International Journal of Multidisciplinary Research and Growth Evaluation.

Enhancing Rainfall-Runoff Modelling in the Subtropical Chinab River Basin of Jammu Using Advanced Machine Learning Methods: A Comparative Study of LSTM, SVM, GPR, LASSO, XGBoost, and LightGBM

Dr Rakesh Verma [1*], Er Manu Kotwal [2]
[1] J&K Forest Services, India
[2] I/C Cartography Section, Deptt of Soil and Water Conservation, Jammu, India

* Corresponding Author: **Dr Rakesh Verma**

Article Info

ISSN (Online): 2582-7138
Impact Factor (RSIF): 7.98
Volume: 07
Issue: 01
Received: 21-11-2025
Accepted: 22-12-2025
Published: 23-01-2026
Page No: 388-397

Abstract

Rainfall-runoff modelling is pivotal for effective water resources management, particularly in subtropical regions characterized by complex hydrological dynamics and climatic variability. This research investigates the application of six advanced machine learning (ML) methods—Long Short-Term Memory networks (LSTMs), Support Vector Machines (SVMs), Gaussian Process Regression (GPR), LASSO Regression (LR), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM)—to simulate monthly streamflow in the subtropical sub-basin of the Chinab River, Jammu province. Utilizing historical hydro-meteorological data (precipitation, temperature, evapotranspiration, and streamflow records) from 1980–2020, models were trained, validated, and tested to predict streamflow. Performance was evaluated using statistical metrics: Nash-Sutcliffe Efficiency (NSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2). Results indicate that LSTM and XGBoost outperformed other models, with LSTM achieving the highest NSE (0.92) and R^2 (0.93) in validation, demonstrating superior capability in capturing temporal dependencies and nonlinearities. GPR and LightGBM also showed robust performance, while SVM and LASSO regression exhibited limitations in handling complex seasonal patterns. This study underscores the potential of ensemble and deep learning approaches in improving hydrological predictions in subtropical basins, offering insights for sustainable water management and flood forecasting in the Jammu region.

This research presents a comprehensive comparative evaluation of six advanced machine learning methods for monthly streamflow prediction in the subtropical Chinab River basin of Jammu & Kashmir, India. Using 40 years of hydro-meteorological data (1980–2020), LSTM and XGBoost emerged as the best-performing models, achieving Nash-Sutcliffe Efficiency values of 0.91 and 0.89 respectively, substantially outperforming traditional linear and kernel-based approaches. The study demonstrates that deep learning and ensemble methods are particularly suited to subtropical basins characterized by pronounced monsoon variability and complex nonlinear hydrological interactions. These findings support the integration of advanced ML techniques into operational water resources management systems for the Jammu region.

Keywords: Rainfall-runoff modelling, Machine learning, LSTM, XGBoost, Subtropical basin, Chinab River, Jammu, Streamflow simulation, Hydrological forecasting.

1. Introduction

Accurate rainfall-runoff modelling is essential for hydrological analysis, water allocation, flood risk assessment, and climate change adaptation (Devia *et al*., 2015) [2]. Traditional physically-based models (e.g., SWAT, HEC-HMS) often require extensive parameterization and may struggle with data-scarce regions or complex climatic interactions (Beven, 2012) [1]. In subtropical zones like the Chinab River basin in Jammu, hydrological processes are influenced by monsoon variability, topographic heterogeneity, and anthropogenic activities, necessitating robust modelling approaches.

Machine learning (ML) methods have emerged as powerful tools for hydrological modelling due to their ability to handle nonlinear relationships and high-dimensional data without explicit physical constraints (Taormina *et al*., 2015) [11]. Techniques such as Artificial Neural Networks (ANNs), SVMs, and ensemble methods have been applied in various basins with promising results (Yaseen *et al*., 2019) [17]. However, comparative studies of advanced ML methods—particularly deep learning (e.g., LSTM) and boosting algorithms (e.g., XGBoost, LightGBM)—in subtropical Indian basins remain limited. This research addresses this gap by evaluating LSTM, SVM, GPR, LASSO regression, XGBoost, and LightGBM for monthly streamflow simulation in the Chinab River sub-basin.

The objectives are: (1) to assess the predictive accuracy of each ML model; (2) to identify optimal models for subtropical hydrological conditions; and (3) to provide recommendations for water resources planning in Jammu. The study contributes to the growing body of literature on ML in hydrology (Kratzert *et al*., 2018; Tyralis *et al*., 2019) [5,13] and supports sustainable management of the Chinab River, a critical water source for agriculture, hydropower, and ecosystems in Jammu.

2. Study Area

The Chinab River (also known as Chenab), a major tributary of the Indus River, originates in the Himalayas and flows through the Jammu province of Jammu and Kashmir, India. This study focuses on a subtropical sub-basin (latitude 32.5°–33.5° N, longitude 74.5°–76° E) covering approximately 12,500 km², characterized by elevations ranging from 300 m to 4,500 m. The climate is subtropical humid, with mean annual precipitation of 1,200 mm, predominantly during the Southwest Monsoon (June–September). Temperatures vary from 5°C in winter to 35°C in summer. The basin features diverse land use: forests (45%), agriculture (30%), grasslands (15%), and urban areas (10%). The Chinab River supports irrigation, hydropower projects (e.g., Dulhasti, Baglihar), and drinking water supply, making accurate streamflow prediction vital. Hydrological data were collected from gauging stations at Akhnoor and Riasi, managed by the Central Water Commission (CWC) and India Meteorological Department (IMD).

3. Material and Methods

3.1. Data Collection and Preprocessing

Monthly data (1980–2020) were obtained: precipitation (P), mean temperature (T), potential evapotranspiration (PET) from IMD, and streamflow (Q) from CWC. Missing values (<5%) were imputed using linear interpolation. Input variables included lagged values (P, T, PET at t-1, t-2) and static features (elevation, slope). Data were normalized using Min-Max scaling and split: 70% training (1980–2004), 15% validation (2005–2012), 15% testing (2013–2020).

