PatchTSTSpike: A Patch-Based Transformer Framework for Cloudburst-Like Extreme Rainfall Detection Using ERA5 Reanalysis

Ankit Kumar[1], Prashant Kumar*[1], Yukiharu Hiasaki[2], and Rajni Rajni[3]

[1]Department of Mathematics, National Institute of Technology, Delhi-110040, India
[2]Department of Physics and Earth Sciences Faculty of Science, University of the Ryukyus, Japan
[3]Jindal Global Business School, O.P. Jindal Global University, Haryana-131001, India

(*Corresponding Author E-mail: prashantkumar@nitdelhi.ac.in)

Abstract. The occurrence of cloudburst-like heavy rainfall events in the Himalayan region is very rare but highly destructive on the other hand predicting them is even more difficult, These event initiates from very complex atmospheric crossover like change in pressure, temperature, Wind etc. In this paper, we have developed PatchTST, that is, a custom spike aware temporal patch based transformer model[5][11][15]. It is designed to analyze extreme rainfalls using Fifth Generation ECMWF Reanalysis data (ERA5 reanalysis data)[Link] [4][14]. We evaluate the model using variable window length ranging from 30 to 90 days. Among all of these windows, the 30-day look-back window provides the most promising detection performance, getting a $ROC\text{-}AUC$ of 0.6007. While using longer temporal windows, particularly 60 days, slight improvement in precision on the cost of reduced recall, implying a give and take between event detection sensitivity and prediction certainty. Overall, the used framework offers a subtle transformer-based approach for studying spike event like precipitation in mountains or hilly areas and provides a practical base for in-future data-driven research on Himalayan weather and it's extremes.

Keywords: Himalayan Meteorology, Rare-event Detection, Transformer Models, ERA5 Reanalysis, Extreme Rainfall

1 Introduction

Extreme rainfall events in the Himalayan region often cause severe floods and landslides ([1] [2][3]). These events are difficult to model because they occur rarely and show strong seasonal variation. It is very difficult to predict such events as the

© The Author(s) 2026
A. Agnihotri et al. (eds.), *Proceedings of the Conference on Bridging Engineering Disciplines with AI and Machine Learning (BEDAIML 2026)*, Advances in Intelligent Systems Research 209,
https://doi.org/10.2991/978-94-6239-697-5_30

complex interactions of several atmospheric variables like temperature, pressure, wind, moisture. Additionaly the captured sudden occurrences and variability of such parameters leads to a highly imbalanced dataset to work with. Climate variability has further contributed in increading cloudburst like extremes in past years[12][13]. Traditionall machine learning models was used to predict the rainfall but extreme events like cloudburst make the model fail to capture such nonlinearity and spikes[9][10]. Deep learning models such as Long Short-Term Memory improves temporal modelling but struggles with long range dependencies[11]. The low number of events also creates a strong class imbalance[16], which makes prediction challenging. Extreme rainfall events have increased in frequency due to climate variability [12][13].

This paper proposes PatchTSTSpike, a patch-based transformer model for detection of cloudburst like extreme rainfall events using ERA5 Reanalysis data[4]. The problem is as a binary event detection using temporal data. Extreme rainfall days are defined using the threshold of month-wise 99th percentile of regional maximum precipitation. Almost all studies on Himalayan rainfall suggest that sudden change in pressure moves moisture and hills play an important role in producing extreme precipitation [1][2][3]. However, The ERA5 Reanalysis dataset struggles to capture accurately very localized mountain rainfall events ([4] [14]). To solve this problem, we use a more localized spike-detection approach based on extreme rainfall thresholds such as top 1 percentile. This allows proper identification of unusual rainfall conditions to long-term atmospheric records. An important difficulty in this dataset is the severe imbalance between normal and extreme events which is to be studied. To handle this, we use Focal Loss[6] so that rare events are acknowledged during training rather than being normalized, like they do in every other model training [6]. We also apply Sharpness-Aware Minimization (SAM)[7] to improve model stability and generalization [7].

