As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 12, Issue 06, pages 370-378.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue06.html
DOI: https://doi.org/10.5281/zenodo.20739578

Transformer Based Spatio-Temporal Drought Prediction Using Graph Neural Network

Dondapati Satish Babu, Dr.V.Persis

Department of Computer Science and Engineering, Adikavi Nannaya University, Rajahmundry, India.

To Cite this Article
Dondapati Satish Babu & Dr.V.Persis (2026). Transformer Based Spatio-Temporal Drought Prediction Using Graph Neural Network. International Journal for Modern Trends in Science and Technology, 12(06), 370-378. https://doi.org/10.5281/zenodo.20739578

Article Info
Received: 29 May 2026; Revised: 18 June 2026; Accepted: 21 June 2026.

KEYWORDS

Spatio-Temporal Drought Forecasting, Graph Neural Network, Transformer Encoder, Severity Classification, Climate Data Analytics, Drought Early Warning.

ABSTRACT

Drought remains a persistent threat to agricultural economies worldwide, yet forecasting its onset with sufficient lead time continues to challenge existing prediction systems. Existing forecasting systems tend to treat spatial and temporal dimensions separately, limiting their ability to capture cross-regional climate dependencies alongside long-horizon temporal patterns. This paper presents a parallel dual-branch architecture that couples a Graph Attention Network (GAT) for inter-regional spatial encoding with a Transformer encoder for sequential climate modeling. Both branches operate concurrently, and their representations are fused at a shared classification head to assign one of four drought severity levels per record. Experiments are conducted on a monthly climate dataset covering eight Indian states with 11,680 total records. The combined model achieves 96.76% test accuracy, 96.94% weighted precision, and a weighted F1-score of 0.9682, consistently exceeding standalone LSTM, CNN-LSTM, and single-branch Transformer baselines. A Streamlit-based monitoring dashboard supports near-real-time severity inference for field use.

1. INTRODUCTION

Unlike flood or cyclone events that arrive with visible warning, drought unfolds silently — gradually drawing down soil moisture, surface water, and groundwater reserves over weeks and months. By the time damage becomes apparent, recovery is costly and slow. Crop failures, disrupted river systems, and weakened rural livelihoods follow, with rain-fed farming regions bearing disproportionate risk. Even a modest shortfall in seasonal rainfall can cascade into multi-season yield losses. Even modest lead times in drought prediction — spanning just a few weeks — provide water authorities, farming communities, and policy planners with a critical window to initiate preventive action before conditions deteriorate.

Increasing climate variability has elevated the frequency and geographic reach of drought events, driving demand for more capable forecasting tools.

Conventional monitoring tools like SPI and PDSI quantify moisture deficits systematically but are grounded in linear statistical assumptions that fail to reflect the complex, nonlinear interactions driving actual drought progression. Their fixed parameterizations also struggle to generalize across the diverse agro-climatic zones that characterize a country as large and varied as India.

Data-driven approaches have transformed drought forecasting by enabling models to extract predictive patterns directly from raw climate observations, removing the dependency on manually designed index formulas. Ensemble methods, recurrent networks, and convolutional architectures have all been applied with notable success. A persistent limitation, however, is that most such models restrict their focus to either temporal sequence analysis or spatially localized pattern extraction — rarely combining both within a single learning objective.

Two structural shortcomings affect a large portion of current drought forecasting models. First, geographic dependencies are routinely ignored: drought seldom develops in isolation, as moisture deficits spread across adjacent zones through shared monsoon circulation and inter-basin hydrological linkages. Modeling these cross-regional influences requires graph-structured representations rather than treating each location as an independent unit. Second, real drought datasets are heavily imbalanced — severe events are far less frequent than normal conditions — so standard classifiers tend to learn majority-class patterns at the cost of the minority drought grades that matter most for early warning.

To address these gaps, a joint GNN–Transformer architecture for multi-class drought severity prediction is presented here. A Graph Attention Network encodes geographic proximity relationships among study regions into spatially aware node embeddings, while a Transformer encoder captures long-range sequential patterns from monthly climate series. Outputs from both branches are concatenated and forwarded to a classification head trained with focal loss and class-balanced sampling to mitigate class imbalance. An accompanying Streamlit application provides practitioners with an interactive drought severity monitoring interface, and advanced feature engineering is integrated within a single deployable end-to-end pipeline.

