



# Comparative analysis of multiple conventional neural networks for landslide susceptibility mapping

Bilal Aslam<sup>1,3</sup> · Adeel Zafar<sup>1</sup> · Umer Khalil<sup>2</sup>

Received: 2 December 2021 / Accepted: 14 August 2022 / Published online: 29 August 2022  
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

## Abstract

In landslide susceptible mountainous regions, the precondition for avoiding and alleviating perilous dangers is the susceptibility mapping of the landslide. In northern Pakistan, landslides due to vigorous seismic zones, monsoon rainfall, extremely sheer slopes, and unfavorable geological conditions present a considerable threat to the mountain areas. This study targets and advances the research in mapping landslide susceptibility in northern Pakistan (Mansehra and Muzaffarabad districts). The central objective of the analysis is to analyze different convolutional neural network (CNN) frameworks and residual network (ResNet) that were constructed by developing distinct data representation algorithms for landslide susceptibility assessment and compare the results. This study considers sixteen landslide conditioning factors related to the incident of landslides centered on the literature review and geologic attributes of the pondered area. The marked historical landslide positions in the deliberated area were arbitrarily split into training and testing datasets, with the earlier containing 70% and the former having 30% of the total datasets. Several commonly exploited measures were used to validate the CNN architectures and ResNet by comparing them with the most prevalent machine learning (ML) and deep learning (DL) techniques. The outcomes of this study revealed that the proportions of regions having very high susceptibility in all the landslide susceptibility maps of the ResNet model and CNN models are considerably alike and less than 20%, which implies that the CNN models are significantly helpful in managing and preventing landslides as to the orthodox techniques. Moreover, the suggested CNN architectures and ResNet attained greater or similar prediction accuracy than other orthodox ML and DL techniques. The values of OA (overall accuracy) and MCC (Matthew's correlation coefficient) of proposed CNNs and ResNet were greater than those of the optimized SVM (support vector machine) and DNN (deep neural network).

**Keywords** Northern Pakistan · Deep learning · Machine learning · Convolutional neural network · Landslide susceptibility

---

✉ Adeel Zafar  
adeel.zafar@riphah.edu.pk

Extended author information available on the last page of the article

## 1 Introduction

Landslides are one of those devastating natural disasters that cause property and human life loss in many countries worldwide (Maqsoom et al. 2021). It is critical to assess landslide-prone areas to manage and mitigate landslide disasters (Hong et al. 2016a). Assessing landslide susceptibility is a vital way to identify the most vulnerable regions to landslides and the critical areas for preventing and managing landslides (Guzzetti et al. 2006, Flentje et al. 2007); thus, it is a valuable method and way to avoid landslides (Guzzetti et al. 2012).

Different statistical-based methods were frequently employed for mapping landslide susceptibility in areas around the world. The existing studies on landslide susceptibility analysis primarily utilized the following conventional regression analysis methods: logistical regression (LR) (Martinović et al. 2016, Tanyas et al. 2019, Nefeslioglu and Gorum 2020), fisher discriminant analysis (FDA) (He et al. 2012, Gupta et al. 2018, Pham and Prakash 2019a, Wang et al. 2020), index of entropy models (Youssef et al. 2015, Wang et al. 2016b, Mondal and Mandal 2019), frequency ratio (Lee et al. 2015, Li et al. 2017, Mandal et al. 2018b), weights of evidence (WOE) (Wang et al. 2016a, Tsangaratos et al. 2017, Bacha et al. 2018, Pamela et al. 2018).

In 25 earthquake-induced landslide (EQIL) areas, Tanyas et al. (2019) studied 64 EQIL inventories and implemented LR along with a slope unit for assessing landslide susceptibility. In one earthquake event area (Wenchuan event), the model obtained the area under curve (AUC) accuracy of 0.88. Nefeslioglu and Gorum (2020) used LR to reveal the landslide hazards through landslide susceptibility assessment of Melen Dam reservoir and achieved better results of landslide spatial probability. Gupta et al. (2018) employed an information value technique for computing the weights of different influencing factors and afterward utilized the weights in FDA and binary LR to assess landslide susceptibility. The prediction precision of the practiced methods was evaluated through Heidke skill score, and the obtained scores of FDA and LR were 0.89 and 0.90, correspondingly.

Furthermore, for evaluating landslide susceptibility, Barella et al. (2019) contrasted seven statistical techniques: likelihood ratio, Bayesian model, landslide density, discriminant analysis, information value, WOE, and LR, and in the accuracy assessment, the WOE technique outperformed the other methods. A landslide inventory is typically utilized as the predictive variable by the orthodox regression analysis techniques, and statistical regression models are established to project the possibility of landslide incidence. Nevertheless, the factor selection and weigh assignment processes are partial to some extent, so these techniques partly depend on expert experience (Kanungo et al. 2011, Gupta et al. 2018).

Moreover, conventional machine learning (ML) and ensemble techniques have also been used to evaluate landslide susceptibility. These techniques include support vector machine (SVM) (Tien Bui et al. 2019, Yu et al. 2019, Pham et al. 2019b), artificial neural network (ANN) (Chen et al. 2017a, Polykretis and Chalkias 2018, Sevgen et al. 2019, Aslam et al. 2022), genetic algorithm-SVM (Ramachandra et al. 2013, Niu et al. 2014) and random forest (RF) (Kim et al. 2018, Sevgen et al. 2019, Aslam et al. 2022). Dou et al. (2020) used four techniques, namely SVM, SVM-Bagging, SVM-Stacking, and SVM-Boosting, for evaluating landslide susceptibility. The SVM-Stacking model showed the poorest functioning while the SVM-boosting model topped all the models. Aslam et al. 2022 assessed and compared the proficiencies of six advanced ML techniques for mapping landslide susceptibility, including linear discriminant analysis, quadratic discriminant analysis, ANN, naive Bayes, multivariate adaptive regression spline, and RF, and discovered that ANN outperformed other techniques.

Zhao and Chen (2020) implemented four ensemble approaches for modeling landslide susceptibility: bagging-functional trees, functional trees, dagging-functional trees, and rotation forest-functional trees. They observed that the bagging-functional trees with an AUC accuracy of 0.804 outperformed the other techniques. Tien Bui et al. (2019) utilized an SVM classifier in combination with four ensemble techniques, namely rotation forest (RF), bagging (BA), random subspace (RS), and AdaBoost (AB), to formulate novel ensemble models for evaluating landslide susceptibility and discovered that the RS-SVM model is superior in the prediction accuracy as compared to the models of SVM, RF-SVM, AB-SVM, and BA-SVM. The orthodox ML techniques can assist the comprehensive analysis of numerous conditioning factors and comparatively illustrate the nonlinear correlation among the conditioning factors and landslide susceptibility satisfactorily; hence, they can obtain high accuracies (Bui et al. 2016).

More recently, various deep learning (DL) techniques have been practiced for evaluating landslide susceptibility. However, this research area is still going through many developments and advancements as presently, more research is focused on DL techniques to determine landslide susceptibility. Wang et al. (2020) assessed landslide susceptibility using recurrent neural networks (RNNs) in Yongxin County, China, and indicated that the RNN system is helpful for mapping landslide susceptibility. Wang et al. 2019 utilized convolutional neural networks (CNNs) for mapping landslide susceptibility in Yanshan County, China, showing that CNNs are better than conventional ML methods.

Compared to traditional ML techniques, DL techniques have more sophisticated structures (Ronoud and Asadi 2019). They thus are more capable of explaining a complex nonlinear problem, such as a landslide system (Cao et al. 2019). Moreover, due to the competent study approach, a DL technique can achieve superior generalization capabilities to an orthodox ML technique (Duo et al. 2019, Liu et al. 2019). Therefore, several researchers have integrated the DL techniques with ML techniques to enhance the performance of ML techniques. For instance, Aslam et al. 2021 recently combined a CNN with three traditional ML classifiers, SVM, LR, and RF, to assess landslide susceptibility of Mansehra and Muzaffarabad districts in northern Pakistan. The authors established that the integration of CNN with RF, SVM, and LR could efficiently enhance the prediction performances of the ML classifiers. Nevertheless, the accuracy of a technique for evaluating landslide susceptibility is highly dependent on a specific area, and the accuracy of a technique may differ when used for various areas, which still needs to be illustrated (Lagomarsino et al. 2017).

Moreover, several studies (Wang et al. 2019, Wang et al. 2020a, b, Yi et al. 2020, Youssef and Pourghasemi 2021, Aslam et al. 2021, Aslam et al. 2022) have established that performing feature selection plays a vital role in landslide susceptibility assessment as it helps to remove redundant features and only to retain the valuable features. Yi et al. (2020) conducted a study for mapping landslide susceptibility utilizing multiscale sampling strategy and CNN in Jiuzhaigou region, China, and used information gain ratio and multicollinearity analysis for shortlisting useful features for the modeling process. Youssef and Pourghasemi (2021) used RF techniques to assess the importance of individual conditioning factors while mapping landslide susceptibility for Abha Basin, Asir Region, Saudi Arabia using ML algorithms. Aslam et al. (2022) have also used RF for feature selection when mapping the landslide susceptibility in Pakistan. Wang et al. 2019 have used gain ratio (GR) and multicollinearity for selecting useful features when performing landslide susceptibility mapping in China. Moreover, CNN has also been used for choosing useful features by Wang et al. (2020a, b) and Aslam et al. (2021) during the landslide susceptibility mapping in China and Pakistan, respectively.

The geographical attributes, monsoon rains, and earthquakes (Dai et al. 2002, Crosta 2004) triggered by the active faults vigorously endanger the mountainous regions of Pakistan to landslide problems. This fact was truly witnessed during the October 8, 2005, Kashmir earthquake that jolted the region of Kashmir and northern areas of Pakistan. The region affected by this earthquake was more than 30,000 km<sup>2</sup>, and it caused thousands of landslides, resulting in 25,500 fatalities (Kamp et al. 2008, Owen et al. 2008). Various researchers have carried out different studies evaluating the induced landslides due to this earthquake (Owen et al. 2008, Saba et al. 2010, Shafique, van der Meijde et al. 2016) and to map the landslide susceptibility of the October 8, 2005, earthquake-affected or adjacent areas (Kamp et al. 2008, Basharat et al. 2016, Bibi et al. 2016, Torizin et al. 2017, Khan et al. 2019). However, these studies, which targeted October 8, 2005, earthquake-affected or adjacent areas, have some limitations as they have primarily used conventional decision-making and quantitative techniques, which are not as accurate as orthodox ML or novel DL techniques. Recently, Aslam et al. (2021) and Aslam et al. (2022) made an effort to map landslide susceptibility in the Mansehra and Muzaffarabad districts in northern Pakistan and fill the existing research gap. In contrast to the previous studies, Aslam et al. (2021), and Aslam et al. (2022) used conventional ML and novel DL techniques and achieved promising results. Moreover, Aslam et al. (2021), and Aslam et al. (2022) used a broad range of landslide conditioning factors and well-known feature selection techniques to select useful features for assessing landslide susceptibility. However, several researchers are presently exploring multiple DL techniques to assess landslide susceptibility; as this field is progressing day by day, the authors efforted to extend the research in this direction further.

The present study aims to further explore the potential of different DL techniques by performing landslide susceptibility mapping for the Mansehra and Muzaffarabad districts in northern Pakistan using multiple CNN frameworks and residual network (ResNet). The current work has subsequent developments to the existing work on landslide susceptibility mapping in the area. First, it uses different advanced DL techniques to obtain high precision in landslide susceptibility assessment. It compares these techniques with orthodox DL and ML techniques to establish the most superior model for the desired objective. The used novel DL techniques were ResNet and different architectures of CNN. In contrast, for comparison purposes, the used orthodox DL technique was DNN, and conventional ML techniques were SVM and LR. Second, it uses two feature selection techniques to select the more valuable features and eliminate redundant features. All these developments are significant contributions of the present study to the current work in the proposed area.

## 2 Study area

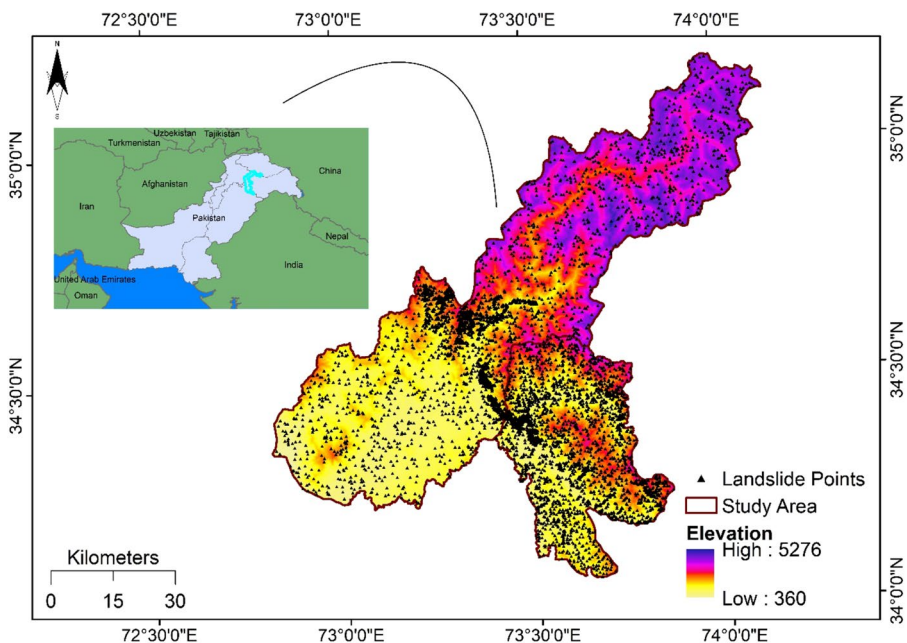
The study area is enclosed by the administrative borders of the Mansehra and Muzaffarabad districts and encompasses an area of around 6220 km<sup>2</sup>. The Mansehra district is situated in Khyber Pakhtunkhwa province, whereas the Muzaffarabad district is located in Pakistan's administrated territory of Azad Jammu and Kashmir. The epicenter of the October 8, 2005, earthquake was detected in the vicinity of Muzaffarabad city, the regional capital of Azad Jammu and Kashmir. On the southern side, the area is characterized by hilly relief. On the southwestern side, primarily by plane to mountainous areas and by an elevated mountain range area on the eastern side. Most prominently, the mountainous areas around the Kaghan and Naran valleys in the Mansehra district are incredibly elevated. The

area’s elevation ranges from around 360 m in the plains to approximately 5276 m in the hills. The map of the study area showing the elevation of the area and the past landslide points is presented in Fig. 1.

