Extreme Rainfall Forecasting for Flood Risk Assessment in North Sumatra Using Singular Value Decomposition and LSTM-Based Deep Learning Models

Reinhard Alfonzo Hutabarat – 13524056
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan Ganesha 10 Bandung
Email: reinhardalfonso@gmail.com, 13524056@std.stei.itb.ac.id

***Abstract*—North Sumatra is highly vulnerable to hydrometeorological disasters driven by extreme tropical rainfall, which remains challenging to predict using conventional numerical models. This study proposes a hybrid Deep Learning approach combining Singular Value Decomposition (SVD) and Long Short-Term Memory (LSTM) to enhance forecasting accuracy and computational efficiency. SVD was employed to reduce the high dimensionality of 20 years of GPM IMERG satellite data (2005–2024), extracting dominant spatiotemporal climate modes. The LSTM network then modeled the temporal evolution of these features to project future anomalies. The model successfully predicted a "Very High" rainfall intensity for November 2025, identifying critical flood hotspots in Medan and Sibolga. These projections are consistent with regional synoptic patterns observed in neighboring provinces and aligned with global atmospheric dynamics, specifically La Niña and Negative IOD conditions. The study concludes that the SVD-LSTM framework provides a robust and efficient tool for early flood risk assessment in complex tropical terrains.**

***Keywords*—Extreme Rainfall, Flood Forecasting, SVD, LSTM, Deep Learning, North Sumatra.**

I. Introduction

Late 2025 marked a devastating period for North Sumatra, as a series of extreme weather events triggered flash floods and landslides across the Bukit Barisan mountain range. This catastrophe underscores that hydrometeorological risks depend critically on the ability to predict the primary hazard: extreme rainfall. Unfortunately, conventional Numerical Weather Prediction (NWP) models often struggle to capture the variability of tropical rainfall in this region, which is highly dynamic, localized, and heavily influenced by complex topography. Conventional models often lack the resolution to resolve these short-lived, high-intensity convective processes without immense computational power.

To address these limitations, Long Short-Term Memory (LSTM) algorithms have emerged as a superior data-driven alternative for modeling time-series dependencies. However, directly applying LSTM to high-dimensional spatial data, such as satellite rainfall imagery, presents significant challenges, specifically high computational loads and the risk of overfitting due to the millions of parameters involved.

This study proposes integrating Singular Value Decomposition (SVD) into the LSTM architecture to solve this dimensionality issue. SVD serves to reduce data dimensionality by extracting dominant spatial patterns and filtering out noise, ensuring that the input to the LSTM model is concise while retaining essential features. By developing this hybrid SVD-LSTM model, this research aims to produce an extreme rainfall forecasting method for North Sumatra that achieves high accuracy in capturing spatiotemporal dynamics while remaining computationally efficient, providing a foundation for more reliable flood early warning systems.

II. Theoretical Basis

A. Atmospheric and Rainfall Patterns in North Sumatra

North Sumatra sits in a strategic position flanked by two major oceans (the Indian and the Pacific) and two continents (Asia and Australia), making it an integral part of the Maritime Continent climate system. Rainfall characteristics in this region are highly complex, influenced by the interaction of atmospheric phenomena across various spatiotemporal scales. Generally, Aldrian and Susanto (2003) classify the Indonesian archipelago into three primary rainfall patterns. Western and central North Sumatra fall under Monsoonal Pattern and Equatorial Pattern.

The Equatorial pattern is dominant from the northwest coast to the central part of North Sumatra, this pattern is characterized by a bimodal rainfall distribution, typically occurring in March–May and October–December. These peaks are associated with the meridional movement of the Inter-Tropical Convergence Zone (ITCZ) as it crosses the equator twice a year. Consequently, this region tends to remain wet year-round with no indistinct dry season.

Although the monsoon pattern is more dominant in Java and Nusa Tenggara, the influence of the Asian Monsoon and Australian Monsoon remains evident in Sumatra. The Asian Monsoon, which carries moist air masses from the South

China Sea and the Indian Ocean, typically intensifies rainfall during the December–February period.

However, this rainfall variability is significantly modulated by inter-annual climate variability phenomena:

- El Niño-Southern Oscillation (ENSO): An ocean-atmosphere interaction in the Equatorial Pacific. generally, El Niño phases are associated with dry conditions (droughts) in Indonesia, while La Niña triggers increased rainfall. However, studies indicate that the rainfall response in northern Sumatra to ENSO is neither as strong nor as consistent as in eastern Indonesia. This is due to the more dominant local influence of the Indian Ocean.
- Indian Ocean Dipole (IOD): An oscillation of sea surface temperatures (SST) in the Indian Ocean. A Negative IOD phase, characterized by warmer SST in the waters west of Sumatra compared to eastern Africa, triggers strong convection and significant rainfall increases along Sumatra's west coast. Major flood events often occur when a Negative IOD phase coincides with La Niña or a strong Asian Monsoon.

B. Physical Mechanisms of Orographic Rain in Bukit Barisan

A primary determinant of high rainfall in North Sumatra is the presence of the Bukit Barisan mountain range. This volcanic chain spans 1,700 km from Aceh to Lampung, with an average elevation of 1,000–3,000 masl. This topographic structure creates a potent orographic effect.