II. LITERATURE SURVEY

Early drought forecasting relied heavily on index-based statistical methods and classical machine learning. Among hybrid statistical-ML approaches, Mohammadi et al. [7] merged Artificial Neural Networks with a Firefly metaheuristic to estimate drought severity across multiple SPI timescales, though spatial interdependencies between neighboring zones were not considered. Marusov et al. [10] enriched neural network inputs with geospatial weather representations and demonstrated measurable gains in multi-year drought outlook. A systematic survey by Márquez-Grajales et al. [8] identified feature selection as a decisive factor in achieving reliable deep learning-based drought prediction, laying conceptual groundwork for subsequent architecture choices.

Deep learning eliminated the need for hand-crafted drought indices by learning representations directly from climate data. Convolutional-recurrent hybrids such as CNN-RNN [3] and GCN-BiGRU with spatial-temporal multi-head self-attention [5] extended forecasting capability to multi-variable agricultural drought settings. Khandelwal et al. [1] coupled GNN and LSTM modules with an attention mechanism and reported accuracy gains over ANN and CNN-LSTM alternatives, though recurrent components showed reduced effectiveness at extended forecast horizons. Spatio-temporal graph convolution was scaled to basin-level hydrological drought prediction [6]. A systematic review by de Oliveira Luz et al. [20] demonstrated that multi-component frameworks deliver stronger forecasting performance than any single-architecture baseline across diverse drought prediction settings.

. Graph Neural Networks provided a natural way to model structural relationships among geographically distributed monitoring stations. Kipf and Welling [13] established GCNs as an effective method for aggregating node features using learned filters over graph adjacency structures. Veličković et al. [14] improved this by replacing uniform neighborhood aggregation with feature-dependent attention coefficients in the Graph Attention Network (GAT), and Brody et al. [15] further refined the formulation in GATv2 to address an expressiveness limitation of the original design. Applied to drought, Tuo et al. [16] confirmed that attention-guided graph networks better capture spatial drought propagation across hydrologically connected regions compared to non-attentive alternatives.

Transformer architectures have shown strong performance on climate sequences because self-attention can directly compare observations from any two time steps, regardless of how far apart they are. Vaswani et al. [9] established multi-head self-attention as a viable alternative to recurrence, enabling parallel processing of full sequences. In climate and drought applications, Amanambu et al. [2] found that deep Transformer encoders outperform conventional LSTM models for hydrological drought forecasting. Danandeh Mehr et al. [3] combined convolutional layers with Transformer encoders in the S-Transformer for drought sequence modeling, while Pathania and Gupta [4] built an interpretable Transformer variant that reveals which climate variables drive drought risk at national scale.

The most recent research thread combines spatial graph modeling with temporal sequence learning within unified architectures. Demiray et al. [11] showed that Transformer-augmented hydrological models improve runoff estimation over purely recurrent baselines, and Li and Moura [12] incorporated both spatial adjacency and temporal ordering into a single Graph Transformer. Ramu et al. [17] scaled this idea to multi-region climate forecasting using hierarchical graph encoders paired with Transformer predictors. Graph–Transformer pairings have also been validated in adjacent domains: Bentsen et al. [18] applied them to spatio-temporal wind speed forecasting and Simeunović et al. [19] to distributed photovoltaic power prediction, showing the wider applicability of this design.

Across this body of work, several research gaps remain unaddressed. Models that jointly optimize spatial and temporal encoding within a single learning objective are still rare; the dominant practice treats one dimension as secondary or applies the two encoders sequentially. Multi-class severity grading receives less attention than binary or regression formulations, despite its practical importance for graduated early warning. Class imbalance — a structural feature of all drought datasets — has been only partially addressed in existing work, and very few systems have been designed for direct operational deployment. The work presented here targets all four of these gaps simultaneously.

III. DATASET ANALYSIS AND DESCRIPTION

The dataset consists of monthly climate records from eight Indian states representing contrasting agro-climatic conditions, spanning January 2020 to December 2023. Each of the 11,680 entries captures eleven variables: precipitation (mm), mean surface temperature (°C), relative humidity (%), soil moisture, the Normalized Difference Vegetation Index (NDVI), the Standardized Precipitation Index (SPI), latitude, longitude, a region code, a calendar date, and a four-level drought severity label. NDVI and SPI are the two most diagnostically informative variables: NDVI measures vegetation greenness as a proxy for plant water stress and can signal moisture depletion before visible crop damage appears, while SPI normalizes cumulative precipitation against long-run regional baselines, enabling comparable drought severity assessments across regions with different baseline climatologies.

