Hybrid deep learning model for flood frequency assessment and flood forecasting

Rajendra P. Pandey | Meena Desai | Rajnesh Panwar

Abstract The most common and persistent natural hazard to people across the globe is flooding. The frequency of floods in a given place is defined as the likelihood and intensity of floods occurring there within a certain period. Examining historical flood data and using techniques are often used to determine the likelihood that a flood of a certain size would occur in a specific location. The method of flood prediction involves making forecasts on the frequency and severity of flooding. It may be influenced by a number of factors, including the topography, river flow, soil moisture content, and the period of rainfall. In this research, we provide a novel Cat Swarm Optimized Spatial Adversarial Network (CSO-SAN) technique for predicting and assessing flood frequency. This technique simulates the yearly greatest flow at the river Mahanadi measurement sites at Andhiyarkore, Bamanidhi, Baronda, and Kurubhatta over 60 years. The CSO-SAN model is adapted for the flood forecasting component to predict the frequency and size of future floods. The model incorporates real-time data from various sources, such as meteorological predictions and information on river flow, to anticipate the probability and severity of upcoming floods. Compared to other conventional statistical techniques and forecasting models, the CSO-SAN model outperformed them in tests conducted on the Mahanadi river basins. The model offers a viable method for improving the precision of flood frequency evaluation and flood forecasting, with significant advantages for managing and reducing flood risk.

Keywords: deep learning, flood frequency, floods forecasting, CSO-SAN

1. Introduction

As one of the most devastating natural hazards, flooding has a negative effect on both the health of populations and their financial well-being. Efficient flood prevention is urgently needed since severe weather conditions have raised in many places of the globe, probably due to shifts in climate scenarios (Dong et al. 2020). A reliable and successful method for flood forecasting is a key component of planning and controlling flood occurrence. A reliable long-term flood forecasting model produces the smallest amount of inaccuracy, which also enables decision-making on optimal reservoir management to maximize operating efficiencies and the lowest amount of flood hazards and provides enough time for creating an appropriate disaster mitigation plan (Alipour et al. 2020). A structured procedure for developing flood maps has been made in the European Union (EU) as a result of the EU Flood Directive (2007/60/EC), which is specific about what types of floods, what flood features, and how often flooding ought to be assessed when evaluating the risk of flooding. The EU is becoming more interested in flooding that results directly from heavy rainfall, often called pluvial floods, due to an increase in natural catastrophes (David and Schmalz 2020).

Additionally, pluvial floods have recently been simulated by combined hydrological-hydraulic techniques using rain-on-grid addressing with time-dependent 2-dimensional models due to a highly interrelated rainfall-runoff process, rapid advancements in mathematical technology, and the accessibility of excellent quality topographic data. However, due to the absence of actual data necessary to validate models, it is still challenging to provide an accurate flood forecast in tiny ungauged basins (Annis et al. 2020). Human activities greatly impact the environment as society and the economy advance. Due to its naturally fragile ecosystem and susceptibility to long-term improper human activity, the loess area experiences significant loss of water and soil erosion. The loess area has developed into a site regularly endangered by flash flood catastrophes due to complicated topographical and geological characteristics and frequent storm drains (Anaraki et al. 2021). This has caused a lot of academic curiosity since storm flooding in the Loess region is very common and precise.

The constraints of traditional flood frequency analysis and forecasting techniques include the presumption of stable data and the incapability to capture intricate nonlinear interactions. The Deep Learning (DL) model has been suggested for flood frequency assessments and forecasting to overcome these constraints. To increase the efficacy and accuracy of flood prediction, the model combines the benefits of both DL and conventional techniques (Samantaray et al. 2021). Hybrid DL
algorithms can precisely and quickly anticipate flood occurrences by concurrently examining different data sources. By examining real-time data from sensing and social media feeds to pinpoint regions of need and organize rescue activities, hybrid deep learning models may help with disaster response and making predictions. To locate damaged regions and gauge the success of relief activities, they may also assist with following disaster reconstruction efforts by examining satellite images (Ren et al 2019).