Orographic Rain Mechanism occurs when horizontal air mass flow is subjected to forced lifting by topographic barriers. As the air rises, it undergoes expansion and adiabatic cooling. Since cold air has a lower capacity to hold water vapor, the vapor condenses rapidly to form vertical Cumulonimbus clouds capable of producing heavy precipitation.

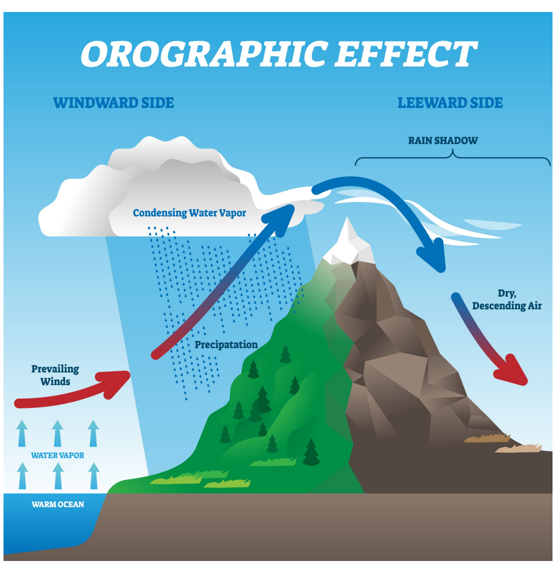

Fig. 1. Illustration of the Orographic Effect. [Source: https://www.shutterstock.com/id/image-vector/orographic-effect-vector-illustration-labeled-weather-1544982272]

Studies on rainfall characteristics in the Bukit Barisan region reveal a clear asymmetry between the western (windward) and eastern (leeward) slopes:

- Windward: Facing the Indian Ocean directly. When westerly zonal winds are dominant, which has a high probability of occurrence, especially during the Asian Monsoon , moist air masses from the ocean directly impact the mountains. Observational data indicates that the highest probability of heavy rainfall in this region occurs during westerly wind deceleration patterns or convergence conditions. Areas such as Sibolga, Central Tapanuli, and Mandailing Natal are located in this zone, rendering them highly susceptible to flash floods. During the November 2025 flood event, this mechanism was exacerbated by a tropical cyclone that strengthened the westerly flow (Westerly Burst) toward Sumatra's west coast, maximizing the orographic lifting effect.
- Leeward: After passing the mountain crest, the air mass descends and undergoes adiabatic warming, causing clouds to evaporate and rainfall to decrease. However, in Sumatra, this effect is not as dry as in subtropical regions due to the high background humidity of the tropics.

C. GPM IMERG

Given the limitations of rain gauge networks in inaccessible mountainous regions, satellite-based estimation data is a vital solution. This research utilizes data from the Global Precipitation Measurement (GPM) mission, an international consortium led by NASA and JAXA.

The specific product used is IMERG (Integrated Multi-satellite Retrievals for GPM). IMERG is an advanced fusion algorithm that combines data from:

- Passive Microwave (PMW) Satellites: Provide accurate physical precipitation estimates but with low temporal resolution.
- Infrared (IR) Satellites: From geostationary (GEO) satellites providing high temporal resolution cloud cover data (every 30 minutes). Morphing algorithms are used to project the movement of PMW rain systems based on cloud motion vectors derived from IR data.

This study employs the IMERG Final Run (V07) product. Unlike the Early Run (4-hour latency) or Late Run (14-hour latency), the Final Run is produced with a latency of approximately 3.5 months. This delay allows for the integration of global monthly rain gauge data from the Global Precipitation Climatology Centre (GPCC) for bias correction. This process establishes the IMERG Final Run as satellite data for historical research and model training.

Data Specifications:

- Spatial Resolution: $0.1° \times 0.1°$.
- Temporal Resolution: 30 minutes.

D. Singular Value Decomposition

To handle the massive spatiotemporal dimensions of GPM satellite imagery, Singular Value Decomposition (SVD) is employed. In meteorological and oceanographic literature, this technique is often used interchangeably with Empirical Orthogonal Function (EOF) analysis or Principal Component Analysis (PCA).

The primary objective of SVD/EOF is to reduce data complexity by identifying coherent spatial patterns that explain the largest variance in the dataset. Suppose we have a rainfall anomaly data matrix X of size $M \times N$, where M is the number of locations (spatial grid points) and N is the number of observation time steps.

Data Transformation (Flattening) before applying SVD, the 3D spatiotemporal data (Latitude × Longitude × Time) must be reshaped into a 2D matrix. As illustrated below, the spatial grid at each time step is flattened into a column vector, forming the input matrix X:

Step 1: Flattening Spatial Grids to Vectors

$$\mathbf{x}_t = \text{flatten}\left(\begin{bmatrix} \text{lat}_1, \text{lon}_1 & \cdots & \text{lat}_1, \text{lon}_n \\ \vdots & \ddots & \vdots \\ \text{lat}_m, \text{lon}_1 & \cdots & \text{lat}_m, \text{lon}_n \end{bmatrix}\right) \longrightarrow \begin{bmatrix} x_1 \\ \vdots \\ x_M \end{bmatrix}$$

Step 2: Constructing Matrix X

$$X_{(M\times N)} = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_N \\ | & | & & | \end{bmatrix}$$

Where M is the total number of spatial points (pixels) and N is the number of time steps.

