CNN-LSTM-attention deep learning model for mapping landslide susceptibility in Kerala, India

Cheng Chen [1], Lei Fan[1*]

[1] Department of Civil Engineering, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China – (Cheng.chen19@student.xjtlu.edu.cn, leifan@xjtlu.edu.cn)

KEY WORDS: Landslide susceptibility; Deep learning; CNN, LSTM; Attention mechanism.

ABSTRACT:

As a typical type of natural disaster, landslides may result in injuries to humans, threats to property security, and economic loss. As such, it is important to understand or predict the probability of landslide occurrence at potentially risky sites. Using typically machine learning (ML) to estimate landslide susceptibility based on a landslide inventory and a set of factors that impact the occurrence of landslides is a common practice. However, in landslide susceptibility assessment, existing DL-based neural network methods use a fully connected layer to optimize the selection of factors, which limits their efficiency in extracting features of those contributing factors. In response to those problems, this study proposed a CNN-LSTM model with an attention mechanism (AM) to avoid the complex optimization of input factors while the same or even better prediction accuracy can be achieved. To compare our method with the existing ones, the historical landslide inventory and the remote sensing data of Kerala, India were used to produce the input variables needed in the methods considered. The results show that our method produced more accurate results, compared to those existing neural network methods (e.g. CNN, LSTM and CNN-LSTM).

1. INTRODUCTION

As a common natural hazard, landslides are characterized by the downward movement of a mass of soils and rocks (Leynaud et al. 2017). They may result in varied degrees of infrastructure destructions and human injuries (Peduto et al. 2021), in addition to a variety of potential damaging consequences on the social and environmental sustainability. Prediction of landslides is essential for landslide prevention and control, as it provides the basis for determining whether to take actions to mitigate the probable negative effects of landslides. In this regard, landslide susceptibility mapping (LSM) is frequently considered (Reichenbach et al. 2018), where the categorization and the geographical distribution of a potential landslide are evaluated. It is also an effective and proactive method for delineating landslide-prone zones.

Many data-driven approaches have been used to generate LSM, ranging from simple to complex mathematical models. Many studies were conducted on the landslide genesis to determine the mathematical or the statistical links between the contributing factors and the probability of landslide occurrence. It is widely accepted that machine learning (ML) more efficiently build the nonlinear relationship between landslides occurrence and contributing factors (Wei et al. 2022).

The statistical prediction models for LSM have widely been explored in the field. To improve the accuracy of LSM, deep learning (DL) as extended ML is capable of learning more complex and hidden features by a hierarchical analysis of features. Therefore, DL frameworks have recently been applied to the landslide susceptibility assessment to efficiently extract more features from the mapping of landslide contributing factors and to enhance the accuracy of the prediction. DL models are less likely to have the defects of ML models such as local optimum and overfitting. A convolutional neural network (CNN) is a variant of multilayer perceptron, consisting of one or more convolutional, max-pooling and fully connected layers. It benefits from local connections, shared weights and the use of the multiple layers to identify the information that is correlated and invariant to locations within local groups of values (Sameen et al. 2020). Convolutional neural networks (CNNs) have demonstrated their capabilities of automatically extracting a large number of robust and invariant features (Romero et al. 2015). Wang et al. (2019) introduced a CNN into LSM for the first time and achieved better LSM results than those obtained using SVM and MLP. Sameen et al. (Sameen et al. 2020) used a DL classifier for LSM based on the feature extraction by 1D-CNN. Pham et al. (2020) used swarm intelligence algorithm for parameter optimization in the CNN-based LSM model to improve the LSM prediction accuracy. Fang et al. (2020) integrated a CNN with three conventional machine learning classifiers and conducted a case study in Yongxin County, China. Xiao et al. (Xiao et al. 2018) also applied long short time memory (LSTM) to landslide susceptibility. LSTM is a special recurrent neural network (RNN) architecture that inherits the sequence learning advantage of RNN and is able to learn time-series data with long time dependence. With its memory block structure, the LSTM model can determine whether the rules learned at the previous time step are useful or not, and then determine whether the learned rules should be passed to the next time step or discarded. Thus, the prediction accuracy is not affected by errors at some previous data.

