Landslide risk zoning in Ruijin, Jiangxi, China

Xiaoting Zhou [1], Weicheng Wu [1*], Ziyu Lin [1], Guiliang Zhang [2], Renxiang Chen [2], Yong Song [2], Zhiling Wang [2], Tao Lang [2], Yaozu Qin [1], Penghui Ou [1], Wenchao Huangfu [1], Yang Zhang [1], Lifeng Xie [1], Xiaolan Huang [1], Xiao Fu [1], Jie Li [1], Jingheng Jiang [1], Ming Zhang [1], Yixuan Liu [1], Shanling Peng [1], Chongjian Shao [1], Yonghui Bai [1], Xiaofeng Zhang [3], Xiangtong Liu [4] and Wenheng Liu [1]

[1] Key Laboratory of Digital Lands and Resources and Faculty of Earth Sciences, East China University of Technology, Nanchang, 330013 Jiangxi, China

[2] 264 Geological Team of Jiangxi Nuclear Industry, Ganzhou, Jiangxi, China

[3] School of Geophysics and Measurement-Control Technology, East China University of Technology, Nanchang, 330013 Jiangxi, China

[4] Faculty of Geomatics, East China University of Technology, Nanchang, 330013 Jiangxi, China

* Correspondence: Wuwc030903@sina.com/ wuwch@ecut.edu.cn; Tel.: +86-18970847278

Abstract: Landslides are one of the major geohazards threatening human society. This study was aimed at conducting such a hazard risk prediction and zoning based on an efficient machine learning approach, Random Forest (RF), for Ruijin, Jiangxi, China. Multiple geospatial and geo-environmental data such as land cover, NDVI, landform, rainfall, stratigraphic lithology, proximity to faults, to roads and to rivers, depth of the weathered crust, etc., were utilized in this research. After pre-processing, including digitization, linear feature buffering and value assignment, 19 hazard-causative factors were eventually produced and converted into raster to constitute a 19-band geo-environmental dataset. 155 observed landslides that had truly taken places in the past 10 years were utilized to establish a vector layer. 70% of the disaster sites (points) were randomly selected to compose a training set (TS) and the remained 30% to form a validation set (VS). A number of non-risk samples were identified in low slope (< 1-3°) areas and also added to the TS and VS in the similar percentage. Then, RF-based classification algorithm was employed to model the probability of landslide occurrence using the above 19-band dataset as predictive variables and TS for training. After performance evaluation, the RF-based model was applied back to the integrated dataset to calculate the probability of the hazard occurrence in the whole study area. The predicted map was evaluated versus both TS and VS and found of high reliability in which the Overall Accuracy (OA) and Kappa Coefficient (KC) are 91.49% and 0.8299 respectively. In terms of the risk probability, the predicted map was further zoned into different risk grades to constitute landslide risk map. Modeling results also revealed the order of importance of the 19 causative factors, and the most important ones are the proximity to roads, slope, May-July rainfall, NDVI and elevation. We hence conclude that the RF algorithm is able to achieve the risk prediction with high accuracy and reliability, and this study may provide an operational methodology for geohazard risk mapping and assessment. The results of this study can serve as reference for the local authorities in prevention and early warning of landslide hazard.

Keywords: Landslide; Hazard factor quantification; Random Forest; Risk zoning

1. Introduction

Landslides are frequent natural disasters, which pose a serious threat to traffic, property and safety of people (Wu and Ai, 1995; Nadim et al., 2006; Assilzadeh et al., 2010; Froude and Petley, 2018). Petley (2012) stated that the damage caused by non-seismic landslides is severe around the

43 world, especially in Asia, and we used to underestimate the toll caused by this kind of disaster.
44 Ruijin County in Jiangxi, South China is such an area where landslide calamity constitutes a serious
45 threat and problem to human society. According to the 264-Geological Team (of the Geological
46 Bureau of Jiangxi Nuclear Industry), landslides have caused damage to 104 residential buildings,
47 and made 2319 people homeless in the past decades. Affected by landslide disaster, the construction
48 and use of the newly-built campus of No. 6 Middle School in Ruijin County have been suspended.
49 The uncertainty and suddenness of disaster constitute potential threats to human daily life (Nadim et
50 al., 2006; Froude and Petley, 2018). The recognition of potential landslide-prone areas acts as an
51 essential part in hazard early warning and aiding decision-makers in land use planning and resource
52 management, as well as reducing losses caused by disasters (Aleotti and Chowdhury, 1999; Wu et
53 al., 2016; Arabameri et al., 2020).

54   In the past decades, a number of studies about landslide risk prediction and assessment have
55 been conducted (Montgomery and Dietrich, 1994; Guzzetti et al., 1999; Aleotti and Chowdhury,
56 1999; Ayalew and Yamagishi, 2005; Ruffff and Czurda, 2008; Arabameri et al., 2017). These
57 studies have proposed a variety of landslide risk prediction and assessment methods, e.g.,
58 knowledge-based, physical, and data-driven methods (Corominas et al., 2014; Reichenbach et al.,
59 2018; Li et al., 2017). Actually, advantages and limitations exist in each approach, for example,
60 knowledge-based and physical methods are mostly intuitive but qualitative or half quantitative
61 (Corominas et al., 2014; Goetz et al., 2015; Li et al., 2017), while data-driven methods are
62 quantitative, yet require powerful computing capacity for big data processing. On the whole,
63 data-driven methods seem more promising for a higher prediction accuracy than other methods, and
64 thus shall be more suitable for landslide risk assessment in areas where there is insufficient
65 geotechnical data (Guzzetti et al., 1999; Corominas et al., 2014; Furlani and Ninfo, 2015; Li et al.,
66 2017; Zhu et al., 2019).