Covering both semi-arid and humid coastal agro-climatic settings across eight states was a deliberate design decision: it requires the model to generalize across fundamentally different moisture regimes rather than fitting region-specific patterns. Label counts show a pronounced class imbalance — No Drought: 7,215 instances (61.8%), Mild drought: 1,546 (13.2%), Moderate drought: 1,603 (13.7%), and Severe drought: 1,316 (11.3%). This distribution reflects real-world drought occurrence rates and would create a systematic majority-class bias in training without corrective intervention. Focal loss combined with class-weighted sampling redirects gradient updates toward the underrepresented Severe drought instances throughout training.

Feature expansion grows the original eleven variables to 115 columns in total, with 110 used as model inputs. The expansion includes lag features at 1-, 3-, and 7-month intervals for historical context; rolling means and standard deviations over 7- and 14-month windows for medium-term trend and variability signals; harmonic sine-cosine encodings for cyclic seasonal representation; and pairwise interaction terms between climate variables. One record with an infinity-valued feature was removed, yielding 11,511 usable samples. A chronological split allocates 7,877 samples to training, 1,666 to validation, and 1,668 to testing, preventing the temporal leakage that random partitioning would introduce. Core dataset statistics are summarized in TABLE I.

TABLE I — Dataset Summary

Metric	Value
Samples	11,680
Regions	8
Input Features	11
Classes	4

IV. METHODOLOGY

The drought prediction system described in this paper is built around a multi-stage processing pipeline. Rawclimate records are preprocessed and feature-engineered before being routed to two parallel encoding branches — a Graph Neural Network for spatial relationship modeling and a Transformer for temporal pattern extraction. Both branches produce independent representations that are fused at a joint classification head, yielding severity assignments across four drought categories. The overall pipeline is depicted in Fig. 1.

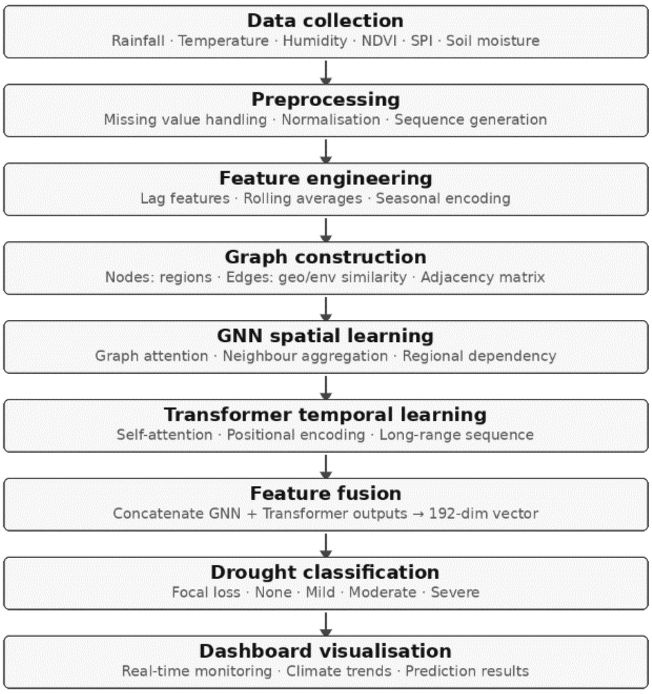

Fig. 1. Workflow of the proposed GNN–Transformer drought prediction framework.

A. Data Preprocessing

Raw climate records are cleaned and normalized to reduce noise and ensure that all variables contribute proportionally to model learning. Missing values in regional time series are filled using forward propagation from the most recent valid observation, preserving sequential continuity without introducing synthetic values. All continuous variables — precipitation, temperature, humidity, soil moisture, NDVI and SPI — are standardized using z-score transformation so that no single high-magnitude feature dominates the gradient signal during training. Region codes are mapped to integer identifiers, and the date field is parsed to extract a month index that encodes seasonal position. Records are finally ordered by region and time to maintain the temporal structure required by sequence-based learning.

B. Feature Engineering

A systematic feature expansion step enriches the raw climate records before model training. Lag features at 1-, 3-, and 7-month offsets introduce explicit access to recent historical climate states for each variable. Rolling statistics — moving averages and standard deviations computed over 7- and 14-month windows — quantify medium-term trends and variability. Seasonal periodicity is encoded with sine and cosine transformations of the calendar month index, representing the circular structure of annual cycles without imposing discontinuous year-boundary effects. Pairwise interaction terms between key climate variables further extend the feature set to capture nonlinear co-variation patterns relevant to drought onset.