The Indus River, with its two major tributaries, the Kunhar River and the Siran River, is the primary drainage of the Mansehra district. The Indus River crosses the Mansehra district from north to south along the western edge. In contrast, the Jhelum River is the primary drainage of the Muzaffarabad district and has two tributaries, specifically the Kunhar River and the Neelum River. Creating deep antecedent valleys, these rivers flow westward before gushing southward beside wider valleys to the Indo-Gangetic Plain. During the monsoon spell in the summer, the temporary variations in the river discharge regime often trigger sporadic instant floods in the area.

The area’s weather conditions are pretty variable, owing to its topography. The area has a subtropical highland climate. During the summer, the weather is warm, while it is cold in the winter. Nevertheless, the northeastern side of the Mansehra district, primarily the Kaghan valley, has cold weather in summer and extremely cold in winter due to heavy snowfall in the mountains. In this area, the warmest month of June is the highest in the Muzaffarabad district, with average maximum and minimum conditions of 37.6 and 22.1 and 35 and 21 °C, respectively, in the Mansehra district. For the county of Muzaffarabad and Mansehra, January is the coolest month, with a high and a minimum mean temperature of 15.9 and 3.2 °C, respectively.

The rainfall in the area is significant; what is more, it receives rainfall even through the driest month. The average yearly rainfall in the Muzaffarabad district is about 1500 mm, whereas the Mansehra district receives between 1400 and 1800 mm rainfall on average per year. Out of the average yearly rainfall, almost one-third happens through the monsoon season, which lasts from the end of June until late August. This extensive rainfall often



**Fig. 1** Map of the study area with past landslide points and elevation of the area

causes severe flooding and landslides, particularly debris flows. Through the winter, the elevated mountainous areas receive rainfall as snow. The area receives little rain during the spring, but the slopes get plentiful surface water through snowmelt. This infiltrating water increases the groundwater table height, thus resulting in erosion.

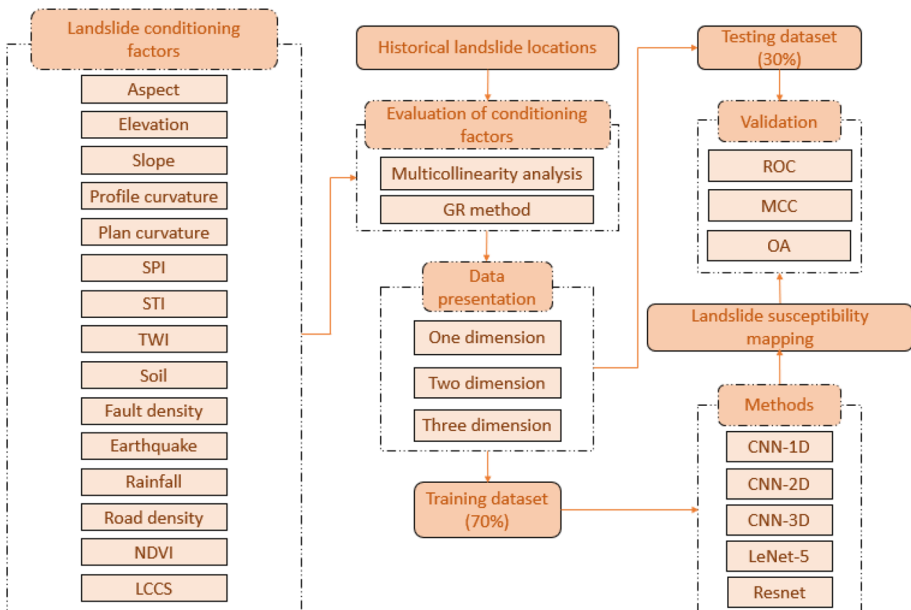
Geologically this region contains the Hazara-Kashmir Syntax, which borders on the Main Thrust and is an area of significant reduction and elevation of crustaceans (Kazmi and Jan 1997; Hussain and Yeats 2009). Overall, the terrain steepness, along with extreme rainfall through monsoon spells, constant swift river slitting, intermittent earthquake jolting, and anthropogenic impacts as destabilizing of slopes during the construction process, make this area extremely prone to slope failures.

### 3 Materials and methods

The detailed methodology is presented with the help of a flow chart in Fig. 2 and is explained in the following sections.

### 4 Landslide inventory mapping

A landslide inventory map comprises past and previously occurred landslide locations (Pham et al. 2016). It contains information regarding landslides’ type, locality, movement, and physical attributes (Rosi et al. 2018). Moreover, it is an important data source for spatial prediction of landslides (Pham et al. 2016) and for understanding the association



**Fig. 2** Flowchart of the methodology for landslide susceptibility mapping through the proposed CNN framework

between landslide incidence and the association with influencing factors (Jiao et al. 2019). High-resolution remotely sensed images could be visually inspected for interpreting landslide inventory mapping (Tsangaratos et al. 2017, Pham et al. 2017b, Pham et al. 2017c). Therefore, for this analysis, LANSAT-8 images were taken to prepare the region's landslide inventory map with 3251 previous locations. All of these landslides represent the geo-environmental situation of landslide zones. 2276 locations (70%), including the outstanding 975 (30%) locations, were arbitrarily selected for training and testing purposes. The splitting was based on previous studies from the literature (Wang et al. 2019, Wang et al. 2020a, b, Yi et al. 2020, Youssef and Pourghasemi 2021, Aslam et al. 2021, Aslam et al. 2022).

Moreover, 3251 non-landslide locations were randomly selected to maintain the class balance and were split into the same proportions (70 and 30%) for particular purposes. The polygon outlines were drawn for clearly visible landslides on the satellite imagery. These polygons were then used for the measurement of the space and the validation of the final susceptibility maps. The polygons were transformed and displayed as points on the study area map shown in Fig. 1 because they had unusually low visibility due to the limited size of the inventory map.

## 5 Landslide conditioning factors

In order to map the susceptibility to landslides, it is typically presumed that future landslides would take into account the same environmental factors which have caused previous landslides. These landslide environmental factors are essential to mapping the susceptibility to landslides. This study uses thematic maps of several landslide factors that influence the susceptibility of landslides in the considered region. In addition to the environmental conditions and available data for the area under discussion, the conditioning factors in the current analysis have been based on the literature review and existing landslide susceptibility studies in the area under consideration (Kamp et al. 2008, Owen et al. 2008, Basharat et al. 2016, Bibi et al. 2016, Shafique et al. 2016, Torizin et al. 2017, Tsangaratos et al. 2017, Pham et al. 2017b, Khan et al. 2019, Fang et al. 2020a, b, Wang et al. 2020). The selected landslide conditioning factors include road density, stream power index (SPI), topographic wetness index (TWI), slope, fault density, elevation, lithology, sediment transport index (STI), aspect, profile curvature, earthquake, normalized difference vegetation index (NDVI), land cover classification system (LCCS), soil data, and plan curvature.

ASTER GDEM (DEM) data with a spatial resolution of 30 m were used to obtain geomorphological factors of elevation, slope, SPI, TWI, aspect, profile curvature, STI, and plan curvature. Data from LANDSAT-8 satellite images with a spatial resolution of 30 m were used to obtain the maps of NDVI and LCCS. The factor of NDVI was obtained using the Infrared (IR) and Red (R) bands in the following formula:  $NDVI = (IR - R) / (IR + R)$ , (Hong, Pourghasemi et al. 2016b). The annual average rainfall data for the last decade was acquired from the Pakistan Meteorological Department (PMD) for preparing the rainfall map. The seismicity map of Pakistan was used for the earthquake map.

The geological and topographical factors were obtained from the geological and topographical maps of Pakistan correspondingly. The fault density and lithology factors were obtained using the scale of 1:500,000 from the geological maps. The road density factor was defined using the scale of 1:200,000 from the topographical maps. The soil data were acquired from

the Soil Survey of Pakistan. All the prepared maps of the considered factors were adapted to the same scale as the obtained DEM (30 m of spatial resolution) for further assessment.

Moreover, we have tried our utmost to eliminate all the possible error-causing factors by taking protective measures such as for the datasets which are geo-temporal and have a seasonal effect, such as rainfall (rainfall data for a decade) and NDVI (NDVI data for one year); we have taken the mean to reduce the ambiguity. Additionally, for those datasets which are standard, we have used the best available datasets to minimize the error. All these efforts were in the wake of obtaining efficient prediction results and producing reliable susceptibility maps.

## 6 Selection of conditioning factors

The selection of conditioning factors as various conditioning factors affect landslides is one of the main steps in landslide susceptibility mapping. Therefore, it is important to choose suitable features in order to yield a comprehensible landslide susceptibility map. The modeling process might become intricate owing to the extreme dimensionality linked with the training and testing datasets. In the meantime, poor modeling precisions always occur due to the curse of dimensionality. The process of selecting features is useful as it uses recognized selection techniques and provides high-quality information. In addition, the unnecessary and surplus characteristics of the selected features can be eliminated during this stage. In the current study, two feature selection techniques, namely GR and multicollinearity analysis, are employed. The details about these techniques are provided next.

### 6.1 GR

For selecting an ideal subset to enhance the prediction performance for mapping the landslide susceptibility, the feature selection technique of GR (Dash and Liu 1997) was utilized. A feature at every node of the decision tree is selected by using the information gain measure. Subsequently, to avoid bias, GR, which is an augmentation of information gain, is anticipated. For better understanding, the GR method is presented followingly. Assuming that  $S$  is a training data plus  $n$  is the overall dataset, then the projected statistics are offered by the following comparison.

$$H(S) = - \sum_{i=1}^n p_i \log_2(p_i)$$

where  $p_i$  signifies the possibility that class  $C_i$  holds a sample, the feature  $F$  has  $m$  values, and the mean entropy related to it is provided by the following equation.

$$E(F) = - \sum_{i=1}^n p_i H(S)$$

then the information gain on feature  $F$  is

$$\text{Gain}(F) = H(S) - E(F)$$

The prospective information attained by distributing  $S$  into  $m$  parts relating to  $m$  outcomes on feature  $F$  is represented by the split information, and it can be obtained by using the following formula.



$$\text{Split Info}_A(S) = - \sum_{i=1}^m X_i \left( \frac{S_i}{S} \right) \log_2 \left( \frac{S_i}{S} \right)$$

Lastly, the GR is characterized by the following equation:

$$\text{Gain Ratio} = \frac{\text{Gain}(F)}{\text{Split Info}_A(S)}$$

The important association of conditioning factors with landslide incidence is revealed by the average merit (AM) derived from this technique. A factor is regarded as an irrelevant feature if the AM value is equal to or less than 0, and it should be excluded from the further modeling process prediction. On the contrary, the remaining factors are suitable to be used for further assessment.

## 7 Multicollinearity analysis

In order to measure the relationship between the landslide conditioning factors, multicollinearity analysis was employed. Multicollinearity is a statistical phenomenon that exhibits an excessive correlation among two or more predictor variables in a multiple regression model (O'Brien 2007). This research uses the multicollinearity of conditioning factors to classify resistance (TL) and inflation variance factor (VIF). Assume that the  $x = (x_1, x_2, \dots, x_n)$  symbolizes a particular separate variable set and that the  $R_j^2$  is the code for the independent  $j$ th variable  $x_j$  that is reverted to all of the remaining variables of the model. The VIF value is computed by utilizing the following relation:

$$\text{VIF} = \frac{1}{1 - R_j^2}$$

The extent of linear relation among independent variables is represented by the TL value, which is the multiplicative inverse of the VIF value. If a factor exhibits a VIF value of more than 10 or a TL value of less than 0.1, it exhibits multicollinearity and ought to be eliminated from the further modeling process.

## 8 Methods

Prior to employing DL methods for landslide susceptibility mapping, all the conditioning factors were initially assembled collectively, so the whole considered region can be regarded as a “multi-channel image.” Every single factor signifies a particular channel, and the final objective is to categorize every pixel in this “image.” The primary stage is to build and train an original DL model. The input layer of an original DL model comprises numerous neurons. Each neuron indicates a landslide conditioning factor. A DL can automatically extract valuable features from the original data, and then these features are used to get the final modeling output. It is a two-stage process of feature extraction and classification. The structure of the model is first built and is trained by exploiting the initial training dataset. The trained model is then used for modeling purposes.

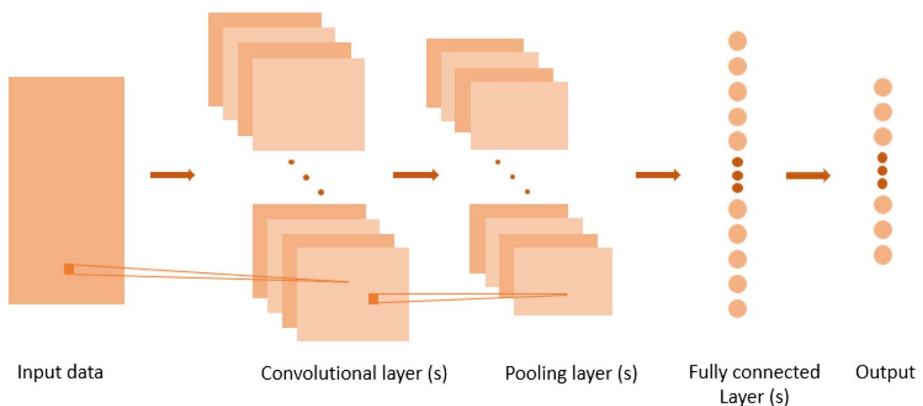
## 9 ResNet

Convolutional networks utilized by the computer vision fraternity have been increasing deeper every year since Krizhevsky et al. (2012) suggested AlexNet in 2012. Here, depth describes the sum of layers in a network, while width describes the sum of kernels of each layer. Recently proposed ResNets get state-of-the-art performance and permit training of exceedingly deep networks up to over 1000 layers (He et al. 2016). ResNet is the deepest network in the literature, with 1202 trainable layers (He, Zhang et al. 2016). Tiny images in the CIFAR-10 dataset were used to train this 1202-layer ResNet (Krizhevsky and Hinton 2009). The image size here is vital, as it indicates that the size of subsequent feature maps is comparatively small, which is crucial in practice to train exceptionally deep models. Like highway networks, ResNets make use of identity shortcut connections that facilitate the movement of information through layers devoid of attenuation that would be instigated by various stacked nonlinear transformations, producing enhanced optimization (Srivastava et al. 2015). Shortcut connections in ResNets are not gated, and untransformed input is constantly spread. He et al. 2016 have established an awe-inspiring empirical performance of ResNets.