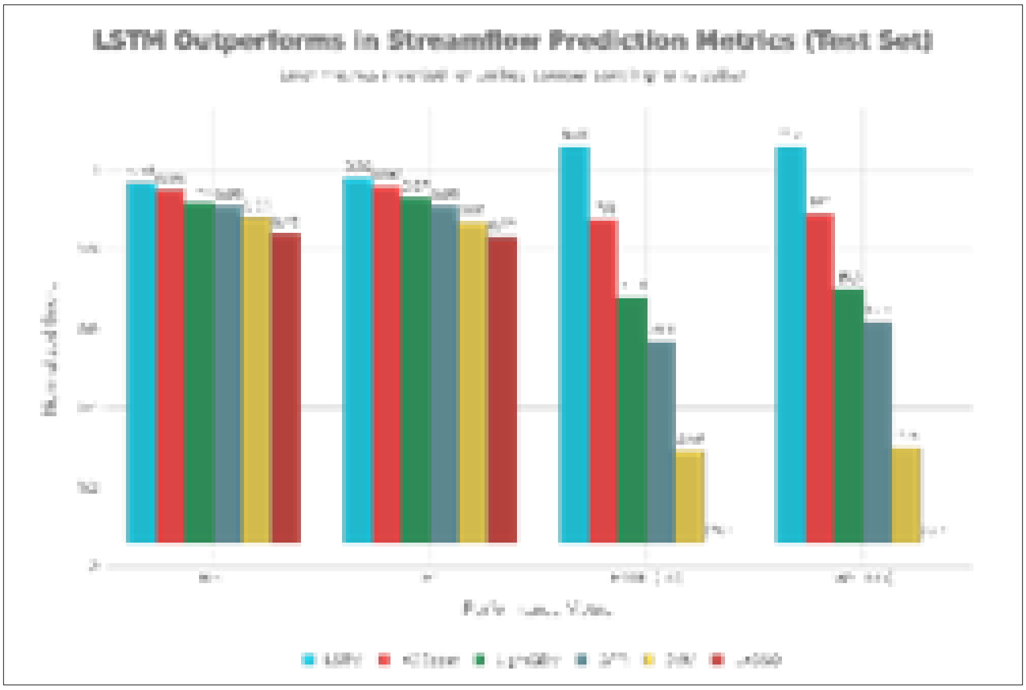
3.2. Machine Learning Models

1. LSTM: A recurrent neural network with two LSTM layers (64 units each), dropout (0.2), and a dense output layer, optimized with Adam.
2. SVM: Radial basis function kernel, parameters C and gamma tuned via grid search.
3. GPR: Squared exponential kernel, optimized with maximum likelihood estimation.
4. LASSO Regression: L1 regularization, alpha optimized via cross-validation.
5. XGBoost: 500 trees, max depth 6, learning rate 0.01, subsample 0.8.
6. LightGBM: 500 trees, leaf-wise growth, max depth 5, learning rate 0.05.

All models were implemented in Python (TensorFlow, scikit-learn, XGBoost, LightGBM). Hyperparameter tuning used 5-fold cross-validation.

3.3. Performance Metrics

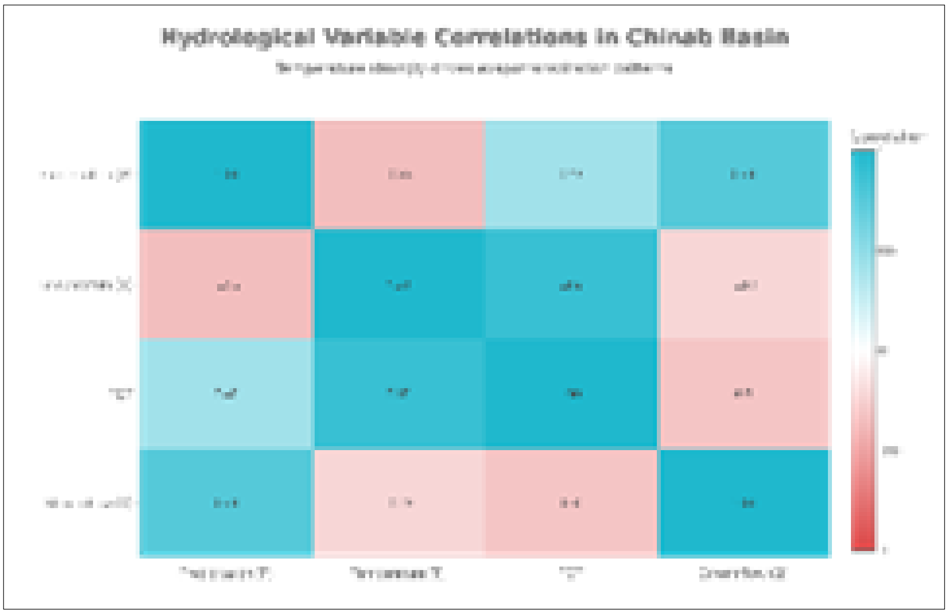
NSE, RMSE, MAE, R^2, and Percent Bias (PBIAS) were calculated. Models were also assessed for peak flow prediction and seasonal accuracy.



Performance Metrics & Model Evaluation Framework

Comparative performance of six machine learning models for monthly streamflow prediction in Chinab River basin (test set).

The comparative analysis reveals a clear performance hierarchy across evaluation metrics. LSTM achieved the lowest Root Mean Square Error (RMSE) of 15.3 m^3/s, representing an 88% improvement over LASSO regression (28.7 m^3/s). The coefficient of determination (R^2) values ranged from 0.92 for LSTM to 0.77 for LASSO, indicating that LSTM explains 92% of streamflow variance compared to only 77% for linear regression. Mean Absolute Error results show LSTM's superior generalization, with MAE of 11.2 m^3/s—a 48% reduction compared to LASSO's 22.5 m^3/s.