Despite the aforementioned progress made by the state-of-the-art DL approaches on LSM, a big challenge in their implementation is the selection and combination of the input contributing factors, which may be categorised into topography, geology, hydrology and land-cover conditions. Atkinson and Massari (1998) used generalised linear modelling to build LSM, and found that the importance of land cover and concavity of slope increased when the other factors were added. Gaidzik and Ramírez-Herrera (2021) compared the performance of LSM using empirically selected contributing factors with that using all contributing factors, and found that the former outperformed in terms of the overall prediction accuracy. These suggest that

https://doi.org/10.5194/isprs-annals-X-3-W1-2022-25-2022

the selection of input factors is essential for the prediction accuracy of LSM.

To solve the aforementioned issues, this paper proposes a CNN-LSTM landslide susceptibility mapping model with the attention mechanism. CNN is used to effectively extract the certain features (i.e. abstract concepts for describing the raw images) from each contributing factor map. The attention module is added to assign different weights to contributing factors to optimize the usage effectiveness of feature maps for improved prediction performance. LSTM works as a decoder to predict the probability of landslide occurrence. The maps of topographical, geological, hydrological and land-cover contributing factors were produced using remote sensing images of Kerala, India, and were used to test our model. The performance of our model was compared.

2. STUDY SITE AND DATA

2.1 Study area

Kerala has been selected as the evaluation location for our method. As seen in Figure 1, it is situated in the southwestern portion of the Indian peninsula on the windward slopes of the Western Ghats and the eastern shore of the Arabian Sea (Hao et al. 2020). This area has a climate typical of the tropics, with seasonal monsoon features. Kerala's bedrock is heavily eroded, resulting in a thick, poorly-consolidated soil layer (mostly clay) across a large portion of the territory (Sajinkumar et al. 2011). In terms of geomorphology, its eastern portion consists of rocky mountains with deep valleys and plateaus, whilst its western shore consists of plains (Vishnu et al. 2019). The Western Ghats are governed by old faulted escarpments situated on the plateau and can have very steep slopes prone to slope collapses (Kuriakose et al. 2009). The maps of the research region (shown in Fig. 1) were generated using Esri's ArcMap software.

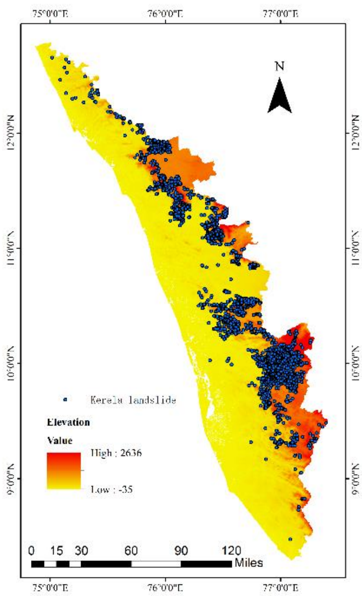


Fig. 1. Study area

2.2 Study data and landslide contributing factors

Kerala's historical landslide inventory was compiled by the Indian Space Research Organisation's National Remote Sensing Centre (NRSC) and the Geological Survey of India. It is accessible on the NRSC website (NRSC 2018). The dataset included a total of 4,728 identified landslides at the time it was consulted for this investigation. As seen in Fig. 1, these landslide locations are represented by a scattering of blue dots. Google Earth (Gorelick et al. 2017) and the Indian Institute of Remote Sensing (Ramasamy et al. 2021) were used to get remote sensing photos of the research region. NASA's SRTM Digital Elevation provided the DEM map with a spatial resolution of 30 meters.

Contributing factors are involved in the formation of LSM (Sameen et al. 2020). In this study, fifteen contributing factors were considered, including topography factors (i.e. altitude, aspect, slope, plan curvature and profile curvature), geological factors (i.e. lithology, distance to faults), land-cover factors (i.e. land use, distance to roads, normalized difference vegetation index (NDVI)), and hydrological factors (i.e. distance to stream, rainfall (Lee and Talib 2005), sediment transport index (STI), stream power index (SPI), topographic wetness index (TWI)). The maps of the aforementioned landslide contributing factors can readily be derived, which are essentially the input data to LSM.

