AI-enabled Landslide Recognition for Effective Public Health Risk Management

Nikita Sharma[1], Sharyu Bhushan Ikhar[1]
[1]*Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India*
[2]*Research Scholar, Department of CS & IT, Kalinga University, Raipur, India*

KEYWORDS

Landslides Recognition, Construction Activities, Remote sensing (RS), Rat Swarm (RS) optimization, Satellite

ABSTRACT

The frequency of hazardous landslides has increased worldwide as a result of increased heavy rainfall events and increased human construction activities. The assessment of landslide vulnerability is an essential and effective method for preventing landslides. To resolve these issues, this paper develops a novel Rat Swarm integrated Random Forest (RSRF) method to manage the risk evolved through the landslide. The LISS-III satellite dataset of remote sensing (RS) images was gathered for the study. The preprocessing is performed by employing the z-score normalization to standardize the data images. The rat swarm (RS) optimization enhances the feature selection in the landslide by efficiently forecasting relevant assessment and the random forest (RF) employed to improve the classification accuracy in the landslide recognition process. Various existing methods are utilized for the comparison performance with the proposed RSRF techniques. The outcome shows that the proposed RSRF method improved more significantly than all other existing techniques in terms of area under the curve (AUC (93.3%)), F1-score (85.5%), log-loss (40.4%), and recall (97%).

1. Introduction

Landslide disaster risk reduction (DRR) is not an exception, and it appears that global efforts and methods have evolved from delivering technical assessments of landslides to increase the involvement of various stakeholders, among the research and the public sector that is highly valued for its collaborative assignment [1]. The landslides represent a serious risk to human safety and property as they occur to cause rock and dirt to slide down a slope's weak area caused by natural or man-made factors [2]. The process of making decisions called risk assessment involves evaluating the possibility of landslide-related death against the resources available and perceptions of appropriate risk [3]. Risk management frequently employs a quantitative framework to evaluate both individual and public risks [12]. Assessing risk involves assessing the perceived level of danger and displayed risks of landslides with resources available to determine that intervention is required [4]. Landslides are one of the greatest hazards associated with nature and can result in significant loss of lives and assets, along with impairments to the natural environment and its services. Disasters known as landslides affect people and result in numerous deaths worldwide [5]. To manage the public health risks associated with landslides, the study aims to establish a novel technique called RSRF [9].

The remaining part of this study is as follows: Phase 2: Literature Review, Phase 3: Methodology, Phase 4: Results and Discussions, and Phase 5: Conclusion.

2. Literature Review

The phrase Residual Risk Assessment (RRA) referred to an innovative method for determining the risk that exists after precautions have been implemented in the study of Frazier et al. [13] and it contributed to the classification. The outcomes showed that the social and medical system risk differed spatially throughout the research region, demonstrating that the health systems' residual risk and management capacities significantly varied [6]. The development of a data-driven tool that could be applied instantly to landslide risk assessment, estimating the probability of fatality depending on the severity of landslides was provided in the article of Pollock and Wartman [7]. They discovered that human involvement was the main cause of death among destruction levels of around 1-6 meter. The development with response to the requirements, the research of Davis and Gandía [8] presented an innovative collaborative risk communication framework. By elaborating the ideas and the related procedure, it advanced the areas of risk communication and disaster management by promoting cooperative risk communication efforts throughout many social and regional environments. Based on the examination of existing research, the author of Puente-Sotomayor et al. [14] addressed landslide risk reduction employing a comprehensive conceptual framework. It indicated to supplement the environmental dimension that frequently illustrated the physical condition of risk with the fundamental causes of landslide risk by examining the economic and socio-cultural features and the historical and regulations that support the earlier aspects [15]. To illustrate an area affected by the possibility of landslide occurrence, the study by Modugno et al. [10] provided the Geographic information system (GIS)-based multi-scale method. The findings emphasized the value of a multi-scale approach and geomorphologic factors such as cover of land, topographic humidity, and local climate conditions possessed more explanation power at the sub-regional level, biological variables remained able to identify areas that were significant areas for landslides at the national level.

2. Methodology

This section, the study region and data preprocessing techniques are explored and Figure 1 depicts the flow of the proposed methodology.