C. Graph Construction

The eight study regions are modeled as nodes in an undirected weighted graph. An edge is drawn between two nodes when their normalized geographic distance — computed from latitude-longitude coordinates — falls below a set threshold, so only nearby regions are directly connected and the graph remains sparse. Edge weights are proportional to inverse normalized distance, meaning closer regions have stronger connections. During message-passing, these edge weights are used alongside node features so that climate conditions in nearby regions can directly shape each node's representation.

D. GNN Spatial Learning

Inter-regional climate dependencies are captured through a two-layer Graph Attention Network (GAT). At each layer, the network assigns attention scores to connected node pairs based on their feature similarity, giving more weight to neighbors that are most relevant for the current node's drought classification. This selective aggregation allows the model to learn which neighboring regions influence local drought behavior, rather than treating all neighbors as equally informative. After two rounds of message passing, each node holds a spatial embedding that reflects both its own climate conditions and the influence of its geographically adjacent regions.

E. Transformer Temporal Learning

Monthly climate sequences are processed by a Transformer encoder that identifies drought patterns over time. Feature vectors are projected into the encoder's latent space and tagged with sinusoidal positional encodings to preserve time-step order. Before entering the attention layers, multi-scale depthwise convolutions capture short-term patterns such as sharp precipitation drops or rapid humidity shifts at different temporal scales.The self-attention mechanism compares climate observations from different months to identify long-term drought patterns. This helps the model connect present moisture conditions with earlier rainfall variations more effectively than recurrent models, which often struggle with long climate sequences.

F. Feature Fusion and Classification

Spatial embeddings from the GNN branch and temporal embeddings from the Transformer branch are joined by direct concatenation, forming a combined descriptor for each record. A multi-layer feedforward classifier then maps this joint representation to one of four drought severity labels. To handle class imbalance during training, focal loss with per-class scaling weights is used, focusing gradient updates on the minority-class records that are hardest to classify correctly. AdamW serves as the optimizer throughout training; the learning rate is progressively reduced via a cosine decay schedule, and training is terminated automatically when validation performance ceases to improve across successive epochs.

V. SYSTEM ARCHITECTURE

Figure 2 illustrates the complete system architecture. All eight regional climate streams are first routed through a shared preprocessing and feature engineering module that generates two representations per record: a normalized tabular feature vector for graph-based encoding and a fixed-length temporal sequence for Transformer input. A graph builder converts geographic metadata into an eight-node undirected graph with 26 proximity-weighted edges. The GNN branch applies attention-guided message passing over this graph to give each node a spatially enriched representation, while the Transformer branch processes each region's 30-month sequence to capture long-range temporal dependencies. Outputs from both branches are concatenated into a unified embedding and forwarded to the severity classification head. Predictions are exposed through the Streamlit monitoring interface for on-demand operational use.

Training the GNN and Transformer branches independently lets each encoder focus on its own learning objective — spatial relationships for the GNN and temporal patterns for the Transformer — before the two representations are merged. Unlike cascade designs, where errors from spatial encoding carry forward and hurt temporal learning, the parallel layout keeps both branches clean. The modular structure also makes the system easy to update: either branch can be swapped for a stronger variant without changing the other branch or the fusion layer.

