TIME-SERIES FORECASTING AND ANOMALY DETECTION FOR EARLY WARNING OF GLACIAL LAKE OUTBURST FLOODS

*Ayush Kumar Singh, Shakti Sharma, Lalita Chaudhary *, Abhishek Singh, Harshita Dhawan, Diya Gupta*

Bennett University, India

ABSTRACT

Glacial Lake Outburst Floods (GLOFs) are a significant climate-induced hazard, requiring innovative solutions for timely prediction and mitigation. This study presents a multi-faceted Early Warning System (EWS) for GLOFs, integrating satellite data, time-series analysis, and deep learning. Using Google Earth Engine (GEE), a curated dataset of key parameters—snow reflux, temperature, precipitation, lake size, and water level—was preprocessed to ensure accuracy and consistency. Seasonal and trend decomposition, coupled with anomaly detection via convolutional autoencoders (CAEs), identified deviations in meteorological patterns indicative of potential GLOF events.

A time-series anomaly detection and forecasting model captured short and long-term trends in critical variables. Validated against historical GLOF events in Ronti Lake and South Lhonak Lake, the prototype achieved strong prediction performance with RMSE of 0.0218 and R^2 value of 0.696. Additionally, 30-day forecasts of parameters such as snow reflux and water levels provided actionable insights into alarming trends.

This study underscores the effectiveness of developing a scalable and transparent Early Warning System for GLOFs. While promising, further refinement is needed to incorporate additional variables and expand coverage to other glacial lakes, paving the way for real-time decision-making and disaster risk reduction in vulnerable regions.

Index Terms— Glacial Lake Outburst Floods , Deep Learning, Time Series prediction

1. INTRODUCTION

Glacial Lake Outburst Floods (GLOFs) are unexpected, high-magnitude events occurring when a glacier lake, often formed by melting of nearby glaciers and accumulation of water behind moraine dams, posing severe threats to ecosystems, infrastructure, and human lives in glacial regions[1]. For instance, the glacial lakes in the Himalayan region, formed over the past five decades due to rapid warming of 0.15°C to 0.60°C per decade [2], pose significant transboundary risks, necessitating regular monitoring, early warning systems, and mitigation measures to prevent loss of lives and infrastructure [3]. This research introduces a comprehensive Early Warning System designed to predict and mitigate GLOF risks by integrating advanced satellite data analysis, time-series modelling[4], and deep learning techniques. The system is specifically tailored for high-risk areas such as Ronti Lake [5] and South Lhonak Lake [6], where glacial lake volume expansion and snow retreat are critical precursors to such events. As per records, the most severe GLOF scenario occurs during a moraine overtopping failure, generating a huge flood discharge of 6064.6 m^3/s and releasing water volume of 25.7 million m^3 [7, 8]. Since the unpredictable nature of the event makes them a growing concern, understanding and timely prediction of GLOFs is crucial for disaster management in vulnerable areas.

The methodology uses curated datasets from Google Earth Engine (GEE), a cloud based platform for easier access to and processing geospatial datasets [9], including thermodynamic and meteorological parameters such as snow reflux, temperature variations, precipitation, lake size, and water level. Rigorous preprocessing steps, including data cleaning, interpolation, normalization, and outlier removal, ensure consistency and reliability. Promising results from the prototype demonstrate its capability to reconstruct past GLOF events, forecast critical variables, and provide real-time alerts. The integration of anomaly detection with predictive modeling establishes a robust framework for enhancing GLOF early warning capabilities, while ongoing improvements in model accuracy and variable inclusion aim to further refine hazard predictions.

1.1. Problem Formulation

The aim of this study is to develop a deep learning-based model and framework for time-series prediction and anomaly detection in meteorological data, specifically targeting GLOFs. The model is designed to identify deviations in key parameters such as wind speed, snow reflux, temperature, precipitation, lake size, and water level, which are critical for predicting GLOF events[10]. The framework incorporates recurrent and convolutional neural networks to capture the spatial and temporal patterns in the data.