The SVD theorem states that any real matrix X can be factored into:

$$X = U\Sigma V^T$$

Where:

- U(Spatial Matrix): An orthogonal matrix of size $M \times M$. The columns of $U(u_1, u_2, \ldots, u_M)$ are referred to as Left Singular Vectors or EOFs. These vectors represent stationary spatial patterns (maps) of variability modes. The property of spatial orthogonality($u_i^T u_j = \delta_{ij}$)ensures that each mode describes a spatially uncorrelated pattern.

$$U = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_M \\ | & | & & | \end{bmatrix}$$

- Σ(Singular Matrix): A diagonal matrix of size $M \times M$ with non-negative diagonal elements $(\sigma_1, \sigma_2, \ldots)$. These values are called Singular Values and are arranged in descending order($\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$). The square of the singular value (σ_k^2) s directly related to the total variance of the data explained by the k-th mode.

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k \end{bmatrix}$$

- V^T(Temporal Matrix): The transpose of an orthogonal matrix of size($N \times N$). The rows of V^T are referred to as Right Singular Vectors or, in the context of data projection, relate to Principal Components (PC). These vectors($v_1, v_2, \ldots, v_N$) represent the time evolution (time series) of each respective EOF spatial mode.

$$V^T = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_N \\ | & | & & | \end{bmatrix}$$

The data reconstruction relationship for a rank-k approximation (Truncated SVD) is:

$$X \approx \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

E. Long Short-Term Memory (LSTM)

Once the data is reduced to Principal Component (PC) time series, the next step is modeling temporal dynamics using Long Short-Term Memory (LSTM). LSTM is an evolution of Recurrent Neural Networks (RNN) specifically designed to handle sequential data with long-term dependencies.

The main weakness of standard RNNs is the vanishing gradient problem, where error gradients backpropagated through time tend to approach zero exponentially as the sequence length increases. This causes RNNs to fail in learning correlations between today's rainfall and atmospheric precursors occurring weeks or months prior. LSTM addresses this by introducing a memory cell structure (C_t) protected by gate mechanisms, as illustrated in Figure 2.

The information flow in an LSTM unit at time t is governe by three main stages corresponding to the colored sections in the diagram:

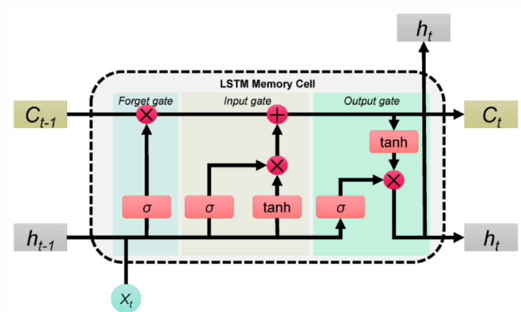

Fig. 2. Architecture of an LSTM Memory Cell showing the flow of data through Forget, Input, and Output gates. [Source: https://mlarchive.com/deep-learning/understanding-long-short-term-memory-networks/]

- Forget Gate (f_t): Located in the cyan block on the left side, this gate decides which information from the previous cell state (C_{t-1}) is no longer relevant. The inputs h_{t-1} and x_t are passed through a sigmoid layer (σ), outputting a value between 0 (completely forget) and 1 (completely keep).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

- Input Gate (i_t): Shown in the beige middle block, this section determines what new information will be stored. It consists of two layers: a sigmoid layer (σ) that decides *which* values to update, and a tanh layer that creates a vector of new candidate values ($\tilde{C}_t$).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

- Cell State Update (C_t): Represented by the top horizontal line running through the diagram. The old cell state C_{t-1} is first multiplied point-wise ($\times$) by the forget factor f_t, and then added point-wise ($+$) with the new scaled input values ($i_t \odot \tilde{C}_t$). This direct linear path is key to avoiding vanishing gradients.

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

- Output Gate (o_t) and Hidden State (h_t): Located in the green block on the right. This gate controls what part of the cell state makes it to the output (h_t). The cell state is put through $\tanh$ (to push values between -1 and 1) and multiplied by the output of the sigmoid gate.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \odot \tanh(C_t)$$

With this architecture, LSTM is capable of capturing complex non-linear temporal patterns, such as seasonal rainfall oscillations or time lags between atmospheric triggers and extreme rainfall events.

F. Integration of the Hybrid SVD-LSTM Model

The hybrid SVD-LSTM model combines the spatial feature extraction advantages of SVD with the temporal prediction capabilities of LSTM. The rationale for this coupling is efficiency and stability.