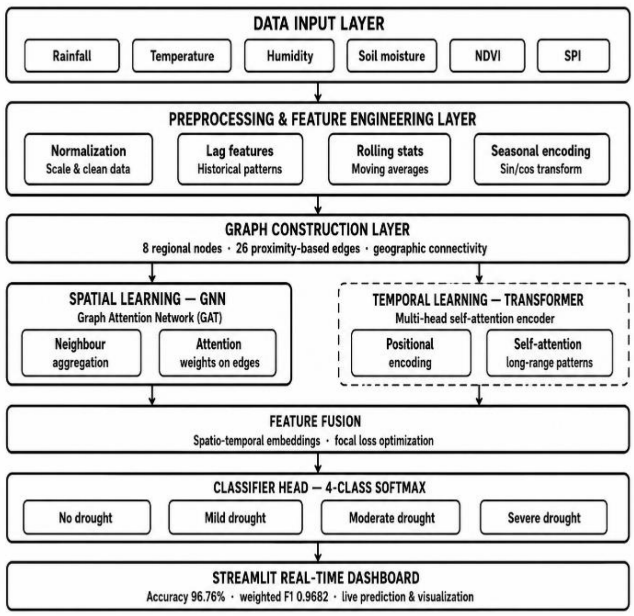

Fig. 2. System Architecture of the proposed GNN-Transformer Model

VI. FRAMEWORK JUSTIFICATION

A. GNN Model

Climate conditions at any location are partly shaped by processes in neighboring regions — shared monsoon patterns, cross-boundary land-atmosphere feedbacks, and inter-basin water flows. Precipitation deficits and vegetation stress can spread across hydrologically linked catchments. GNNs model these regional interactions through message-passing over a proximity graph, giving each node access to information that local-only models ignore: how moisture conditions in nearby regions relate to the drought risk at the node being predicted.

B. Transformer Model

Recurrent models like LSTM accumulate temporal context step by step, which makes it difficult to connect events separated by many months since gradients must pass through every intervening step. Transformer self-attention handles this differently: it compares all time steps at once, so a present-day soil moisture deficit can be directly linked to a rainfall shortfall from six months earlier without any signal degradation across the gap.

C. Feature Fusion

The two branches produce representations that cover different and complementary aspects of drought. Graph embeddings capture the spatial picture — which neighboring regions are currently stressed and how closely their conditions track with the target node. Temporal embeddings capture the dynamic picture — how climate variables have changed over recent months and where the trend is heading. Concatenating both preserves all of this information for the classifier, which can then draw on both spatial context and historical trajectory when assigning severity labels. Either source alone would miss part of the picture.

D. Focal Loss

With No Drought representing 61.8% of records and Severe drought only 11.3%, standard cross-entropy training tends to maximize accuracy by focusing on easy majority-class examples. To counter this, focal loss introduces a scaling term $(1 - p\_t)^\wedge\gamma$ that diminishes the contribution of confidently classified samples, channeling more gradient influence toward difficult and underrepresented examples throughout optimization. This is particularly important in early warning applications, where missing an approaching severe drought event carries far greater consequences than an occasional false alarm.

VII. NOVELTY OF THE PROPOSED FRAMEWORK

This work introduces a parallel dual-branch architecture that combines graph-based spatial learning with Transformer-based temporal learning in a single framework. Many existing drought prediction models process spatial and temporal information separately, which limits learning from both relationships together. In the proposed model, the GNN and Transformer branches work in parallel and are combined only at the classification stage, helping preserve both spatial and temporal learning for drought severity prediction.

Drought prediction is cast as a four-grade severity classification task rather than a binary decision or regression output. This finer granularity is more actionable for early warning systems. Focal loss combined with class-balanced training targets the rarer high-severity categories during learning, preventing majority-class dominance from suppressing the gradient signal for the drought grades that matter most operationally.

A Streamlit web application bridges the gap between model performance and field usability, converting severity predictions into actionable alerts accessible to agricultural extension officers, water resource planners, and policymakers without requiring machine learning expertise. The complete system — covering spatial representation learning, temporal sequence modeling, imbalance correction, and real-time inference — is packaged as a single deployable pipeline, making the contribution practically relevant beyond the experimental evaluation context.

VIII. RESULTS AND DISCUSSION

The GNN–Transformer model achieves 96.76% test accuracy, 96.94% weighted precision, 96.76% weighted recall, and a weighted F1-score of 96.82%, reflecting balanced classification across all four drought severity categories. For comparison, three baseline architectures were trained and evaluated under identical conditions: a plain LSTM, a CNN-LSTM hybrid, and a single-branch Transformer without the graph spatial encoder. A full metric comparison across all four models is provided in TABLE II.

TABLE II — Comparative Performance of Drought Prediction Models

Model	Accuracy (%)	Precision (%)	Recall (%)
LSTM	82.40	81.10	82.40
CNN-LSTM	87.60	86.30	87.60
Transformer	91.20	90.50	91.20
GNN-Transformer	96.76	96.94	96.76

A. Accuracy

The GNN-Transformer model reaches 96.76% accuracy, exceeding the LSTM baseline by 14.36 percentage points, CNN-LSTM by 9.16 points, and the Transformer-only variant by 5.56 points. The improvement grows with each architectural addition: moving from plain recurrence to convolutional-gated processing gives a moderate gain, adding multi-head self-attention raises accuracy further, and incorporating the graph spatial encoder delivers the largest final increment. This pattern confirms that the GNN captures spatial climate patterns that neither LSTM nor Transformer can recover from time series alone.