2 Data and Preprocessing

2.1 Study Region

The Himalayan region is characterized by surprisingly uneven terrains and strong monsoon-driven rainfall [1][2][3][13]. These conditions generally leads to extreme event such as flash floods and landslides and cloudburst itself. In this paper, the entire Himalayan domain is treated as a single aggregated region. Indian Himalayan region of Arunachal, Assam, Himachal Pradesh, Jammu & Kashmir, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, Tripura, Uttarakhand is taken into

consideration. The study region in this research work is plotted in the Fig. 1. Our main target here is to detect those rare extreme events.

2.2 Predictor Set and Data Integrity

Predictor variables are obtained from the daily ERA5 (Climate Data Store)[4][14] regional time series, It provide a consistent multi-decadal atmospheric trends and variations [4]. Five predictors are used: regional mean precipitation (tp_region_mean_mm), near-surface temperature (t2m_region), relative humidity (r_region), wind component (u_region), and total cloud cover (tcc_region). To avoid information leakage, the regional maximum precipitation (tp_region_max_mm) is not used as an input feature. It is reserved only for constructing the target spike labels.

2.3 Spike Label Construction

Extreme rainfall detection is evaluated as a binary classification task, that is, Either a heavy rainfall or not, using setting a threshold on maximum daily precipitation series (tp_region_max_mm). As we know rainfall patterns vary across seasons in the Himalayas, fixed thresholds are avoided in model training. Instead, spike events are defined using the month-wise 99th percentile threshold ($Q_{0.99}^{(m)}$).The distribution of rainfall and percentile thresholds are shown in Fig. 2 and the observed spike events in the dataset are visualized(see Fig. 2**)**. A day y_t is labeled as a rare event $(y_t = 1)$ if the maximum precipitation surpasses the threshold for that month respectively. Otherwise, it is labeled as a non-spike $(y_t = 0)$. This is how we have defined extreme rainfalls in our paper.

Table 1. Class distribution in the training split

Class	Count	Fraction
Non-Spike(0)	16,430	0.9886
Spike(1)	189	0.0114

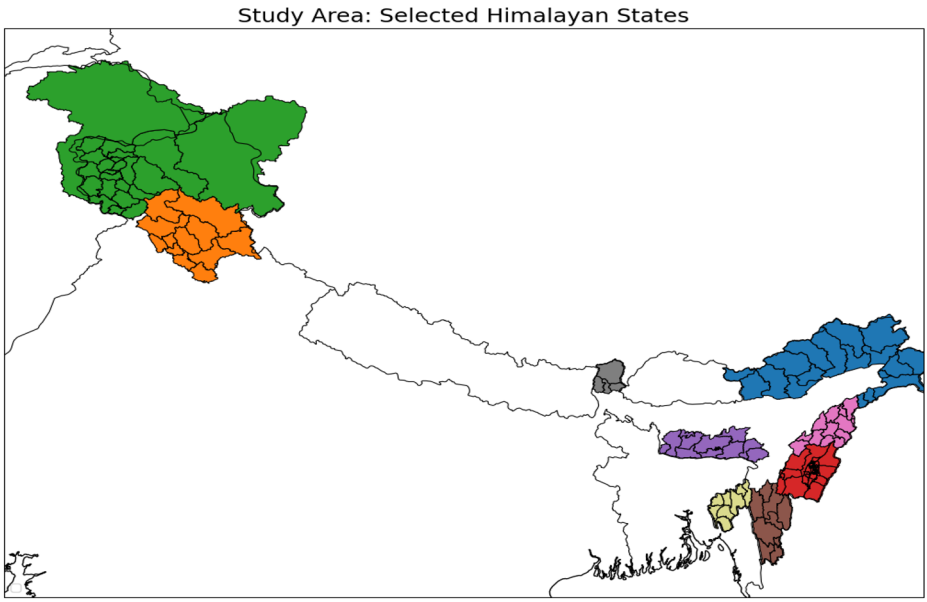

Fig. 1. Study region covering the Himalayan domain used in this study.