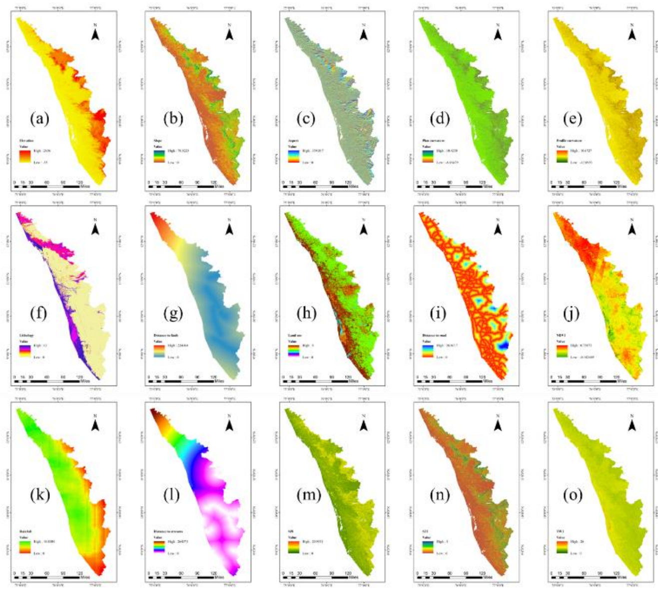


Fig. 2. Maps of the contributing factors considered: (a) altitude, (b) slope, (c) aspect, (d) plan curvature, (e) profile curvature (f) lithology, (g) distance to faults, (h) land use, (i) distance to road, (j) *NDVI*, (k) rainfall, (l) distance to stream, (m) *SPI*, (n) *STI* and (o) *TWI*.

3. METHODOLOGY

3.1 The CNN-LSTM-attention deep learning model proposed

In our method, a network model combining CNN and LSTM is developed, and meanwhile an attention mechanism is added to enhance the useful information. Its overall structure is shown in Fig. 3. The maps of contributing factors are derived from the remote sensing images and pre-processed as the inputs to our model. As shown in Fig. 3, the CNN framework is used as an encoder to take the input maps to extract features. Specifically, this framework is constructed by a deep residual block, which is a simple but extremely effective network structure that adds a skip connection to the simple forward propagation (Kwon 2021). The framework consists of three one-dimensional convolutional layers, three normalization layers, three ReLU (Rectified Linear Units) activation functions and a maximum pooling layer. Since the inputs are continuous spatial sequences, one-dimensional convolutional layers are used. To fully utilise the spatial information of the contributing factor maps, continuous

convolution, activation functions and pooling layers are used to extract features from those maps and to deepen the network.
After the encoder, the input data are transformed into higher-dimensional feature maps. Since the feature maps extracted by the convolution and the pooling operations do not change their order, the feature maps are fed directly into LSTM. A LSTM network processes series data through the gating mechanism, including forget gates, input gates and output gates. They can control the removal or the addition of information for forgetting and remembering, respectively. The feature maps are converted to the corresponding hidden states by the LSTM network. After that, the LSTM hidden layer states are used to schedule attention weights to each contributing factor map. The attention weights affect the input and the output of the LSTM cells, where the channels with higher weights will have more information being fed into the LSTM network. The attention weights can be adjusted dynamically to improve the effectiveness of data extraction of the network during training. Its final output is a prediction of the probability of landslide occurrences at individual pixel locations.

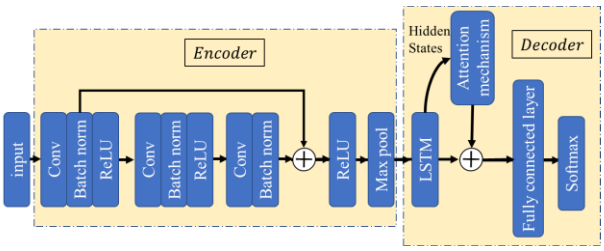


Fig 3. Overall structure of our method

3.2 Component structure of CNN

The basic CNN always has an input layer, a convolutional layer, an activation function, a pooling layer, a fully connected layer, and an output layer. The CNN acts as an encoder to extract features and captures the salient feature attributes from the dataset, enabling it to differentiate without the need for manually driven complex rules. The purpose of the convolution operation is to extract different features of the input layer. More convolutional layers enable iterative learning of more complex representations from low-level features. Pooling is a form of down-sampling that is used to reduce the dimensionality of the feature maps without changing the depth. As the most commonly used operation among different pooling methods, MaxPooling can retain strong features and eliminate weak features to reduce the model complexity and avoid overfitting. The activation function ReLU of the convolution layer is an exponential linear unit, which can speed up the convergence and improve the robustness of the model.