*Email ID: Lalita.chaudhary19@gmail.com

2. DATASET

The dataset used for developing the Early Warning System (EWS) for GLOFs comprises several fields of meteorological data extracted and processed using Google Earth Engine, focusing primarily on the geographical coordinates of Ronti Lake. Key data fields include:

- **WS10M**: Wind speed at 10 meters
- **PS**: Surface pressure
- **PRECTOTCORR**: Corrected precipitation total
- **T2MDEW**: Dew point temperature at 2 meters
- **ALLSKY SFC LW DWN**: All-sky surface longwave downward radiation
- **WS50M RANGE**: Wind speed range at 50 meters
- **Evaporation Rate**: Calculated manually based on meteorological equations
- **Sensible Heat Flux**: Calculated manually to estimate heat exchange at the surface

The dataset's granularity and accuracy were optimized for anomaly detection and analysing temporal patterns and spatial variations in meteorological parameters, to identify the early warning indicators linked to GLOFs. Each feature (as described in the dataset description) was preprocessed to ensure consistency and reliability. Preprocessing steps included [11]:

- **Data Cleaning**: Dealing with missing values and outliers.
- **Normalization**: Scaling the data using `StandardScaler` to standardize the range of features.
- **Sequence Creation**: Transforming the time-series data into fixed-length sequences for input into the model.

3. METHODOLOGY

3.1. Model Architecture

The proposed model is a type **reconstruction convolutional autoencoder model** that detects anomalies in timeseries data. The model consists of an **encoder-decoder** structure [12] with skip connections to enhance feature extraction and reconstruction. Below is a detailed breakdown of the architecture:

3.1.1. Encoder

- **Convolutional Blocks**:
  - Two 1D convolutional layers with 64 and 32 filters, respectively, followed by batch normalization, ReLU activation, and dropout for regularization.
  - Skip connections are added to preserve spatial information.
- **LSTM Blocks**:
  - Two LSTM layers with 64 and 32 units, respectively, compiled with batch normalization and dropout.
  - A bottleneck LSTM layer with 16 units compresses the temporal information.

3.1.2. Decoder

- **LSTM Blocks**:
  - Two LSTM layers with 32 and 64 units, respectively, followed by batch normalization and dropout.
  - Skip connections from the encoder LSTM layers are added to improve reconstruction.
- **Transposed Convolutional Blocks**:
  - Two 1D transposed convolutional layers with 32 and 64 filters, respectively, followed by batch normalization, ReLU activation, and dropout.
  - Skip connections from the encoder convolutional layers are added to enhance spatial reconstruction.
- **Final Reconstruction**:
  - A 1D convolutional layer with a single filter produces the final output, reconstructing the input sequence.

3.1.3. Custom Loss Function

The model is trained using a custom loss function that incorporates **Mean Squared Error** and **Mean Absolute Error** to balance the emphasis on large and small errors:

$$\text{Custom Loss} = \text{MSE} + 0.5 \times \text{MAE}$$

3.2. Training Process

Model training was done using the following setup:

- **Adam Optimizer**, learning rate of 0.001.
- 64 **Batch Size**
- 50 **Epochs** with early stopping to combat overfitting.
- **Callbacks**:
  - **Early Stopping** to monitor validation loss and restore the best weights if no improvement is observed for 5 epochs.
  - **Learning Rate Reduction** reduces the learning rate by a factor of 0.5 if validation loss plateaus for 3 epochs.

3.3. Anomaly Detection

The detection of anomalies, a critical challenge extensively explored across various research fields [13], is achieved by comparing the model's reconstructions with the input sequences. The steps include:

- **Reconstruction Error**: Compute the MSE between input and reconstructed sequences.
- **Thresholding**: Define an anomaly threshold as:

  $$\text{Threshold} = \mu_{\text{train loss}} + 3 \times \sigma_{\text{train loss}}$$

  where $\mu_{\text{train loss}}$ and $\sigma_{\text{train loss}}$ are the mean and standard deviation of the training loss, respectively.
- **Anomaly Identification**: Flag sequences with reconstruction errors exceeding the threshold as anomalies.