1) Dimensionality Reduction & Denoising: Satellite rainfall data contains thousands of pixels. Training an LSTM directly on every pixel (pixel-wise) is computationally expensive and prone to noise. SVD compresses this information into a few principal components (e.g., 10–20 PCs) representing >90% of the signal variance, while discarding high-frequency components that often represent noise.
2) Focused Prediction: The LSTM only needs to learn patterns from a small number of clean and orthogonal PC time series. This accelerates training convergence. Recent studies, demonstrate that coupled models like SVD-LSTM consistently outperform single models in accuracy metrics (RMSE, MAE).
3) Spatial Reconstruction: The predicted PC results from the LSTM are projected back into spatial space using the EOF vectors (U). Since EOFs represent physical atmospheric patterns, the reconstructed results automatically maintain realistic spatial structures, avoiding the spatial discontinuity issues often found in independent pixel-wise predictions.

III. Research Methodology

A. Computational Environment and Software

This research was conducted using the Python programming language within a cloud computing environment. Python was selected for its extensive ecosystem of data science libraries. Large-scale remote sensing data integration was facilitated by the Google Earth Engine (GEE) API, enabling petabyte-scale data processing without the need for local raw file storage.

Key libraries utilized include:

- Google Earth Engine: For efficient acquisition and preprocessing of GPM IMERG satellite data.
- NumPy: Employed for high-performance matrix manipulations, specifically linear algebra operations for SVD decomposition.
- TensorFlow / Keras: An industry-standard Deep Learning framework used to construct and train the LSTM neural network.
- Cartopy: A cartographic library used to visualize prediction results onto projection-accurate geospatial maps.

```python
import ee
import geemap
import numpy as np
import matplotlib.pyplot as plt
import cartopy.crs as ccrs
import cartopy.feature as cfeature
import tensorflow as tf
import random
import warnings
import pandas as pd

from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense
from tensorflow.keras.callbacks import EarlyStopping
```

Fig. 3. Library Implementation.

B. Data Acquisition and Study Area

Rainfall data was retrieved from the GPM IMERG Final Run (V07) Level 3 Monthly product. This product was selected as it represents the most advanced global precipitation estimate, fusing microwave and infrared sensors with gauge-calibrated corrections.

- Study Area: The domain is bounded by coordinates $97° - 101°$ E and $0° - 4.5° N$, encompassing the entire physiography of North Sumatra.
- Data Period: A 20-year period was selected (January 1, 2005 – December 31, 2024). This duration is critical for capturing long-term climate cycles, such as the El Niño-Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD), which typically have recurrence intervals of 3–7 years.

```python
# Area: Sumatera Utara
# [West, South, East, North]
bounds = [97.0, 0.0, 101.0, 4.5]
roi_sumut = ee.Geometry.Rectangle(bounds)
scale_meters = 11132

dataset = ee.ImageCollection("NASA/GPM_L3/IMERG_MONTHLY_V07") \
    .select('precipitation') \
    .filterDate('2005-01-01', '2024-12-31') \
    .filterBounds(roi_sumut)
```

Fig. 4. Dataset filtering with GPM IMERG V07.

C. Data Preprocessing

This stage transforms raw satellite data into a format suitable for mathematical analysis.

1) Unit Conversion: The original GPM IMERG data is provided in precipitation rate (mm/hr). For hydrological and flood analysis, this was converted into monthly accumulated rainfall($mm/month$). A conversion factor of 730.5 was applied, representing the average number of hours in a month (24 hours $\times$ 365.25 days/12 months).
2) Spatiotemporal Matrix Construction: Satellite data exists as a 3-dimensional cube (Latitude, Longitude, Time). The SVD algorithm requires a 2-dimensional matrix input. Consequently, the spatial dimensions were flattened so that matrix X has dimensions ($M \times N$), where M is the total number of spatial pixels and N is the number of months (240 months).
3) Anomaly Calculation: Flood prediction models must focus on extreme deviations rather than normal seasonal patterns. Therefore, the climatological mean (the average monthly rainfall pattern over 20 years) was calculated and subtracted from the observed data:

$$X_{anomaly} = X_{observed} - X_{seasonal\_mean}$$

This step removes static seasonal bias, forcing the LSTM to learn the dynamics of actual climate anomalies.

```python
HOURS_PER_MONTH = 730.5
data_matrix = data_matrix * HOURS_PER_MONTH
rows, cols, time_steps = data_matrix.shape
X_flat = data_matrix.reshape(rows * cols, time_steps)
num_years = time_steps // 12
X_reshaped_year = X_flat[:, :num_years*12].reshape(rows*cols, num_years, 12)
climatology = np.mean(X_reshaped_year, axis=1)

climatology_tiled = np.tile(climatology, (1, num_years + 1))[:, :time_steps]
X_anomaly = X_flat - climatology_tiled
```

Fig. 5. Data Preprocessing Implementation.

D. Spatial Decomposition (SVD)

The anomaly matrix was deconstructed using Singular Value Decomposition (SVD). SVD decomposes complex data into three component matrices: U (Spatial Patterns), σ (Signal Strength), and V^T (Temporal Coefficients).

$$X = U \cdot \sigma \cdot V^T$$

This study retained the top k=50 principal modes. This selection was based on Scree Plot analysis, where the first 50 modes explained the majority of the data variance (90%) while discarding trailing modes associated with random noise. By reducing thousands of pixels to just 50 primary features, the computational load on the LSTM is significantly reduced without losing essential spatial information.

```python
U, S, Vt = np.linalg.svd(X_anomaly, full_matrices=False)
varians = np.round((S ** 2) / np.sum(S ** 2) * 100, 2)
k = 50
U_k = U[:, :k]
S_k = np.diag(S[:k])
V_k = Vt[:k, :]
```

Fig. 6. Spatial Decomposition Implementation.