(Zanchetta and Coulibaly 2020) summarized and analyzed recent developments in understanding the environmental variables that precede severe rainfall events, developing monitoring systems for pertinent hydro meteorological variables, and operationalizing weather and hydrological simulations for the forecasting of flash floods. Most study has concentrated on developing techniques for data from multiple sources, assimilation, and integration, and putting together strategies for uncertainty assessments because there has been exponential growth in data availability and computing capacity. The rainfall criterion may significantly reduce the computational complexity of anticipating floods using a data-driven method. (Ke et al 2020) employed a rainfall criterion for identifying flood vs. non-flood occurrences based on Machine Learning (ML) methodologies, applied to an instance of Shenzhen city in China, to prepare the communities against regular pluvial flood events, particularly in a changing environment. Increasing atmospheric temperatures have caused a drop in Eastern Hemlock (EH) in the Northeastern United States, which has led to an increase in the water supply of the watershed. (Knighton et al 2019) assessed the likelihood that a shift in the flooding regime would result from an EH loss. Using stable isotope measurements of stream, soil, and plant xylem water, they first examined how the root system regulates plant hydraulics in EH and American Beech. (Sit and Demir 2019) aimed to investigate how well Artificial Deep Neural Networks (ADNN) anticipates floods. The research offers a dataset that focuses on the connection of measurements on river systems but offers algorithms that may be used to anticipate the stream stage. Additionally, it demonstrates how neural networks may improve current models by incorporating data and be particularly beneficial in time-series prediction, such as in flood disasters.

(Ding et al 2020) discussed the necessity for flood forecasting, and an understandable Spatio-Temporal Attention Long Short-Term Memory model (STA-LSTM) based on LSTM and attention mechanisms were proposed. They built the model using LSTM and flexible attention mechanisms, normalized the data using the Max-Min approach, chose the hyper parameters using the variable management method, and then trained the models using the Adam algorithm. To forecast severe flooding in a tiny highland watershed with varied lead times, (Wu et al 2019) suggested a powerful artificial intelligence-based model called a Support Vector Regression (SVR) model created from the idea of statistical learning. The physiological concept of reaction time establishes the delays associated with the algorithm's input factors. The equation accepts the postponed average amounts of rainfall and runoff as input parameters. In a hilly watershed in China, 69 flash flood occurrences were gathered from 1984 to 2012 and utilized for model training and evaluation. Quantitative Precipitation Forecasts (QPFs) based on longer-term mathematical weather forecasts or short-term infrared extensions are used for actual time flood forecasting is one of the more effective methods in this respect. They created a novel real-time mixing strategy in (Yoon 2019), to increase the precision of rainfall predictions for hydrological purposes. Each method has unique benefits and drawbacks. They examined the hydrologic application of six QPFs utilized in Seoul, South Korea, for predicting urban flooding. The Long Short-Term Memory (LSTM) neural network was used in (Liu et al 2020), to model rainfall-runoff correlations for catchments with various climatic variables. Several streams in China with different climatic zones were used to evaluate the LSTM approach. To confirm the LSTM model's superiority in terms of historical prediction issues, the Recurrent Neural Network (RNN) was chosen for evaluation. To assess the viability of this new approach, the outcomes of LSTM were also compared with those of the Xinjiang model (XAI), a popular process-based model. In this study, we present a novel Cat Swarm Optimized Spatial Adversarial Network (CSO-SAN) technique for calculating the frequency of flooding and executing flood forecasting.