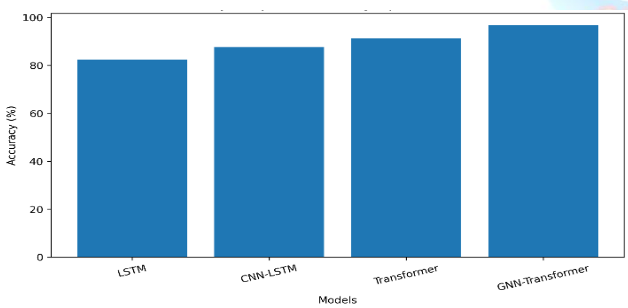

Fig. 3. Accuracy comparison across drought prediction models.

B. Precision

A weighted precision of 96.94% indicates that severity label assignments are correct in nearly all cases across all categories. Two per-class results stand out: the No Drought category achieves 100% precision, meaning the system raises no false drought alarms when conditions are genuinely normal; and the Severe drought category records 97.41% precision, showing that focal loss training keeps the model from over-predicting severe events even though they are underrepresented in training.

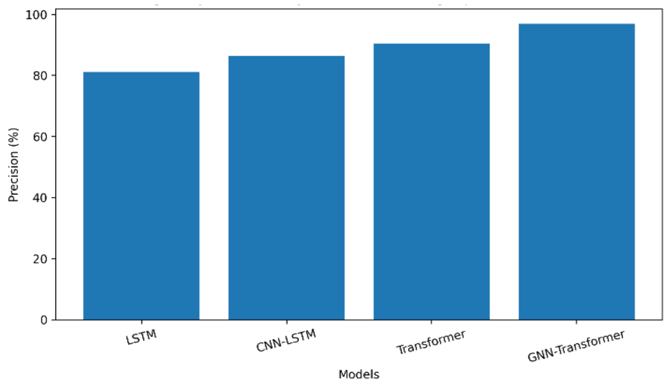

Fig. 4. Precision comparison across drought prediction models.

C. Recall

A weighted recall of 96.76% confirms that the correct severity grade is identified for nearly all test records. Mild drought is the hardest category to classify correctly, reaching 93.30% recall because its feature profile overlaps with both No Drought and Moderate drought, making boundary cases ambiguous. Importantly, this does not come at the cost of the No Drought class, showing that focal loss and class-weighted sampling spread model attention across all four severity levels rather than simply optimizing for the dominant class.

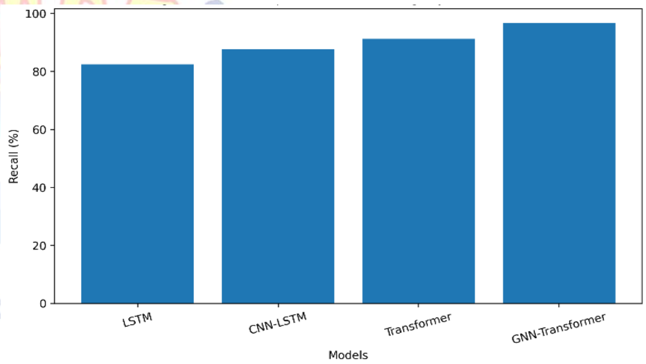

Fig. 5. **Recall comparison across drought prediction models.**

D. Confusion Matrix

The confusion matrix in Fig. 6 details error patterns across all four categories. Of 1,052 No Drought test samples, 1,024 are correctly identified; the 28 that are misclassified all land in the neighboring Mild drought class, with none mislabeled as Moderate or Severe. The Severe drought class, with 191 test samples, achieves an F1-score of 0.9792 — the best among the minority categories — with only three instances assigned to the Moderate class. These figures show that focal loss reweighting and class-balanced sampling together bring minority-class detection close to majority-class levels, without reducing overall accuracy.