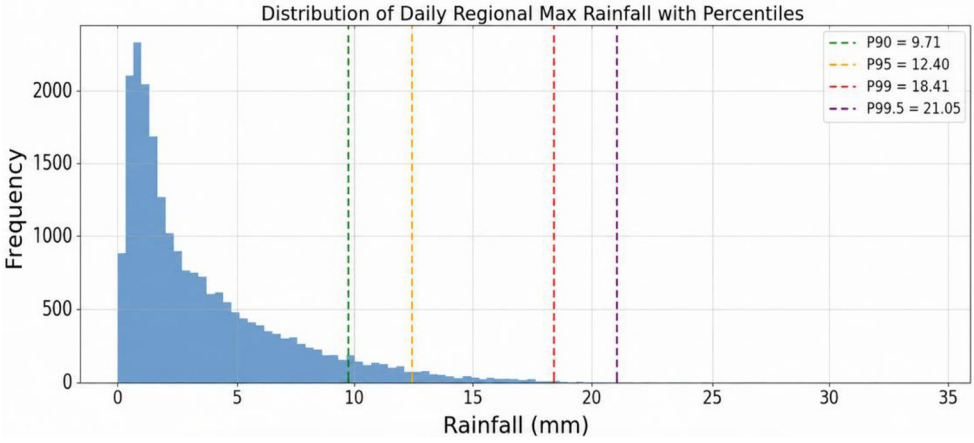

Fig. 2. Distribution of regional maximum daily rainfall with key percentile thresholds.

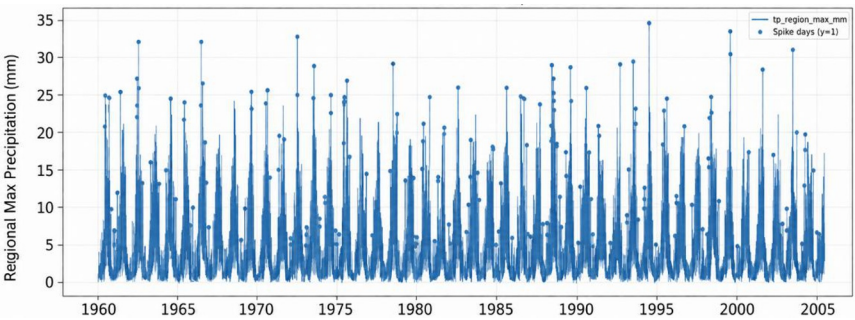

Fig. 3. Regional maximum daily precipitation time series with spike-day markers derived from month-wise P99 thresholding.

3 Methodology

3.1 Sequence Construction

The spike-detection problem is formulated as a binary classification task using daily regional atmospheric predictors. For a sequence length L, the model input is a tensor of multiple variable.

$$x \in \mathbb{R}^{L \times F} \tag{1}$$

where $F = 5$ represents the number of predictor variables used in the non-leaky setup. Each sequence contains L consecutive days of mean of total precipitation i.e., tp_region_mean_mm, t2m_region, r_region, u_region, and tcc_region. The target corresponds to the binary spike label associated with the prediction time step.

To study the effect of temporal context, we evaluate four sequence lengths:

$$L \in \{30,45,60,90\} \tag{2}$$

That means a period of 30 days, 45, days, 60 days and 90 days are taken into account one by one and models demonstarted different behavior, As Indian climate remains static for as long as 30 to 45 days the period of 90 days is more prone to show variability and thus non-reliable for analysis.

These setup allow comparison between shorter and longer temporal windows under the same training setup.

Every sequence is generated separately within the subsequent training, validation and test splits. It effectively ensures that no temporal leakage occurs between partitions.

3.2 PatchTSTSpike Architecture

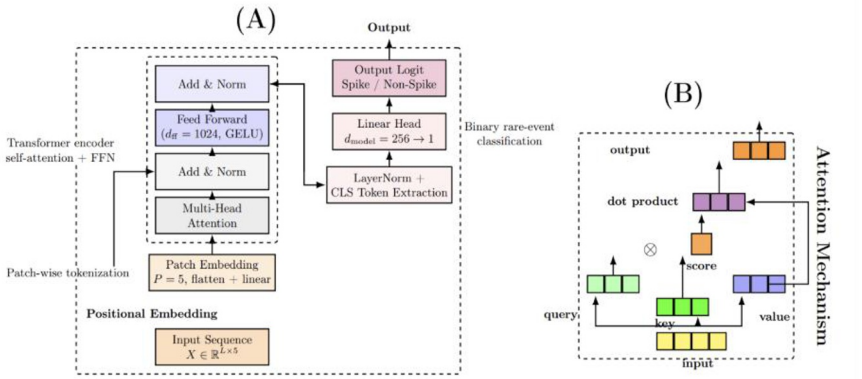

Fig. 4. Architecture of the PatchTSTSpike model. (A) The multivariate daily input sequence. (B) Self-attention mechanism used within each encoder layer.