### 9.1 CNN

CNNs are proficient in visual recognition and have many coalescents, maximum pooling, and fully linked layers. A CNN model is a class of feedforward NN (artificial neurons of this NN respond to a fraction of the adjacent elements), demonstrating vigorous execution in visual image assessment (Girshick 2015). This indicates that a CNN is a variant of a multilayer perceptron containing one or more convolution, max pooling, and fully connected layers (Shin et al. 2016). The structure of a conventional CNN is presented in Fig. 3. A conventional CNN always contains input, convolutional, max pooling, fully connected, and output layers. The input layer is basically an  $m \times n$  matrix, and each element in this matrix has a feature value.

Consequently, the input parameter can be characterized as a 2D plot. There are several convolutionary units in every convolutionary layer. In order to optimize the parameters of every unit, a backpropagation algorithm is used. Overall usage aims to attain



**Fig. 3** The generalized architecture of CNN

various input layer characteristics (Sharif Razavian et al. 2014). The initial convolution layer could give only some low-level characteristics such as corners, edges, and lines. Other convolutionary layers can learn more complex representations from these low-level characteristics. Pooling is a grave operation in the CNN strategy (Szegegy et al. 2015). The dimensionality of feature maps is reduced without modifying their depth. This reduces sampling. Max pooling is the most common manipulation in different approaches to pooling. It seeks to separate the feature maps and generate the maximum value for each region into several quadrangular areas. In addition, the dimensionality of the data can be continuously reduced, thereby reducing calculation cost and the number of parameters. This avoids the problem of overfitting. The fully connected layer arranges the achieved representations, and the output layer produces classification results to reduce the loss of feature information. The position of detected characteristics to other characteristics should be noted to be more significant than the actual position, which is unique to CNN.

## 10 The proposed CNN architectures

This work seeks to create a CNN system to map landslide susceptibility. The present landslide data may not be well adapted to the CNN architectures because of the various representations that may influence the susceptibility outcomes. Consequently, different architectures of CNN were designed to match multiple data representations. The following subparagraphs describe the four different ways of data representation for constructing CNN architectures. The CNN and the data representation algorithms 1D, 2D, and 3D are referred to as CNN-1D, CNN-2D, and CNN-3D.

## 11 LeNet-5

LeNet-5 is a CNN structure that can be used to recognize a handwritten digit and can solve several visual problems effectively (LeCun et al. 1995). However, LeNet-5 cannot be used directly to measure landslide susceptibility. In this study, we, therefore, proposed multiple CNN architectures that would evaluate landslide susceptibility and afterward compare it to the landslide susceptibility from the orthodox LeNet-5 CNN structure.

A commonly employed LeNet-5 consists of eight layers (LeCun et al. 1998). Given an input data as  $n \times n$ , followed by a convolutional layer C1 with six  $(n - 4) \times (n - 4)$  feature maps and a  $5 \times 5$  neighborhood in the input map is linked to every element in the feature maps. Having six  $\frac{(n-4)^2}{2^2}$  feature maps, a max pooling layer S2 is for subsampling. The kernel size of max pooling is  $2 \times 2$ , and every element in the feature maps is linked to the preceding layer. With sixteen  $\frac{(n-12)^2}{2^2}$  feature maps, C3 is another convolutional layer, and each element of the layer is linked to a  $5 \times 5$  neighborhood in the former layer. With sixteen  $\frac{(n-12)^2}{4^2}$  feature maps, S4 is another layer that performs the max pooling process. F5 and F6 are fully connected layers with correspondingly 120 and 84 neural units. Lastly, two neural units are produced in the output layer to show the binary classification outcomes as landslide and non-landslide. The architecture of LeNet-5 for  $n = 24$  is shown in Fig. 4.

### 11.1 CNN-1D

The input data can be considered an example of mapping landslide susceptibility such that each pixel possesses many landslide conditions. As a consequence, each input data grid cell with a length defined by the number of conditioning factors is described by a column vector. Each element of this vector also refers to a landslide conditioning function. A 1D CNN structure has been created to analyze landslide susceptibility based directly on landslide conditioning factors' knowledge. One of any convolution, max pooling, and completely linked layers includes the 1D CNN architecture. When  $n$  landslide conditioning factors are present in the data input, the convoluted layer of  $m$  to 1 size  $N$  kernels filters the data input, then  $N$  feature vectors of length  $(n - m + 1)$ . This layer is often used. Each element in the vector function is connected in the input vector to an  $m$  of the 1 neighborhood. The size of the max pooling layer is  $a \times 1$ , and the consequent outcomes are composed of  $N$  vectors having a length of  $\left\lceil \frac{n-m+1}{a} \right\rceil$ . To represent the extracted features, the fully connected layer having  $k$  neural units comprehends the preceding layer. Lastly, two neural units are produced in the output layer to solve a binary classification problem. For  $n = 15$ ,  $N = 20$ ,  $m=3$ ,  $a = 2$  and  $k=50$ , the architecture of CNN-1D is illustrated in Fig. 5.

### 11.2 CNN-2D

The CNN technique was efficient in the processing of images. Still, the start should be by transforming a 1D input grid cell (vector) composed of several features into a 2D matrix in order to use this technique for mapping landslide susceptibility. The number of landslide conditioning factors was compared in the present study with characteristic values of each factor, and the highest of both was the 2D matrix dimension. For example, neither factor exceeds the overall number of factors by its characteristic values (16). Consequently, for each grid cell, we generated a matrix of 16 to 16. Figures 6 and 7 show the conversion handling of a 1D grid (vector) into a 2D matrix. Explicitly the item value in this matrix is assigned a value of 1 for each column vector. The rest of the vector unit values are given 0 for the associated characteristic value.

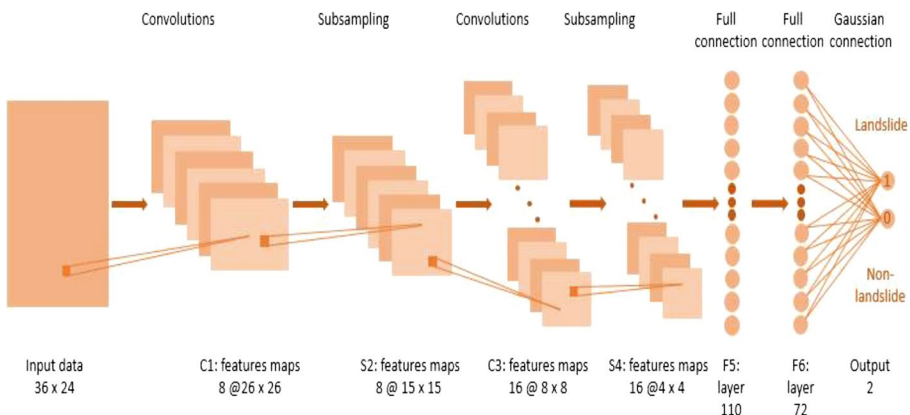


Fig. 4 Architecture of LeNet-5

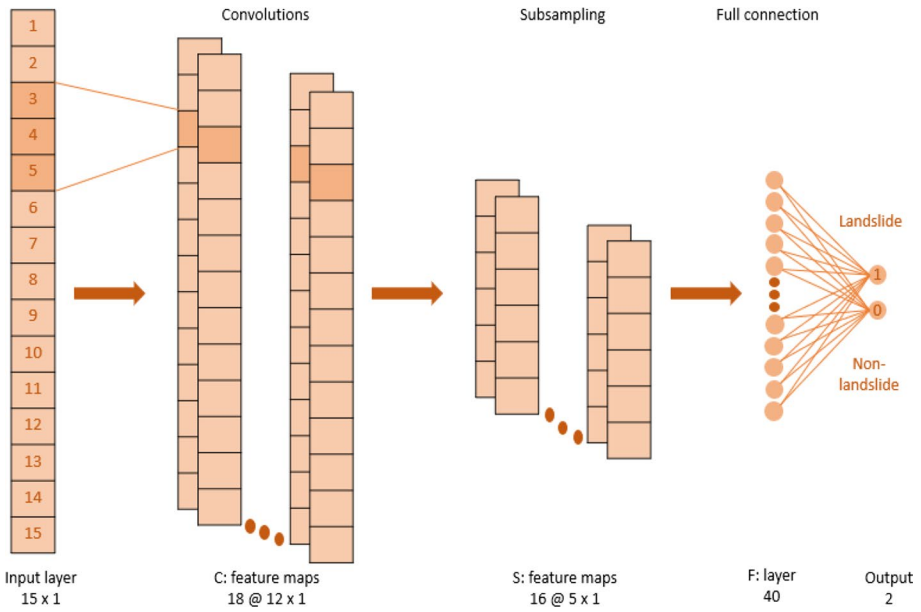


Fig. 5 Architecture of CNN-1D

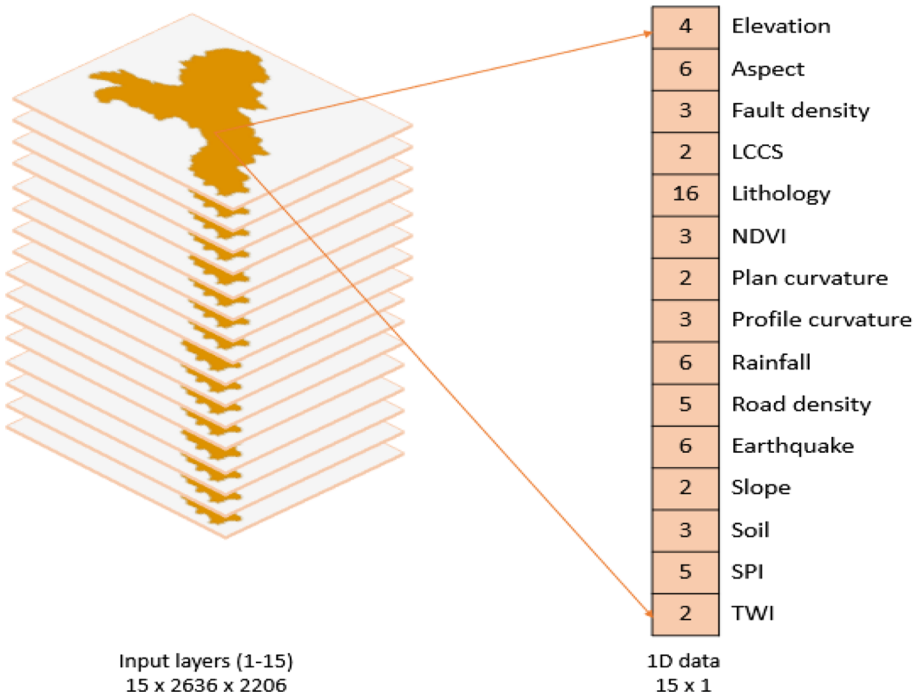


Fig. 6 Input data for conversion into a 2D data form

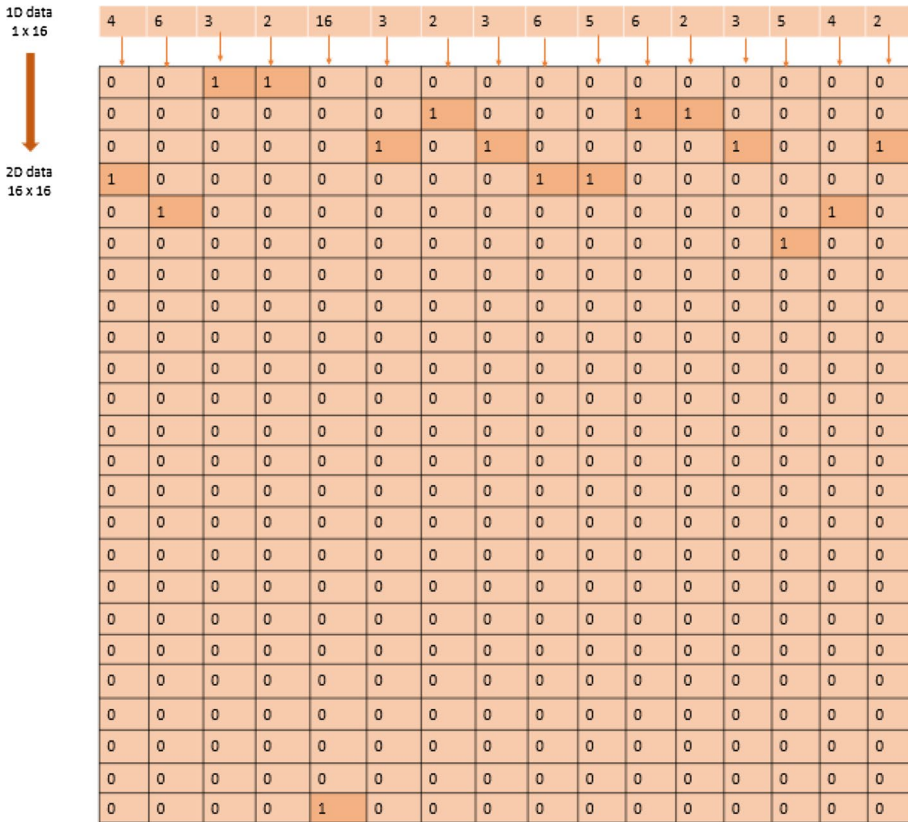


Fig. 7 The conversion of a 1D grid cell to a 2D matrix

The 2D CNN structure proposed consists of one convolutionary and one max pooling layer. A dropout layer follows every layer of convolution. A transformation from  $n$  to  $n$  matrix is supposed to take place from every cell unit in the grid landslide. In the initial convolutionary layer, the input data are tested using  $N$  kernels  $m$  to  $m$ , and thus this layer includes  $20(n - m + 1)$  corresponding maps.  $(n - m + 1)$  function maps. Each grid cell in the function maps is associated with the  $m$  to the area around the input map. The dropout manipulation briefly leaves the NN groups, depending on a particular probability, through the training of CNN. This process resolves the over-fitting problem and enhances classification precisions. A drop manipulation is employed following every convolutional phase. The max pooling layer has a size of  $a \times a$ . Consequently, the outcomes of this layer contain  $N$  matrices with a size of  $\left[ \frac{n-m+1}{a} \right] \times \left[ \frac{n-m+1}{a} \right]$ , followed by a dropout manipulation, were subsequently assigned to the second convolutional layer having  $M$  kernels with a size of  $m \times m$ . The outcomes of this convolutional layer contain  $M$  matrices having a size of  $\left[ \frac{n-(a+1)(m-1)}{a} \right] \times \left[ \frac{n-(a+1)(m-1)}{a} \right]$ . These outcomes were later uninterruptedly assigned to the second max pooling layer with  $M$  kernels having a size of  $a \times a$ . This max pooling layer produces resultant  $M$  feature maps having a size of  $\left[ \frac{n-(a+1)(m-1)}{a^2} \right] \times \left[ \frac{n-(a+1)(m-1)}{a^2} \right]$ . The extracted features were reorganized by a fully connected layer having  $k$  neural units

following the previous layer. Ultimately, the binary classification outcomes in the form of “landslide” and “non-landslide” are indicated by the produced two neural units by the output layer. The architecture of CNN-2D when  $n=24$ ,  $m=3$ ,  $N=20$ ,  $M=15$ ,  $a=2$  and  $k=78$  is presented in Fig. 8.