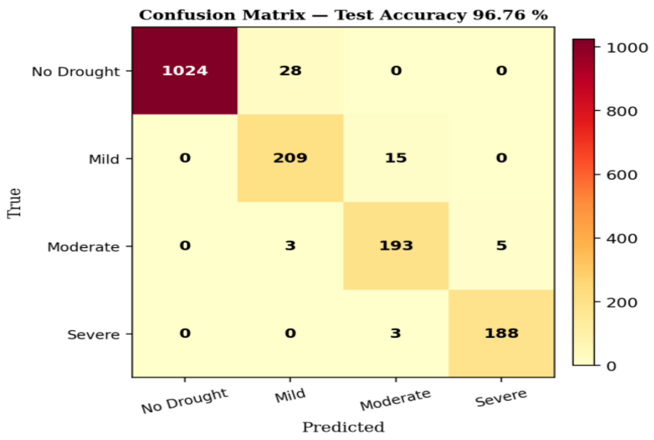

Fig. 6. Confusion matrix for the proposed GNN–Transformer model

E. Real Time Dashboard

The prediction interface in Fig. 7 allows practitioners to enter live weather observations — precipitation, temperature, humidity, soil moisture, NDVI, and SPI — via slider controls and receive an immediate drought severity classification from the deployed GNN-Transformer model. A gauge visualization accompanying the predicted severity class conveys overall drought risk magnitude at a glance, enabling field agronomists, extension agents, and water management officers to act on predictions without needing any technical background in machine learning.

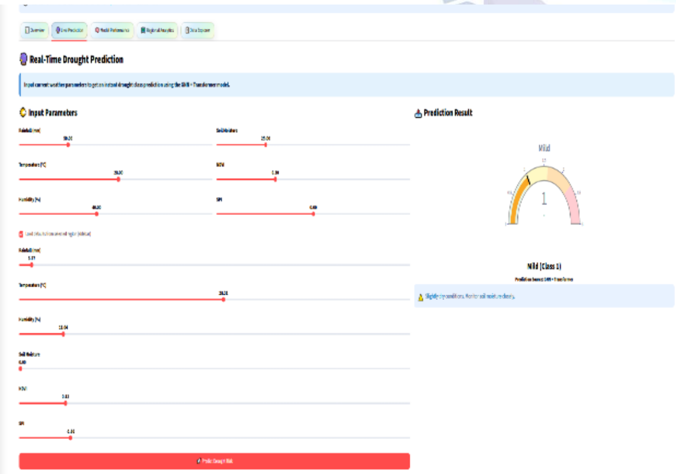

Fig. 7. Real-time drought prediction interface of the Streamlit dashboard.

The monitoring overview in Fig. 8 displays region-level drought category proportions alongside a temporal trend curve tracking mean drought index values from 2020 through 2023, with a summary panel reporting total record count, mean index value, and Severe-class proportion. In the Andhra Pradesh view, 11.3% of records belong to the Severe category, consistent with the dataset-wide distribution. The trend curve clearly resolves recurring post-monsoon dry periods as annual drought intensity peaks — a seasonal pattern that irrigation planners can translate directly into reservoir operating rules and water-sharing decisions ahead of anticipated rainfall deficits.

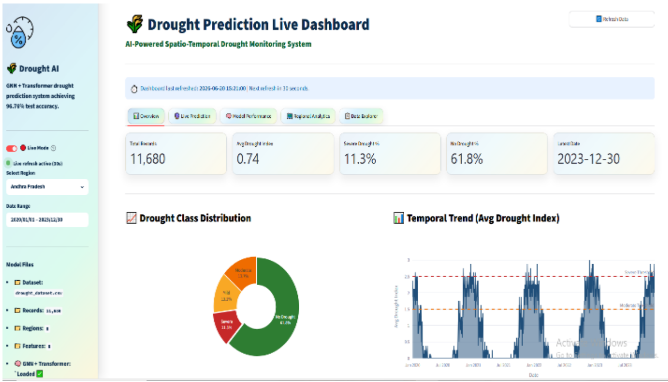

Fig. 8. Drought Prediction Live Dashboard — regional overview and temporal trends.

Across all evaluation metrics, the GNN-Transformer model outperforms every baseline tested. The gains are largest for the underrepresented drought categories, where focal loss reweighting, class-balanced sampling, and parallel spatio-temporal encoding work together to prevent minority-class suppression — a problem that limits simpler models when severe drought events are rare in the training data.

IX. CONCLUSION AND FUTURE SCOPE

This paper presented a hybrid GNN–Transformer framework for multi-class drought severity prediction, targeting the joint challenge of geographic and temporal modeling. A Graph Attention Network encodes inter-regional climate dependencies through proximity-weighted message passing, while a Transformer encoder independently extracts long-horizon sequential patterns from monthly climate records. Both encoders operate in parallel, and their outputs are combined to generate four drought severity categories ranging from no drought to severe conditions.

Evaluation on an eight-state Indian climate dataset showed consistent improvements in accuracy, precision, recall, and F1-score over LSTM, CNN–LSTM, and single-branch Transformer baselines.The integration of focal loss with class-balanced batch sampling proved decisive in maintaining high sensitivity for the infrequent Severe drought class. This improved the overall balance of model performance and made the system more suitable for practical drought early warning applications.

