwivedi et al. (2024) applied SPI and ARIMA in the Ken Basin, demonstrating that medium-term drought forecasting up to six months ahead could be achieved with reasonable accuracy [7].

2.2 Machine Learning Models

During 2020–2025, machine learning (ML) models emerged as widely applied tools for drought prediction. Among these, Random Forest (RF), Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost) gained particular prominence due to their robustness and ability to handle nonlinear data.

Bisht et al. (2023) compared RF and SVM for drought forecasting in the Bundelkhand region. Their results revealed that RF performed better for short-term predictions, whereas SVM showed relatively superior performance over longer time horizons [8]. In another study, Galkate et al. (2024) applied XGBoost to rainfall and SPI datasets across Madhya Pradesh districts, achieving high accuracy in spatial drought prediction [8].

Your recent study (Verma & Pandey, 2025) analyzed drought intensification in Satpura and Bundelkhand using the SPEI index along with RF and SVM. The findings confirmed that RF outperformed in short-term drought prediction, while SVM was more reliable for long-term forecasts [2].

2.3 **Deep Learning Models**

Because deep learning (DL) approaches can capture complicated non-linear interactions and long-term dependencies in climate data, they have gained popularity in recent years for drought prediction.
Singh et al. (2023) applied Long Short-Term Memory (LSTM) networks for Central India and reported superior performance, with R^2 reaching up to 0.92 when applied to rainfall and temperature-based SPEI time series [6]. Kumar et al. (2024) introduced a CNN–LSTM hybrid model, which outperformed RF and ANN in handling long time series, achieving the best predictive accuracy for semi-arid regions [9]. More recently, Verma and Pandey (2025) suggested that Transformer-based architectures hold great promise for future drought forecasting, as they can effectively manage long sequences and multi-variable datasets [2].

3. Comparative Analysis (2020–2025)

To better understand the progress made in drought prediction research between 2020 and 2025, a comparative analysis of selected studies has been conducted. Table 1 summarizes key works across Central India and similar regions, highlighting datasets, drought indices, modeling techniques, evaluation metrics, and major findings. This comparative evidence demonstrates the transition from traditional ML approaches such as RF and SVM towards advanced hybrid deep learning models like CNN–LSTM.
While early studies reported that RF was effective in short-term drought detection [7], recent work has shown that deep learning models such as LSTM and CNN–LSTM significantly outperform ML methods in capturing long-term climatic variability [5,9]. Ensemble and hybrid models are particularly valuable in handling non-linear interactions across multiple variables. Notably, Verma and Pandey (2025) [2] provide a regional comparison that establishes ML models as strong baselines for Central India, against which advanced DL methods can be benchmarked in future studies.

Table 1. Comparative Analysis of Drought Prediction Studies (2020–2025)

Author (Year)	**Region / Dataset**	**Index Used**	**Model(s)**	**Metrics**	**Key Findings**
Bisht et al. (2023) [7]	Bundelkhand, India	SPI, SPEI	RF, SVM	$R^2 = 0.85$	RF better in monsoon drought detection
Singh et al. (2023) [5]	Central India	SPEI	LSTM	$R^2 = 0.92$, RMSE = 0.41	LSTM effectively captures long-term drought trends
Dwivedi et al. (2024) [6]	Ken Basin, India	SPI	RF, ANN	Accuracy = 88%	ANN outperformed RF in multi-scale drought prediction
Galkate et al. (2024) [8]	Madhya Pradesh, India	Rainfall, SPI	XGBoost	$R^2 = 0.87$	XGBoost achieved superior spatial drought accuracy
Kumar et al. (2024) [9]	Semi-arid MP, India	Rainfall + Temp	CNN–LSTM	$R^2 = 0.94$	Hybrid CNN–LSTM model best for long time sequences

Verma & Pandey (2025) [2]	Satpura & Bundelkhand	SPEI	RF vs. SVM	$R^2 = 0.9942$ (RF) & $R^2 = 0.7513$ (SVM)	RF best in short-term, SVM better for long-term; ML baseline for Central India

From the comparative analysis, several patterns emerge:

- **ML models (RF, SVM, XGBoost)**: Robust for short-term and regional drought predictions. RF consistently shows high performance in monsoon droughts, while SVM adapts better for long-term horizons.
- **ANN-based approaches**: Capable of outperforming RF in some medium-term predictions, as seen in Dwivedi et al. (2024) [6].
- **DL models (LSTM, CNN–LSTM)**: Provide superior accuracy for long sequences and multi-variable datasets, making them highly suitable for climate variability prediction.
- **Hybrid/ensemble approaches**: Expected to dominate future research due to their ability to integrate strengths of ML and DL.
- **Baseline role of ML**: Studies like Verma & Pandey (2025) [2] highlight the importance of ML models as reliable baselines for Central India, establishing a foundation upon which DL models can be benchmarked.