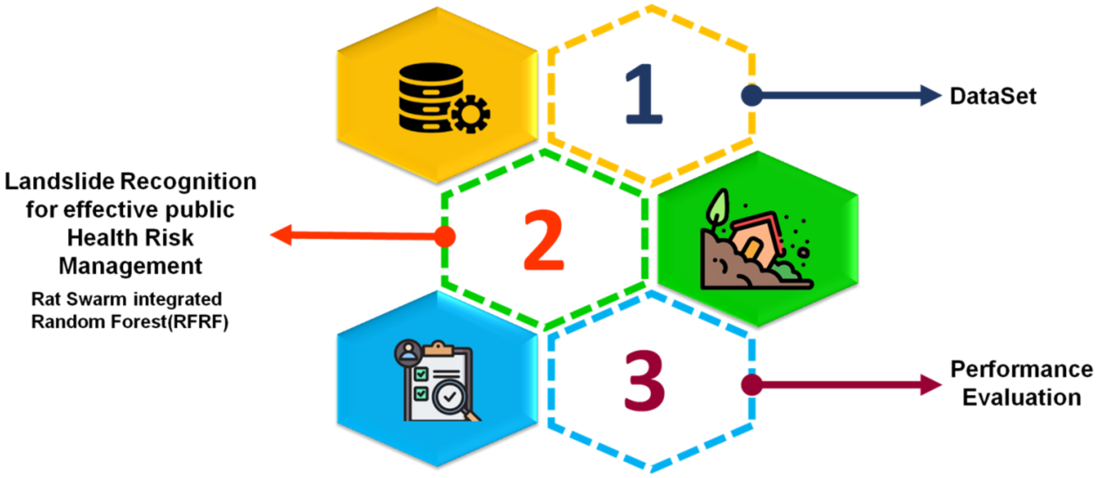

Figure 1 Proposed Methodology

Data Preparation and Preprocessing

We gathered the dataset from Kaggle (https://www.kaggle.com/datasets/kkhandekar/lanslide-recent-incidents-india). The following columns are included in the dataset. Title-The landslides incident's title. An account of the landslide occurrence is called a landslide incidence.

Landslide Recognition Employing Rat Swarm Integrated Random Forest (RSRF)

For the recognition of landslides for effective public health risk management, we have combined the rat swarm (RS) optimization algorithm with the random forest (RF) machine learning (ML) method for appropriate prediction.

Rat Swarm (RS) Optimization

The RS optimization is a metaheuristic algorithm that is based on the hunting and following behavior of rats. Swarms of male and female rats are native species. Certain animals can die as a result of the often very aggressive behavior of rats. It uses mathematical modeling of rats' aggressive and following behaviors for optimization. Following an evaluation by an objective function, the randomized set is iteratively constructed by examining the aggressive and resulting behaviors of rats. Based on the initial RS approach, the starting positions of feasible solutions are selected at random within the search area. These positions are denoted in Equation (1).

$$b_i = b_{imin} + rand * (b_{imax} - b_{imin}), i = 1,2, \dots, n \quad (1)$$

Where the lower and upper boundaries of the i^{th} variable are represented by b_{imax} and b_{imin}, and n is the total number of individuals. The remaining search agents can discuss their positions with the top search agents discovered currently. The following example, Equation (2), has been presented to demonstrate that rats attack using bait and to calculate the actual location of the rats as they move in advance.

$$\overrightarrow{s_y}(b+1) = |\overrightarrow{s_p},(b) - \vec{s}|$$
(2)

The most optimal solution found away is $\overrightarrow{s_p}$, wherein $\overrightarrow{s_y}(b+1)$ updates its locations with rats. s obtained from the preceding equation by applying the following Equation (3).

$$\vec{s} = n * \overrightarrow{s_y}(b) + C * \left(\overrightarrow{s_p}(b) - \overrightarrow{s_y}(b)\right)$$
(3)

The parameters n and C are determined by applying the following Equations (4 and 5), where $\overrightarrow{s_y}(b)$ indicates the locations of the rats.

$$n = X - b * \left(\frac{X}{Iter_{max}}\right), b = 1, 2, 3, \ldots, Iter_{max}$$
(4)

$$C = 2 * rand$$
(5)

Random Forest (RF)