67   Owing to the heterogeneity in geological and environmental conditions, the scale and
68 mechanism of landslides may differ from one place to another (Cao et al., 2019). This makes the
69 hazard prediction complicated and requires a consideration of the hazard-causative factors as many
70 as possible while dealing with such risk assessment. Recently, remote sensing (RS) and Geographic
71 Information System (GIS) have been taking an significant part in the study of disaster risk zoning
72 (Grizer et al., 2001; Wu et al., 2003; Wang et al., 2005; Lai et al., 2019; Chang et al., 2020). RS
73 technique can provide not only multitemporal and time-series spatial information of large and even
74 inaccessible areas over span of decades but also timely pre- or post-hazard spatial data easily
75 (Youssef et al., 2009; Wasowski et al., 2015; Arabameri et al., 2020). Therefore, RS is an effective
76 tool for hazard monitoring and assessment. To be precise, satellite images can provide the
77 environmental factor layers (e.g., topography, land cover and anthropogenic activities) which can be
78 used for landslide risk prediction and modeling (Pradhan et al., 2010; Arabameri et al., 2020). The
79 other main intrinsic geological and meteorological hazard-inducing factors are also fundamental and
80 essential for this purpose.

81   In the past years, machine learning techniques including artificial intelligence and deep
82 learning have gained a momentum in geospatial big data processing. For example, data-driven
83 algorithms such as Support Vector Machines (SVM), Random Forests (RF), and Artificial Neural
84 Networks (ANN) have been well applied in land resource mapping (Wu et al., 2016), prediction of
85 soil salinity (Wu et al., 2018) and of ore mineralization in geological field, and shown a superior

86 performance to the traditional approaches (Huang, 2018; Qin et al., 2018; Achour and Pourghasemi,
87 2019; Dou et al., 2019; Sameen et al., 2020). Comparing with other machine learning approaches,
88 the RF algorithm has clear advantages, i.e., it does not require to normalize and discretize the data,
89 and is less sensitive to outliers and runs faster than SVM (Breiman, 2001; Wu et al., 2016; Zhang et
90 al., 2017). Landslide causative factors often present nonlinear relationships (Corominas et al., 2014;
91 Zhu et al., 2019). RF algorithm is able to catch such nonlinear features among the factors and also
92 to prevent overfitting (Breiman, 2001; Goetz et al., 2015; Arabameri et al., 2020).

93   In view of the reliable prediction result obtained from regression and classification with the RF
94 algorithm (Wu et al., 2016; Wu et al., 2018), the objective of this study is to employ this algorithm,
95 one of the data-driven methods, to model the landslide risk taking Ruijin County as an example. As
96 RF algorithm has been rarely applied to landslide study, one specific objective of this research is to
97 find out an operational RF-based approach for this kind of geohazard zoning and mapping.

98 **2. Data and Methods**

99 *2.1 The study area*

100   Ruijin County is located in the southeast of Jiangxi Province, China, extending from 115° 41'
101 10" to 116° 21' 49" E in longitude and from 25° 32 '15" to 26° 17' 45" N in latitude, covering an
102 area of about 2436 km^2 (Fig. 1). From the view of topography, the elevation of the study area varies
103 from 70 to 1211 m with a mean of 324 m while the slope from 0 to 65° with an average of 15°.
104 Hydrologically, the main rivers are Meijiang, Mianjiang and Jiubao Rivers running through the
105 study area as sub-tributaries of the Gongshui River watershed. The study area belongs to the
106 subtropical humid climate zone and is characterized by four distinct seasons, sufficient rainfall and
107 long frost-free period. Heavy rainfall often occurs from March to July, accounting for more than
108 50 % of the annual rainfall with amount of about 1663.5 mm, an average of the period from 1968 to
109 2017. The annual mean temperature is 21.54 ℃ and July is the hottest month of the year with a
110 mean temperature of 28.8 ℃.

111   The hot and humid weather leads to severe weathering of rock mass giving rise to formation of
112 a thick weathered crust in which most landslides take places. From the human side, artificial cutting
113 on slope for infrastructure construction (such as roads and highways) and housing development
114 provoke instability of the crust mass causing landslides.
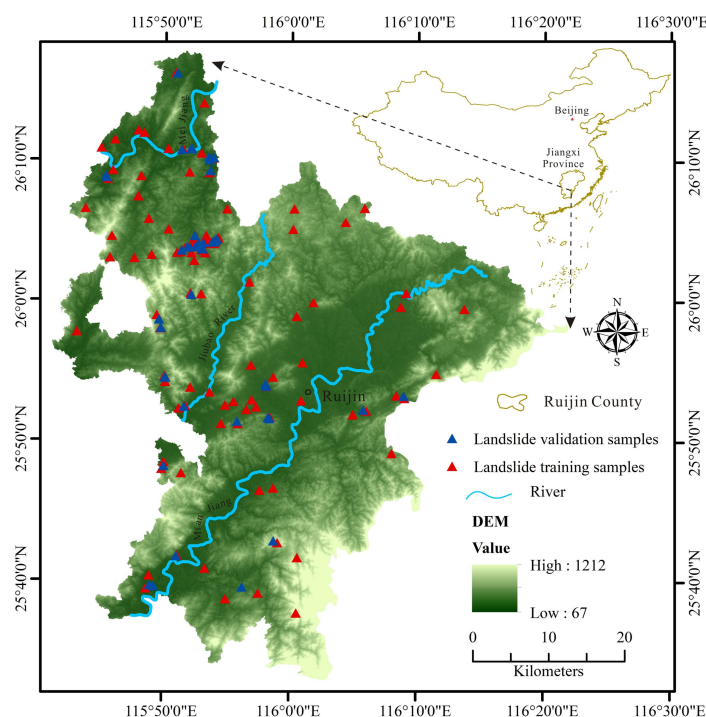
Fig. 1 Location of the study area, Ruijin County, Jiangxi, China and location of the training and validation
sites of landslides in the study area.