### 11.3 CNN-3D

The input data for the pondered region can be represented by a size 3D matrix ( $c=n+n$ ), where  $n$  means the row and column of every data layer and  $c$  refers to the number of landslide conditioning factors. Figure 9, for example, shows the 3D data representation of each grid cell neighborhood with a scale of  $7 \times 7$ .

We designed a 3D CNN architecture under these conditions to obtain details and the spatial relation of the conditions of landslide incidence. The 3D CNN network contained one convolutional layer of  $N$  kernels with a size of  $m \times m \times m$ , one max-pooling layer, and one fully connected layer. With a size of  $(c - m + 1) \times (n - m + 1) \times (n - m + 1)$ , the convolutional layer has  $N$  feature maps, given the  $c \times n \times n$  input data. Every grid cell is linked to an  $m \times m \times m$  neighborhood in the input data. The subsequent hidden layer with a size of  $a \times a$  is a max-pooling layer. Consequently, the outcomes from the max-pooling layer have  $N$  feature maps with a size of  $\left[\frac{c-m+1}{a}\right] \times \left[\frac{n-m+1}{a}\right] \times \left[\frac{n-m+1}{a}\right]$ . After the max-pooling layer, there is a fully connected layer with  $k$  neural units, and it pursues the former layer to learn the extracted features. Ultimately, two neural units are positioned in the output layer to signify “landslide” and “non-landslide” prediction. The architecture of CNN-3D for  $c=15$ ,  $n=7$ ,  $N=20$ ,  $m=3$ ,  $a=2$ , and  $k=78$  is presented in Fig. 10.

## 12 The related parameters

The parameter settings greatly influence the performance of prediction/classification techniques. In the present study, we also incorporated related parameters in the proposed CNN architectures to improve the performance. This effort was made to enhance the prediction performance of the models and achieve superior susceptibility maps. The effects of loss and activation functions are highly important for CNN architecture. The effects of any

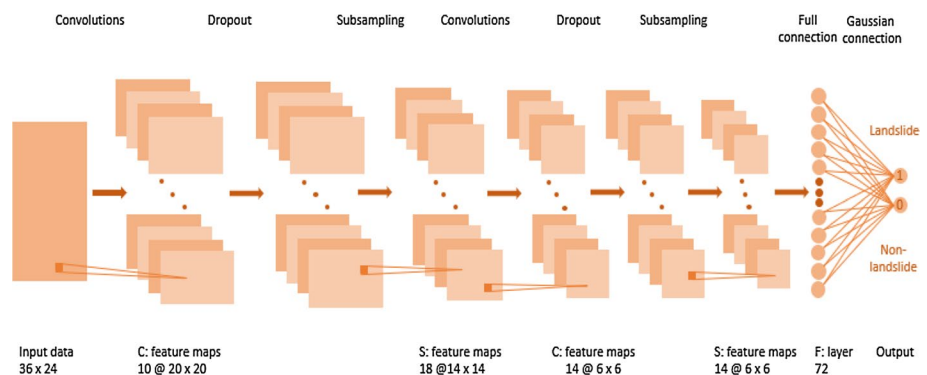


Fig. 8 Architecture of CNN-2D

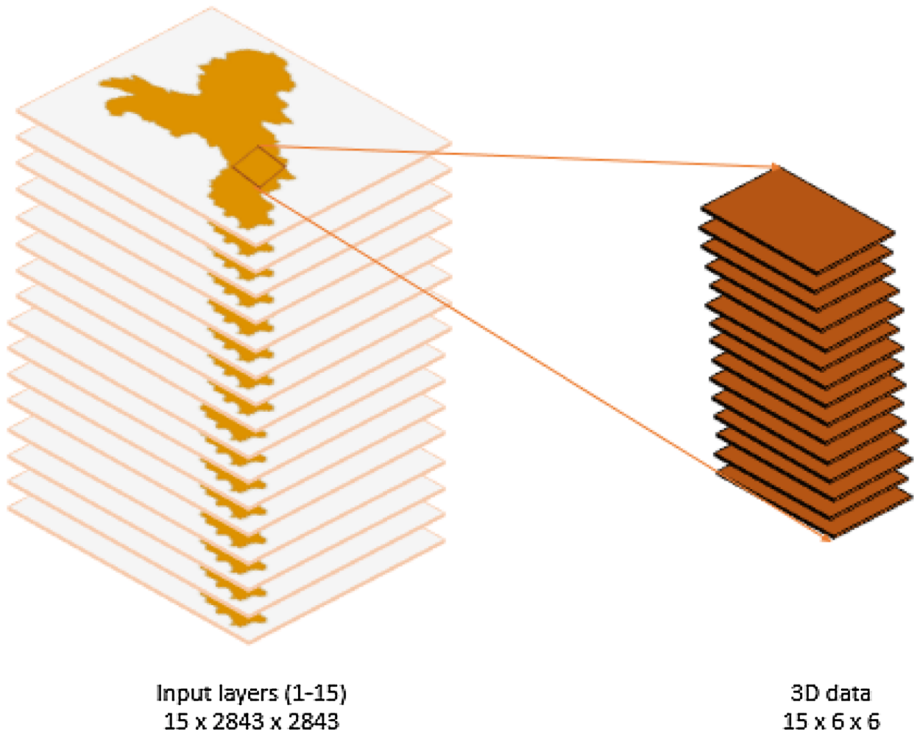


Fig. 9 3D data form

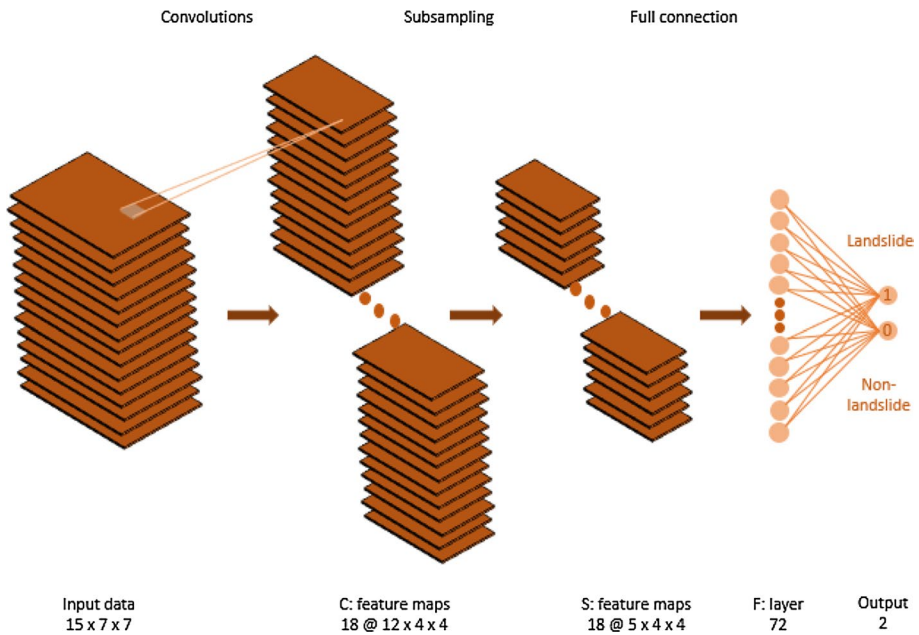


Fig. 10 Architecture of CNN-3D



successor layer are a linear feature of its predecessor layer as far as the ANN techniques are concerned. But the actual scenario through this linear association is challenging to embody (Huang and Babri 1998). The activation function approach has been used to overcome the appropriate problem of the output data. It can effectively form nonlinear ties from linear ties using predefined (nonlinear) activation functions (Dahl et al. 2011).

The current research incorporated the CNN architectures with the rectified linear unit (ReLU) function (Dahl et al. 2013). The ReLU function, with two major benefits, is among CNN's most popular and effective activation functions. The first benefit is that this function facilitates overcoming the dilemma of gradient disappearance. Secondly, it is more economical as well as useful for training prediction techniques compared to the additional activation functions (Maas et al. 2013). The loss function was a categorical cross-entropy function, and its optimizer was an advanced adaptive gradient (AdaGrad) algorithm (Anthonopoulos, Christodoulidis et al. 2016). The AdaGrad technique can limit the learning rate and use various learning rates per iteration for individual learning parameters (Duchi et al. 2011). The adaptive moment estimation (Adam) (Kingma and Ba 2014) was used as an optimizer for ResNet. It is a first-order gradient-based algorithm of stochastic objective functions based on adaptive estimates of lower-order moments. The posterior probability was produced for each grid cell using the softmax function (Lawrence, Giles et al. 1997). The prediction output is greatly improved by dropping manipulation since the neural network units corresponding to a given possibility can be momentarily abandoned during training (Hinton et al. 2012). Explicitly, the dropout manipulation obliges a nerve unit, in addition to overfitting between secret units, to work with other arbitrary neural units (Srivastava et al. 2014). Furthermore, this manipulation will improve the generalization of prediction techniques (Dahl et al. 2013).

### 13 Model evaluation methods

The performance of the suggested methodology was evaluated using the measurements of OA and ROC (Tsangaratos and Ilia 2016, Pham et al. 2017b, Chen et al. 2017c). The proportion of the number of grid cells, either landslide or non-landslide that were classified accurately (represented by *a*) to the entire grid cells (represented by *b*) gives the OA value, which is calculated using the following formula:

$$OA = \frac{a}{b} \times 100\%$$

A higher OA value embodies superior classification precision. For evaluating the functioning of landslide prediction techniques, the ROC curve is a standard practiced approach (Bradley 1997). The TP (true positive) rate, which is described as “sensitivity,” is plotted against the FP (false positive) rate, which is described as “100-specificity,” at different threshold values to produce the curve. Furthermore, for quantitatively evaluating the performance of landslide susceptibility mapping techniques, the measurement of AUC has been applied broadly (Wang et al. 2017; Mandal and Mandal 2018a). In particular, if the AUC value is in the proximity of 1, the prediction approach is deemed outstanding (Tsangaratos et al. 2017, Pham et al. 2017a, Zhu et al. 2018).

The Matthews correlation coefficient (MCC) (Matthews 1975) has also been practiced in various ML techniques, even though the two assortments are of very distinct volumes. The following expression characterizes the MCC:

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

where the number of non-landslide and landslide locations classified precisely is represented by TN (true negative) and TP, and the number of non-landslide and landslide locations classified wrongly are represented by FN (false negative) and FP, respectively. Additionally, this measure is a correlation coefficient among the observed and predicted classes. Usually, the MCC value equal to 1 portrays a perfect prediction for the final result. On the other hand, the MCC value of 0 and  $-1$  signifies an arbitrary prediction and a complete disparity between the prediction and observation.

Furthermore, the significant difference among suggested techniques was evaluated using the Chi-square test (Kuncheva 2004). It is based on a former proposition that the techniques used to map landslide susceptibility have no substantial change (Tallarida and Murray 1987). For the validation purpose, the Chi-square and p-values were designated and computed. Generally, a p-value less than 0.05 and a Chi-square value greater than 3.841 indicates that between the two techniques, there exists a significant difference (Pham et al. 2017a).

## 14 Results and analysis

The training set designated was used to determine the predictive potential of the landslide conditioning factors by analyzing multicollinearity and using the GR method. The results of the landslide conditioning factor multicollinearity study are summarized in Table 1. The VIF value of SPI and STI is observed to be 9.772 and 9.43, which are very near the threshold value of 10. Thus, their contribution to the model is quite less. As far as the GR method is concerned, the factors with higher weights are additionally meaningful to the practiced techniques, while factors with weights equivalent to zero cannot contribute to the modeling process; therefore, they should be omitted from further analysis. Figure 11 shows the AM value of the respectively conditioning factor. The land use factor has the highest AM value of 0.0525 compared to the other factors, which means that it is comparatively significant as the other factors.