3.3 Component structure of LSTM

LSTM is special type of RNN, which can avoid the vanishing gradient problem and accelerate convergence (Bahad et al. 2019). As for the basic structure of LSTM, there are three cells (input gate, forget gate and output gate) responsible for regulating in or out of the memory cell. The main function of the input gate is to add information to the memory unit. The forget gate controls whether to remember or delete the information from the previous step. The output gate is responsible for providing useful information to the subsequent memory block. The data processing flow in the LSTM unit is shown in Fig. 5

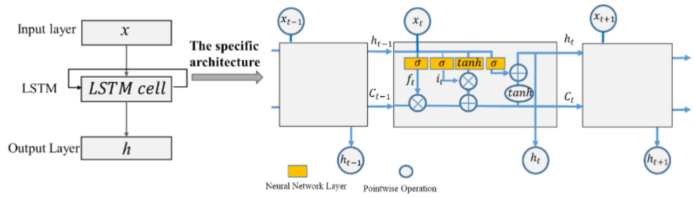


Fig. 4. Basic LSTM cells.

The LSTM unit is specifically explained in Fig.5. The input x (a complex feature extracted from the contributing factors map, which is the output from the CNN), the hidden state from the previous state h_{t-1} and the previous cell state c_{t-1} are input through the LSTM unit to get the output gate. The output gate is a value between 0 and 1 where the value 0 represents the complete deletion of information and the value 1 represents the complete retention of information. Next, the input gate uses two values x_t and h_{t-1} to consider storing new information in the new cell state t. At the same time, the value through the input modulation gate has a hyperbolic tangent activation function, so that the output value ranges from -1 to +1, which reflects the amount of information to be forgotten. Subsequently, the old cell state c_{t-1} is updated to the new cell state c_t by multiplying the output of the old cell state and the forget gate, followed by adding the multiplication of the input gate and the input modulation gate output. After that, the output gate takes the input value and updates the old hidden states. Finally, the cell state c_t goes through the hyperbolic tangent function, and the output of the output gate is multiplied by the new hidden state h_t. The equations associated with this calculation processing are given in Eq. (1) – Eq. (6). By updating the learnable parameters such as the coefficient matrices and the bias vectors, the model can learn more data information through the LSTM unit.

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

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, l_t\right] + b_i\right) \tag{2}$$

$$\tilde{c}_t = \tanh\left(W_c \cdot \left[h_{t-1}, l_t\right] + b_c\right) \tag{3}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \tag{4}$$

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, l_t\right] + b_o\right) \tag{5}$$

$$h_t = o_t \circ \tanh\left(c_t\right) \tag{6}$$

where σ is the sigmoid function; tanh is the hyperbolic tangent function; $\circ$ is the pointwise multiplication operator from Hadamade product; i_t, f_t, o_t, $\tilde{c}_t$, c_t and l_t are input gate, forget gate, output gate, input modulation gate, cell state, and hidden state at time t, respectively; W_i, W_f, W_o, W_c are the coefficient matrices; b_i, b_f, b_o, b_c are the bias.

3.4 Attention mechanism

The attention mechanism prefers to selectively learn the inputs and to correlate the inputs with the output sequence at the output of the model. In more details, the prediction results in the output sequence depend on which input items are selected.
The output from the hidden layer of LSTM is represented as the input of the attention to obtain the distribution of the initial attention weights. In the decoding phase, the attention mechanism uses the attention weights represented by Eq. 7 to select the relevant parts from the hidden vectors. The hidden information between the front and the back layers is split over time and useful outputs are generated from the sequence using

Eq. 8. Probability is forwarded to the fully connected layer (FCN) for predicting the probability of landslide occurrences. The weights represent the importance of the state parameters. These hidden states are aggregated into a vector representation C by the attention computed using Eq.7 to Eq. 9.

$$e_i = u \tanh(wh_i + b) \quad (7)$$

$$\alpha_i = \frac{exp(e_i)}{\sum_i exp(e_i)} \quad (8)$$

$$C = \sum_i \alpha_i h_i \quad (9)$$

where e_i indicates the probability distribution of attention at i-th moment; u and w denote the weighing coefficients; b is the bias coefficient; tanh is hyperbolic tangent function; α_i is the summation weight obtained by the normalization of the alignment coefficient and C is the weighted feature.