To model multivariate atmospheric sequences, we use a transformer-based classifier called PatchTSTSpike[5][11][15]. The model reduces temporal redundancy by converting the input sequence into fixed-length temporal patches[5].
For an input sequence

$$X \in \mathbb{R}^{L \times F} \tag{3}$$

that sequence is divided into contiguous chunks of length $P = 5$

Let,

$$N = \frac{L}{P} \tag{4}$$

be the number of such chunks also called a patch. Each patch is flattened and sent directly into an embedded space:

$$z_i = W_p \, \mathrm{vec}\left(X_{(i-1)P+1:iP}\right) + b_p, \qquad z_i \in \mathbb{R}^d \tag{5}$$

Where W_p and b_p are learnable parameters and d is the embedding dimension.
The patch embeddings are combined with a learnable classification token and positional embeddings:

$$\mathbf{H}_0 = [\mathbf{z}_{\mathrm{cls}}, z_1, z_2, \ldots, z_N] + \mathbf{E}_{\mathrm{pos}} \tag{6}$$

The sequence is then processed by stacked Transformer encoder layers consisting of multi-head self-attention, residual connections, feed-forward networks, Gaussian Error Linear Unit(GELU) activation, dropout, and layer normalization[5][11][15]. The self-attention operation is defined as following equation,

$$\text{Attn}(Q,K,V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt[2]{d_h}}\right)V, \quad (7)$$

where Q, K, and V are the query, key and the value projections[11].
After the encoder stack, the transformed token is extracted and passed through a linear layer to produce a scalar logit for spike classification. The main architecture settings are: embedding dimension (d_model) is 256 attention heads (n_heads) is 8 encoder layers (n_layers) are 4 feed-forward dimension (d_ff) is 1024 dropout is 0.1 and patch length (patch_len) is 5.
Overall architecture and flow diagram of the proposed PatchTSTSpike model is shown (see Fig. 4.)[5][15].
Although inspired by patch-based transformer models for time-series data[5][11][15], the architecture here is adapted specifically for rare rainfall spike detection rather than long-horizon forecasting.

3.3 Training and Evaluation

Model training uses focal loss[6] to address the strong class imbalance[16], with validation Precision-Recall Area Under Curve (PR-AUC) used for model selection [6]. Optimization is performed using SAM[7] with AdamW, some learning rate, weight decay of 10^{-4}, batch size of 256, and 20 training epochs [7]. Mixed precision is disabled to maintain numerical stability. Experiments are conducted for sequence lengths

$$L \in \{30,45,60,90\} \quad (8)$$

When moving between window lengths, compatible weights are transferred while positional embeddings are reinitialized.