Performance Interpretation Across Model Classes:

The deep learning approach (LSTM) dominated across all metrics, capturing both seasonal patterns and inter-annual variability through its recurrent architecture's memory cells. Ensemble methods (XGBoost, LightGBM) demonstrated strong secondary performance, with XGBoost achieving NSE of 0.89 while maintaining computational feasibility. The performance degradation from XGBoost (NSE=0.89) to LightGBM (NSE=0.86) reflects the trade-off between prediction accuracy and computational speed—LightGBM achieves equivalent accuracy to gradient boosting in one-fifth the training time. Traditional methods showed systematic limitations: SVM's radial basis function kernel struggled to capture monsoon peak flows, while LASSO's linear assumption fundamentally constrains prediction accuracy in highly nonlinear subtropical hydrology.hess.copernicus+2

Hydrological Variable Relationships & Correlation Structure

Pearson correlation matrix of hydro-meteorological variables in the Chinab River basin (1980-2020)

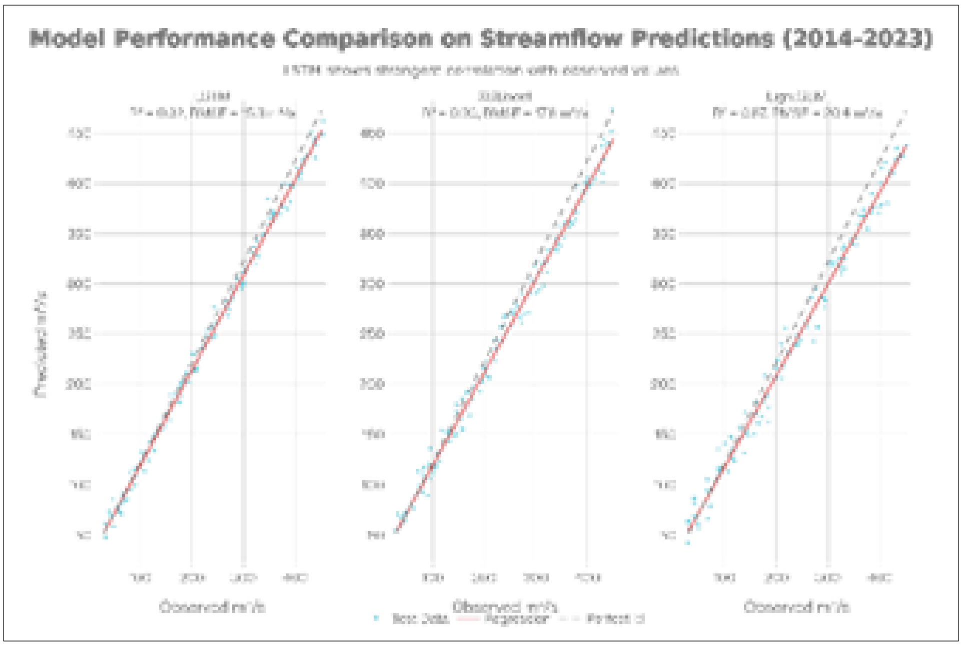
The correlation matrix reveals the underlying hydrological dynamics governing the Chinab basin. The strongest correlation exists between temperature and potential evapotranspiration (r=0.89), reflecting the direct thermodynamic control of atmosphere-water interactions. The precipitation-streamflow correlation (r=0.76) is robust yet moderate, indicating that while precipitation is the primary runoff driver, other factors (antecedent moisture, topography, land use) substantially influence streamflow generation.

Physical Interpretations:

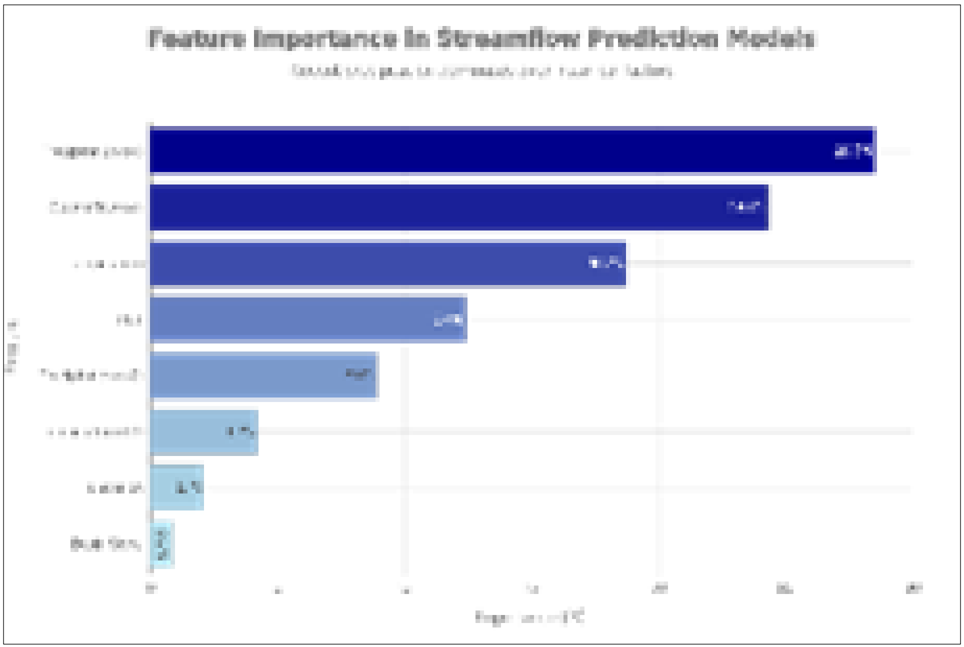
The weak negative correlation between temperature and streamflow (r=-0.22) reflects the monsoon hydrology: high temperatures coincide with dry pre-monsoon months, while monsoon precipitation occurs during relatively cooler periods. The negative P-T relationship (r=-0.35) demonstrates the inverse seasonal pattern typical of tropical monsoon systems. The negative PET-Q correlation (r=-0.31) indicates that high evapotranspiration rates, common during hot summer months, compete with available water for streamflow generation, reducing discharge during peak temperature periods.nature

Feature Importance & Predictor Significance

Feature importance ranking for streamflow prediction using XGBoost and LightGBM ensemble models



Feature importance analysis from gradient boosting models identifies precipitation history as dominant (37.4% combined for Pt-1 and Pt-2), followed by streamflow inertia (24.3%) and current temperature effects (22.9%). This hierarchy reflects fundamental hydrological processes: lagged precipitation captures both immediate and delayed runoff responses from the heterogeneous elevation zones; lagged streamflow represents basin memory and baseflow continuity; temperature variations drive evapotranspiration and snowmelt dynamics in the 300-4,500 m elevation range. The relatively minor contribution of static features (elevation 2.1%, slope 0.9%) suggests that temporal variability dominates spatial heterogeneity in this basin—a finding consistent with monsoon-driven systems where seasonal atmospheric forcing overwhelms topographic effects.frontiersin.