2.2 Data and Processing Procedures

For landslide risk assessment, it is unavoidable to deal with non-digital geo-environmental data
such as geological strata, faults, rivers as they are essential for this purpose. It is hence necessary to
convert them into digital and meaningful values so that they can be incorporated as quantitative
variables for landslide risk modeling. The global methodological procedure includes data
pre-processing, digitization, linear features buffering, rational numeric value assignment to
descriptive factors and buffers, risk modeling and validation, and finally accuracy assessment.
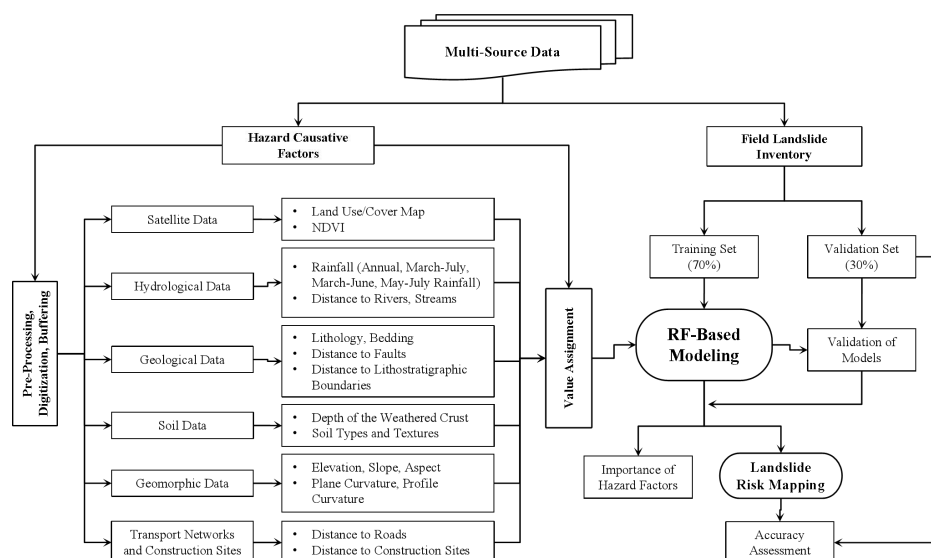These procedures are presented in a flowchart shown in Fig. 2.

Natural Hazards
and Earth System
Sciences
Discussions
Open Access
EGU

Fig. 2 Methodological flowchart.

125

126   2.2.1 Geospatial and geo-environmental data

127   2.2.1.1 Satellite data

128   1) Landsat imagery: Landsat 5 TM images of October and November 2006-2010, and Landsat

129 8 OLI images dated May 2017 and Sept 2019 were obtained from the USGS data server

130 (https://glovis.usgs.gov). After atmospheric correction using COST model (Chavez, 1996; Wu,

131 2003 and Wu et al., 2013), Landsat 8 images were used for land cover mapping using the approach

132 proposed by Wu et al. (2016) and Landsat 5 data for deriving the averaged multiyear autumn NDVI

133 (Fig. 3a). In general, vegetation can help soil hold water content and improve its mechanical

134 properties through root system to stabilize slopes. Thus, landslides may arise more likely in

135 unvegetated areas rather than in vegetated ones (Montgomery et al., 2000; Reichenbach et al., 2014).

136 However, this will be completely different when slope is cut or excavated because of road

137 construction or housing development.

138   2) Very high resolution images: those are available on Google Earth as a complementary

139 source of ground-truth data. The road and river networks were also extracted from Google Earth

140 (Fig. 3b, 4d).

141   According to the principle of RF algorithm, we shall use two types of samples for modeling as

142 input variables: one is the locality of landslides that have taken places and the other is the stable

143 areas where landslides are unlikely to occur (Frattini et al., 2010; Depicker et al., 2020; Arabameri,

144 2020). Identified on Google Earth, the stable areas are places where slope is less than five degree,

145 e.g., water bodies, urban areas, and cultivated land. Landslides with an area greater than 900 m^2 (1

146 Landsat pixel) that had been overlooked during the field observation were also identified and

147 delineated on Google Earth.

148   2.2.1.2 Hydrological data

149   1) Rainfall: Rainfall is the main factor triggering landslides (Monsieurs et al., 2018; Depicker

150 et al., 2020). Depicker et al. (2020) stated that rainfall condition was the direct cause of many