1.1. Problem statement

For authorities to plan for and lessen the effects of flood occurrences, flood frequency assessments, and floods are essential for flood management. However, since hydrological systems are intricate and dynamic, it may be difficult to determine the frequency of flooding and estimate when it will occur. The absence of precise and current data on hydrological factors, including rainfall, river flow, and soil moisture, is one of the major obstacles in assessing flood frequency. As a result, estimating the frequency of flood episodes and precisely modeling their likelihood are challenging tasks. The 'creeping' and sneaky character of flood outbreaks makes it difficult to create reliable systems to send prompt flood alerts. This is because designing early warning systems involves knowledge of several technologies. It seems sense that such an issue might pose a greater problem for underdeveloped countries like Fiji. For this reason, an economical solution with a low investment need for such technologies is preferred for flood forecasting. The novel Cat Swarm will solve these issues Optimized Spatial Adversarial Network (CSO-SAN) flood modeling approach that this study presents.

The other sections of the article are as follows: section 2 discusses methodology, section 3 discusses findings, and the last section gives the conclusion.

2. Materials and Methods

https://www.malque.pub/ojs/index.php/msj
In this paper, a hybrid model called CSO-SAN is proposed to forecast the rising potential of a flood channel network using information obtained from sensors that measure flooding. CSO-SAN was the one that we decided to go with because of its shorter training periods, lower prediction costs, and increased prediction accuracy.

2.1. Dataset

The Mahanadi River is a large peninsular river that flows east in eastern-central India. It has a total size of 1, 41, 589 km and is separated into the delta, center, and upper areas. It makes up around 4.3% of India’s total land area, with most of the basin being in Odisha and Chhattisgarh. Before reaching the Bay of Bengal, the river traverses the remaining portion of the regions of Jharkhand, Maharashtra, and Madhya Pradesh. Near the hamlet of Pharsiya in Raipur, Chhattisgarh, the Mahanadi begins at an elevation of around 442 meters above mean sea level. The distance from its origin to where it converges at the Bay of Bengal is about 851 km, of which 357 kilometers are in Chhattisgarh, and 494 km are in Odisha. It is between coordinates 19o20' and 23o35' N and longitudes 80o30' to 86o50' E. The subtropical climate of this area has maximum temperatures of approximately 45°C in May and lowest temperatures of about four °C in winter from December to January. During the monsoon season, which lasts from June to October, the average annual rainfall is 1200 mm. Over the Mahanadi basin, the amount of rainfall is unequal. As shown in Figure 1, the gauging sites in Andhiyarkore, Bamanidhi, Baronda, and Kurubhatta are where the present research is being carried out (Sahoo et al 2019). Natural occurrences like high rain during the monsoon, notably in the basin’s upper region, which almost generally occurs every two to three years, generate overbank flow, leading to severe floods. The monsoonal months of May to October, from 1960 to 2019, were the period for which each month’s floodwater discharge was gathered from IMD, Raipur, and Chhattisgarh, India.

![Mahanadi River Basin](image)

**Figure 1** Proposed field of research.

2.2. Cat Swarm Optimization (CSO)

Cat Swarm Optimization (CSO) may be used to estimate flood frequency and anticipate floods by enhancing the parameters of mathematical models that express the likelihood of flood occurrences.

The Log-Pearson Type III (LPIII) distribution, often used to simulate the likelihood of flood episodes, may have its parameters optimized using CSO, for instance, in evaluating flood frequency. The subsequent expression describes the LPIII distributions:

\[ F(x) = 1 - (1 - P)^T \] (1)

Where T, the return time, is the typical number of years among floods of similar or larger size, P is the exceedance possibility of the flood, and F(x), is the total distribution function of flood magnitude x.

The location variable(μ), scale variable(σ), and shape variable(γ) of the LPIII distribution may all be optimized using the CSO method. CSO may increase the accuracy of the flood frequency prediction model by determining the ideal parameter values that best suit the existing data.