Two parts of the enhanced random forest technique include the RF classifier method that consists of many decision trees. RF separates the data into smaller portions and creates the tree's branches. The outcome is a tree containing leaf and decision nodes at every level. Each determined component's relevance is shown by several branches in the decision node and the leaf node maintains the significance of the result of the individual's potential circumstances. By employing many classifier decision trees, the possibility that one tree of decisions could be inadequate to predict the value property has been removed. The RF connects the results from many trees to generate the final result. For the RF, approximation error, and confidence estimations, the functions in margins were represented by Equations (6-8). In the instance, an ensemble of classifiers can be illustrated by the values $h_1(B), h_2(B), \ldots, h_k(B)$, and the input information is found in the vectors$Y$, X. That is the accurate expression of the margin.

$$MG(B,A) = yv_k I(h_k(B) = A) - max_{j \neq A} yv_k I(h_k(B) = j)$$
(6)

The indicator function is represented by $I(.)$. These values represent the generalization error.

$$PE *= P_{B,A}\big(ME(B,A)\big) < 0$$
(7)

The possibility is explained by using the B, A dimension. Every collection of trees in a RF has more classifiers as $h_k(B) = H(B, \Theta_k)$. Based on the effective equation of large numbers and forest structure, the probability $PE *$ determines Equation (8).

$$P_{B,A}\big(P_\Theta(h(B,\Theta) = A) - max_{j \neq A} P_\Theta(h(B,\Theta) = j) < 0\big)$$
(8)

The weighted algorithm provides an updated technique to improve the model after the landslide recognition. This is one of the most efficient techniques in ML. Below is the Equation (9) to describe the adjustment technique.

$$H(b) = sign\left(\sum_{t=1}^{t} \alpha_t h_t(B)\right)$$
(9)

Considering that for each $t = 1, \ldots., t$, $(B_1, A_1), \ldots, (B_M, A_M)$, for $B_i \in b, A_i \in a = \{-1, +1\}$. Begin from $D_1(i) = \frac{1}{M}$. WRF employs the use of the distributions (D_t) after inadequate workouts. The normalization variable in this instance is represented by Z_t. The result appears to be the following Equation (10).

$$D_{t+1(i)} = \frac{D_t(i)}{Z_t} \times \left\{E^{-\alpha t} \ \ if \ \ h_t(B_i) = A_i \ E^{\alpha t} \ \ if \ \ h_t(B_i) \neq A_i = \frac{D_t(i)exp(-\alpha t A_i h_t(B_i))}{Z_t}\right\}$$
(10)

RF technique provides accurate predictions by preventing the model's overall variance from increasing significantly. Where, X is the random number and C is a selected value in the interval $[0, 2]$. Whereas $Iter_{max}$ is the maximum iteration that can be obtained, b indicates the most recent iteration of the optimization process. The proposed RSRF an innovative method improves the recognition of landslides that is essential to manage public health assessment. It makes utilization of RS optimization abilities and RF capability of prediction to accurately identify the regions that have landslide possibilities. It facilitates immediate actions and reduces the health risk and it is an efficient, innovative technique for preparation and landslide mitigations.

3. Results and discussion

The system utilizes an Intel core CPU, 16GB RAM and python 3.9.13 as programming language. Landslide location features were employed in this study to validate the obtained landslide assessment. The algorithms that were utilized were evaluated for performance. It assessed the performance parameter, including recall, AUC (area under the curve), log-loss, and F1-score, to evaluate the forecasting accuracy of various methods like “LSTM (long short-term memory), RNN or recurrent neural networks and DNN (deep neural networks)” [11].

The objective of the research is to increase the consistent utilization of AUC as outcome predictors in public health by providing illustrative instances of various risk factors. The more reliable the predictor and that is the way AUC is calculated in Equation (11). Log loss is one of the performance metrics and it is utilized in classification operations to calculate the reliability of possible predictions. Equation (12) shows the log-loss calculation. Figure 2 (a and b) illustrates the outcomes of AUC and log loss.

$$AUC = \int_0^1 f(x)dx$$
(11)

$$Log - loss = -\frac{1}{N}\sum_{i=1}^{N}[y_i log(p_i) + (1 - y_i)log(1 - p_i)]$$
(12)