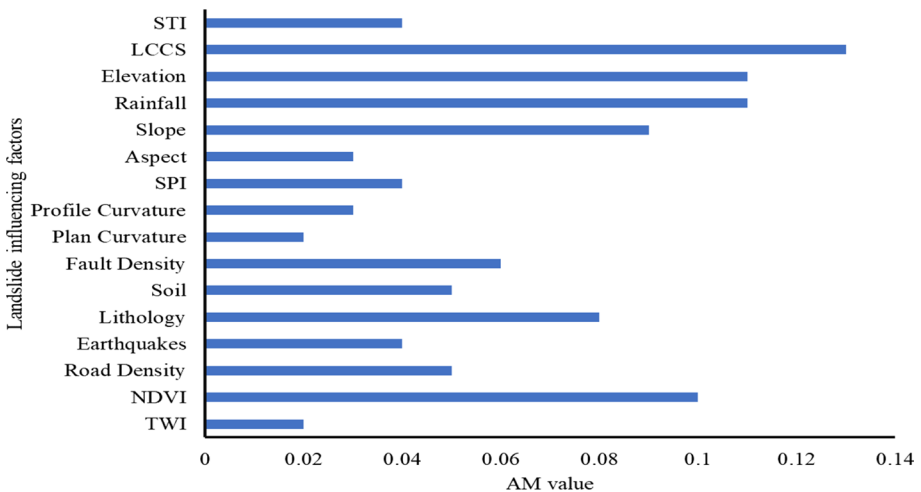
The AM values of aspect, lithology, elevation, slope, rainfall, soil, NDVI, and SPI are between 0.0474 and 0.0151. Furthermore, the AM values of plan curvature, fault density, road density, profile curvature, and TWI are positive but less than 0.01, which implies that their contribution to the models is very little. All the outstanding landslide conditioning factors with more than zero AM values participate in the models. Still, their contribution is very small, and they must be considered for the modeling process because they still have impact on landslides have high frequency ration as far as landslide is concern (Table 2). The reclassified maps of all the landslide conditioning factors are presented in Fig. 12a–p.

## 15 Model validation and comparison

The trial-and-error method was used to optimize all the parameter settings to develop the CNNs. Table 3 displays the parameter settings for all CNNs and ResNet. Every grid cell in the zone in question was designated a susceptibility index after the formulation of the models through training datasets. The weights obtained were then allocated for each factor

**Table 1** Outcomes of multicollinearity analysis

Landslide conditioning factors	Statistics	
	TOL	VIF
Elevation	0.262	3.95
Aspect	0.948	1.126
Fault density	0.878	1.213
LCCS	0.708	1.495
Lithology	0.789	1.346
NDVI	0.713	1.485
Plan curvature	0.582	1.813
Profile curvature	0.739	1.435
Rainfall	0.573	1.752
Earthquake	0.811	1.177
Road density	0.835	1.144
Slope	0.293	3.191
Soil	0.334	2.816
SPI	0.085	9.772
STI	0.079	10.65
TWI	0.426	2.228



**Fig. 11** AM value of each landslide conditioning factor obtained through the GR method

class, and the final susceptibility maps were drawn up in an ArcGIS setting for each model. The natural breaking technique has been used to reclassify the higher vision indices in five very low, medium, medium, and very high classes. The obtained susceptibility maps for landslides from different CNN frameworks and ResNet are shown in Fig. 13.

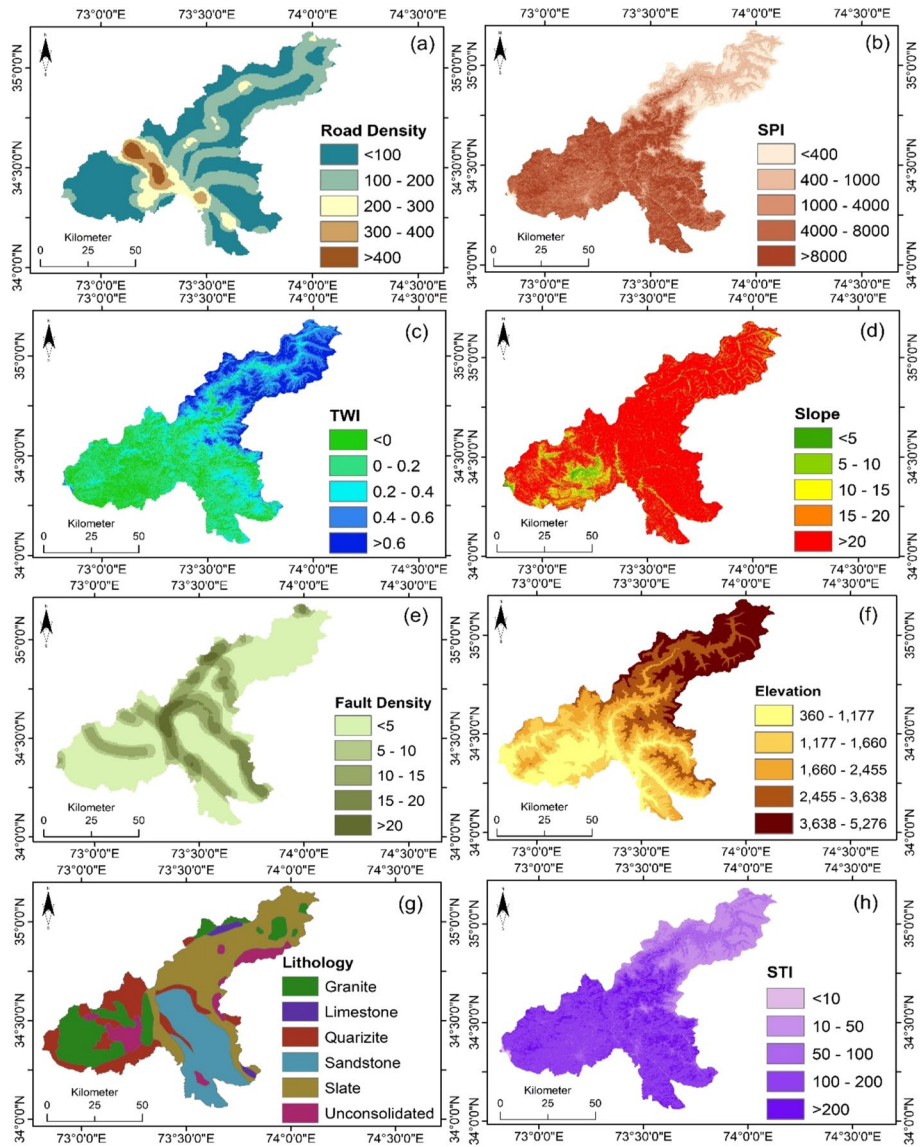
In contrast, for each of the models, the distribution of percentages for each of the susceptibility classes is shown in Fig. 14. All the maps show that the northeast portion of both districts was classified as high and very highly susceptible areas. The outcomes of

**Table 2** Spatial relationship between past landslides and different classes of landslide conditioning factors using the FR model

Factor	Class	No. of landslides	Percentage of Landslide	No. of pixels in domain	Percentage of domain	FR
Road density	< 100	163	0.05	5,344,183	0.43	0.116279
	100–200	455	0.14	2,858,517	0.23	0.608696
	200–300	748	0.23	1,739,967	0.14	1.642857
	300–400	845	0.26	1,367,117	0.11	2.363636
	> 400	1,040	0.32	1,118,550	0.09	3.555556
SPI	< 400	358	0.11	2,237,100	0.18	0.611111
	400–1000	260	0.08	2,609,950	0.21	0.380952
	1000–4000	585	0.18	1,367,117	0.11	1.636364
	4000–8000	878	0.27	3,231,367	0.26	1.038462
	> 8000	1170	0.36	2,982,800	0.24	1.5
TWI	< 0	228	0.07	3,231,367	0.26	0.269231
	0–0.2	358	0.11	1,864,250	0.15	0.733333
	0.2–0.4	488	0.15	2,982,800	0.24	0.625
	0.4–0.6	1138	0.35	869,983	0.07	5
	> 0.6	1040	0.32	3,479,933	0.28	1.142857
Slope	< 5	98	0.03	745,700	0.06	0.5
	5–10	228	0.07	372,850	0.03	2.333333
	10–15	358	0.11	1,118,550	0.09	1.222222
	15–20	618	0.19	497,133	0.04	4.75
	> 20	1951	0.6	9,694,100	0.78	0.769231
Fault density	< 5	358	0.11	5,841,317	0.47	0.234043
	5–10	293	0.09	2,361,383	0.19	0.473684
	10–15	683	0.21	1,739,967	0.14	1.5
	15–20	845	0.26	1,367,117	0.11	2.363636
	> 20	1073	0.33	1,118,550	0.09	3.666667
Elevation	360–1177	650	0.2	3,231,367	0.26	0.769231
	1177–1660	780	0.24	2,734,233	0.22	1.090909
	1660–2455	585	0.18	2,485,667	0.2	0.9
	2455–3638	845	0.26	1,864,250	0.15	1.733333
	3638–5276	390	0.12	2,112,817	0.17	0.705882
Lithology	Granite	325	0.1	1,615,683	0.13	0.769231
	Limestone	228	0.07	621,417	0.05	1.4
	Quartzite	358	0.11	1,491,400	0.12	0.916667
	Sandstone	878	0.27	2,982,800	0.24	1.125
	Slate	1105	0.34	4,722,767	0.38	0.894737
	Unconsolidated	358	0.11	994,267	0.08	1.375
STI	< 10	943	0.29	2,858,517	0.23	1.26087
	10–50	748	0.23	2,112,817	0.17	1.352941
	50–100	683	0.21	1,864,250	0.15	1.4
	100–200	260	0.08	3,728,500	0.3	0.266667
	> 200	618	0.19	1,864,250	0.15	1.266667

**Table 2** (continued)

Factor	Class	No. of land-slides	Percentage of Land-slide	No. of pixels in domain	Percentage of domain	FR
Aspect	< 70	325	0.1	497,133	0.04	2.5
	70–140	845	0.26	3,355,650	0.27	0.962963
	140–210	293	0.09	2,485,667	0.2	0.45
	210–280	1463	0.45	4,225,633	0.34	1.323529
	> 280	325	0.1	1,864,250	0.15	0.666667
NDVI	< 0	715	0.22	2,609,950	0.21	1.047619
	0–0.3	813	0.25	1,615,683	0.13	1.923077
	0.3–0.6	683	0.21	1,367,117	0.11	1.909091
	0.6–0.9	585	0.18	2,485,667	0.2	0.9
	> 0.9	455	0.14	4,349,917	0.35	0.4
Profile curvature	< - 1	845	0.26	2,858,517	0.23	1.130435
	- 1 – -0.5	455	0.14	1,118,550	0.09	1.555556
	- 0.5–0	1463	0.45	6,338,450	0.51	0.882353
	0–0.5	130	0.04	1,242,833	0.1	0.4
	> 0.5	358	0.11	869,983	0.07	1.571429
Earthquake	High	3251	1	12,428,333	1	1
	LCCS	Bare area	618	0.19	372,850	0.03
Soil data	Built up	228	0.07	869,983	0.07	1
	Agriculture	455	0.14	1,864,250	0.15	0.933333
	Forest	813	0.25	2,112,817	0.17	1.470588
	Range lands	1008	0.31	6,338,450	0.51	0.607843
	Snow	130	0.04	869,983	0.07	0.571429
	Clay	325	0.1	372,850	0.03	3.333333
	Clay loam	553	0.17	4,349,917	0.35	0.485714
Rainfall	Loam	1073	0.33	4,847,050	0.39	0.846154
	Sandy clay loam	683	0.21	1,615,683	0.13	1.615385
	Sandy loam	585	0.18	994,267	0.08	2.25
	Water bodies	33	0.01	248,567	0.02	0.5
	715–790	455	0.14	2,734,233	0.22	0.636364
	790–874	845	0.26	1,491,400	0.12	2.166667
Plan curvature	874–935	683	0.21	1,864,250	0.15	1.4
	935–1084	618	0.19	3,355,650	0.27	0.703704
	1084–1331	650	0.2	2,982,800	0.24	0.833333
	< 0	845	0.26	2,858,517	0.23	1.130435
	0–0.2	358	0.11	1,118,550	0.09	1.222222
Plan curvature	0.2–0.4	1723	0.53	6,089,883	0.49	1.081633
	0.4–0.6	228	0.07	1,367,117	0.11	0.636364
	> 0.6	98	0.03	994,267	0.08	0.375



**Fig. 12** Thematic maps of landslide conditioning factors: road density (a), SPI (b), TWI (c), slope (d), fault density (e) elevation (f), lithology (g), STI (h), aspect (i), NDVI (j), profile curvature (k), earthquake (l), LCCS (m) soil data (n), rainfall (o) and plan curvature (p)

ResNet, CNN-1D, and CNN-2D were very much alike. In Fig. 13b, the high and very high-class concentration is more intense than the concentration in other obtained maps. The high and very high classes in Fig. 13a and c, and 13d were quite similar and more intense than the SVM model in Fig. 13 f. However, in the outcome of CNN-3D in Fig. 13e, the high and very class in the deliberated area was more than the outcome of the SVM model in Fig. 13 f. The result of DNN in Fig. 13 g demonstrated that the very high and