151 shallow landslides. Daily rainfall data from Jan 2008 to Dec 2013 were obtained from 40
152 meteorological stations in Ruijin County and its adjacent areas. As landslides mainly occurred in
153 March to July, especially in June and July but without recorded occurrence time, our intention was
154 to investigate which months of rainfall may best reveal its role in landslide events. Thus, apart from
155 the mean annual rainfall, March-June rainfall, May-July rainfall and March-July rainfall of these six
156 years were also aggregated and gridded into raster with 30 m pixel size using Inverse Distance
157 Weighting (IDW) approach.
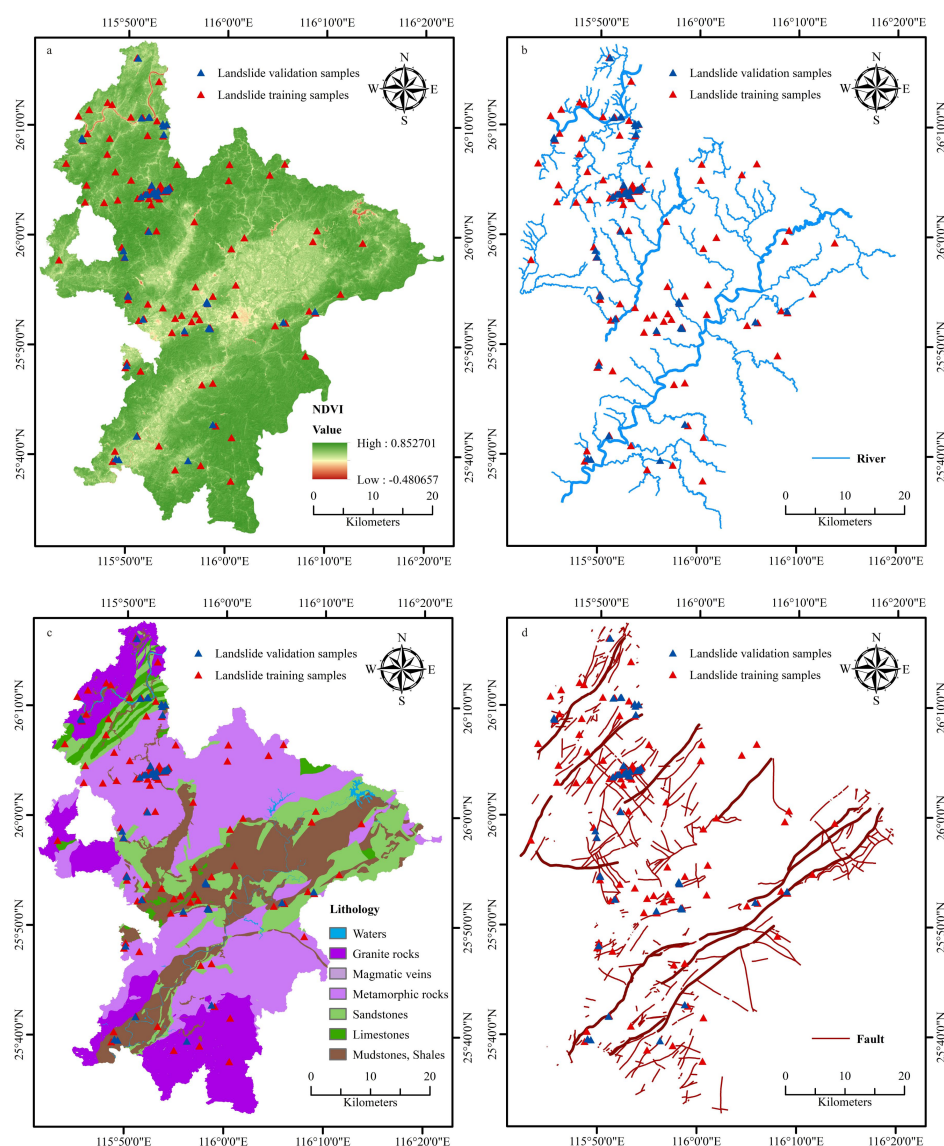
158 **Fig. 3** Landslide hazard factors: (a) NDVI, (b) Rivers, (c) Lithology, and (d) Faults.

Natural Hazards
and Earth System
Sciences
Discussions

159   2) River network: The influence of rivers on the occurrence of landslides is reflected by the
160   proximity to, or rather, the distance to them (Chen et al., 2018a; Cao et al., 2019; Arabameri et al.,
161   2019). Thus, the rivers were vectorized from Google Earth (Fig. 3b) and buffered into belts with an
162   interval of 30, 60, 90, 120 and 150 m, respectively, for streams and 60, 120, 180, 240 and 300 m
163   respectively for the main rivers. Then, these buffers were assigned values in terms of their
164   propensity or their importance in the event of landslide. For example, for the main rivers buffers of
165   0-60, 60-120, 120-180, 180-240 and 240-300 m were respectively assigned with 20, 15, 10, 5 and 1
166   while for streams, buffer zones of 0-30, 30-60, 60-90, 90-120 and 120-150 m respectively with 10,
167   8, 6 4 and 1. This implies that the closer to the river the higher propensity of landslide.

168   Finally, these buffers are converted to raster data with 30 m cell size using the "Polygon to
169   Raster" tool as proposed by Wu et al. (2018).

170   2.2.1.3 Geological and geomorphic data

171   1) Geological strata and formations: Geological strata were extracted from 1/50,000
172   Geological Map. Except for Ordovician, Silurian, Triassic and Tertiary, the strata of other
173   geological periods are mostly exposed. In terms of texture and composition, the lithology of
174   different strata in the study area can be divided into 113 classes. To facilitate the geohazard analysis,
175   these lithological classes were further aggregated into six main categories: (1) granitic rocks, (2)
176   magmatic veins, (3) metamorphic rocks, (4) sandstone, (5) limestone, (6) mudstone and shales as
177   shown in Fig. 3c. Just based on lithology and in absence of faults and joints, granitic massif would
178   possess the highest resistance while mudstone the lowest to landslides. Hence, from (1) to (6), the
179   propensity is likely to increase and they were respectively assigned values 1, 2, 3, 5, 7 and 10.

180   According to the field observation, landslide events occurred frequently on the boundaries
181   between two formations, especially, between the Quaternary sediments and other strata. Therefore,
182   the lithostratigraphic boundary factor was also obtained by buffering and rasterization, then added
183   to analyze the landslide risk.

184   2) Faults: This kind of geological structure has a prominent effect on the stability of rock mass
185   (Smets et al, 2016; Delvaux et al., 2017). There formed the spectacular thrust nappe structure which
186   was characterized by strong faulting activity in the study area. Such structure is accompanied with a
187   series of faults and folds, which tend to be the landslide-prone areas, e.g., the fragile belts related to
188   fold hinges, fracture zones and joints. As a matter of fact, the proximity to fault plays a role in such
189   hazard events, i.e., the closer to the fault, the higher risk of landslide may exist. For this reason, the
190   faults in the study area (Fig. 3d) were divided into three groups in terms of scale, i.e., big faults if
191   their length is > 10-20 km, medium faults if they are 2-10 km, and small faults if they are < 2 km.
192   The big faults were buffered into five zones of 0-120 m, 120-240 m, 240-360 m, 360-480 m and
193   480-600 m, and were respectively assigned with values of 20, 15, 10, 5 and 1. For the medium ones,
194   they were also buffered into five zones of 0-60 m, 160-120 m, 120-180 m, 180-240 m and 240-300
195   m with assigned values of 10, 8, 6, 4 and 1. The small faults were again buffered into five zones of
196   0-30 m, 30-60 m, 60-90 m, 90-120 m and 120-150 m and respectively assigned with 5, 4, 3, 2 and 1.
197   These fault buffers were gridded into raster layer of 30 m in resolution.