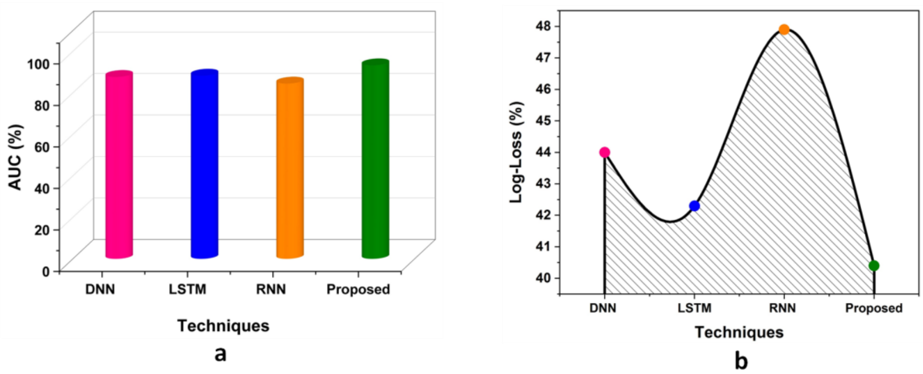

Figure 2 Comparison Outcomes of (a) AUC and (b) Log-loss

Recall frequently referred to as sensitivity or the proportion of pertinent cases in a dataset, indicates the effectiveness of the model recognizes all applicable instances. The total amount of positive occurrences is divided by the total amount of true positives (TP) to determine the overall number. Equation (13) expresses the recall. The F1-score is the harmonic measure of recall and precision. It is most useful for circumstances where a dataset combines these two measures and has a different class distribution. Decisions in recall and precision are represented by an enhanced F1-Score. Equation (14) is utilized to calculate the F1-score. In Figure 3 (a and b), the comparison outcomes of recall and F1-score is demonstrated. Table 1 depicts the outcomes of all the methods.

$$Recall = \frac{TP}{TP+FN}$$
(13)

$$F1 - score = 2 * \frac{Precision*Recall}{Precision+Recall}$$
(14)

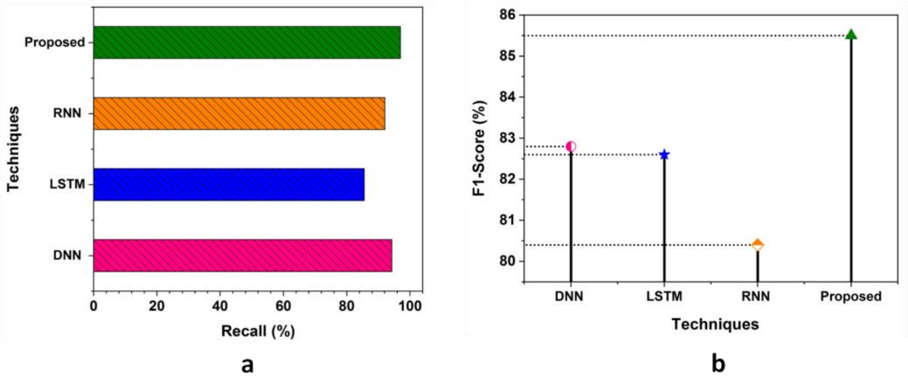

Figure 3 Comparison Outcomes of (a) F1-score and (b) Recall

Table 1 Comparison of Results of Proposed and Existing Techniques

Techniques	**AUC (%)**	**F1-score (%)**	**Log-loss (%)**	**Recall (%)**
DNN	87.7	82.8	44.0	94.3
LSTM	88.3	82.6	42.3	85.6
RNN	84.5	80.4	47.9	92.1
Proposed	**93.3**	**85.5**	**40.4**	**97**

4. Conclusion and future scope

The frequency of landslides has grown due to an increase in both human construction and the occurrence of severe rainfall events on a worldwide basis. The crucial and successful approach to preventing landslides is assessing the risk of landslides. To solve these issues, the study developed the novel RSRF technique to manage the risk that has emerged as an outcome of the landslide. For the investigation, remote sensing (RS) images from the LISS-III satellite collection were gathered. The z-score normalization is used in the preprocessing to normalize the gathered images. To compare the performance of the proposed RSRF techniques, a variety of existing approaches are explored. In terms of AUC (93.3%), F1-score (85.5%), log-loss (40.4%), and recall (97%), the results demonstrate that the proposed RSRF method outperformed all other existing methods. Limitations of the proposed technique include the need for high-quality satellite data and the possibility of over-fitting with complicated characteristics. Further research could focus on integrating more data sources.