Similarly, CSO may be utilized in flood forecasting to improve the weights and biases of neural networks for predicting upcoming flood occurrences. The formula that follows may be employed when representing a neural network:

\[ y = f(Wx + b) \] (2)

Where W is the weight matrix, b is the bias vector, f is the activation function, y is the anticipated flood size, and x is the input data (for example, rainfall river flow).
By reducing the difference between the predicted and actual flood occurrences, the CSO method may improve the neural network’s weights and biases. CSO may increase the efficacy of the flood forecasting system by determining the ideal weights and biases values.

2.2. Spatial Adversarial Network (SAN)

Spatial Adversarial Networks (SANs) are a sort of deep learning architecture that may be used for flood frequency evaluation and prediction by creating synthetic flood data and increasing the accuracy of predictive models. This can be accomplished using SAN ability to improve the quality of predicting systems.

The generator network and a network of discriminators comprise the two primary parts of SAN. Simulated flood data produced by the generator network is supplied into the discriminator network with actual flood data. The discriminator network has been trained to differentiate between actual and false flood data. The generator network is then upgraded to provide more accurate flood data. The procedure is continued until there is no difference between the created and actual flood data.

The deep neural network (DNN) is a form of neural network composed exclusively of direct and inverted convolutional layers. In essence, neural networks combine many “neurons” or processing units to calculate the (sometimes complicated) correlations among input, x, and output, y, data vectors. The traditional DNN design also stacks additional layers of neurons. The neuron’s usual expression is provided by

\[ g(y) = e(y, z + a) \]  

(3)

Where \( g(\cdot) \) stands for the neuron’s scalar output, and \( e(\cdot, \cdot) \) is a recognized quadratic expression as the “activation function,” \( e(\cdot, \cdot) \) stands for the exponential outcome, \( z = [z_1, ..., z_M] \) is a group of masses with the same number of dimensions \( M \) as \( y \), and \( a \) is the bias related to the neuron. A DNN has to be “learned” or taught to be helpful. Every neuron has its own adjusted \( z \) and parameters. Throughout learning so the DNN can carry out a particular job as effectively as feasible. The descent of gradients can maximize \( z \) and \( a \) when \( e(\cdot) \) is distinguishable. The rectified linear unit (ReLU), sigmoid function, and hyperbolic tangent function are examples of common \( e(\cdot) \) forms.

SAN can produce artificial flood data that accurately mimics the geographic patterns of flood occurrences. This may be especially helpful when historical data are absent or when the data that is available does not accurately reflect current or projected flood trends.

The convolutional layer is the main component of the Convolutional Neural Network (CNN) architecture. Due to the convolutional operator’s explicit consideration of the spatial arrangement in the input data, it has been widely used to estimate flood frequency. The convolutional layer, \( h \), is created using a succession of k=1 when the input data is a 2-D picture;

\[ g_{l, u}(Y_{x,u}) = e \sum_{j=1}^{M_l} \sum_{i=1}^{M_j} Z_{l,j}^{j,i} Y_{v} + j, u + i + a_l \]  

(4)

During flood frequency, SAN may be used to create artificial flood data added to the existing data to increase the precision of statistical models like the LPIII distribution. By doing so, you may lessen the uncertainty around the available data and help the flood frequency estimation algorithm become more accurate.

To train neural networks to anticipate future flood occurrences, SAN may be used to create artificial flood data. This may aid in boosting the flood forecasting model’s accuracy, especially when historical data are absent, or the available data does not accurately reflect current or potential flood trends. SAN is a potential method for estimating flood frequency and predicting floods that may assist in increasing the precision and dependability of predictive models. By combining it with additional strategies like Cat Swarm Optimization or assimilation of approaches, its efficacy may be further increased.

2.3. Cat Swarm Optimized Spatial Adversarial Network (CSO-SAN)

For the evaluation and forecasting of flood frequency, the Cat Swarm Optimized Spatial Adversarial Network (CSO-SAN) is a hybrid deep learning system that combines Cat Swarm Optimization (CSO) with Spatial Adversarial Network (SAN).