198   3) Depth of the weathered crust, soil type and texture: Weathering is the process converting
199   rocks into soils to constitute the weathered crust of our land surface. Landslides mostly take place in
200   this crust in which soil texture seems of significant impact on (Kitutu et al., 2009) and the
201   variability of soil types and depths of the crust play a part in the occurrence of such events (Fan et

202   al., 2016). Because different soil types and textures have different sand percentage, grain sizes and
203   porosity affecting the permeation of rain water. If liquidized by penetrated water, the crust bottom
204   (soil/rock interface) may serve as slip surface of landslides as friction and resistance from the
205   underlying rocks are reduced by this process. As soon as it has reached certain threshold, landslide
206   occurs. Thus, the crust depth, i.e., the depth of the slippery soil/rock interface, is a plausible
207   indicator of landslide volume and scale.

208   Here the data of soil types were obtained from the Bureau of Jiangxi Coal Geology and the
209   sand percentage (%), in which high sand percentage (low percentage of clay but high porosity)
210   seems favorable for permeation of rain water and provoking landslide event, was considered as an
211   indicator of soil contribution. Hence, soils with sand percentage > 40%, 30-40%, 20-30%, 10-12%,
212   5-10% and 0-5%, were respectively assigned with values of 10, 8, 6, 4, 2 and 1. At last soil
213   proneness map was converted into raster of 30 m resolution.

214   The depth data of the weathered crust were obtained from the field 1282 measurements. In
215   assumption that all the ridges have a crust of 0.5m in depth, these field observed depths were
216   interpolated using Kriging approach, then converted into raster layer of 30m resolution.

217   4) Geomorphic data: Slope (angle) is a key driver of landslides and a triggering angle threshold
218   of 28° - 38° was reported by Fan et al.(2016); at the same time, elevation, aspect, plane curvature
219   and profile curvature may also contribute to the occurrence of the hazards (Corominas et al., 2014;
220   Guzzetti et al., 2005; Galli et al., 2008; Pourghasemi and Kerle, 2016; Cao et al., 2019). The
221   ASTER GDEM data with a spatial resolution of 30 m, jointly developed by METI of Japan and
222   NASA of the United States, were obtained for Ruijin County from the Geospatial Data Cloud
223   (http://www.gscloud.cn/) and used to derive elevation, slope, aspect, plane and profile curvatures
224   (Fig. 1, Fig. 4a, 4b).

225   2.2.1.4 Land use/cover, transport system and construction sites

226   Using the classification approach proposed by Wu et al. (2016), land cover mapping was
227   achieved for Ruijin with an accuracy of 90.99%. The main land cover type is forests (54.25%),
228   followed by shrub/woodlands (29.33%), croplands (6.65%), artificial areas (urban, villages, road
229   and other infrastructures, 5.36%), barelands (1.45 %) and waters (1.41%) (Fig. 4c). Forests cover
230   hills and mountains, artificial areas and croplands are mainly distributed in lowlands (valleys) with
231   low slope. For risk modeling purpose, forest cover was considered of low proneness and assigned a
232   value of 1-2. On the contrary, unvegetated hilly slopes and barelands were regarded of high
233   propensity and assigned a value of 10, while zero-slope croplands, urban and water-bodies were
234   treated as non-risk (zero probability) areas. At the same time, NDVI can be used as an indicator of
235   vegetation greenness and abundance, indirectly representing the development degree of the root
236   system. For barelands, woodlands and forests, NDVI shall be a good indicator of propensity of
237   landslide.

238   Road construction is one of the important human activities leading to slope failure
239   (García-Rodríguez et al., 2008; Cao et al., 2019). Similarly, housing development along two sides
240   of the roads or on the brink of hills by cutting slopes constitutes also an important factor making the
241   slope massif unstable. The influence of road on landslide is also reflected by the distance to them
242   (Chen et al., 2018a; Cao et al., 2019; Arabameri et al., 2020). The road system (Fig. 4d) was also
243   assigned values the same as was done for rivers and faults.
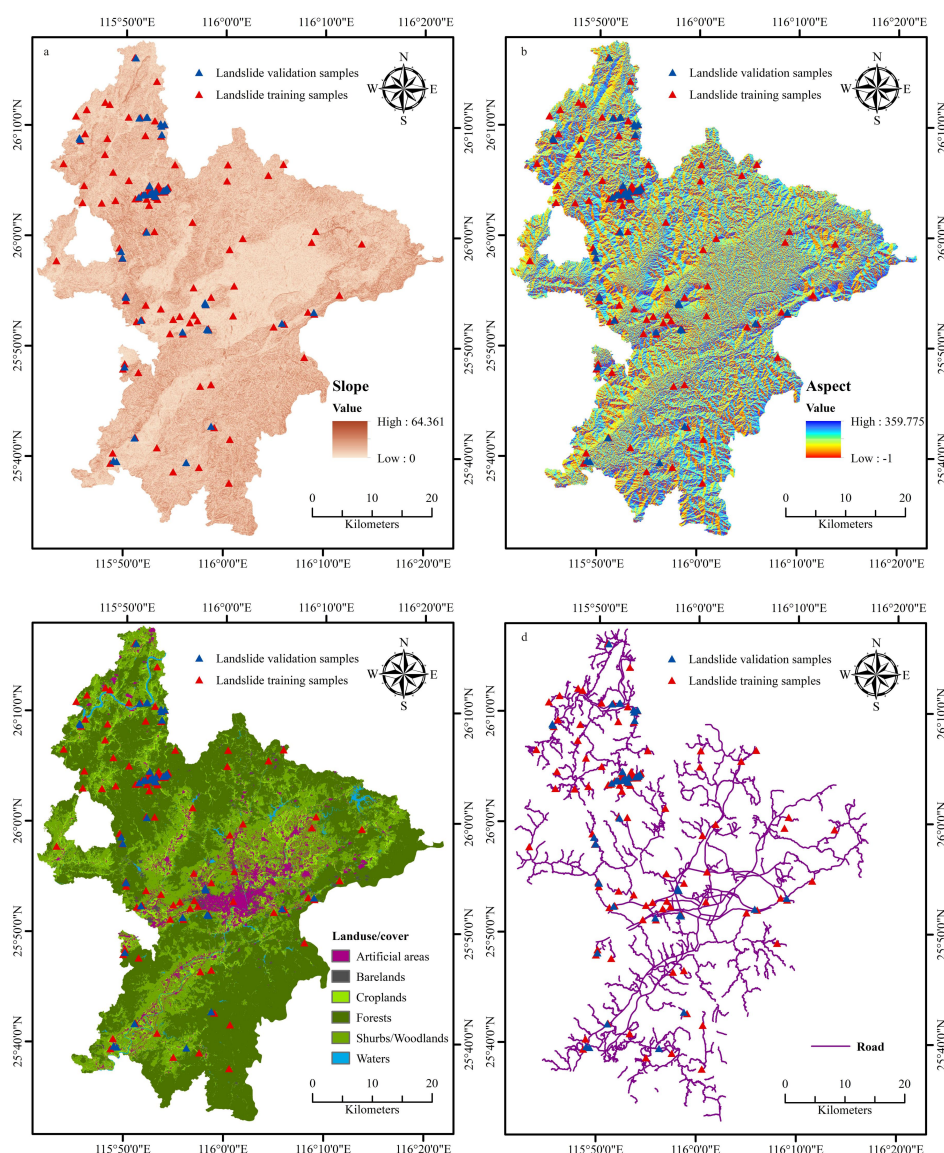
Fig. 4 Landslide hazard factors: (a) Slope, (b) Aspect, (c) Land use/cover, and (d) Roads.