E. Temporal Modeling with LSTM

The temporal coefficients (V^T) derived from SVD, representing the evolution of rainfall pattern intensity over time, served as the model input. Network Architecture Design:

1) Sliding Window Method: Data was prepared with a look-back period of 12 months. This parameter was chosen because the annual hydrological cycle is periodic, rainfall conditions in the coming month are strongly influenced by the trend of the preceding annual cycle.
2) Hidden Layer: An LSTM layer with 64 neuron units was employed. This number provides sufficient memory capacity to capture complex non-linear patterns without causing overfitting.
3) Optimizer: The Adam optimizer was utilized for its ability to adaptively adjust the learning rate, accelerating model convergence.
4) Loss Function: Mean Squared Error (MSE) was used as the loss function, as it is sensitive to outlier errors, making it suitable for predicting extreme values.

```python
model = Sequential()
model.add(LSTM(64, input_shape=(look_back, k)))
model.add(Dense(32, activation='relu'))
model.add(Dense(k))
model.compile(loss='mse', optimizer='adam')
```

Fig. 7. Temporal Modeling with LSTM Implementation.

F. Recursive Forecasting and Spatial Reconstruction

The primary challenge in predicting for the year 2025 is the absence of future input data. To address this, a Recursive Prediction strategy was implemented:

1) The model predicts January 2025 using data ending in December 2024.

2) The prediction for January is then treated as ground truth and appended to the input to predict February 2025.
3) This process repeats iteratively until December 2025.

The final step is Spatial Reconstruction. The LSTM predictions are abstract coefficients. To obtain interpretable flood maps, these coefficients are projected back using the spatial patterns (U) and singular weights (S), and the climatological mean is added back:

$$X_{final\_forecast} = (U \cdot S \cdot V^T_{predicted} + X_{climatology})$$

The result is a precipitation map in units of $mm/month$, visualizing the location and intensity of potential future flooding.

```
forecast_steps = 12
curr_seq = pc_scaled[-look_back:]
forecast_scaled = []

for _ in range(forecast_steps):
    input_seq = curr_seq.reshape(1, look_back, k)
    pred = model.predict(input_seq, verbose=0)
    forecast_scaled.append(pred[0])
    curr_seq = np.vstack((curr_seq[1:], pred))

forecast_pcs = scaler.inverse_transform(np.array(forecast_scaled))
raw_mode1 = V_k.T[:, 0]

dates_hist = pd.date_range(start='2005-01-01', periods=len(raw_mode1), freq='MS')
dates_future = pd.date_range(start='2025-01-01', periods=12, freq='MS')

train_predict = model.predict(X_train, verbose=0)
train_predict_original = scaler.inverse_transform(train_predict)
train_mode1 = train_predict_original[:, 0]
dates_train = dates_hist[look_back:]

X_recon_anomaly = np.dot(U_k, np.dot(S_k, forecast_pcs.T))
climatology_2025 = climatology[:, :12]
X_final_2025 = X_recon_anomaly + climatology_2025

rata_rata_bulanan = np.mean(X_final_2025, axis=0)
idx_puncak = np.argmax(rata_rata_bulanan)
nama_bulan = calendar.month_name[idx_puncak + 1]
```

Fig. 8. Recursive Forecasting and Spatial Reconstruction Implementation.

IV. RESULT AND ANALYSIS

A. Decomposition of Dominant Spatiotemporal Patterns

The fundamental step in this study involves reducing the dimensionality of complex rainfall data to extract pure climate signals.

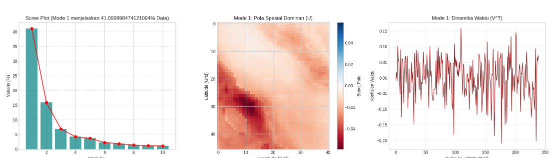

Fig. 9. SVD decomposition results: Scree plot, spatial pattern (U), and temporal dynamics (V^T) of the dominant mode (Mode 1).

The Scree Plot analysis (Left Panel) indicates that Mode 1 plays a critical role, explaining 41.1% of the total variability in rainfall data across North Sumatra, suggesting that nearly half of the rainfall behavior in this region is governed by a single dominant pattern. Furthermore, the spatial analysis (Center Panel) reveals that the spatial weight map (U) features a concentration of deep red hues indicating high intensity, extending vertically parallel to the Bukit Barisan Mountain Range. This confirms that the primary rainfall pattern in North Sumatra is orographic, where moist air masses from the Indian Ocean are uplifted by the mountains, undergo condensation, and precipitate heavily on the western slopes such as Sibolga, Tapanuli, and Mandailing Natal. Meanwhile, the temporal analysis (Right Panel) demonstrates that the temporal coefficient graph (V^T) exhibits rapid fluctuations embedded with a cyclical trend, representing the dynamic interaction between seasonal oscillations (Monsoons) and interannual climate disturbances such as ENSO and IOD.