4. RESULTS

The proposed EWS for glacial lake outburst floods (GLOFs) integrates time-series analysis with satellite data to address this critical hazard. The system focuses on snow retreat cover and increasing glacial lake volumes as key indicators, leveraging data from flood-prone areas such as Ronti Lake and South Lhonak Lake. A functional prototype was developed using a dataset curated via Google Earth Engine (GEE), incorporating thermodynamic parameters such as snow reflux, temperature variations, precipitation, lake size, and water level. The data underwent preprocessing, including cleaning, interpolation, outlier removal, and normalization, to ensure consistency. Time series data was structured with lag features to capture historical trends, and seasonal and trend decomposition highlighted cyclical and long-term patterns.

If the reconstruction loss for a sample exceeds this threshold value [14], it can be inferred that the model is observing a pattern with which it is unfamiliar. This sample will be designated as an anomaly.

Metric	Training	Testing	Interpretation
RMSE	0.06	0.0218	Low prediction error
MSE	0.3	0.000476	Low squared error
R^2	-	0.696	Strong correlation

Table 1. *Performance Metrics for Time-Series Prediction*

The model trained on historical data using 90-10 train-test split and rolling forecast cross-validation to preserve temporal dependencies. The prototype demonstrated promising results in forecasting past GLOF events, achieving RMSE of 0.06 and MSE of 0.3, indicating accurate hazard prediction. A separate test yielded RMSE of 0.0218, MSE of 0.000476, and R^2 value of 0.696, meaning the model is able to explain approximately 69.6 variance in the target variable (shown in table 1). Although 30.4 of the variance remains unexplained, these results are encouraging for a prototype, as it captures a significant portion of underlying patterns. The trained AR model successfully forecasted key parameters, such as snow reflux and water level, for the next 30 days, enabling the identification of alarming trends[15]. These predictions were integrated into an early warning system to provide real-time alerts for potential GLOFs.

5. CONCLUSION

The EWS for Glacial Lake Outburst Floods (GLOFs) represents a tangible and data-driven solution to a pressing climate-induced hazard. Although the prototype has shown promising results: achieving RMSE of 0.0218 and explaining 69.6 variance in target variable, it is essential to recognize its limitations. The unexplained 30.4 variance highlights the complexity of GLOF dynamics and the need for further refinement. This is not merely a theoretical exercise; the system has been rigorously tested using real-world data from Ronti Lake and South Lhonak Lake, and its predictions have been validated against historical GLOF events.

What sets this system apart is its integration of satellite data, thermodynamical parameters, and advanced time series modeling, all processed through Google Earth Engine (GEE). However, the true credibility of this work lies in its transparency: the preprocessing steps, including data cleaning, interpolation, and normalization, are meticulously documented, and the model's performance metrics are openly shared.

While the results are encouraging, this is not the final word. The prediction power of the model can be further improved by incorporating additional variables, including seismic activity or glacier movement patterns, and by expanding the dataset to include more glacial lakes. The proposed integration of Digital Twin technology for 3D simulations is not just aspirational, but a logical next step to improve real-time decision making.

6. REFERENCES

[1] S. Harrison, J. S. Kargel, C. Huggel, J. Reynolds, D. H. Shugar, R. A. Betts, A. Emmer, N. Glasser, U. K. Haritashya, J. Klimeš, L. Reinhardt, Y. Schaub, A. Wiltshire, D. Regmi, and V. Vilímek, "Climate change and the global pattern of moraine-dammed glacial lake outburst floods," *The Cryosphere*, vol. 12, no. 4, pp. 1195–1209, 2018.

[2] Rayees Ahmed, Gowhar Farooq Wani, Syed Towseef Ahmad, Mehebub Sahana, Harmeet Singh, and Pervez Ahmed, "A review of glacial lake expansion and associated glacial lake outburst floods in the himalayan region," *Earth Systems and Environment*, vol. 5, no. 3, pp. 695–708, 2021.

[3] Nazir Ahmed Bazai, Peng Cui, Paul A Carling, Hao Wang, Javed Hassan, Dingzhu Liu, Guotao Zhang, and Wen Jin, "Increasing glacial lake outburst flood hazard in response to surge glaciers in the karakoram," *Earth-Science Reviews*, vol. 212, pp. 103432, 2021.