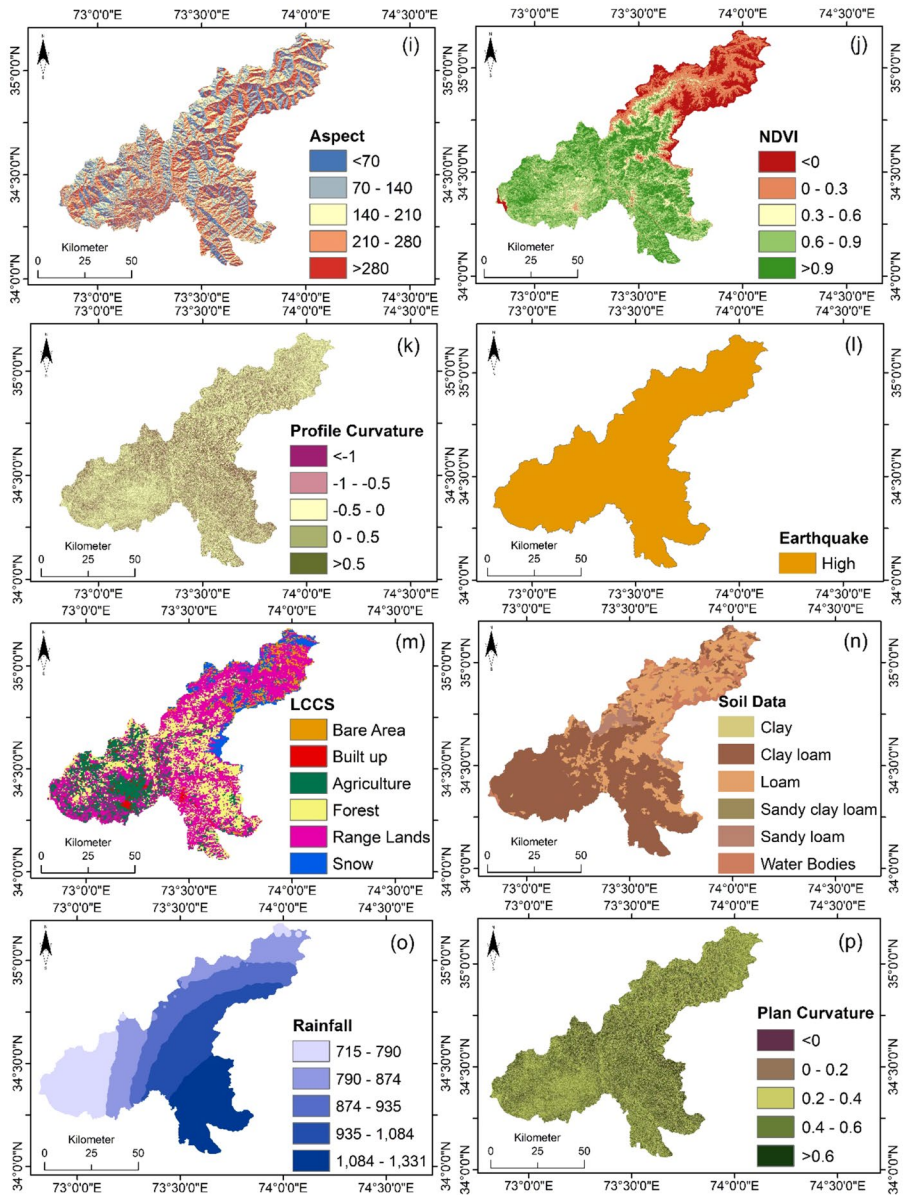


Fig. 12 (continued)

high susceptible zones in the study area are less than those in the maps obtained through ResNet, LeNet-5, CNN-1D, CNN-2D, CNN-3D, and SVM but are more than LR. The map obtained through LR has the least proportion of high and very high class, as seen from Fig. 13 h.

For the proposed multiple CNN frameworks and the ResNet technique, the OA and MCC values are listed in Table 4. The OA values of ResNet were higher than that of the

**Table 3** Parameter settings of the CNNs

Model	Parameter settings					
	Convolutional kernel size	Max pooling kernel size	Number of iterations	Activation function	Optimizer	Dropout rate
CNN-1D	3 × 1	2 × 1	300	ReLU	AdaGrad	0.4 and 0.3
CNN-2D	3 × 3	2 × 2	100	ReLU	AdaGrad	
CNN-3D	3 × 3 × 3	2 × 2 × 2	300	ReLU	AdaGrad	
LeNet-5	5 × 5	2 × 2	15	ReLU	AdaGrad	
ResNet	5 × 5	2 × 2	20	ReLU	Adam	

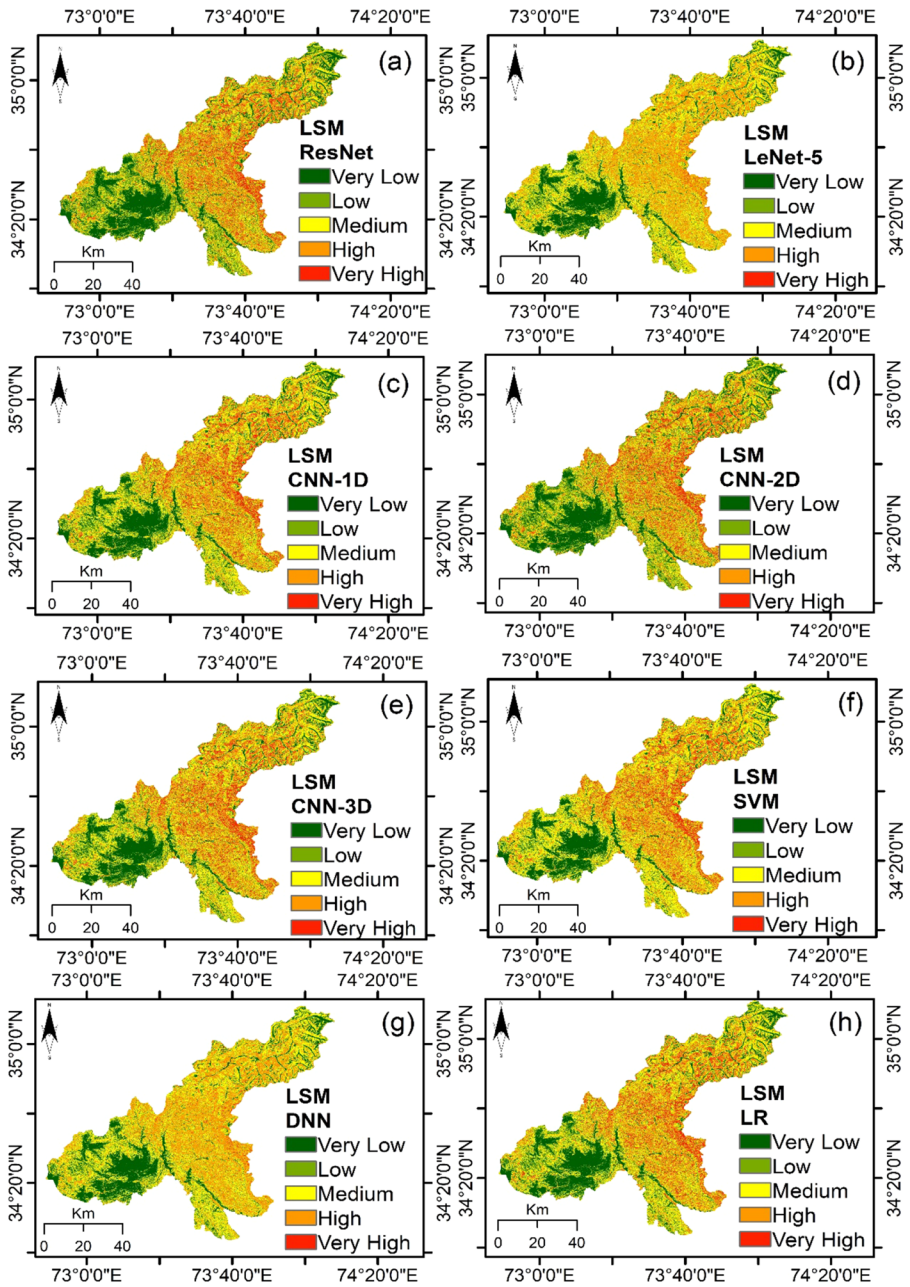
suggested CNN frameworks. In particular, the OA value (83.52%) of ResNet was higher than other methods, which is almost 1% greater than that of CNN-3D (82.54%), which has the second-highest OA value. However, CNN-2D obtained an OA value of 81.71%, followed by CNN-1D with an OA value of 79.32%. Moreover, ResNet also obtained the highest MCC value (0.596) compared to other techniques. Secondly, it was CNN-3D which obtained the highest MCC value (0.582), followed by CNN-1D, LeNet-5 plus CNN-2D with MCC values of 0.535, 0.526, and 0.514, correspondingly.

The obtained ROC curves using the testing set for all the used CNN frameworks are presented in Fig. 15. It can be witnessed from the figure that the CNN-3D method has the maximum AUC value of 0.874, which specifies that it has a superior predictive capacity to the other CNN methods. Moreover, the CNN-1D and CNN-2D methods attained comparable AUC values of 0.871 and 0.872, respectively, whereas the LeNet-5 technique attained the AUC value of 0.864. The significant difference between the proposed prediction techniques was evaluated using a Chi-square test. A Chi-square value greater than 3.841, whereas a significant level value ( $p$ ) less than 0.05 indicates a significant difference among the prediction techniques. For the different proposed CNNs and ResNet, the Chi-square values plus the significant levels are listed in Table 5. It shows that ResNet and all the CNNs are quite distinct as their Chi-square and the significant level values comply with the threshold conditions.

For further validation of the efficacy of the proposed models, the best performing model (ResNet) in the preceding experiments was chosen to be compared with different common ML and DL techniques. DNNs are typically feedforward networks, where data flow from the input layer to the output layer without looping back (CireşAn et al. 2012). First, a simulated neural unit map is generated by DNN, then links these neural units by assessing their weights. Subsequently, a probability between 1 and 0 is produced by multiplying the input data with the weights. The chosen DNN has a network architecture composed of five layers containing four hidden, fully connected layers. The number of neural units in the four hidden layers is 50, 30, 20, and 10. The prediction outcomes in the output layer are obtained with two neural units, corresponding to landslide and non-landslide units. For comparison with an ML technique, the SVM classifier with a radial basis function kernel was utilized. For SVM, the  $\gamma$  ( $2^{-9}$ ) and optimal  $C$  ( $2^7$ ) were obtained through five-fold cross-validation varying from  $2^{-5}$  and  $2^{-15}$  to  $2^{15}$  and  $2^5$ , correspondingly. The landslide susceptibility maps obtained from SVM and DNN methods are presented in Fig. 12f g, respectively.

The OA as well as the MCC values of the orthodox DL and ML techniques are listed in Table 6. ResNet achieved the highest OA value of 83.52%. However, CNN-3D achieved an OA value of 82.54%, almost 4% more than the optimized SVM (78.43%). Followingly,





**Fig. 13** Landslide susceptibility maps for different CNN and orthodox ML and DL methods: ResNet (a), LeNet-5 (b), CNN-1D (c), CNN-2D (d), CNN-3D (e), SVM (f), DNN (g), and LR (h)

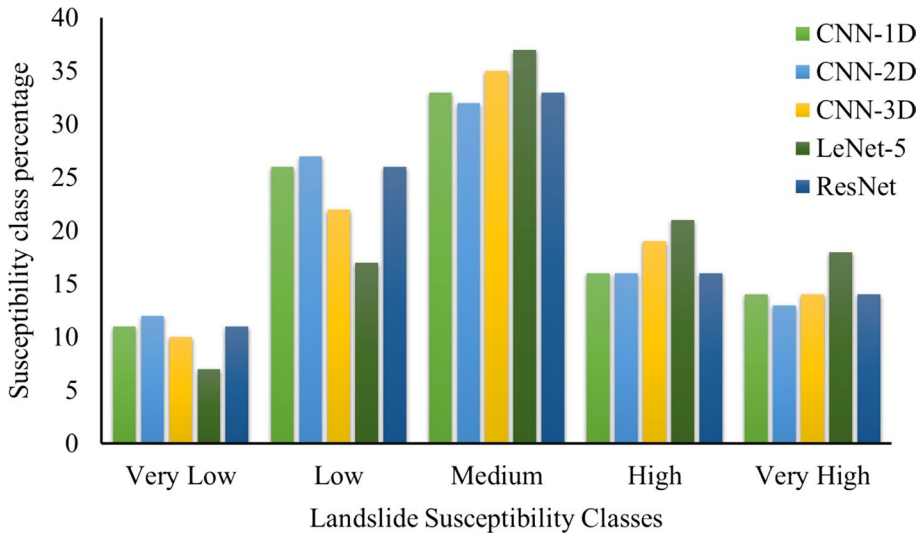


Fig. 14 Proportions of various landslide susceptibility classes for different CNNs

Table 4 Performance comparison of different methods

Model	OA value (%)	MCC
CNN-1D	79.32	0.535
CNN-2D	81.71	0.514
CNN-3D	82.54	0.582
LeNet-5	78.12	0.526
ResNet	83.52	0.596

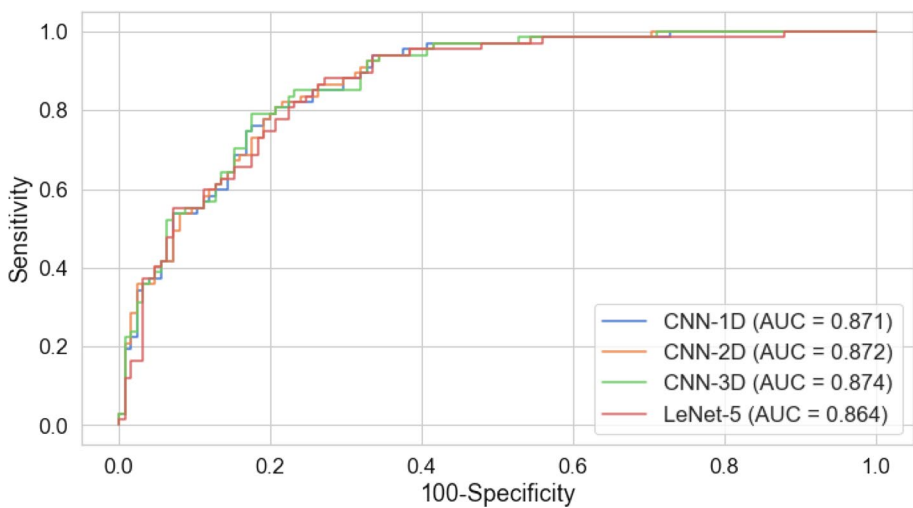


Fig. 15 ROC curves for different CNNs using the validation set

**Table 5** Chi-square values and significant levels of different CNNs

Comparative pairs	Chi-square value	P-value	Significance level
CNN-1D vs. CNN-2D	204.55	<0.0001	Yes
CNN-1D vs. CNN-3D	193.74	<0.0001	Yes
CNN-1D vs. LeNet-5	265.87	<0.0001	Yes
CNN-1D vs. ResNet	212.44	<0.0001	Yes
CNN-2D vs. CNN-3D	194.54	<0.0001	Yes
CNN-2D vs. LeNet-5	276.43	<0.0001	Yes
CNN-2D vs. ResNet	248.52	<0.0001	Yes
CNN-3D vs. LeNet-5	247.76	<0.0001	Yes
CNN-3D vs. ResNet	213.55	<0.0001	Yes
LeNet-5 vs. ResNet	195.85	<0.0001	Yes

it is LeNet-5 plus DNN, which have quite comparable OA values of 78.12 and 77.3%, correspondingly. Moreover, ResNet also accomplished the highest MCC value (0.596). However, as compared to the optimized SVM, CNN-3D also accomplished a considerably greater MCC value. The obtained MCC value of CNN-3D is 0.514, followed by the LeNet-5 (0.526), DNN (0.511), and SVM (0.427).