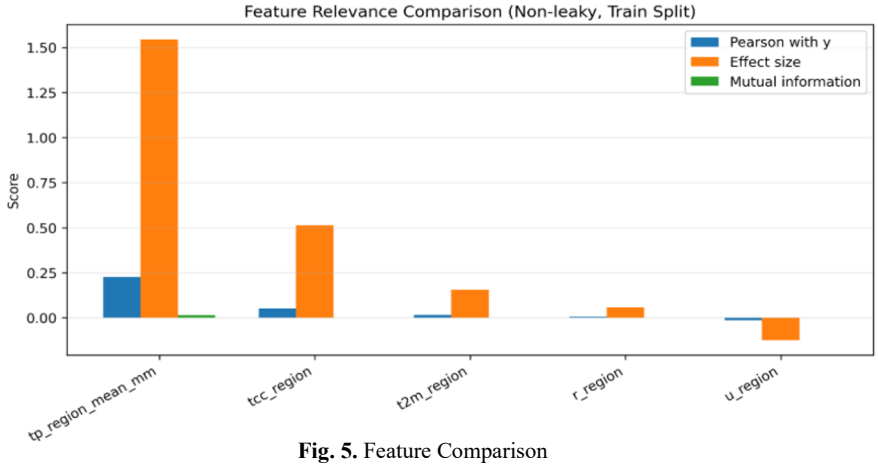

Fig. 5. Feature Comparison

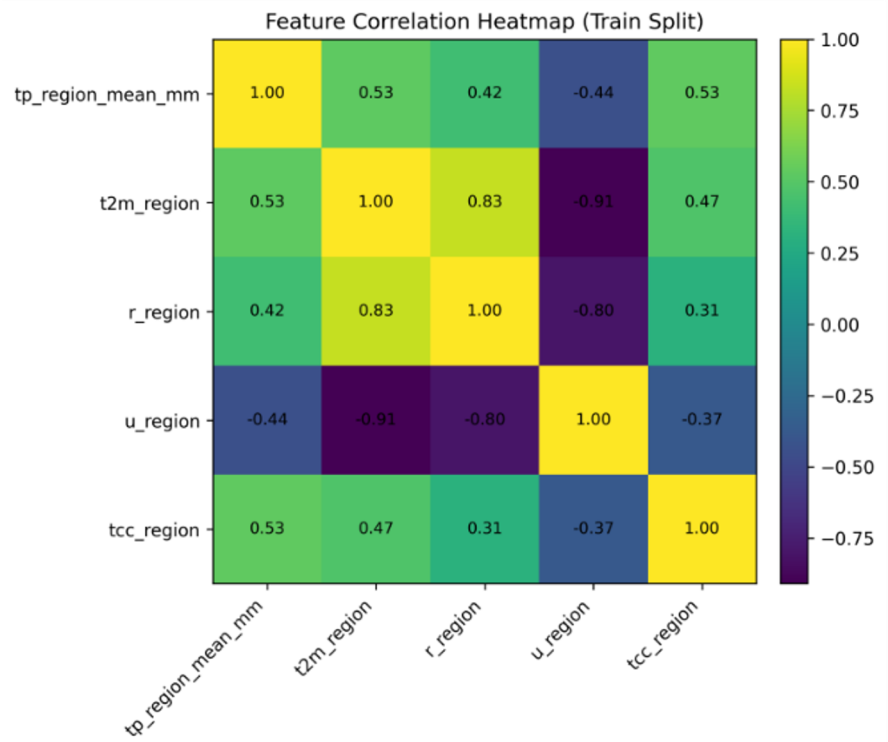

Fig. 6. Correlation Heatmap

Table 2. Feature relevance table

Feature	**Pearson with y**	Effect Size	Mutual Info
tp_region_mean_mm	0.225925	1.543890	0.015254
tcc_region	0.052988	0.513079	0.001032
t2m_region	0.015824	0.156718	0.000389
r_region	0.006299	0.058178	0.000000
u_region	-0.012699	-0.123356	0.000000

A good validation result is evaluated using Receiver Operating Charfacteristics Area Under Curve (ROC-AUC), PR-AUC, F1-score, Precision and Recall. This way ensures that performance differences is observed due to the effect of temporal dependencies rather than changes in data splits or training configuration. The variables used are theoretically interrelated with events such as cloudburst or simply precipitation, Although rainfall is dependent on such variable but the extent of

dependency may vary due to landscape and floral differences. While the correlation suggests some are strongly depended and some are not focus is maintained on strongly correlated parameters to maintain the highest accuracy possible from the variability of dataset. Feature relevance comparison and correlation are shown in Fig. 5 and Fig. 6 respectively.

4 Results and Discussion

4.1 Feature Relevance Analysis

Using Pearson correlation we have depicted correlated variables, standardized mean separation, and mutual information in Table 2. Mean precipitation (tp_region_mean_mm) shows the strongest relationship with Cloudburst, followed by cloud cover (tcc_region).

This shows that heavy dew or moisture and cloud covers are important factors related with extreme rainfall conditions. Temperature (t2m), relative humidity (r), and wind (u) also contributed in the analysis but show weaker correlation. The distribution of key features across spike and non-spike are shown in Fig. 7.