B. Deep Learning Model Performance

Following the extraction of dominant patterns, the LSTM model was trained to learn the temporal evolution of these features.

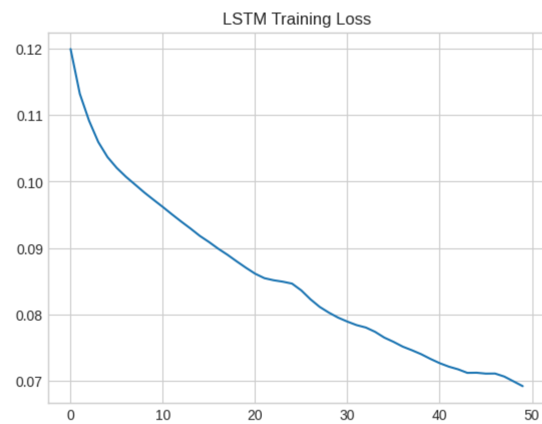

Fig. 10. LSTM training loss curve showing model convergence over 50 epochs.

The Training Loss graph demonstrates excellent model convergence. The loss value decreases steadily as epochs (learning iterations) increase, stabilizing at a low value (¡ 0.07). This proves that the constructed LSTM architecture is capable of learning long-term temporal dependencies without suffering from overfitting or underfitting. The model successfully captures non-linear patterns from historical data (2005-2024) to project conditions for 2025.

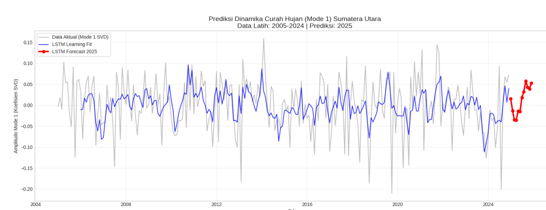

Fig. 11. Forecasted temporal coefficient (PC-1) for 2025, indicating a peak anomaly in late year.

The forecasting results for 2025 (Red Line) exhibit an extreme upward trend in the final quarter of the year. The sharp spike in the graph during November and December 2025 indicates the presence of a strong weather anomaly that is projected to trigger rainfall significantly above the normal average.

C. Spatial Projection of Flood Risk in November 2025

Based on the spatial reconstruction from the LSTM predictions, the following maps illustrate the potential rainfall distribution within the study area.

- North Sumatra

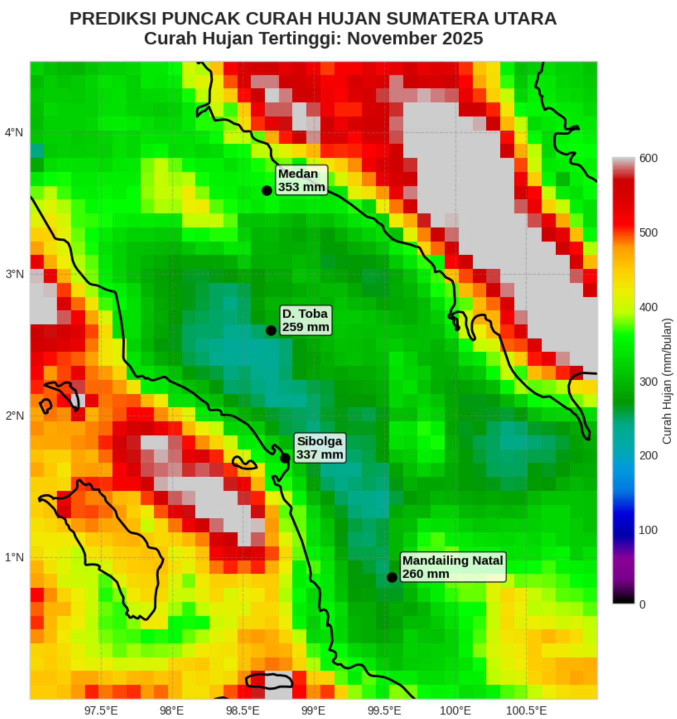

Fig. 12. Spatial projection of extreme rainfall distribution in North Sumatra for November 2025.

The prediction map indicates that North Sumatra is poised to experience rainfall in the "Very High" category during November 2025, with distinct spatial characteristics across the province. In the West Coast and Highland regions, cities such as Sibolga (337 mm) and Mandailing Natal (260 mm) are situated in deep green to yellow zones, signaling high rainfall intensities primarily driven by orographic effects. Concurrently, on the eastern coast, the urban center of Medan is projected to receive rainfall amounting to 353 mm/month. This figure is particularly critical given Medan's low-lying topography and the susceptibility of the Deli watershed to overflow. Rainfall exceeding the 300 mm threshold in such densely populated areas serves as a strong indicator for potential urban and riverine flooding.