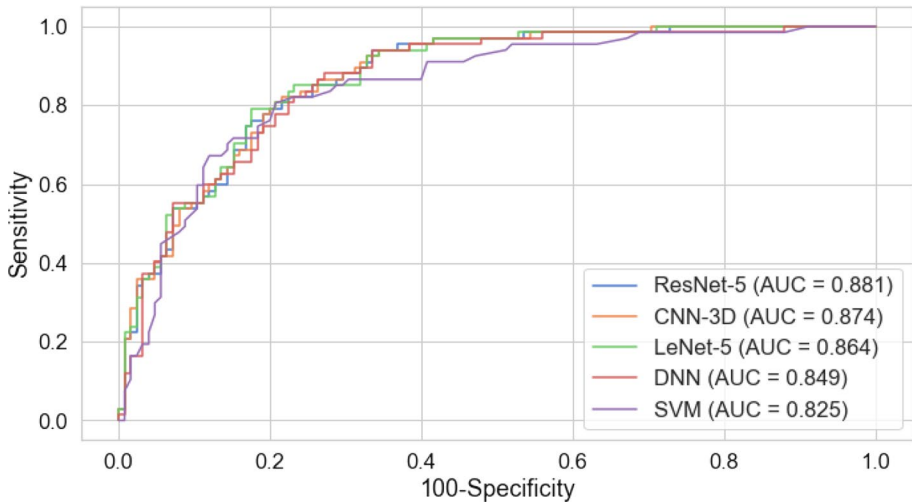
For comparison, the ROC curves using the validation set for ResNet, CNN-3D, LeNet-5, DNN, and SVM are shown in Fig. 16. The predictive power as per the AUC of ResNet was observed to be better than the optimized SVM and DNN. However, CNN-3D obtained an AUC value greater than SVM and DNN but less than ResNet. Moreover, the CNN-3D, and LeNet-5, correspondingly attained AUC values of 0.881, 0.874, and 0.864, whereas DNN and SVM likewise achieved AUC values of 0.849 and 0.825.

## 16 Discussion

The susceptibility mapping of landslides is of immense importance as it assists in visually examining landslide-prone areas. Previously, various models have been established for mapping the landslide susceptibility of a specific region, and their functioning has been contrasted. Yet, the prediction precision of these developed techniques is still under discussion (Akgun 2012, Martinović et al. 2016, Chen et al. 2017b). Hence, it is essential to explore novel techniques and approaches for assessing landslide susceptibility (Bui et al.

**Table 6** Performance comparison of different methods

Model	OA value (%)	MCC
CNN-3D	82.54	0.582
ResNet	83.52	0.596
LeNet-5	78.12	0.526
DNN	77.63	0.511
SVM	78.43	0.427



**Fig. 16** ROC curves for CNNs, DL, and ML methods using the validation set

2016). In landslide susceptibility evaluation studies, significant progress has been accomplished using orthodox regression analysis or ML techniques. Comparatively, simple structure prediction models are constructed through these techniques.

Lately, numerous ML techniques have been utilized for the spatial prediction of landslides and compared with each other for a given area. These techniques include SVM (Chen et al. 2018), LR (Tsangaratos and Ilia 2016), decision tree (Chen et al. 2017c), and ANN (Chen et al. 2017a). Moreover, various ensemble techniques have been built and used for the same purpose. These include rotation forest (Pham, Shirzadi et al. 2018), bagging (Pham et al. 2017a), and AdaBoost (Hong, Liu et al. 2018). The raw data are not processed outstandingly by the orthodox ML methods because they have a limited ability to do so (LeCun, Bengio et al. 2015).

To some extent, these models have flaws in defining the intricate nonlinear landslide systems and in avoiding the overfitting problem. Modern DL techniques such as ResNet, CNNs, and DNN have essential improvements to the above problems compared to the orthodox methods. Two significant characteristics of DL techniques are their nonlinear and multilayer structures, which enable them to depict an intricate nonlinear landslide system influenced by several conditioning factors. Nevertheless, the DL techniques have a robust enhancement over ML techniques and play a progressively significant role in natural language processing, image processing, and computer vision. The DL techniques are smart enough to automatically study the representation from the raw data required for prediction. Thus, it is favorable to discover the possibility of using potent DL techniques for assessing landslide susceptibility (Wang, Fang et al. 2019).

Furthermore, valid algorithms, such as “batch” or “dropout,” are incorporated in DL techniques to prevent overfitting effectively. Thus, DL techniques are likely to enhance the assessment precision of landslide susceptibility. Nevertheless, comparatively limited studies performed landslide susceptibility assessments using DL techniques.

The pondered area (Mansehra and Muzaffarabad districts) for this study is in the northern part of Pakistan. As this region has suffered numerous landslides due to seismic activities and extensive rainfall and is expected to experience these types of events in the future;

thus, it is vital to map landslide susceptibility for preventive measures. In the current study, an effort has been made to address this issue with the assessment and comparison of ResNet and multiple CNN architectures. The DL techniques: ResNet and CNN have superior fitting power, and they are comparatively effective in feature extraction, which enhances the prediction ability of the models, thus generating efficient susceptibility maps. Moreover, we used multiple activation functions and optimizers to enhance the prediction results and found that the models showed superior performance with ReLU (activation function) and AdaGrad (optimizer). Thus, we incorporated these functions into the models for better prediction results. These techniques were also compared with DNN, LR, and SVM classifiers to compare and show the performance differences.

The purpose of comparing DL and ML techniques was to establish the best-performing model among the multiple proposed conventional and modern models. Therefore, the best-performing CNN and the ResNet models were compared with the state-of-the-art ML technique (SVM) and DL technique (DNN) to establish the superiority of the best-performing model. Among the two used ML techniques, SVM was selected for comparison purposes because several studies have established its superiority over other conventional ML techniques for landslide susceptibility mapping (Wu et al. 2016, Tien Bui et al. 2019, Yu et al. 2019, Pham et al. 2019b).

Landslides are highly complex and organized progressions through many environmental and topographical variables known as landslide conditioners. Sixteen conditioning factors were analyzed based on previous research, available evidence, and environmental conditions in the region to map landslide susceptibility in the current investigation. Each considered factor was translated into spatially delimited layers or maps of 30 m to 30 m grid size, corresponding to the DEM data obtained within an ArcGIS setting. In order to map the landslide susceptibility, the choice of an effective terrain mapping unit is also essential. A standard grid cell model was used to assess landslide susceptibility because it is the most common approach to spatially represent this form of a dataset (Tsangaratos and Ilia 2016).

The predictive ability of all the conditioning factors must be assessed before investigating landslide susceptibility. For this purpose, multicollinearity analysis was used to assess correlations among the factors considered. Moreover, the GR technique ranked these factors based on their importance. The outcomes of the multicollinearity analysis disclosed that the STI has intense multicollinearity and ought to be eliminated from the additional process. On the contrary, the outcomes of the GR technique revealed that land use and NDVI have greater AM values as compared to the other factors, demonstrating that these two factors are additionally crucial for landslide incidence. NDVI can precisely display the surface vegetation coverage. It can be seen from Fig. 12j that those regions where the NDVI value is less have a higher potential for landslides and fall under the class of high and very high susceptibility in produced susceptibility maps (Fig. 13). Furthermore, it must be noted that in mountainous areas like the one under consideration, landslides, in addition to external forces, occur continuously because of precipitation, even on slopes overlaid with considerable vegetation.

After the feature selection process, the selected features were then used to build the multiple CNN architectures and ResNet model. Moreover, various data representations transformed from initial landslide data are introduced as well as incorporated in the suggested CNN architectures. The spatial information can be efficiently extracted by CNN through local connections and can considerably decrease the required network parameters by distributing weights. The local link can easily be used by the architecture of CNN-1D, which enables this CNN to learn the more convoluted representations from factor vectors progressively. As CNN-2D at first demonstrated outstanding performance in visual image

analysis, we transformed every 1D factor vector into a 2D matrix to adequately obtain the important hidden features. Although the CNN-1D and CNN-2D can be considered the same regarding the convolution operation, and their outcomes are also very similar, the point of considering the two different designs and calling them two different models is to have more results for comparison purposes. Moreover, we intended to maintain the order rather than jumping directly to CNN-3D after CNN-1D. The only difference lies mainly in the data representation instead of the mechanism of landslide modeling. Besides learning factor representations, the CNN-3D also gets local spatial information.

The results disclosed that the total percentages of high and very high classes are approximately comparable for CNN-1D, CNN-2D, and ResNet. Moreover, the high and very high classes of the ResNet, CNN-1D, CNN-2D, as well as CNN-3D are comparatively less than the LeNet-5 method but greater than SVM, DNN, and LR. ResNet accomplished higher corresponding values than the OA plus the MCC values of the projected CNNs. Furthermore, the ResNet and CNN-3D obtained superior prediction performance in the succeeding experiments compared to the well-known DNN and SVM classifiers. Lastly, using the testing dataset, the AUC value (0.881) turned out to be the highest for ResNet, followed by CNN-3D (0.874), which shows that ResNet and the proposed 3D structure can effectively improve the prediction performance, and they may prove to be a good technique for upcoming investigations. Therefore, it can be asserted that the suggested techniques are more feasible for managing and preventing landslides.

The outcomes of this study exhibited that ResNet is better than the proposed CNN frameworks. However, CNN-3D is better than the conventional DL technique, namely DNN, and the orthodox ML techniques, namely LR and SVM, which indicates that the suggested data representation forms of CNNs might be favorable vigorous approaches for mapping the landslide susceptibility. Different CNN frameworks for mapping the susceptibility to landslides have also been used in Yanshan County, China, by Wang, Fang et al. 2019. The results showed that the CNN frameworks performed better as compared to orthodox ML and DL techniques. Finally, the proposed CNN architectures offer a new way of handling raw landslide data in the present analysis. Moreover, the produced landslide susceptibility maps through the employed techniques can essentially be used as a guide by the planners and policymakers to avert and alleviate the landslide risk by positing future land use zones appropriately and ascertaining and establishing alleviation urgencies for the endangered areas.

## 17 Conclusion

The core objective of this work was to examine the application of different convolutional neural network (CNN) frameworks and residual network (ResNet) for mapping landslide susceptibility in northern Pakistan (Mansehra and Muzaffarabad districts). The considered area is among those regions in the world which are severely prone to landslides due to the geological and environmental settings of the region. The practiced frameworks turned out to be a valuable approach, and they can be utilized for other regions around the globe with comparable attributes. The ResNet and intended CNNs were validated centered on the evaluation of sixteen conditioning factors that were obtained from several supplementary sources. The ResNet and the proposed CNNs were employed to produce landslide susceptibility maps of the deliberated area and were contrasted to the orthodox deep learning (DL) and machine learning (ML) techniques of DNN and SVM, and LR, respectively.

Different objective measures such as OA, MCC, ROC, and AUC were utilized to confirm the outcomes. The obtained landslide susceptibility maps using ResNet and the suggested CNNs are more effective for managing and preventing landslides as compared to those from orthodox techniques, as confirmed by the investigational outcomes. The prediction results of the ResNet model outperformed all the other used models as per OA, MCC, and AUC values. However, the CNN-3D model was found to be better among the proposed CNNs. Thus, these models can be applied to generate definitive susceptibility maps. Lastly, when using a DL technique such as ResNet or the multiple CNN frameworks in the present study, the prediction precisions of susceptibility maps can be efficiently enhanced by following two strategies during the construction of the model's architecture by including dropout manipulation and by selecting an activation function. In conclusion, ResNet and CNNs are very favorable for the spatial prediction of landslides and can be used for future research. Moreover, these techniques can also be compared with more efficient DL techniques in the future for mapping landslide susceptibility.

**Funding** The authors have not disclosed any funding.

**Data Availability** The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Declarations

**Conflict of interest** The authors declare no competing interests pertinent to the context of this work.

**Ethical approval** The study does not include any human subjects. Hence, no ethical approval was required for this study.