3.5 Model evaluation criteria

To assess the accuracy of the results produced by the models considered, the following set of evaluation criteria are considered in our investigation. True Positive (TP) and True Negative (TN) represent the number of correctly predicted landslide and non-landslide, respectively. False Positives (FP) and False Negatives (FN) represent the number of mis-predicted landslide and non-landslide. The statistical metrics such as accuracy, recall, Precision, ROC curve and AUC, can be calculated using the following four metrics (i.e. TP, TN, FP and FN).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (10)$$

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

$$Precision = \frac{TN}{TN+FP} \quad (12)$$

$$F1_{score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (13)$$

$$Kappa = \frac{P_{Acc} - P_e}{1 - P_e} \quad (14)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (15)$$

where, P_{Acc} represents the classification accuracy; P_e represents the random correct rate.

3.6 Experiments descriptions

The performance of our model (i.e. the CNN-LSTM with attention mechanism) for LSM using the Kerala dataset is presented. The effects of the selection of input data (i.e. the contributing factor maps) on the prediction results of our model are explored. The Kerala dataset had recorded a total number of 4728 identified landslides. As LSM is a binary classification task that needs positive and negative samples, non-landslide data with the same proportion of identified landslide is randomly selected. The Kerala dataset was randomly divided into three subsets for training (70%), validation (15%) and testing (15%), respectively. The training dataset was used to train the models considered, while the validation dataset was used to optimize the parameters of those models. Finally, our model and the reference models were evaluated using the test dataset, the performances of which were compared. The experiments were implemented using TensorFlow, which are open source software libraries that use data flow graphs and have widely been used DL. Five-fold cross validation was used to objectively characterize the performance of the models considered, and the mean value of the five-fold cross validation results is used to represent model evaluation criteria.

3.7 Generation of landslide susceptibility map

On the basis of prior research, there are two primary ways to create maps of landslide susceptibility. One method calculates the susceptibility value (i.e., the probability of landslide occurrence generated by the model) of random location points in the study area, and then uses inverse distance weighted (IDW) spatial interpolation to assign values to the entire study area based on a small number of location points. Using the Jenks natural break method, the susceptibility values were then divided into five classes: very high, high, medium, low, and very low. Another approach determined the susceptibility values of every pixel in the research region. The choice between the two approaches is mostly determined by the research area and computer performance. Due to the size of the research region and the lack of available computing resources, the random location points-based approach is used in this study.

4. RESULTS

4.1 Effects of different combinations of contributing factors on model performance

Comparing the different DL-based models, LSTM performed the worst and CNN performed better than LSTM. Our model achieved the best performance with a final test accuracy of 98.30%.

CNN and LSTM have a contributing factors extraction method related to their network structure, their features goes from local to abstract level (Taherkhani et al. 2018). For the DL-based model, the attention mechanism can be used to select contributing factors by reinforcing the channel weights in the network. Due to the end-to-end characteristics of deep learning models, it is more reliable to rely on the network's own weights.

Model	Training accuracy	Testing accuracy	AUC	Recall	Precision	F1 score	Kappa	MCC
CNN	0.9326	0.9288	0.9518	0.9101	0.9121	0.9111	0.8211	0.8424
CNN-LSTM	0.9668	0.9638	0.959	0.9489	0.9569	0.9529	0.8406	0.8417
CNN-LSTM Attention	0.9851	0.983	0.9836	0.9713	0.9719	0.9716	0.871	0.8913

Table 1. Results of the reference models for the three feature combination scenarios.

As the proposed model used attention mechanism to assign the channel weights for adjusting of the selection of contributing factors, its effectiveness of the proposed model is validated compared with CNN and CNN-LSTM models. The learning curves of CNN, CNN-LSTM, and our model are shown in Fig. 5 – Fig.7, respectively. It can be seen from the training curves that CNN and CNN-LSTM were able to reach early convergences during model training. It benefited from their simpler network structures, making them have fewer training parameters for an early convergence. Conversely, our model was in a slower convergence rate due to the lack of convergence of the attention mechanism in the early stage of training (shown Fig. 7). However, the advantage of the attention mechanism is that it weighed the target data and effectively focused on the valid data for prediction, and as such it led to better prediction accuracy in the end.