There were no landslide accidents recorded in the study area caused by earthquakes, so the earthquake factor was not considered in this study.

2.2.1.5 Integrated geo-environmental dataset

The occurrence of landslide is a result of the combined action of all the causative geo-environmental factors (Corominas et al., 2014; Zhu, et al., 2019). In this study, all these factors that may contribute to the occurrence of landslide will be considered for risk modeling. The above processed raster layers namely geological strata, proximity to faults, to lithostratigraphic boundaries, to roads and to to rivers, depth of the weathered crust, soil types and texture, elevation, slope, aspect,

9

plane curvature, profile curvature, land use/cover, NDVI, multiyear mean annual rainfall, multiyear mean May-July rainfall, multiyear mean March-June rainfall, multiyear mean March-July rainfall were incorporated into a 19-band dataset in Datum WGS 84 and Projection UTM 50 by Layer Stacking function. These raster layers were considered as hazard-causative factors or independent hazard predictors.

2.2.2 Field data

First-hand data were obtained from the field observation during the Geological Hazard Survey Campaign in Ruijin County on a scale of 1/50,000 by the 264 Geological Team of Jiangxi Nuclear Industry in the recent years. The realistic landslides that had occurred in the past ten years were investigated and recorded. Some landslides had been missed during the field observation were digitized on Google Earth as a complement to the former. In total, 155 historical landslide cases were collected. The landslide sites ranged from 20 m^2 to 64000 m^2 in size. Most of the landslides are small in scale, i.e., less than 900 m^2 in study area. To obtain the optimal spatial presentation of samples for RF modeling, the landslide points with areas less than 900 m^2 were buffered with radius of 30 m and then rasterized into pixels with size of 30 m (Wu et al., 2018), and those larger than 900 m^2, a direct rasterization from vector was conducted. These cases were assigned a value of 1, indicating that the events of landslide have truly taken places, i.e., the probability is 1.

As mentioned above, the non-risk areas (low-slope (< 1-5°) valleys, plain, urban and water-bodies) were integrated into the field dataset as zero-risk area, i.e., the risk probability is 0.

At last, we randomly selected 70 % of the landslide samples (109 cases) plus 70 % of the non-risk zones to constitute a training set (TS) and used the remained ones (45 cases, 30 %) as a validation set (VS).

2.3 Landslide risk modeling

Among the machine learning algorithms, Wu et al (2016 and 2018) found that RF and support vector machines (SVM) performed equally well in classification, better than artificial neural networks (ANN), but RF performed best in regression prediction. Hence, RF classification algorithm was selected for geohazard risk modeling in this study. The overall procedure was summarized in Fig. 2 and the detail on modeling, validation and accuracy assessment is given in the following sections.

2.3.1 RF classification of risk probability

RF classification, based on growing decision trees, is an ensemble of tree classifiers that allow to conduct classification by predicting the probability of a given pixel to be classified into the target class through majority voting. The key technique of this algorithm lies in that a bootstrap sample of the TS is used to build each tree, and a stochastic selection of the input variables is searched to determine the best split for each node. Meanwhile, RF algorithm uses out-of-bag (OOB) estimates to define the generalization error and the importance of each variable (Breiman, 2001). RF will not overfit if the number of decision trees (NT) increases to certain level. Thence, NT should be large enough to reduce the OOB error of classification to a stable level in the training process. It is to mention here that instead of classification of land cover types, we employ this algorithm to classify the probability of risk and non-risk for each pixel.

10

2.3.2 Application of RF algorithm

In this study, RF classification was conducted within EnMap-Box which is a package particularly developed to process and analyze image data (Waske et al., 2012). While conducting RF modeling, we kept the combined 19-band dataset as input variables with 109 landslides used as TS and 45 as VS. Some key parameters of RF classification include the Impurity Function, the Stop Criteria (for node splitting), the Number of randomly selected Features (or Number of Variables) at each node and Number of Trees (NT) with classification and regression algorithm (Wu et al., 2016, 2018).