## References

- Akgun A (2012) A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at İzmir. *Turk Landslides* 9(1):93–106
- Anthimopoulos M et al (2016) Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging* 35(5):1207–1216
- Aslam B et al (2021) Development of integrated deep learning and machine learning algorithm for the assessment of landslide hazard potential. *Soft Comput* 25(21):13493–13512
- Aslam B et al (2022) Comparison of multiple conventional and unconventional machine learning models for landslide susceptibility mapping of Northern part of Pakistan. *Environ, Develop Sustain*: 1–28
- Bacha AS et al (2018) Landslide inventory and susceptibility modelling using geospatial tools, in Hunza-Nagar valley, northern Pakistan. *J Mt Sci* 15(6):1354–1370
- Ballabio C, Sterlacchini S (2012) Support vector machines for landslide susceptibility mapping: the Staffora River Basin case study Italy. *Math Geosci* 44(1):47–70
- Barella CF et al (2019) A comparative analysis of statistical landslide susceptibility mapping in the south-east region of Minas Gerais state. Brazil. *Bull Eng Geol Environ* 78(5):3205–3221
- Basharat M et al (2016) Landslide susceptibility mapping using GIS and weighted overlay method: a case study from NW Himalayas Pakistan. *Arabian Journal of Geosciences* 9(4):1–19
- Bibi T et al (2016) Landslide susceptibility assessment through fuzzy logic inference system (FLIS). *The International Archives of Photogrammetry. Remote Sens Spat Inform Sci* 42:355
- Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn* 30(7):1145–1159
- Bui DT et al (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* 13(2):361–378

- Cao J et al (2019) Urban noise recognition with convolutional neural network. *Multimedia Tools and Applications* 78(20):29021–29041
- Chen W et al (2017a) A GIS-based comparative study of Dempster-Shafer, logistic regression and artificial neural network models for landslide susceptibility mapping. *Geocarto Int* 32(4):367–385
- Chen W et al (2017b) GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naïve-Bayes tree, and alternating decision tree models. *Geomatics Nat Hazards Risk* 8(2):950–973
- Chen W et al (2017c) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* 151:147–160
- Chen W et al (2018) A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. *Bull Eng Geol Environ* 77(2):647–664
- CireşAn D et al (2012) Multi-column deep neural network for traffic sign classification. *Neural Netw* 32:333–338
- Crosta GB (2004) Introduction to the special issue on rainfall-triggered landslides and debris flows. *Eng Geol* 3(73):191–192
- Dahl GE et al (2011) Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Trans Audio Speech Lang Process* 20(1):30–42
- Dahl GE et al (2013) Improving deep neural networks for LVCSR using rectified linear units and dropout. In: 2013 IEEE international conference on acoustics, speech and signal processing, IEEE
- Dai F et al (2002) Landslide risk assessment and management: an overview. *Eng Geol* 64(1):65–87
- Dash M, Liu H (1997) Feature selection for classification. *Intell Data Anal* 1(3):131–156
- Demir G (2019) GIS-based landslide susceptibility mapping for a part of the North Anatolian Fault Zone between Reşadiye and Koyulhisar (Turkey). *Catena* 183:104211
- Dou J et al. (2020) Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed. *Japan Landslides* 17(3):641–658
- Duchi J et al (2011) Adaptive subgradient methods for online learning and stochastic optimization. *J Mach Learn Res* 12(7)
- Duo Z et al (2019) Oceanic mesoscale eddy detection method based on deep learning. *Remote Sensing* 11(16): 1921
- Fang Z et al (2020) Integration of convolutional neural network and conventional machine learning classifiers for landslide susceptibility mapping. *Computers & Geosciences*: 104470
- Flentje PN et al (2007) Guidelines for landslide susceptibility, hazard and risk zoning for land use planning.
- Girshick R (2015) Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision
- Gupta SK et al (2018) Selection of weightages for causative factors used in preparation of landslide susceptibility zonation (LSZ). *Geomatics Nat Hazards Risk* 9(1):471–487
- Guzzetti F et al (2006) Estimating the quality of landslide susceptibility models. *Geomorphology* 81(1–2):166–184
- Guzzetti F et al (2012) Landslide inventory maps: New tools for an old problem. *Earth Sci Rev* 112(1–2):42–66
- He S et al (2012) “Application of kernel-based Fisher discriminant analysis to map landslide susceptibility in the Qinggan River delta, Three Gorges China. *Geomorphology* 171:30–41
- He K et al (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition
- Hinton GE et al (2012) Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580
- Hong H et al (2016a) “GIS-based landslide spatial modeling in Ganzhou City China. *Arab J Geosci* 9(2):112
- Hong H et al (2016b) Landslide susceptibility assessment in Lianhua County (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models. *Geomorphology* 259:105–118
- Hong H et al (2017) A hybrid fuzzy weight of evidence method in landslide susceptibility analysis on the Wuyuan area China. *Geomorphology* 290:1–16
- Hong H et al (2018a) “Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *Catena* 163:399–413
- Hong H et al (2018b) Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *Catena* 163:399–413
- Huang G-B, Babri HA (1998) Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded non-linear activation functions. *IEEE Trans Neural Networks* 9(1):224–229



- Hussain A, Yeats RS (2009) Geological setting of the October 8 2005 Kashmir earthquake. *J Seismolog* 13(3):315–325
- Jiao Y et al (2019) Performance evaluation for four GIS-based models purposed to predict and map landslide susceptibility: a case study at a World Heritage site in Southwest China. *Catena* 183:104221
- Kamp U et al (2008) GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region. *Geomorphology* 101(4):631–642
- Kanungo D et al (2011) Combining neural network with fuzzy, certainty factor and likelihood ratio concepts for spatial prediction of landslides. *Nat Hazards* 59(3):1491
- Kazmi AH, Jan MQ (1997) *Geology and tectonics of Pakistan*. Graphic publishers
- Khan H et al (2019) Landslide susceptibility assessment using Frequency Ratio, a case study of northern Pakistan. *Egypt J Remote Sens Space Sci* 22(1):11–24
- Kim J-C et al (2018) Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang Korea. *Geocarto Int* 33(9):1000–1015
- Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980
- Krizhevsky A et al (2012) Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 25:1097–1105
- Krizhevsky A, Hinton G (2009) Learning multiple layers of features from tiny images
- Kuncheva L (2004) *Combining pattern classifiers methods and algorithms*. John Wiley and Sons. Hoboken
- Lawrence S et al (1997) Face recognition: a convolutional neural-network approach. *IEEE Trans Neural Networks* 8(1):98–113
- LeCun Y et al (2015) Deep Learn. *Nat* 521(7553):436–444
- LeCun Y et al (1995) Comparison of learning algorithms for handwritten digit recognition. In: International conference on artificial neural networks, Perth, Australia
- LeCun Y et al (1998) Gradient-based learning applied to document recognition. In: *Proceedings of the IEEE* 86(11): 2278–2324
- Lee M-J et al (2015) Forecasting and validation of landslide susceptibility using an integration of frequency ratio and neuro-fuzzy models: a case study of Seorak mountain area in Korea. *Environ Earth Sci* 74(1):413–429
- Li L et al (2017) A modified frequency ratio method for landslide susceptibility assessment. *Landslides* 14(2):727–741
- Liu Y et al (2019) Intelligent wind turbine blade icing detection using supervisory control and data acquisition data and ensemble deep learning *Energy Sci Eng* 7(6): 2633–2645
- Maas AL et al (2013) Rectifier nonlinearities improve neural network acoustic models. *Proc. icml*
- Mandal SP et al (2018b) Comparative evaluation of information value and frequency ratio in landslide susceptibility analysis along national highways of Sikkim Himalaya. *Spat Inform Res* 26(2):127–141
- Mandal S, Mandal K (2018a) Modeling and mapping landslide susceptibility zones using GIS based multivariate binary logistic regression (LR) model in the Rorachu river basin of eastern Sikkim Himalaya India. *Model Earth Syst Environ* 4(1):69–88
- Maqsoom A et al (2021) Landslide susceptibility mapping along the China Pakistan Economic Corridor (CPEC) route using multi-criteria decision-making method. *Model Earth Syst Environ*:1–15
- Martinović K et al (2016) Development of a landslide susceptibility assessment for a rail network. *Eng Geol* 215:1–9
- Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim et Biophys Acta (BBA)-Protein Struct* 405(2):442–451
- Mondal S, Mandal S (2019) Landslide susceptibility mapping of Darjeeling Himalaya, India using index of entropy (IOE) model. *Appl Geomatics* 11(2):129–146
- Nefeslioglu HA, Gorum T (2020) The use of landslide hazard maps to determine mitigation priorities in a dam reservoir and its protection area. *Land Use Policy* 91:104363
- Niu R et al (2014) Susceptibility assessment of landslides triggered by the Lushan earthquake, April 20, 2013, China. *IEEE J Sel Top Appl Earth Observations Remote Sens* 7(9):3979–3992
- Owen LA et al (2008) Landslides triggered by the October 8 2005 Kashmir earthquake. *Geomorphology* 94(1–2):1–9
- O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. *Qual Quant* 41(5):673–690
- Pamela P et al (2018) Weights of evidence method for landslide susceptibility mapping in Takengon, Central Aceh, Indonesia. In: *Proceedings of the IOP conference series: Earth & Environmental Science*, Prague, Czech Republic
- Pavelsky TM, Smith LC (2008) RivWidth: A software tool for the calculation of river widths from remotely sensed imagery. *IEEE Geosci Remote Sens Lett* 5(1):70–73

- Pham BT et al (2016) “Evaluation of predictive ability of support vector machines and naive Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS. *J Geomat* 10:71–79
- Pham BT et al (2017a) A novel ensemble classifier of rotation forest and Naïve Bayer for landslide susceptibility assessment at the Luc Yen district, Yen Bai Province (Viet Nam) using GIS. *Geomatics. Nat Hazards Risk* 8(2):649–671
- Pham BT et al (2017b) Landslide susceptibility assessment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. *Theoret Appl Climatol* 128(1–2):255–273
- Pham BT et al (2017c) “Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena* 149:52–63
- Pham BT et al (2018) A hybrid machine learning ensemble approach based on a radial basis function neural network and rotation forest for landslide susceptibility modeling: a case study in the Himalayan area, India. *Int J Sedim Res* 33(2):157–170
- Pham BT et al (2019b) A novel intelligence approach of a sequential minimal optimization-based support vector machine for landslide susceptibility mapping. *Sustainability* 11(22):6323
- Pham BT, Prakash I (2019a) Evaluation and comparison of LogitBoost Ensemble, Fisher’s Linear Discriminant Analysis, logistic regression and support vector machines methods for landslide susceptibility mapping. *Geocarto Int* 34(3):316–333
- Polykretis C, Chalkias C (2018) Comparison and evaluation of landslide susceptibility maps obtained from weight of evidence, logistic regression, and artificial neural network models. *Nat Hazards* 93(1):249–274
- Ramachandra T et al (2013) Prediction of shallow landslide prone regions in undulating terrains. *Disaster Adv* 6(1):54–64
- Ronoud S, Asadi S (2019) An evolutionary deep belief network extreme learning-based for breast cancer diagnosis. *Soft Comput* 23(24):13139–13159
- Rosi A et al (2018) The new landslide inventory of Tuscany (Italy) updated with PS-InSAR: geomorphological features and landslide distribution. *Landslides* 15(1):5–19
- Saba SB et al (2010a) Spatiotemporal landslide detection for the 2005 Kashmir earthquake region. *Geomorphology* 124(1–2):17–25
- Saba SB et al (2010b) Spatiotemporal landslide detection for the 2005 Kashmir earthquake region. *Geomorphology* 124(1–2):17–25
- Sevgen E et al (2019a) A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression. ANN and random forest. *Sensors* 19(18):3940
- Sevgen E et al (2019b) A novel performance assessment approach using photogrammetric techniques for landslide susceptibility mapping with logistic regression ANN and random forest. *Sensors* 19(18):3940
- Shafique M et al (2016) A review of the 2005 Kashmir earthquake-induced landslides; from a remote sensing prospective. *J Asian Earth Sci* 118:68–80
- Sharif Razavian A et al (2014) CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops
- Shin H-C et al (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 35(5):1285–1298
- Srivastava N et al (2014) Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 15(1):1929–1958
- Srivastava RK et al (2015) “Training very deep networks.” arXiv preprint arXiv:1507.06228
- Szegedy C et al (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition
- Tallarida RJ, Murray RB (1987) Chi-square test. *Manual of pharmacologic calculations*. Springer, pp 140–142
- Tanyas H et al (2019) A global slope unit-based method for the near real-time prediction of earthquake-induced landslides. *Geomorphology* 327:126–146
- Tien Bui D et al (2019) New ensemble models for shallow landslide susceptibility modeling in a semi-arid watershed. *Forests* 10(9):743
- Torizín J et al (2017) Statistical landslide susceptibility assessment of the Mansehra and Torghar districts, Khyber Pakhtunkhwa Province Pakistan. *Natural hazards* 89(2):757–784
- Tsangaratos P et al (2017) Applying information theory and GIS-based quantitative methods to produce landslide susceptibility maps in Nancheng County China. *Landslides* 14(3):1091–1111

- Tsangaratos P, Iliu I (2016) Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: the influence of models complexity and training dataset size. *CATENA* 145:164–179
- Wang L-J et al (2016a) A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. *Geosci J* 20(1):117–136
- Wang Q et al (2016b) Application of statistical index and index of entropy methods to landslide susceptibility assessment in Gongliu (Xinjiang, China). *Environ Earth Sci* 75(7):599
- Wang Q et al (2017) Integration of information theory, K-means cluster analysis and the logistic regression model for landslide susceptibility mapping in the Three Gorges Area China. *Remote Sensing* 9(9):938
- Wang Y et al (2019) Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County China. *Sci Total Environ* 666:975–993
- Wang G et al (2020a) Spatial prediction of landslide susceptibility based on GIS and discriminant functions. *ISPRS Int J Geo-Information* 9(3):144
- Wang Y et al (2020b) Comparative study of landslide susceptibility mapping with different recurrent neural networks. *Comput Geosci* 138:104445
- Wu Y, Li W, Wang Q, Liu Q, Yang D, Xing M et al (2016) Landslide susceptibility assessment using frequency ratio, statistical index and certainty factor models for the Gangu County, China. *Arab J Geosci* 9(2):1–16
- Yi Y et al (2020) Landslide susceptibility mapping using multiscale sampling strategy and convolutional neural network: a case study in Jiuzhaigou region. *Catena* 195:104851
- Youssef AM et al (2015) Landslide susceptibility mapping at Al-Hasher area, Jizan (Saudi Arabia) using GIS-based frequency ratio and index of entropy models. *Geosci J* 19(1):113–134
- Youssef AM, Pourghasemi HR (2021) “Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. *Geosci Front* 12(2):639–655
- Yu L et al (2019) Landslide susceptibility mapping combining information gain ratio and support vector machines: a case study from Wushan segment in the three gorges reservoir area China. *Appl Sci* 9(22):4756
- Zhao X, Chen W (2020) GIS-based evaluation of landslide susceptibility models using certainty factors and functional trees-based ensemble techniques. *Appl Sci* 10(1):16
- Zhu A-X et al (2018) A comparative study of an expert knowledge-based model and two data-driven models for landslide susceptibility mapping. *CATENA* 166:317–327

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Authors and Affiliations

Bilal Aslam<sup>1,3</sup>  · Adeel Zafar<sup>1</sup> · Umer Khalil<sup>2</sup>

Bilal Aslam  
ba924@nau.edu

<sup>1</sup> Department of Cyber Security and Data Science, Riphah International University, Islamabad Campus, Islamabad, Pakistan

<sup>2</sup> ITC Faculty of Geo-information Science and Earth Observation, University of Twente, Enschede, The Netherlands

<sup>3</sup> School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, United States