The following equations may be used to describe the CSO-SAN algorithm:

Generator network:

\[ z \sim p(z) \]  

(5)

\[ x' = G(z) \]  

(6)

\( G \) is the generator network improved through CSO, \( z \) is a random noise vector, and \( p(z) \) is a probability distribution function.

Network discriminator:

\[ y = D(x) \]  

(7)
\[ y' = D(x') \]  \hspace{1cm} (8)

Where \( y \) and \( y' \) are the outputs of the discriminator network for the actual and synthetic flood data, respectively, and \( D \) is the discriminator network that is improved via CSO.

In summary, the CSO-SAN strategy shows promise for enhancing models for estimating flood frequency and predicting floods. By combining it with additional strategies like physical-based models or data assimilation techniques, its efficacy may be further increased.

3. Results

This section examines the efficacy of various models for predicting floods in the Mahanadi River, including Deep Convolutional Neural Network (DCNN) (Guo et al 2021), Long Short-Term Memory (LSTM) (Le et al 2019), Fully Convolutional Network (FCN) (Nemni et al 2020), and the proposed Cat Swarm Optimized Spatial Adversarial Network (CSO-SAN) model. The classifier output may be stated as follows in the framework of the CSO-SAN approach for calculating flood frequency and flood forecasting: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). These four categories assess the accuracy, precision, recall, and f1-measure of the approaches above, including the CSO-SAN method. To increase the precision and dependability of the model's predictions, the objective is to maximize the number of true positives and negatives while decreasing the number of false positives and false negatives.

3.1. Accuracy

The effectiveness of models for calculating flood frequency and flood forecasting is often assessed using the measure of accuracy. It calculates the model's accuracy as a percentage of the total number of predictions.

\[ \text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \]  \hspace{1cm} (9)

Figure 2 and Table 1 compare the suggested CSO-SAN method's accuracy with other flood forecasting techniques. The findings show that the CSO-SAN approach is more accurate than the current methods.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCNN</td>
<td>91.5</td>
</tr>
<tr>
<td>LSTM</td>
<td>92</td>
</tr>
<tr>
<td>FCN</td>
<td>94</td>
</tr>
<tr>
<td>CSO-SAN [Proposed]</td>
<td>98.3</td>
</tr>
</tbody>
</table>

Table 1 Accuracy outcomes with different techniques.

![Figure 2 Comparative Analysis of Accuracy.](https://www.malque.pub/ojs/index.php/msj)

3.2. Precision

Precision is a measure of how well the approach predicts floods. With fewer false positive predictions (FP) and a high accuracy score, the model properly predicts the presence of floods in most of its positive predictions.

\[ \text{Precision} = \frac{TP}{TP+FP} \]  \hspace{1cm} (10)

Figure 3 and Table 2 compare the suggested CSO-SAN method's precision with other flood forecasting techniques. The findings show that the CSO-SAN approach is more precise than the current methods.
3.3. Recall

Recall, commonly called sensitivity or true positive rate, is a performance statistic utilized in binary classification issues like flood frequency estimation and flood forecasting. It assesses the model's capacity to accurately detect all positive situations: the frequency of flood events.

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

The recall of the proposed CSO-SAN approach is compared with conventional flood forecasting methods in Figure 4 and Table 3. The results demonstrate that the CSO-SAN technique is more effective than the current methods.

### Table 2 Precision outcomes with different techniques.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Precision (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCNN</td>
<td>80.3</td>
</tr>
<tr>
<td>LSTM</td>
<td>85.4</td>
</tr>
<tr>
<td>FCN</td>
<td>89.8</td>
</tr>
<tr>
<td>CSO-SAN [Proposed]</td>
<td>91.4</td>
</tr>
</tbody>
</table>