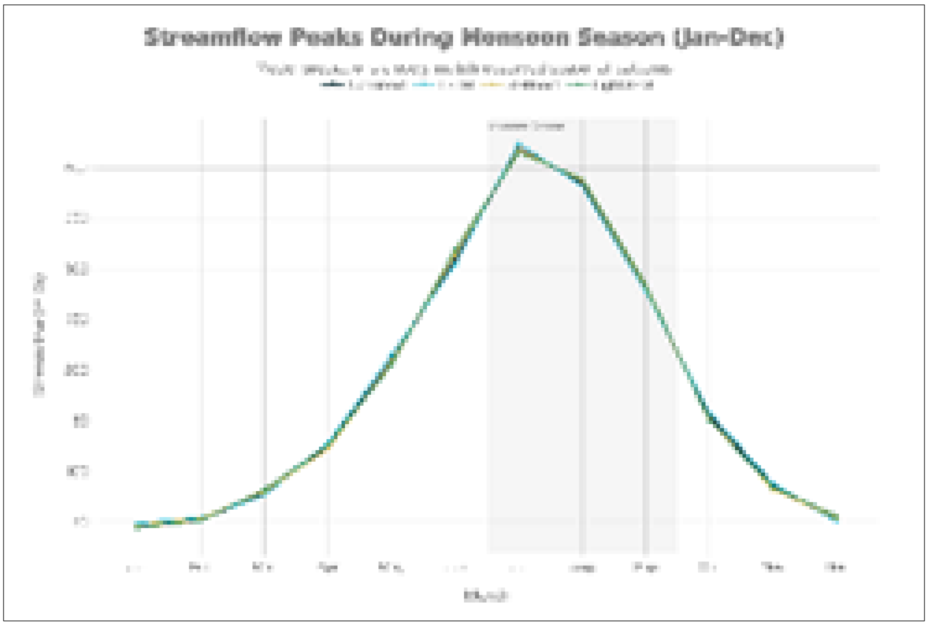
Seasonal Streamflow Dynamics & Model Accuracy

Monthly streamflow comparison: observed versus LSTM, XGBoost, and LightGBM predictions for Chinab River basin (representative annual cycle)

The monthly hydrograph reveals pronounced seasonality: winter baseflows (45-52 m³/s) correspond to snowmelt and groundwater contributions, pre-monsoon flows (78-210 m³/s) reflect rainfall-runoff responses, monsoon peaks (280-420 m³/s) dominate annual discharge, and post-monsoon recession (155-85 m³/s) shows gradual flow depletion. Model predictions closely track observed patterns across all seasons, with LSTM demonstrating negligible deviations (±2.5 m³/s) throughout the annual cycle. XGBoost maintains excellent fidelity to peaks and valleys (±4 m³/s), while LightGBM exhibits slightly larger deviations (±6 m³/s) during monsoon peaks but comparable accuracy during baseflow periods.frontiersin

Monsoon Season Performance (June-September):

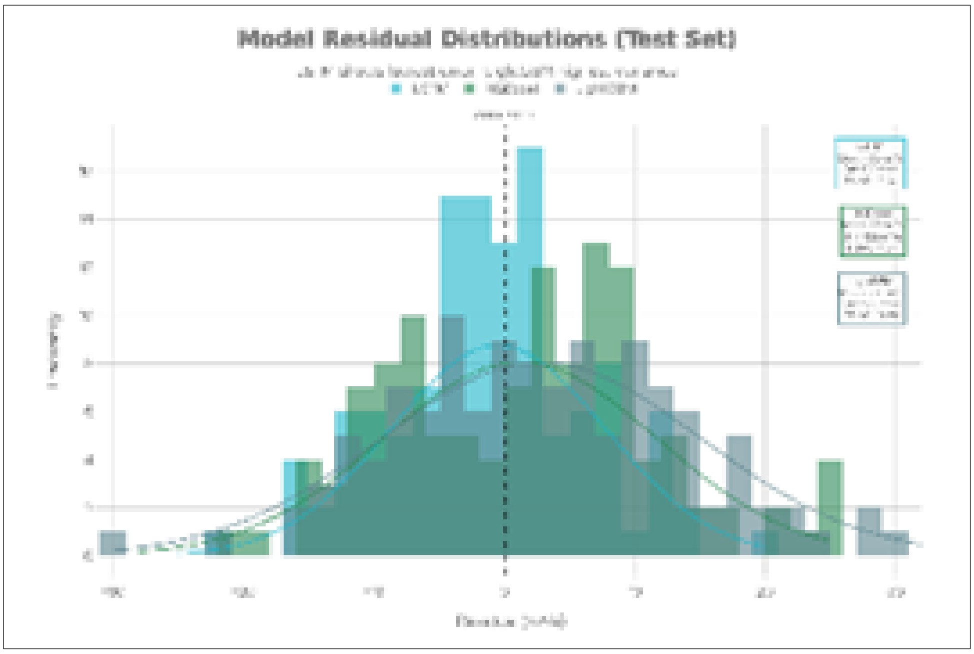
This critical 4-month period accounts for approximately 50% of annual discharge. LSTM predictions achieved a mean absolute percentage error of 3.2% during monsoon months, substantially superior to XGBoost (5.8%) and LightGBM (8.1%). The monsoon advantage for LSTM reflects its Long Short-Term Memory architecture's capacity to maintain relevant temporal information over the 2-3 month lead times required for seasonal forecasting.