Reference

[1] NS. Ruiz-Cortés, and I. Alcántara-Ayala, "Landslide exposure awareness: a community-based approach towards the engagement of children", Landslides, pp. 1501-14, 2020. https://doi.org/10.1007/s10346-020-01391-w

[2] H. Bao, C. Zeng, Y. Peng, S. Wu, "The use of digital technologies for landslide disaster risk research and disaster risk management: progress and prospects", Environmental Earth Sciences, pp. 81(18), 446, 2022. https://doi.org/10.1007/s12665-022-10575-7

[3] Pranjit Kalita, Mitali Barman, Anupama Chetia, Priyanka Goyary, & Komal Kumari. (2024). Mapping the Landscape of

Literature on Use of AI in Libraries: A Bibliometric Analysis Using Scopus Database. Indian Journal of Information Sources and Services, 14(2), 24–27. https://doi.org/10.51983/ijiss-2024.14.2.04

[4] A. Strouth, and S. McDougall, "Societal risk evaluation for landslides: historical synthesis and proposed tools", Landslides, 18(3), pp.1071-1085, 2021. https://doi.org/10.1007/s10346-020-01547-8

[5] P. Arrogante-Funes, A.G. Bruzón, F. Arrogante-Funes, R.N. Ramos-Bernal and R. Vázquez-Jiménez, "Integration of vulnerability and hazard factors for landslide risk assessment", International Journal of environmental research and public health, 18(22), p.11987, 2021. https://doi.org/10.3390/ijerph182211987

[6] Cabra, J. L., Parra, C., & Trujillo, L. (2022). Earprint touchscreen sensoring comparison between hand-crafted features and transfer learning for smartphone authentication. Journal of Internet Services and Information Security JISIS, 12(3), 16-29.

[7] W. Pollock, and J. Wartman, "Human vulnerability to landslides", GeoHealth, 4(10), p.e2020GH000287, 2020. https://doi.org/10.1029/2020GH000287

[8] M.S. Davis, and Á. Gandía, "Principles for collaborative risk communication: Reducing landslide losses in Puerto Rico", Journal of Emergency Management, 19(8), 2021. https://doi.org/10.5055/jem.0547

[9] Kiyomoto, S., Fukushima, K., & Miyake, Y. (2012). Design of Categorization Mechanism for Disaster-Information-Gathering System. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 3(4), 21-34.

[10] S. Modugno, S.C.M. Johnson, P. Borrelli, E. Alam, N. Bezak, and H. Balzter, "Analysis of human exposure to landslides with a GIS multiscale approach", Natural Hazards, 112(1), pp.387-412, 2022. https://doi.org/10.1007/s11069-021-05186-7

[11] JM. Habumugisha, N. M. Chen, Rahman, MM. Islam, H. Ahmad, A. Elbeltagi, G. Sharma, SN. Liza, A. Dewan. "Landslide susceptibility mapping with deep learning algorithms", Sustainability. 14(3), 1734, 2022. https://doi.org/10.3390/su14031734

[12] A. Strouth and S. McDougall, "Individual risk evaluation for landslides: key details". Landslides, 19(4), pp.977-991, 2022. https://doi.org/10.1007/s10346-021-01838-8

[13] T.G. Frazier, E.X. Wood, and A.G. Peterson, "Residual risk in public health and disaster management", Applied Geography, 125, p.102365, 2020. https://doi.org/10.1016/j.apgeog.2020.102365

[14] F. Puente-Sotomayor, A. Egas, and J. Teller, "Land policies for landslide risk reduction in Andean cities", Habitat International, 107, p.102298, 2021. https://doi.org/10.1016/j.habitatint.2020.102298

[15] Llopiz-Guerra, K., Daline, U.R., Ronald, M.H., Valia, L.V.M., Jadira, D.R.J.N., Karla, R.S. (2024). Importance of Environmental Education in the Context of Natural Sustainability. *Natural and Engineering Sciences*, *9*(1), 57-71.