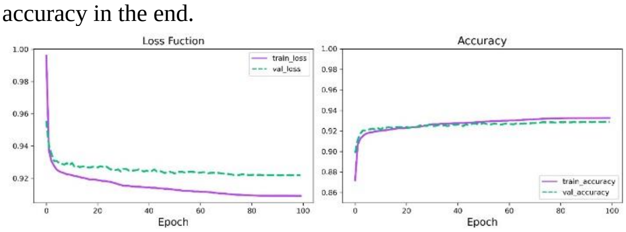


Fig. 5. Learning curve of Loss value and accuracy using CNN

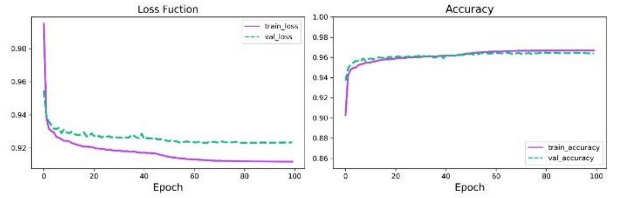


Fig. 6. Learning curve of Loss value and accuracy using CNN with LSTM

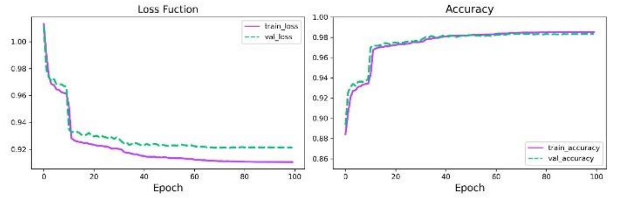


Fig. 7. Learning curve of Loss value and accuracy using CNN LSTM with attention mechanism

4.2 Contributing factors weights from attention mechanisms

The results in Table.1 shows that our proposed model can improve the prediction accuracy with attention mechanism. In more detail, our model more effectively addresses the issue of choosing which contributing factors within the model will be more useful in a single prediction among the DL-based models. It can automatically learn the importance of each contributing factor at each step of the training process and make the more critical components play larger roles by giving larger weights to those contributing factors.

To better understand how the attention mechanism worked, we used the representation α_i in Eq. 8 to describe the attention weights. We randomly selected four samples from the sorted dataset to visualise the process of the attention mechanism in Fig. 8. The values of the weights given to individual contributing factors are shown in Fig. 8, which visualise the degree of influence of different input contributing factors of the model. The most impactful from the first sample are altitude, aspect, TWI, SPI, NDVI and plan curvature. In addition, the weights assigned to the same contributing factor varied from one sample to another. The determination of weights in the attention mechanism was integrated with the model, and the weights of the attention mechanism were sought by updating u,w,b (shown in Eq. 14) during the training process.

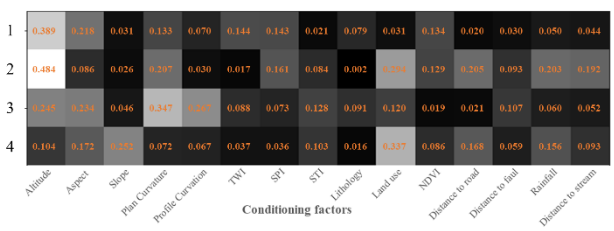


Fig. 8. Attention mechanism weights map of four samples (the horizontal coordinates are Altitude, Aspect, Slope, Plan Curvature, Profile Curvature, TWI, SPI, STI, Lithology, Land use, NDVI, Distance to road, Distance to fault, Rainfall, Distance to stream, in that order. The vertical axis is randomly selected four samples)

To rank the contributing factors, the relative importance of each contributing factor in the proposed model was expressed using the average attention weight value of the whole testing samples. Fig. 9 shows the average attention weights of the proposed model, from which altitude, distance to the road and land use were found to have larger weights (0.429, 0.278, 0.196, respectively). In addition, the attention weight of lithology was small (0.009). The rest of weights ranged between 0.08 and 0.15.