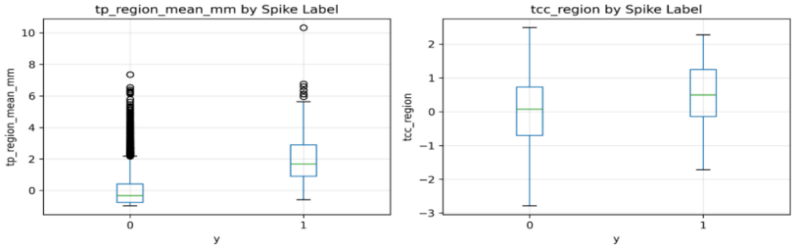

Fig. 7. Boxplots of tp_region_mean_mm and tcc_region

4.2 Temporal Context Performance

We evaluated sliding windows of $30, 45, 60$,and 90 days. Among these settings, the 30-day window shows the strongest overall result, achieving the highest test $PR - AUC$ (0.0162) and $ROC - AUC$ (0.6007). On the other hand, the 60-day window produces higher precision and F1-score, behaving as a low level detector. These results suggest that shorter temporal context ($L = 30$) is better for capturing a wider range of spike events, while longer context ($L = 60$) favors more

confident predictions. Despite the difficulty of the task and the strong class imbalance[16], the results indicate that *PatchTSTSpike* provides meaningful predictive capability for regional rainfall spike detection[5][15]. The model performances for different window length are summarized in Table 3. Also, the precision-recall and ROC curve for the best performing models are shown in Fig. 8 and Fig. 9 respectively.

Table 3. Window-length comparison for PatchTSTSpike.

L	**Val PR-AUC**	**Test ROC-AUC**	**Test PR-AUC**	**Test F1**	**Precision**	**Recall**
30	0.015253	0.600733	0.016226	0.041885	0.022099	0.400000
45	0.059858	0.549670	0.012179	0.037037	0.022727	0.100000
60	0.021638	0.544864	0.015355	0.047619	0.083333	0.033333
90	0.019150	0.576554	0.012570	0.031915	0.017341	0.200000

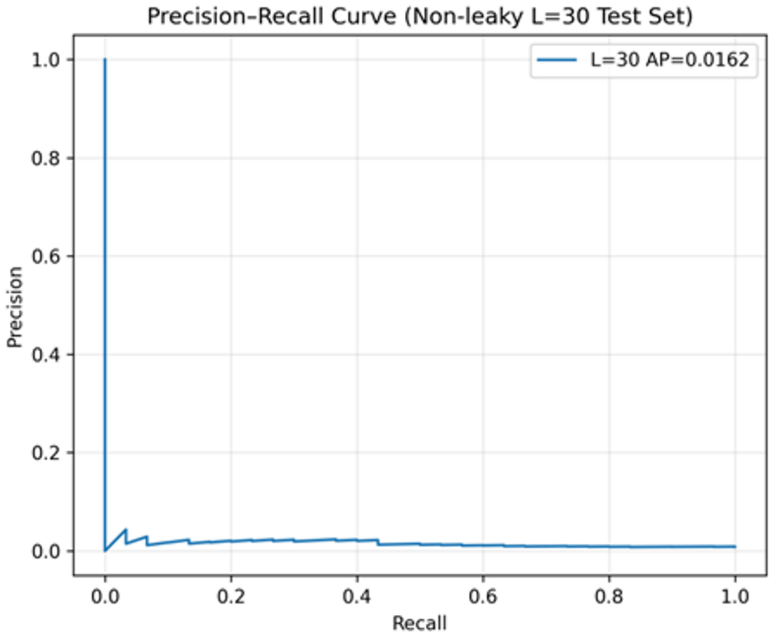

Fig. 8. Precision–recall curve for the selected PatchTSTSpike model.