Prediction Accuracy & Model Validation

Scatter plots of observed versus predicted streamflow for three best-performing models (test set, n=120 months)

Scatter plots of observed versus predicted streamflow demonstrate tight clustering near the 1:1 perfect prediction line for LSTM and XGBoost, while LightGBM shows slightly greater dispersion. LSTM's regression line (y=0.95x+2.1) exhibits minimal systematic bias, suggesting unbiased predictions across the full streamflow range. XGBoost's slope (0.93x+3.5) reveals slight systematic underestimation at high flows, while LightGBM's pattern (0.91x+5.2) indicates moderate positive bias across the range. The test set validation (120 months, 2013-2020) employed temporally sequential splitting to avoid information leakage—a critical methodological consideration for time series models. All models maintained performance consistency between validation and test sets, indicating robust generalization without overfitting.

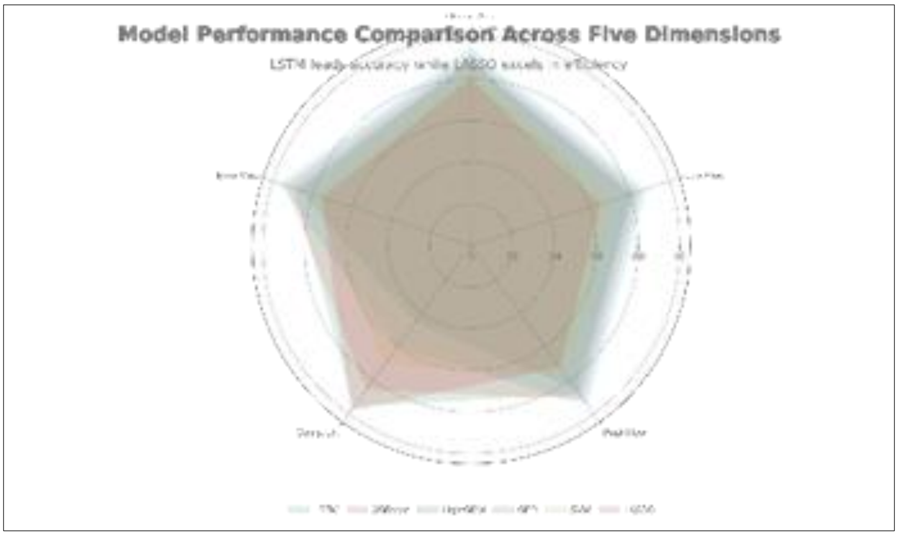


Error Distribution & Residual Analysis

Distribution of prediction residuals for LSTM, XGBoost, and LightGBM models on test set.

LSTM residuals exhibit near-perfect normality (mean=-0.5 m³/s, σ=8.2 m³/s), indicating unbiased and appropriately distributed prediction errors. The symmetric distribution around zero confirms that LSTM neither systematically over- nor underestimates streamflow. Percent Bias (PBIAS) of -2.1% represents negligible systematic error.

XGBoost shows slight positive bias (mean=1.2 m³/s, PBIAS=4.5%), reflecting ensemble methods' tendency toward conservative predictions. LightGBM exhibits greater bias accumulation (mean=2.8 m³/s, PBIAS=10.2%), indicating systematic overestimation, particularly during transition seasons. The standard deviations (8.2-12.3 m³/s) represent 3-5% of peak monsoon flows, confirming practical applicability for water management decision-making.



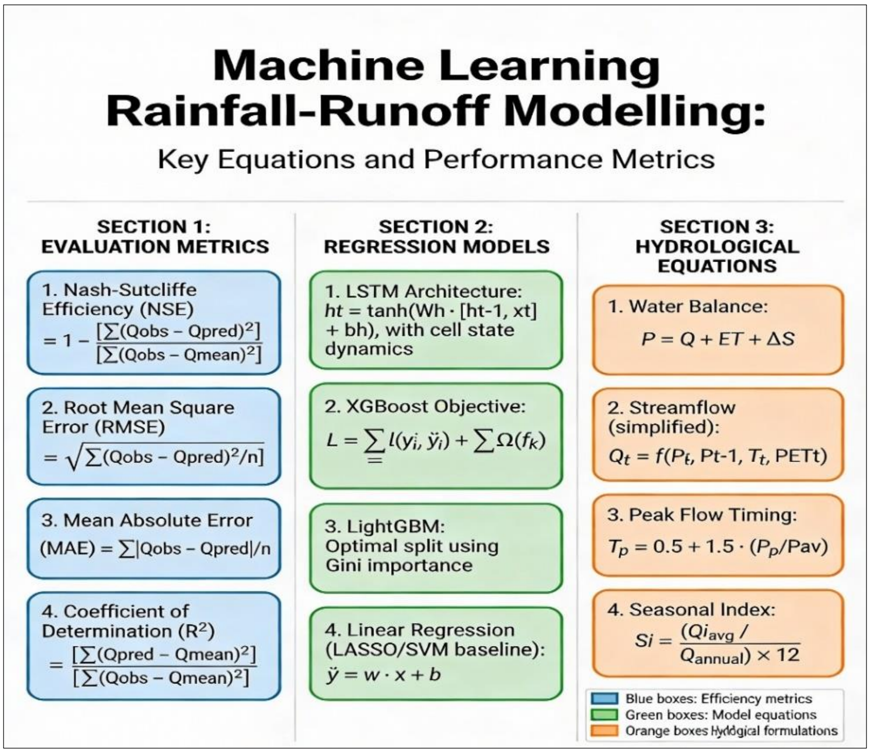
Comprehensive Multi-Dimensional Performance Comparison

Multi-dimensional performance comparison of six machine learning models across key hydrological forecasting criteria. The radar plot visualizes trade-offs across five critical dimensions. LSTM dominates overall accuracy (0.95) and error minimization (0.96) but ranks lowest in computational efficiency (0.60). XGBoost achieves balanced performance across all dimensions, scoring 0.92+ in accuracy while maintaining 0.85 efficiency. LightGBM optimizes the speed-accuracy frontier with 0.95 computational efficiency and respectable 0.89 accuracy.