- Regional Validation

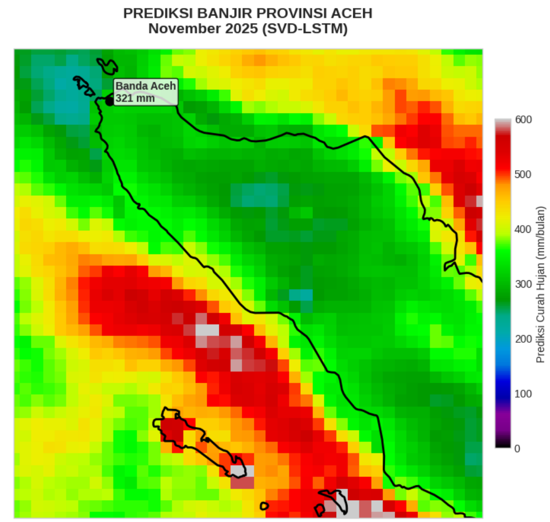

Fig. 13. Regional validation: Predicted rainfall intensity map for Aceh Province.

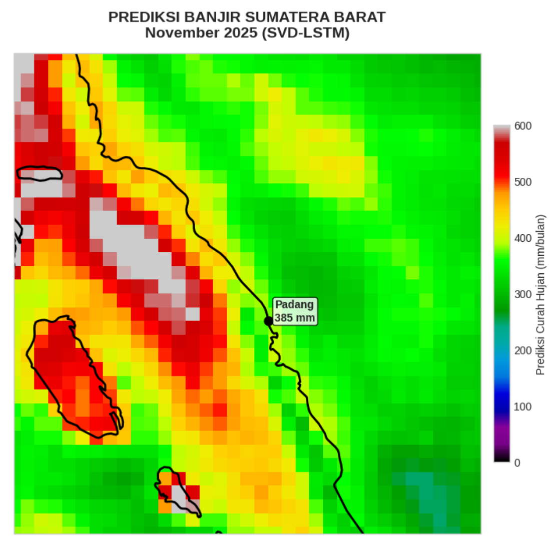

Fig. 14. Regional validation: Predicted rainfall intensity map for West Sumatra Province.

To validate that this prediction is not merely a localized anomaly, a comparative analysis was conducted with neighboring provinces. The results reveal that in Aceh Province, Banda Aceh is projected to reach 321 mm, with extreme rainfall patterns appearing evenly distributed along its western coast. Similarly, to the south, the city of Padang in West Sumatra is predicted to experience extreme rainfall reaching 385 mm. The consistency of these high-intensity patterns (indicated by red-yellow zones) across Aceh, North Sumatra, and West Sumatra provides compelling evidence that the driver of this potential flood event is a synoptic-scale regional phenomenon affecting the entire northwestern sector of Sumatra Island, rather than isolated local showers.

D. Meteorological and Environmental Synthesis

The high rainfall predicted by the SVD-LSTM model aligns with the atmospheric dynamics recorded and forecasted for the period.

1) Influence of Global Phenomena
According to the BMKG weather outlook, global atmospheric conditions in late 2025 highly favor massive rain cloud formation in western Indonesia. Specifically, the Dipole Mode Index (DMI) is recorded at -0.6, indicating a Negative IOD phase where sea surface temperatures in the waters west of Sumatra are warmer than usual, thereby supplying abundant water vapor to North Sumatra. Furthermore, although the La Niña condition is considered weak with a Nino 3.4 index recorded at -0.42, its combination with the Negative IOD creates a constructive interference effect. This interaction significantly increases rainfall potential above normal levels, exacerbating the hydrological risks in the region.
2) Regional Atmospheric Disturbances
The model's prediction is reinforced by the presence of regional weather disturbances. Tropical Cyclone "SENYAR" has been detected in the Malacca Strait, directly drawing air masses and intensifying rainfall in Aceh and North Sumatra. Field observation data recorded extreme daily rainfall in North Aceh (310.8 mm/day) and Medan (262.2 mm/day). This validates the model's prediction placing Medan as a flood hotspot with high monthly accumulation. Furthermore, the activity of Rossby Equatorial Waves and MJO Phase 6 further exacerbates the duration and intensity of the rainfall.
3) Deforestation and Land Use Change
Beyond meteorological factors, the impact of flooding in North Sumatra is significantly aggravated by environmental degradation, particularly deforestation and land use change. Massive deforestation in upstream watersheds, such as those around Lake Toba and the Bukit Barisan range, drastically reduces the soil's infiltration capacity. Consequently, the high rainfall predicted by the model (300-400 mm) is likely to convert directly into rapid and destructive surface run-off rather than being absorbed by the ground. In regions with steep topography like Mandailing Natal and Sibolga, this combination of extreme rainfall and the loss of forest cover sharply increases the risk of flash floods and landslides, a pattern consistent with the historical disaster records of the region.

V. CONCLUSION

This study demonstrates the effectiveness of the hybrid Singular Value Decomposition (SVD) and Long Short-Term Memory (LSTM) approach in modeling and predicting extreme rainfall. The application of SVD proved critical in reducing the dimensionality of complex spatiotemporal rainfall data, where the decomposition of Mode 1 alone accounted for 41.1% of the dominant variability in North Sumatra's rainfall patterns. This dimensionality reduction enabled the LSTM model to operate more efficiently by isolating pure climatic signals from data noise, resulting in stable training convergence and low loss values.