The Gini Coefficient was selected for the Impurity Function. The Stop Criteria was set as the default values which was a Minimum number of samples at a node of 1 or a Minimum impurity calculated based on the Gini index of 0. The Number of randomly selected Features (or Number of Variables) at each node was the square root of all available features. The default value of number of trees was 100 within EnMap-Box. In this study, NT was set to 300 and 500 in order to achieve a better prediction.

After parameterization of RF classification model using the integrated dataset as predictive variables and TS for training and internal validation, the model was applied back to the integrated dataset for landslide prediction, i.e., the probability of landslide occurrence in each pixel. The accuracy of modeling is calculated versus the independent VS.

2.3.3 Importance of variables

The importance of variables in the RF classification can be evaluated by variable substitution method. In other words, it can be measured by calculating the difference of the OOB error before and after value substitution. Factor importance of F_i can be expressed as follows:

$$VIM\ (F_i) = \frac{1}{NT} \sum_t errOOB_t^i - errOOB_t \qquad (1)$$

where NT is the number of trees, $errOOB_t$ is an error for tree t of the forests when all the factors are included, $errOOB^i$ refers to an error after removing the factor F_i and $VIM(F_i)$ is variable importance for F_i. For RF classification and its result produced, a high value indicates that the factor is of high importance and vice versa.

2.3.4 Validation of models

Base on the Confusion Matrix, Precision, Recall, F1 index, Kappa Coefficient (KC) and Overall Accuracy (OA) can be calculated to evaluate the accuracy and performance of landslide risk prediction model (Congalton, 1983).

Generally, both TS and VS can be used to calculate these statistical indices. The evaluation results of TS show the adaptability of the model to the training data, while those of VS reveal the predictive and generalization ability of the model (Tien Bui et al., 2012).

According to previous studies, the smaller the high-risk area predicted by the model, the more historical landslide points concentrated there, which indicates that the model has high predictivity (Cao et al., 2019; Dou et al., 2019). The success rate curve and prediction rate curve can be made respectively by using the landslide risk prediction results of TS and VS. For the study area, the prediction accuracy of landslide risk map can be analyzed and demonstrated by the area under curve (AUC) features (Chung et al., 2003; Yilmaz, 2009; Nicu et al., 2019).

3. Results and Discussion

3.1 Landslide risk zoning

The landslide risk zoning was achieved based on the modeled risk probability when RF modeling was implemented at NT = 300. All pixels were divided into four levels: low-risk (probability < 0.5), medium-risk (probability ≥ 0.5 and < 0.7), high-risk (probability ≥ 0.7 and < 0.9) and extremely high-risk (probability ≥ 0.9). The landslide risk map of Ruijin was hence produced (Fig. 5a).

The predicted results of the landslide-prone areas were largely consistent with the field survey: (1) High-risk zones were mainly linearly distributed along rivers and roads (Fig. 5b). Many landslides were often caused by river undercutting and artificial cutting for road construction and housing. (2) In the central part of the study area, high-risk zones are concentrated in the Quaternary soil layer, or rather, in the weathered crust, especially along the boundaries of lithologic strata (Fig. 5c). The Quaternary unconsolidated soil layer with loose structure provided rich material for landslides. The boundaries of lithologic strata behaved as unstable structural interfaces, which were important factors for landslides. (3) In the granitic massif, there were also obvious high-risk zones distributed along the roads (Fig. 5b, 5c). Weathering accelerated by humidity, high undulating landform and tectonically active settings of the study area, changed the intrinsic properties of the material and reduced the strength of the near-surface rocks.

Table 1 shows that 24.83 km^2 of areas identified as Extremely high risk zones account for 1.02% of the total study area, and High, Medium and Low risk zones take up 227.57 km^2 (9.32%), 472.39 km^2 (19.36%) and 1715.60 km^2 (70.30 %) respectively. Additionally, 94.19% of the field samples, i.e., the realistic landslides, took places in 10.34% of the entire study area, which was categorized as High-risk and Extremely-high risk zones in our risk zone map.

3.2 Number of Trees with RF classification

The selection of NT has a great influence on the accuracy of RF modeling. The performance of classification or regression is poor and the error is large when NT is small. As it grows, the OOB error decreases continuously and reaches eventually a threshold (Breiman, 2001). However, the complexity of RF model is directly proportional to NT. If there are too many decision trees, the operating efficiency will decrease as it becomes more time-consuming and the optimal result may not be obtained. The previous study by Wu et al. (2018) confirmed that in both low (e.g., 100) and high NT (e.g., 1000) cases, the algorithm did not perform well, but it did when NT was set to 300 and 500. .It is clear that the OOB error tends to be stable after NT gets greater than 300 (Fig. 6), or rather, the model accuracy becomes greater than 96%. Hence, 300 was finally used for NT when performing landslide risk modeling.
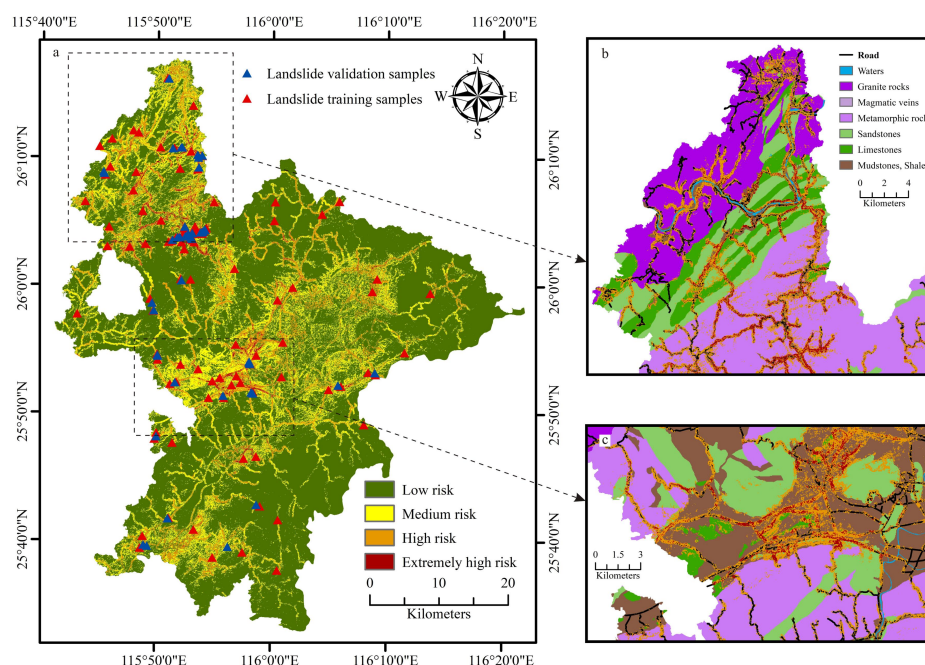
**Fig. 5** The landslide risk zoning map of Ruijin County.