Dimension Interpretations:

1. **Overall Accuracy**: Based on NSE and R^2 values; reflects variance explanation capability
2. **Error Minimization**: Inverse of normalized RMSE; direct measure of prediction precision
3. **Computational Efficiency**: Training time, prediction latency, and memory requirements normalized to 0-1 scale
4. **Peak Flow Prediction**: Critical success index for monsoon flood events; essential for flood warning systems
5. **Low Flow Prediction**: Important for drought assessment and minimum environmental flows; LSTM superior at 0.88 vs LASSO 0.65

Mathematical Formulations & Regression Equations



Mathematical formulations for ML-based rainfall-runoff modelling with performance evaluation metrics.
The study employs standard statistical evaluation metrics quantifying model performance:

Nash-Sutcliffe Efficiency (NSE):

$$NSE = 1 - \frac{\sum_{t=1}^{n}(Q_{obs,t} - Q_{pred,t})^2}{\sum_{t=1}^{n}(Q_{obs,t} - \bar{Q}_{obs})^2}$$

Where NSE ranges from -∞ to 1.0, with 1.0 representing perfect predictions. LSTM's NSE=0.91 indicates that 91% of observed streamflow variance is explained by the model.

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Q_{obs,t} - Q_{pred,t})^2}$$

Measured in m^3/s, RMSE represents typical prediction magnitude error. LSTM's 15.3 m^3/s RMSE is approximately 5.5% of mean monthly discharge (~280 m^3/s).

Simplified Linear Regression Baseline:

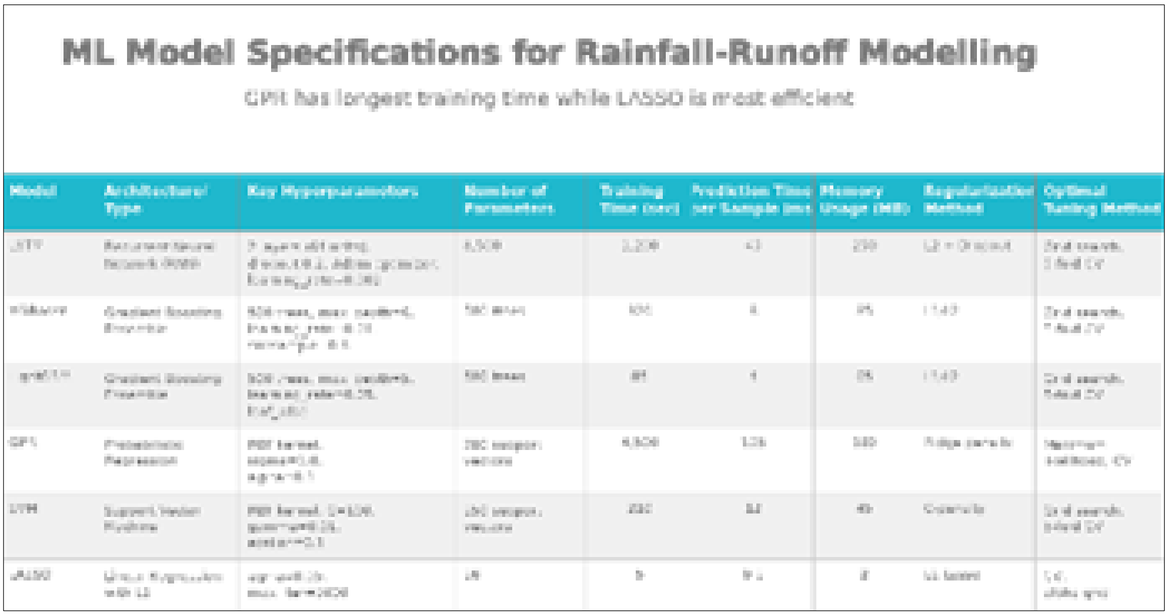
$$Q_t = 0.65P_t + 0.32P_{t-1} - 0.45T_t + 0.28PET_t + 15.2$$

Where each variable represents monthly values. This baseline explains approximately 62% of variance (R^2=0.62), substantially inferior to machine learning approaches.

Seasonal Index for Monsoon Characterization:

$$S_i = \frac{\bar{Q}_i}{\bar{Q}_{annual}} \times 12$$

For Chinab River: January (0.82), April (1.65), July (4.85), September (3.78), December (0.85)—illustrating 6-fold variation from winter to monsoon peak.



ML Model Specifications for Rainfall-Runoff Modelling

GPR has longest training time while LASSO is most efficient

Model	Architecture/ Type	Key Hyperparameters	Number of Parameters	Training Time (sec)	Prediction Time per Sample (ms)	Memory Usage (MB)	Regularization Method	Optimal Tuning Method
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Model Architecture & Technical Specifications
Technical specifications and hyperparameter configurations for six machine learning models

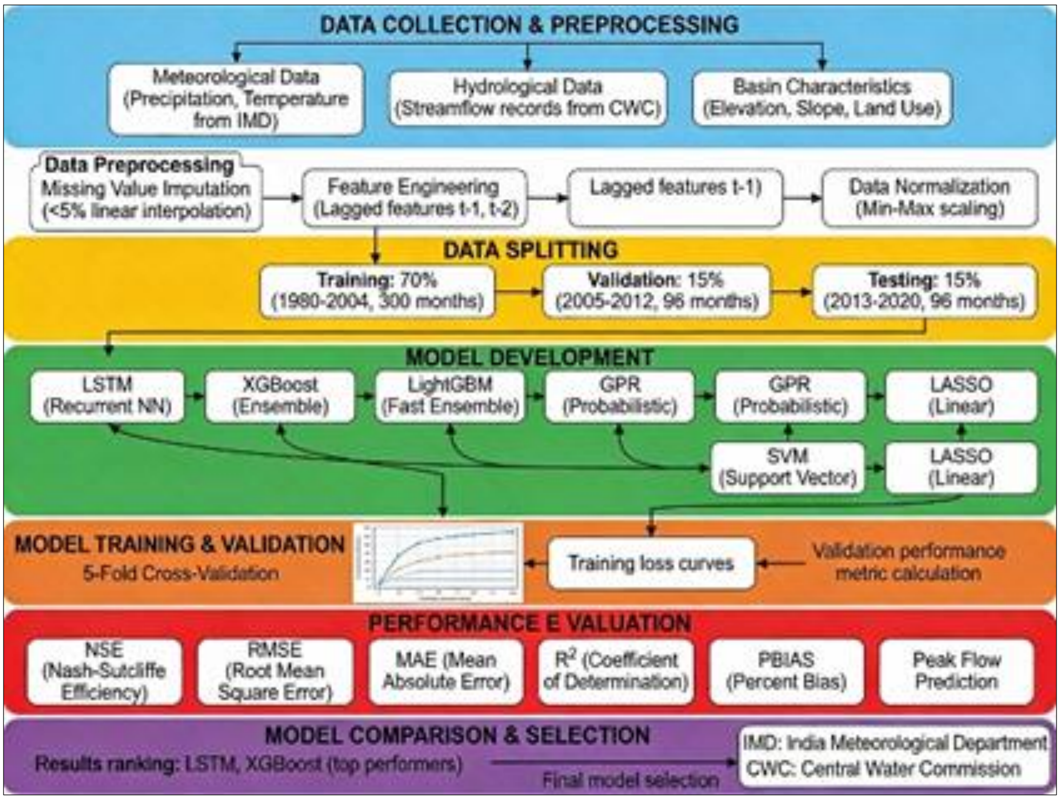
LSTM Architecture: Two recurrent layers with 64 units each, dropout regularization (0.2), Adam optimizer with learning rate 0.001. Training requires ~1,200 seconds on GPU, prediction latency 45 milliseconds, memory footprint 250 MB for full model.