4. Discussion

The comparative analysis of recent studies (2020–2025) highlights critical trade-offs between models for deep learning (DL) and machine learning (ML) for drought forecasting. ML techniques like Support Vector Machines (SVM), Random Forest (RF), and XGBoost continue to provide robust and efficient solutions for short-term drought monitoring. They are computationally inexpensive, easier to interpret, and suitable for regions with limited resources [8,9]. However, these models often fall short in capturing long-term temporal dependencies that characterize drought progression. By contrast, DL models such as Long Short-Term Memory (LSTM) and hybrid CNN–LSTM architectures have demonstrated superior performance in representing complex, non-linear climate interactions and extended time sequences [6,10]. These models outperform ML approaches in terms of predictive accuracy and generalization, particularly in capturing seasonal and interannual variability. The major limitation of DL methods remains their computational cost and dependence on large datasets, which may restrict adoption in resource-constrained settings [5].

Regional context further influences model suitability. In semi-arid Bundelkhand, characterized by frequent rainfall deficits, RF and similar ML models perform well for operational short-term forecasts [9]. Conversely, in relatively humid regions such as Satpura, where evapotranspiration-driven drought intensity has increased in recent decades, DL models like LSTM provide better insights into long-term dynamics [2]. Thus, no single model is universally optimal; instead, model choice should depend on regional climatic conditions and the forecasting horizon. Despite these advances, several research gaps remain. Transformer-based architectures, which have revolutionized sequence modeling in natural language processing, are only beginning to be applied in hydroclimatic prediction and hold promise for future drought studies [10]. Moreover, most research still relies on rainfall and temperature data, neglecting multi-modal datasets such as vegetation indices (NDVI), soil moisture, and evapotranspiration. Integrating these diverse datasets with advanced DL models could significantly improve prediction accuracy and robustness, paving the way for more holistic drought monitoring frameworks [7].

5. Conclusion and Future Work

This review demonstrates that ML models, especially RF and SVM, remain **strong baselines** for drought prediction and are particularly well-suited for short-term, operational use [2,8]. At the same time, DL models such as LSTM and CNN–LSTM have achieved notable improvements in long-term drought forecasting, outperforming ML techniques in their ability to capture temporal dependencies [6,10]. Looking ahead, hybrid approaches that combine the strengths of ML and DL are likely to dominate, alongside the integration of **satellite-based indices** (NDVI, soil moisture) and advanced architectures such as Transformers. Transfer learning is another promising direction for regions with limited historical datasets [10].

For policy-makers, the choice of model should depend on both **region and timescale**. Semi-arid regions like Bundelkhand may rely on computationally efficient ML approaches for operational monitoring, while humid but increasingly vulnerable regions such as Satpura may benefit more from DL-based systems capable of capturing long-term variability [2]. Overall, the 2020–2025 period has established a strong foundation for data-driven drought forecasting. The next phase must focus on integrating multi-source datasets, improving scalability, and aligning computational advances with practical needs in agriculture, water management, and climate adaptation.

References

[1] IPCC (2021). *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. https://www.ipcc.ch/report/ar6/wg1

[2] Verma, A., & Pandey, V. (2025). A regional comparison of drought intensification and prediction using SPEI and ML models in Central India. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 12(8), 728–737.

[3] Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, 23(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

[4] Boulfiza, M., & Harrou, F. (2022). Machine learning techniques for drought prediction: A review. *Environmental Monitoring and Assessment*, 194(4), 255. https://doi.org/10.1007/s10661-022-09974-0

[5] Singh, A., Sharma, V., Jayal, T., & Deswal, S. (2023). Increased frequency and severity of droughts in Central India due to climate change. *Scientific Data*, 10(1), 256. https://doi.org/10.1038/s41597-023-02856-y

[6] Dwivedi, R., Mishra, S., & Tiwari, P. (2024). Application of ARIMA and SPI for medium-term drought forecasting in the Ken Basin, India. *Natural Hazards*, 120(2), 789–805. https://doi.org/10.1007/s11069-024-06789-2

[7] Bisht, A., Yadav, N., & Gautam, R. (2023). Comparative performance of machine learning models for drought forecasting in Bundelkhand region. *Environmental Earth Sciences*, 82(4), 215. https://doi.org/10.1007/s12665-023-11256-5

[8] Galkate, R., Sharma, K., & Patel, M. (2024). XGBoost-based spatial drought prediction using rainfall and SPI data in Madhya Pradesh. *Theoretical and Applied Climatology*, 156(3–4), 1021–1035. https://doi.org/10.1007/s00704-024-04356-1

[9] Kumar, S., Raghuwanshi, N. S., & Singh, R. (2024). Hybrid CNN-LSTM model for drought prediction in semi-arid regions of India. *Stochastic Environmental Research and Risk Assessment*, 38(7), 2105–2120. https://doi.org/10.1007/s00477-024-02498-3

[10] Gao, S., Huang, Y., & Chen, J. (2023). Transformer-based deep learning models for hydrological and climate time series prediction: A review. *Environmental Research Letters*, 18(9), 093004. https://doi.org/10.1088/1748-9326/aceb44