[4] Yuchen Fang, Jiandong Xie, Yan Zhao, Lu Chen, Yunjun Gao, and Kai Zheng, "Temporal-frequency masked autoencoders for time series anomaly detection," in *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, 2024, pp. 1228–1241.

[5] Pratima Pandey, Prakash Chauhan, C. M. Bhatt, Praveen K. Thakur, Suresh Kannaujia, Pankaj R. Dhote, Arijit Roy, Santosh Kumar, Sumer Chopra, Ashutosh Bhardwaj, and S. P. Aggrawal, "Cause and process mechanism of rockslide triggered flood event in rishiganga and dhauliganga river valleys, chamoli, uttarakhand, india using satellite remote sensing and in situ observations," *Journal of the Indian Society of Remote Sensing*, vol. 49, no. 5, pp. 1011–1024, May 2021.

[6] Ashim Sattar, Ajanta Goswami, Anil.V. Kulkarni, Adam Emmer, Umesh K. Haritashya, Simon Allen, Holger Frey, and Christian Huggel, "Future glacial lake outburst flood (glof) hazard of the south lhonak lake, sikkim himalaya," *Geomorphology*, vol. 388, pp. 107783, 2021.

[7] Ashim Sattar, Ajanta Goswami, and Anil V. Kulkarni, "Hydrodynamic moraine-breach modeling and outburst flood routing - a hazard assessment of the south lhonak lake, sikkim," *Science of The Total Environment*, vol. 668, pp. 362–378, 2019.

[8] Caroline Taylor, Tom R Robinson, Stuart Dunning, J Rachel Carr, and Matthew Westoby, "Glacial lake outburst floods threaten millions globally," *Nature Communications*, vol. 14, no. 1, pp. 487, 2023.

[9] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore, "Google earth engine: Planetary-scale geospatial analysis for everyone," *Remote Sensing of Environment*, vol. 202, pp. 18–27, 2017, Big Remotely Sensed Data: tools, applications and experiences.

[10] Sanjaya Gurung, Saroj Dhoj Joshi, and Binod Parajuli, "Overview of an early warning system for glacial lake outburst flood risk mitigation in dudh-koshi basin, nepal," *Sciences in Cold and Arid Regions*, vol. 13, no. 3, pp. 206–219, 2021.

[11] Salvador García, Sergio Ramírez-Gallego, Julián Luengo, José Manuel Benítez, and Francisco Herrera, "Big data preprocessing: methods and prospects," *Big Data Analytics*, vol. 1, no. 1, pp. 9, 2016.

[12] Pankaj Malhotra, Anusha Ramakrishnan, Gaurangi Anand, Lovekesh Vig, Puneet Agarwal, and Gautam Shroff, "Lstm-based encoder-decoder for multi-sensor anomaly detection," 2016.

[13] Haipeng Ji, Xiaoqiang Qiao, Jiang Zhang, Yihang Du, Tao Zhang, and Fayu Wan, "Spectrum anomaly detection under multiple sensors using metric-adversarial learning," *IEEE Communications Letters*, 2025.

[14] Hongsong Chen, Xingyu Li, and Wenmao Liu, "Multivariate time series anomaly detection by fusion of deep convolution residual autoencoding reconstruction model and convlstm forecasting model," *Computers Security*, vol. 137, pp. 103581, 2024.

[15] A. Emmer, S. K. Allen, M. Carey, H. Frey, C. Huggel, O. Korup, M. Mergili, A. Sattar, G. Veh, T. Y. Chen, S. J. Cook, M. Correas-Gonzalez, S. Das, A. Diaz Moreno, F. Drenkhan, M. Fischer, W. W. Immerzeel, E. Izagirre, R. C. Joshi, I. Kougkoulos, R. Kuyakanon Knapp, D. Li, U. Majeed, S. Matti, H. Moulton, F. Nick, V. Piroton, I. Rashid, M. Reza, A. Ribeiro de Figueiredo, C. Riveros, F. Shrestha, M. Shrestha, J. Steiner, N. Walker-Crawford, J. L. Wood, and J. C. Yde, "Progress and challenges in glacial lake outburst flood research (2017–2021): a research community perspective," *Natural Hazards and Earth System Sciences*, vol. 22, no. 9, pp. 3041–3061, 2022.