**Figure 3 Comparative Analysis of Precision.**

### Table 3 Recall outcomes with different techniques.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Recall (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DCNN</td>
</tr>
<tr>
<td>0</td>
<td>65</td>
</tr>
<tr>
<td>1</td>
<td>69</td>
</tr>
<tr>
<td>2</td>
<td>67</td>
</tr>
<tr>
<td>3</td>
<td>65</td>
</tr>
<tr>
<td>4</td>
<td>68</td>
</tr>
</tbody>
</table>

**Figure 4 Comparative Analysis of Recall.**
3.4. F1-measure

The F1 measure is a helpful indicator since it strikes a compromise between recall and accuracy. It may be crucial in datasets with imbalances where one class is much more prevalent than the other. The accuracy and recall of the classification are weighted fundamental means, and the F1-measure is defined as:

\[
F1 - measure = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]  

The recommended CSO-SAN method F1-measure is contrasted with existing flood forecasting methods in Figure 5 and Table 4. The results demonstrate that the CSO-SAN strategy is superior to the other existing approaches in accuracy.

Table 4 F1-measure outcomes with different techniques.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>62</td>
<td>72</td>
<td>82</td>
<td>92</td>
</tr>
<tr>
<td>1</td>
<td>63</td>
<td>73</td>
<td>83</td>
<td>93</td>
</tr>
<tr>
<td>2</td>
<td>65</td>
<td>75</td>
<td>85</td>
<td>95</td>
</tr>
<tr>
<td>3</td>
<td>67</td>
<td>77</td>
<td>87</td>
<td>97</td>
</tr>
<tr>
<td>4</td>
<td>69</td>
<td>79</td>
<td>89</td>
<td>99</td>
</tr>
</tbody>
</table>

![Figure 5 Comparative Analysis of F1-measure.](image)

3.5. Discussion

The current flood warning systems usually track the spread of floods, such as the increase in the river and other bodies of water. However, there is a rising chance to create more complex models that might foretell infrastructure breakdowns during flood events as more sophisticated infrastructure data and methodology become available. The suggested hybrid DL model aims at precisely forecasting flood propagation. The case study in Mahanadi shows how the suggested CSO-SAN paradigm might provide early warning for cascading failure and sequential interruption. The suggested CSO-SAN model also applies to other infrastructure networks, such as the electricity grid and transportation systems, as well as other academic fields, like biology and information science. Its use is not only restricted to flood prediction.

The cascade failure of the system and the access to necessary facilities may therefore be anticipated using the records of component breakdowns over time. The proposed CSO-SAN model performs well in forecasting the flood cascade, although there are still certain areas where it might be improved. The CSO-SAN model is trained in this article using three flood events. However, a bigger training data set may further illuminate the benefit of DL techniques. Therefore, the training set may be increased to produce a more reliable and accurate flood forecast as more flood occurrences are recorded. Second, the construction sector also contributes significantly to the control of urban floods. Third, fresh categorization and period analysis methods are always being created as DL models advance quickly.

4. Conclusion

This paper provides a hybrid DL model called Cat Swarm Optimized Spatial Adversarial Network (CSO-SAN) and it’s testing in a case study of flood prediction carried out in the Mahanadi river basin based on hydrological information obtained at four gauging stations. The CSO-SAN technique is a potential approach for determining flood frequency and predicting floods. The complicated nonlinear interactions between flood variables have been successfully captured by the CSO-SAN approach, making it suitable for forecasting and frequency study of floods. The predictive capability of the framework is
measured at 98.3% accuracy. Because the data set was unbalanced, we used precision, recall, and F-measure statistics to determine which parameters produced the most accurate results. According to the findings, the CSO-SAN model is superior to all of the other models used, which indicates that it may be used to acquire an accurate assessment of the frequency of floods and make predictions. Even though the CSO-SAN technique has shown encouraging results, more can be done. The CSO-SAN approach may be optimized by experimenting with other hyper parameters and structures or by incorporating additional machine learning methods.

Ethical considerations
Not applicable.

Declaration of interest
The authors declare no conflicts of interest.

Funding
This research did not receive any financial support.

References