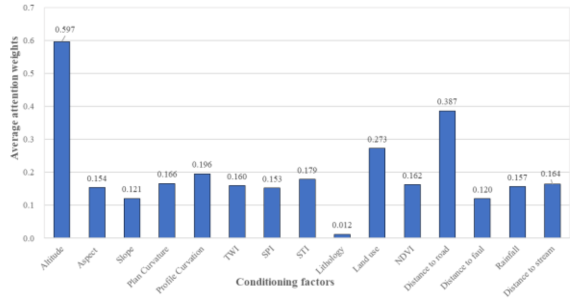
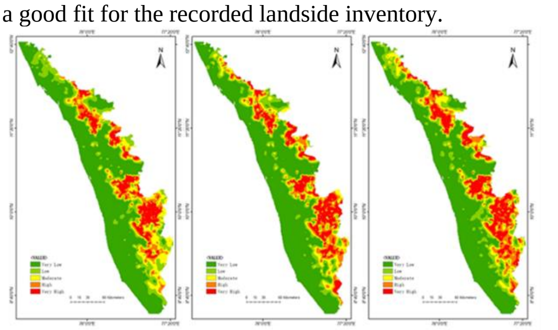


Fig. 9. The average attention weights of the contributing factors in the proposed model

4.3 Landslide susceptibility mapping

In this experiment, each contributing factor map was divided into 155,783,680 grids with a spatial resolution of 30 m. Fig.14 shows the landslide susceptibility map generated by our model and the reference models (i.e. red, orange, yellow, cyan and green are indicated as very high, high, medium, low and very low of the landslide occurrence probability respectively.). Referring to the locations of the historical landslide in Fig.1, the obtained susceptibility maps in Fig.10 showed the area of very high and high probability of the landslide occurrence generally conform to the distribution of the historical landslides. CNN and CNN-LSTM provided a clearer characterisation of the specific extent of landslides. The LSM of our model had a representation of the distribution of landslides, in which the susceptibilities of the southwest and northeast locations were described as regions of high-risk and medium risks, respectively. These are consistent with the landslide information shown in Fig. 1, demonstrating that the LSM created using our model was a good fit for the recorded landside inventory.



(a) CNN (b) CNN-LSTM (c) CNN-LSTM-Attention

Fig.10. Landslide susceptibility maps

5. DISCUSSION

In this case, locations of landslides were concentrated in mountainous areas. This was also related to the topographic variations of the area, with a clear demarcation between plains and mountains, which also related to the digital elevation models. So that the digital elevation models with high temporal resolution as well as spatial resolution will contribute to the prediction accuracy of the LSM. However, due to limitations in data acquisition, the effects of the digital elevation models could not be explored in the paper. Landslides are highly spatially and temporally correlated, but the historical landslide inventory data used in this study cannot cover a wider time span of collection and a precise temporal record of the landslides.

https://doi.org/10.5194/isprs-annals-X-3-W1-2022-25-2022

A range of contributing factors were considered in this study. However, it is unlikely that all the information associated with these contributing factors were collected at the same time. It is more likely that there will temporal differences in that information. This may lead to inconsistency in the information used for the modelling if temporal changes occurred, which would affect the prediction accuracy of landslide susceptibility.

The construction of landslide susceptibility maps relies heavily on the credibility of historical landslide inventories. As such, for modelling, the credibility of the data needs to be strictly considered. This requires adequate field or remote sensing surveys to obtain a detailed regional description of the landslide. However, the landslide inventory used in this study only recorded the point location of the latitude and longitude coordinates.

6. CONCLUSION

An CNN-LSTM-attention network is proposed in this paper for landslide susceptibility mapping. It is characterized by adaptively reassigning weights of the input contributing factor maps using the attention mechanism to achieve better prediction accuracy. Experiments were carried out using the Kerala dataset to test the performances of the proposed model and reference models (CNN and CNN-LSTM). Experimental results show that our proposed model outperforms than the reference models in classification and sensitivity prediction.

ACKNOWLEDGMENT

This research was funded by XJTLU key Program Special Fund, grant number KSF-E-40.

REFERENCES

Atkinson, P. M. and R. Massari (1998). "Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy." Computers & Geosciences 24(4): 373-385.

Bahad, P., P. Saxena and R. Kamal (2019). "Fake news detection using bi-directional LSTM-recurrent neural network." Procedia Computer Science 165: 74-82.