Based on forecasts derived from this reduced data, the model projects rainfall anomalies in the "Very High" category for November 2025 in North Sumatra, identifying critical hotspots in Medan (353 mm), Sibolga, and Mandailing Natal. The consistency of these predictive patterns, which extend regionally to Aceh and West Sumatra provinces, highlights the model's capability to capture synoptic-scale weather phenomena driven by global atmospheric dynamics such as La Niña and Negative IOD. Consequently, the SVD-LSTM framework serves as a reliable and valid early warning instrument for future hydrometeorological disaster mitigation.

VI. APPENDIX

The program source code can be accessed here and a video explanation of this paper can be found here.

VII. ACKNOWLEDGMENT

First of all, the author would like to express gratitude and thanks to God Almighty for His blessings and grace, which have given the author the opportunity to write and complete this paper as well and as thoroughly as possible. The author would also like to thank his family for their continuous support and encouragement during both difficult and happy times. The author would also like to express his deepest gratitude to the lecturer of IF2123 K-01, Dr. Ir. Rinaldi Munir, S.T, M.T., who has always shared his knowledge and taught the author many things that will prepare him for the future. In addition, I would like to thank my classmates in K-01 who have provided various dynamics both inside and outside the classroom. I would also like to thank all of my Artificial Intelligence friends who have helped me brainstorm ideas and paraphrase my poor English. Lastly, thank you to myself for continuing to learn and striving to complete this paper. I truly enjoyed this course so much that my hair kept growing, which signifies the wealth of knowledge I gained from this course.

If you are working on something that you really care about, you don't have to be pushed. The vision pulls you

Steve Jobs

REFERENCES

[1] R. Munir, "Singular Value Decomposition (SVD)," *Informatika STEI ITB*, Des. 19, 2025. [Online]. Available: https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2024-2025/15-Teori-Bilangan-Bagian1-2024.pdf

[2] M. L. Khodra, "01 RNN: What & Why," unpublished lecture notes, *Informatika STEI ITB*, Indonesia, 2025. [Accessed: Dec. 19, 2025].

[3] M. L. Khodra, "04 LSTM: What & Why," unpublished lecture notes, *Informatika STEI ITB*, Indonesia, 2025. [Accessed: Dec. 19, 2025].

[4] Kompas.com, "BMKG Ungkap Penyebab Banjir Sumatera: Curah Hujan Bulanan Tumpah dalam Satu Hari," Nasional Kompas, Dec. 01, 2025. [Online]. Available: https://nasional.kompas.com/read/2025/12/01/15570891/bmkg-ungkap-penyebab-banjir-sumatera-curah-hujan-bulanan-tumpah-dalam-satu. [Accessed: Dec. 19, 2025].

[5] Redaksi Berita, “Banjir Sumatera: Kombinasi Faktor Alam dan Ulah Manusia,” [Online]. Available: https://www.ums.ac.id/berita/teropong-jagat/banjir-sumatera-kombinasi-faktor-alam-dan-ulah-manusia. [Accessed: Dec. 19, 2025].

[6] S. Poornima and M. Pushpalatha, “Prediction of Rainfall Using Intensified LSTM Based Recurrent Neural Network with Weighted Linear Units,” *Atmosphere*, vol. 10, no. 11, p. 668, 2019.

[7] GeeksforGeeks, “Deep Learning: Introduction to Long Short Term Memory,” [Online]. Available: https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/. [Accessed: Dec. 19, 2025].

[8] Agungnoe, “Bencana Banjir Bandang Sumatra, Pakar UGM Sebut Akibat Kerusakan Ekosistem Hutan di Hulu DAS,” *Universitas Gadjah Mada*, Dec. 01, 2025. [Online]. Available: https://ugm.ac.id/id/berita/bencana-banjir-bandang-sumatra-pakar-ugm-sebut-akibat-kerusakan-ekosistem-hutan-di-hulu-das/. [Accessed: Dec. 24, 2025].

[9] NASA., “NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG),” Algorithm Theoretical Basis Document, NASA, 2019. [Online]. Available: https://gpm.nasa.gov/data/imerg.

[10] T. H. Ramadhani, “Bukit Barisan: Tulang Punggung Geografis Pulau Sumatera,” [Online]. Available: https://pantauriau.com/news/detail/43314/bukit-barisan-tulang-punggung-geografis-pulau-sumatera. [Accessed: Dec. 24, 2025].

[11] E. Aldrian and R. D. Susanto, “IDENTIFICATION OF THREE DOMINANT RAINFALL REGIONS WITHIN INDONESIA AND THEIR RELATIONSHIP TO SEA SURFACE TEMPERATURE,” [Online]. Available: https://pure.mpg.de/. [Accessed: Dec. 19, 2025].

STATEMENT OF ORIGINALITY

I hereby declare that this paper is my own writing, not an adaptation, or translation of someone else’s paper, and not plagiarized.

Bandung, 24 Desember 2025

Reinhard Alfonzo Hutabarat, 13524056