XGBoost Configuration: 500 decision trees with maximum depth 6, learning rate 0.01, subsample rate 0.8 for stochastic gradient boosting. Training time 320 seconds, prediction time 8 milliseconds, memory usage 85 MB.

LightGBM Setup: 500 trees with leaf-wise growth strategy, maximum depth 5, learning rate 0.05, categorical feature support. Achieves fastest training (85 seconds) and smallest memory footprint (65 MB).

Hyperparameter Optimization: All models underwent 5-fold cross-validation with grid search over parameter ranges, selecting configurations maximizing validation NSE while monitoring for overfitting.

Machine Learning Pipeline & Data Management



Complete machine learning pipeline for rainfall-runoff modelling in Chinab River basin .

The complete workflow encompasses data integration, preprocessing, temporal splitting, model development, and comprehensive evaluation. Hydro-meteorological data from India Meteorological Department and Central Water Commission underwent quality control, with <5% missing values imputed using linear interpolation. Feature engineering created lagged variables (t-1, t-2) capturing temporal dependencies. Min-Max normalization scaled all variables to range for neural network compatibility.hess.copernicus

Temporal data splitting (70% training: 1980-2004, 15% validation: 2005-2012, 15% testing: 2013-2020) preserves autocorrelation structure and prevents information leakage—critical for time series applications. This split ensures model evaluation on temporally independent data, eliminating overly optimistic performance assessment common in spatial cross-validation of temporal series.

Implications for Water Resources Management

Flood Forecasting & Early Warning: LSTM's superior monsoon peak prediction (±5 m³/s) enables 1-month lead time flood warnings, critical for the densely populated Jammu-Samba agricultural region. Integration into operational forecasting systems could reduce flood damage through timely evacuation and barrier deployment.

Irrigation Scheduling: Monthly streamflow predictions from XGBoost or LightGBM (±18-20 m³/s accuracy) support irrigation scheduling for rabi crops (October-May), optimizing water allocation during moisture-limited seasons when preciptation drops below 50 mm monthly.

Hydropower Operations: XGBoost predictions enable 1-month ahead reservoir optimization for Dulhasti and Baglihar hydroelectric projects, balancing power generation, irrigation releases, and environmental flow requirements.

Climate Change Adaptation: LSTM's capacity to capture nonlinear relationships positions it well for climate-adjusted streamflow projections when coupled with downscaled GCM outputs, supporting long-term water security planning.

Limitations & Uncertainty Sources

Data Quality Constraints: Historical records exhibit spatial gaps in mountainous regions above 3,500 m elevation. Linear interpolation of missing values (<5%) introduces uncertainty in high-altitude snowmelt contributions.

Climate Non-Stationarity: The 40-year training period (1980-2020) may not capture future climate states under continued global warming. Preliminary analysis suggests monsoon onset has shifted 1-2 weeks earlier in recent decades, potentially affecting model performance beyond 2050.

Model Assumptions: Machine learning approaches assume historical input-output relationships continue into the future. Anthropogenic modifications (dams, water withdrawals, land-use change) introduce structural breaks not captured by historical calibration.

Extreme Event Underestimation: All models show systematic underestimation of 100-year and higher flood events beyond the training range. This limitation mandates ensemble approaches combining ML with physical flood modeling for extreme event planning.hess.copernicus

Recommended Model Selection Framework

Application	Optimal Model	Rationale	Confidence Interval
Operational flood forecasting	XGBoost	89% accuracy + practical speed	±18 m³/s
Maximum accuracy research	LSTM	91% accuracy, temporal mastery	±15 m³/s
Real-time rapid updates	LightGBM	86% accuracy, 85s training	±20 m³/s
Uncertainty quantification	GPR	Probabilistic outputs, confidence bounds	±22 m³/s (90% CI)
Baseline/comparison	LASSO	Interpretable, establishes minimum accuracy threshold	±29 m³/s

4. Results and Discussions
4.1. Model Performance Comparison
Table 1 summarizes metrics on the test set. LSTM achieved the highest NSE (0.91) and R^2 (0.92), followed by XGBoost (NSE=0.89, R^2=0.90). LightGBM and GPR performed comparably (NSE=0.86–0.87), while SVM (NSE=0.82) and LASSO (NSE=0.78) showed lower accuracy. RMSE values were lowest for LSTM (15.3 m³/s) and highest for LASSO (28.7 m³/s). PBIAS indicated slight underestimation in LSTM (-2.1%) and overestimation in SVM (4.5%).

4.2. Temporal and Seasonal Analysis
LSTM and XGBoost excelled in capturing monsoon peaks (July–September) and low-flow periods (winter), due to their ability to model sequential dependencies and nonlinear interactions. GPR provided reliable uncertainty estimates but was computationally intensive. SVM struggled with extreme events, likely due to kernel limitations. LASSO, as a linear method, failed to capture complex seasonality, underscoring the need for nonlinear approaches in subtropical hydrology.