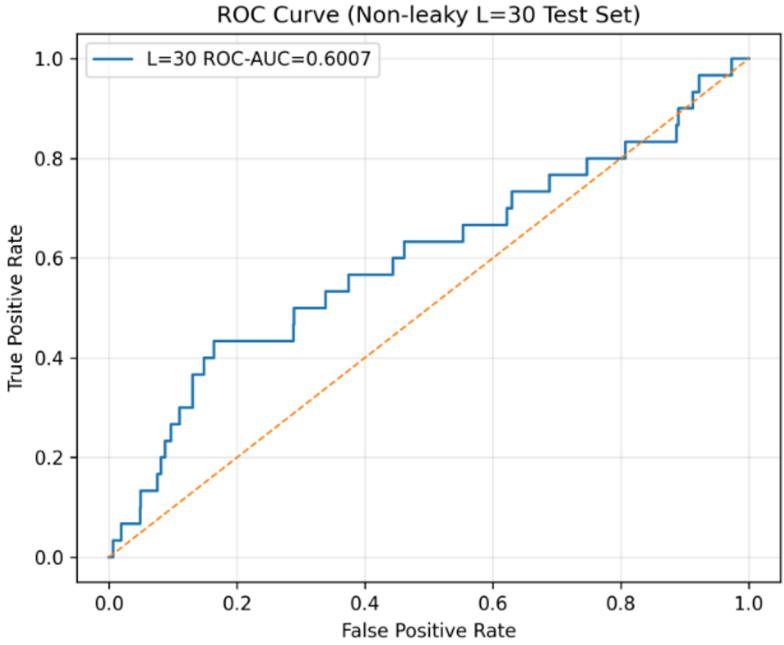

Fig. 9. ROC curve for the PatchTSTSpike model.

5 Conclusion

This study presents a framework for detecting Himalayan rainfall spikes using the custom PatchTSTSpike transformer model[5][11][15]. Extreme events are defined using month-wise 99th percentile thresholds, while atmospheric predictors are kept separate from the precipitation labels to avoid information leakage. The analysis shows that regional mean precipitation and cloud cover provide the strongest signals for identifying spike events, while temperature, humidity, and wind mainly offer supporting context. Experiments with different sequence lengths show that temporal context plays an important role. A 30-day window provides the most balanced performance, while a 60-day window produces fewer but more confident detections. Longer sequences appear to introduce additional noise that weakens the predictive signal.

Future work for our paper can be done by implementing this framework using the spatial information as well as the land–surface variables and snow cover to better capture atmospheric changes and variability.

References

1. Li, H., Beldring, S., Xu, C.-Y., Huss, M., Melvold, K., Jain, S.K.: Precipitation pattern in the Western Himalayas revealed by four datasets. Hydrology and Earth System Sciences 22, 5097–5110 (2018)
2. Chen, Y., Sharma, S., Zhou, X., Yang, K., Li, X., Niu, X.: Spatial performance of multiple reanalysis precipitation datasets on the southern slope of central Himalaya. Atmospheric Research 250, 105365 (2021)
3. Dollan, I.J., Bookhagen, B., Strecker, M.R., Thiede, R.C., Scherler, D.: An assessment of gridded precipitation products over High Mountain Asia. Journal of Hydrology: Regional Studies 54, 101675 (2024)
4. Hersbach, H., Bell, B., Berrisford, P., et al.: The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society 146(730), 1999–2049 (2020)
5. Nie, Y., Nguyen, N.H., Sinthong, P., Kalagnanam, J.: A time series is worth 64 words: Long-term forecasting with transformers. In: International Conference on Learning Representations (2023)
6. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
7. Foret, P., Kleiner, A., Mobahi, H., Neyshabur, B.: Sharpness-aware minimization for efficiently improving generalization. In: International Conference on Learning Representations (2021)
8. Dong, M. Y., L. S. Chen, P. Q. Zheng, and J. S. Pan, titled "Research progress on abrupt intensification of heavy rainfall and super heavy rainfall associated with landfalling tropical cyclones"
9. Sun, Q., et al. “Deep Learning for Precipitation Prediction: A Survey.” IEEE Transactions on Geoscience and Remote Sensing, 2020.
10. Mosavi, A., et al. “Flood Prediction Using Machine Learning Models: Literature Review.” Water, 2018.
11. Vaswani, A., et al. “Attention Is All You Need.” Advances in Neural Information Processing Systems, 2017.
12. Hochreiter, S., Schmidhuber, J. “Long Short-Term Memory.” Neural Computation, 1997.
13. Westra, S., Fowler, H.J., Evans, J.P., Alexander, L.V., Berg, P., Johnson, F., Kendon, E.J., Lenderink, G., Roberts, N.M.: Future changes to the intensity and frequency of short-duration extreme rainfall. Reviews of Geophysics 52(3), 522–555 (2014)
14. Tarek, M., Brissette, F.P., Arsenault, R.: Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modeling over North America. Journal of Hydrology 593, 125878 (2021)
15. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. AAAI Conference on Artificial Intelligence (2021)
16. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering 21(9), 1263–1284 (2009)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.