Fang, Z., Y. Wang, L. Peng and H. Hong (2020). "Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping." Computers & Geosciences 139: 104470.

Gaidzik, K. and M. T. Ramírez-Herrera (2021). "The importance of input data on landslide susceptibility mapping." Scientific Reports 11(1): 19334.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau and R. Moore (2017). "Google Earth Engine: Planetary-scale geospatial analysis for everyone." Remote sensing of Environment 202: 18-27.

Hao, L., A. Rajaneesh, C. Van Westen, K. Sajinkumar, T. R. Martha, P. Jaiswal and B. G. McAdoo (2020). "Constructing a complete landslide inventory dataset for the 2018 monsoon disaster in Kerala, India, for land use change analysis." Earth system science data 12(4): 2899-2918.

Kuriakose, S. L., G. Sankar and C. Muraleedharan (2009). "History of landslide susceptibility and a chorology of landslide-prone areas in the Western Ghats of Kerala, India." Environmental geology 57(7): 1553-1568.

Kwon, S. (2021). "MLT-DNet: Speech emotion recognition using 1D dilated CNN based on multi-learning trick approach." Expert Systems with Applications 167: 114177.

Lee, S. and J. A. Talib (2005). "Probabilistic landslide susceptibility and factor effect analysis." Environmental Geology 47(7): 982-990.

Leynaud, D., T. Mulder, V. Hanquiez, E. Gonthier and A. Régert (2017). "Sediment failure types, preconditions and triggering factors in the Gulf of Cadiz." Landslides 14(1): 233-248.

NRSC. (2018). "NRSC." from https://bhuvan-app1.nrsc.gov.in/disaster/disaster.php?id=landslide_monitor.

Peduto, D., M. Santoro, L. Aceto, L. Borrelli and G. Gullà (2021). "Full integration of geomorphological, geotechnical, A-DInSAR and damage data for detailed geometric-kinematic features of a slow-moving landslide in urban area." Landslides 18(3): 807-825.

Pham, V. D., Q.-H. Nguyen, H.-D. Nguyen, V.-M. Pham and Q.-T. Bui (2020). "Convolutional neural network—optimized moth flame algorithm for shallow landslide susceptible analysis." IEEE Access 8: 32727-32736.

Ramasamy, S., S. Gunasekaran, J. Saravanavel, R. M. Joshua, R. Rajaperumal, R. Kathiravan, K. Palanivel and M. Muthukumar (2021). "Geomorphology and Landslide Proneness of Kerala, India A Geospatial study." Landslides 18(4): 1245-1258.

Reichenbach, P., M. Rossi, B. D. Malamud, M. Mihir and F. Guzzetti (2018). "A review of statistically-based landslide susceptibility models." Earth-Science Reviews 180: 60-91.

Romero, A., C. Gatta and G. Camps-Valls (2015). "Unsupervised deep feature extraction for remote sensing image classification." IEEE Transactions on Geoscience and Remote Sensing 54(3): 1349-1362.

Sajinkumar, K., S. Anbazhagan, A. Pradeepkumar and V. Rani (2011). "Weathering and landslide occurrences in parts of Western Ghats, Kerala." Journal of the Geological Society of India 78(3): 249-257.

Sameen, M. I., B. Pradhan and S. Lee (2020). "Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment." CATENA 186: 104249.

Taherkhani, A., G. Cosma and T. M. McGinnity (2018). "Deep-FS: A feature selection algorithm for Deep Boltzmann Machines." Neurocomputing 322: 22-37.

Vishnu, C., K. Sajinkumar, T. Oommen, R. Coffman, K. Thrivikramji, V. Rani and S. Keerthy (2019). "Satellite-based assessment of the August 2018 flood in parts of Kerala, India." Geomatics, Natural Hazards and Risk 10(1): 758-767.

Wang, Y., Z. Fang and H. Hong (2019). "Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China." Science of the total environment 666: 975-993.

Wei, R., C. Ye, Y. Ge and Y. Li (2022). "An attention-constrained neural network with overall cognition for landslide spatial prediction." Landslides 19(5): 1087-1099.

Xiao, L., Y. Zhang and G. Peng (2018). "Landslide susceptibility assessment using integrated deep learning algorithm along the China-Nepal highway." Sensors 18(12): 4436.