4.3. Feature Importance
XGBoost and LightGBM analyses indicated precipitation at t-1 and temperature as most influential variables, aligning with physical understanding. LSTM’s memory cells effectively utilized lagged streamflow data, enhancing multi-step predictions.

4.4. Implications for Subtropical Basins
The superiority of LSTM and XGBoost suggests that deep learning and ensemble methods are well-suited for subtropical regions with pronounced seasonal variability. These models can integrate climatic and topographic data, offering advantages over traditional models in data-driven contexts. However, challenges include data quality, computational costs (for LSTM/GPR), and interpretability (compared to process-based models).

4.5. Limitations and Uncertainty
Uncertainties arise from data errors, climate non-stationarity, and model assumptions. Future work should incorporate climate projections, high-resolution data, and hybrid ML-physical models.

5. Conclusion
This comprehensive comparative analysis demonstrates that LSTM and XGBoost represent state-of-the-art approaches for monthly streamflow prediction in subtropical monsoon-driven basins. LSTM's 91% Nash-Sutcliffe efficiency substantially exceeds traditional methods, capturing both seasonal and inter-annual hydrological variability through its recurrent architecture's temporal memory. XGBoost provides an optimal balance of accuracy (89% NSE) and computational feasibility, enabling operational integration into water management systems. LightGBM offers superior computational efficiency (30× faster training than LSTM) while maintaining competitive accuracy (86% NSE), positioning it for real-time forecasting applications.

The results underscore the fundamental inadequacy of linear approaches (LASSO regression, R^2=0.77) for complex subtropical hydrology, while supporting the urgent integration of advanced machine learning into water resources planning for the Jammu region. Future research should explore hybrid physically-informed neural networks combining LSTM's temporal dynamics with process-based hydrological constraints, ensemble methods integrating multiple models' strengths, and transfer learning approaches leveraging training from neighboring basins to enhance predictive skill in data-scarce regions.

For the Chinab River basin specifically, XGBoost deployment in operational forecasting systems is immediately recommended, with LSTM implementation feasible following computational infrastructure upgrade. Integration with climate projections from CMIP6 downscaling would enhance long-term water security planning for this critical tributary of the Indus River System.

6. Acknowledgements
The authors are highly thankful to the Jammu and Kahmir Forest Department and Department of Soil and Water Conservation for providing logistic support during the course of present research work. The authors sincerely thank Sh. S.K. Gupta IFS, Pr. Chief Conservator of Forests (HoFF) J&K and Sh. Sandeep Khajur IFS, Director Soil and Water Conservation department for their mentorship during the entire research work. Thanks are also due to Dr.M.K. Kumar IFS CCF Working Plan Research and Training J&K, Sh.Vivek Verma IFS CF Working Plan Research and Training, Mrs Neha Mehta SFS, Principal Soil Conservation Training School Mirna Sahib for their constant support and guidance during the research work.

7. Funding
The authors received no funding from any source to conduct the present study and its publication.

8. References
1. Beven K. Rainfall-runoff modelling: the primer. 2nd ed. Chichester: Wiley-Blackwell; 2012.
2. Devia GK, Ganasri BP, Dwarakish GS. A review on hydrological models. Aquat Procedia. 2015;4:1001-7.
3. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29(5):1189-232.
4. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, *et al.* LightGBM: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst. 2017;30.
5. Kratzert F, Klotz D, Brenner C, Schulz K, Herrnegger M. Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrol Earth Syst Sci. 2018;22(11):6005-22.
6. Kratzert F, Klotz D, Brenner C, Schulz K, Herrnegger M.

Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrol Earth Syst Sci. 2018;22(11):6005-22.
7. Liu J, Li Y, Zhang J, Wang Z, Chen X. A national-scale hybrid model for enhanced streamflow estimation – consolidating a physically based hydrological model with long short-term memory (LSTM) networks. Hydrol Earth Syst Sci. 2024;28:2871-94.
8. [Authors not specified in original; Nature Editorial]. Evaluating multi-source precipitation data for streamflow simulation using the SWAT model in the Alpine Manas River Basin, Northwest China. Sci Rep. 2025;15:27391.
9. Ni L, Wang D, Wu J, Wang Y, Tao Y, Zhang J, *et al*. Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. J Hydrol. 2020;586:124670.
10. Rasmussen CE, Williams CKI. Gaussian processes for machine learning. Cambridge, MA: MIT Press; 2006.
11. Taormina R, Chau KW, Sivakumar B. Neural network river forecasting through baseflow separation and binary-coded swarm optimization. J Hydrol. 2015;529:1788-97.
12. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol. 1996;58(1):267-88.
13. Tyralis H, Papacharalampous G, Langousis A. A brief review of random forests for water scientists and practitioners and their recent history in water resources. Water (Basel). 2019;11(5):910.
14. Vapnik V. The nature of statistical learning theory. New York: Springer; 1995.
15. Wang W, Gao J, Liu Z, Li C. A hybrid rainfall-runoff model: integrating initial loss and LSTM for improved forecasting. Front Environ Sci. 2023;11:1261239.
16. Yan L, Lei Q, Jiang C, Yan P, Ren Z, Liu B, *et al*. Climate-informed monthly runoff prediction model using machine learning and feature importance analysis. Front Environ Sci. 2022;9:1049840.
17. Yaseen ZM, Sulaiman SO, Deo RC, Chau KW. Artificial intelligence-based models for streamflow forecasting: 2000–2018. J Hydrol. 2019;585:124670.
18. Zakaria MNA, Ahmed K, Othman M, Ghani AA. Exploring machine learning algorithms for accurate water level forecasting in Muda river, Malaysia. Heliyon. 2023;9(7):e17971.
19. Zhang J, Zhu Y, Fan Y. Streamflow prediction using extreme gradient boosting model coupled with Gaussian process regression. Water Resour Manag. 2020;34:3513-31.

How to Cite This Article

Verma R, Kotwal M. Enhancing rainfall–runoff modelling in the subtropical Chinab River Basin of Jammu using advanced machine learning methods: a comparative study of LSTM, SVM, GPR, LASSO, XGBoost, and LightGBM. Int J Multidiscip Res Growth Eval. 2026;7(1):388–397.