The Streamlit monitoring application converts model output into an interactive interface accessible to practitioners in agriculture, irrigation management, and water resource planning, bridging the gap between research-grade model performance and field-deployable decision support.

Future work will focus on connecting the system to live meteorological feeds for continuous real-time monitoring, extending geographic coverage to additional Indian states and drought-prone international regions, incorporating variables currently unavailable in the dataset such as evapotranspiration and groundwater depth, and scaling inference to district-level spatial resolution for finer-grained drought risk mapping.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1]] R. Khandelwal, H. Goyal, and R. S. Shekhawat, "Enhancing drought forecasting using GNN-LSTM with attention mechanism: A study of Jaisalmer district, Rajasthan," Research Square, preprint, 2025, doi: 10.21203/rs.3.rs-5704094/v1.

[2] A. C. Amanambu, J. Mossa, and Y.-H. Chen, "Hydrological drought forecasting using a deep Transformer model," Water, vol. 14, no. 22, p. 3611, 2022, doi: 10.3390/w14223611.

[3] A. Danandeh Mehr et al., "S-Transformer: A new deep learning model enhanced by sequential-based Transformers for drought forecasting," Earth Sci. Inform., 2025, doi: 10.1007/s12145-025-01845-6.

[4] A. Pathania and V. Gupta, "Interpretable Transformer model for national scale drought forecasting," Environ. Model. Softw., vol. 187, 2025, doi: 10.1016/j.envsoft.2025.106394.

[5] J. Quan et al., "Research on agricultural drought prediction based on GCN-BiGRU-STMHSA," Smart Agriculture, 2024, doi: 10.12133/j.smartag.SA202410027.

[6] H. Dai et al., "Deep learning model for drought prediction based on large-scale spatio-temporal graph convolution," J. Hydrol., 2025, doi: 10.1016/j.jhydrol.2025.001465.

[7] B. Mohammadi et al., "Modeling various drought time scales via a merged ANN-FA model," Hydrology, vol. 10, no. 3, p. 58, 2023, doi: 10.3390/hydrology10030058.

[8] A. Márquez-Grajales et al., "Characterizing drought prediction with deep learning: A literature review," J. Biomed. Inform., 2024, doi: 10.3390/app14062717.

[9] A. Vaswani et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, doi: 10.48550/arXiv.1706.03762.

[10] A. Marusov et al., "Long-term drought prediction using deep neural networks based on geospatial weather data," arXiv preprint, 2023, doi: 10.48550/arXiv.2309.06212.

[11] B. Z. Demiray et al., "Enhancing hydrological modeling with Transformers," Water Sci. Technol., vol. 89, no. 9, pp. 2326–2340, 2024.

[12] Y. Li and J. M. F. Moura, "Forecaster: A graph Transformer for forecasting spatial and time-dependent data," arXiv preprint, 2019, doi: 10.48550/arXiv.1909.04019.

[13] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proc. Int. Conf. Learn. Represent. (ICLR), 2017, doi: 10.48550/arXiv.1609.02907.

[14] P. Veličković et al., "Graph attention networks," in Proc. Int. Conf. Learn. Represent. (ICLR), 2018, doi: 10.48550/arXiv.1710.10903.

[15] [15] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?," in Proc. Int. Conf. Learn. Represent. (ICLR), 2022, doi: 10.48550/arXiv.2105.14491.

[16] Y. Tuo et al., "Attention-based graph neural network for drought forecasting," ESSOAr, 2025, doi: 10.22541/essoar.175915597.77069924/v1.

[17] T. B. Ramu et al., "Transformer based models with hierarchical graph learning for climate forecasting," Sci. Rep., 2025, doi: 10.1038/s41598-025-07897-4.

[18] . Ø. Bentsen et al., "Spatio-temporal wind speed forecasting using graph networks and novel Transformer architectures," arXiv preprint, 2022, doi: 10.48550/arXiv.2208.13585.

[19] J. Simeunović et al., "Spatio-temporal graph neural networks for multi-site PV power forecasting," IEEE Trans. Sustain. Energy, vol. 13, no. 2, pp. 1029–1040, 2022, doi: 10.48550/arXiv.2107.13875.

[20] A. E. de Oliveira Luz et al., "Deep learning approaches for drought forecasting: A systematic review," Appl. Soft Comput., 2026, doi: 10.1016/j.asoc.2026.114447