Table 1. Distribution of historical landslides within different risk levels

| Landslide Risk Level	Area (km²)	Area Percentages (%)	Number of Historical landslides	Number Percentages (%)
Extremely high	24.83	1.02	99	63.87
High	227.57	9.32	47	30.32
Medium	472.39	19.36	5	3.23
Low	1715.60	70.30	4	2.58

3.3 Importance of hazard factors

For geohazard assessment, it is critical to understand the importance, more concretely, the role of each geo-environmental factor in such disaster event. In terms of the OBB error, Fig. 7 shows the importance of all the hazard factors considered in this research with the first four factors as follows: 1) distance to road, 2) slope, 3) May-July rainfall, 4) NDVI, 5) elevation, and so on.
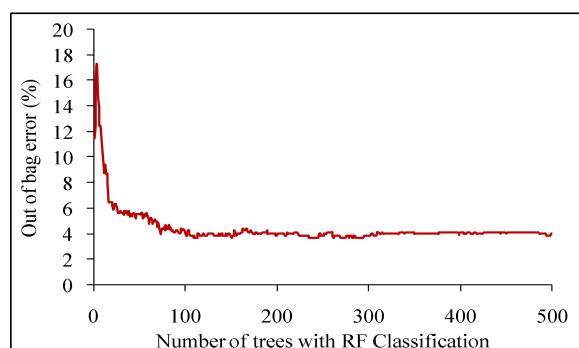
Fig. 6 Out-of-bag (OOB) error plot versus number of trees (NT) within RF modeling.

In the case of Ruijin, the order of importance seems plausible. Stable slope becomes unstable as a result of road construction, i.e., slope cutting or housing development, and naturally, steeper slope has higher propensity to slide due to gravity. May-July rainfall is a triggering factor as it liquidizes the slippery interface when it reaches certain threshold, i.e., the rainfall amount leading to saturation of soil after penetration and starting to flow on the soil/rock interface. Actually, the more rainfall in short time, the higher landslide disaster risk may be developed. Rainfall is thus widely employed as a Weather Indicator of landslides. NDVI, an autumn mean of five-year period and an indicator of vegetation abundance, vigor and root system development of forests and woodlands (herbaceous layer is mostly withered at that time), can largely reflect the stability and instability of the weathered crust. It is hence reasonable that these factors were identified as the most important hazard-causative factors in Ruijin though all others may play also a certain role in the geohazard events.
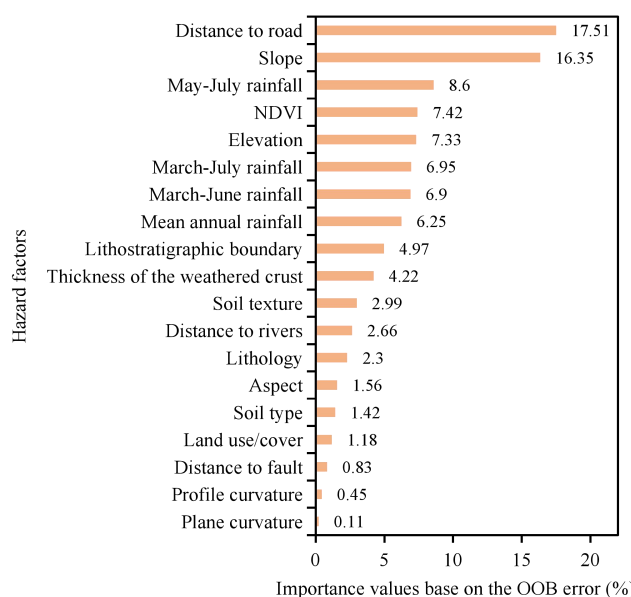
Fig. 7 Importance of the variables (geo-environmental factors) in provoking landslide events.

387 The importance of hazard factors associated with landslides had been also discussed by other
388 authors. The one of Dou et al. (2019) showed that precipitation was the most significant factor but
389 according to those of Arabameri et al. (2017) and Cao et al. (2019), DEM was the most important
390 one. Actually, it is understandable that in different geological environments, the mechanism of
391 landslides may be different and so the importance of hazard factors.

392 *3.4 Validation of the modeling results*

393 Only after being validated, has the model potential to be applied elsewhere. We used five
394 statistical indicators to evaluate the performance of landslide risk model, including Precision, Recall,
395 F1 Index, KC and OA as mentioned above. Against the training dataset, the Precision, Recall, F1
396 Index, KC and OA of the modeled results were 0.9908, 0.9818, 0.9863, 0.9724 and 0.9862,
397 respectively, while they were 0.95, 0.8867, 0.9173, 0.8299 and 0.9149 respectively versus VS
398 (Table 2). Above all, statistical indicators have shown that the RF classification model has good
399 performances in predicting the landslide risk.

400 **Table 2.** Performance of the RF modeling vs training set (TS) and validation set (VS)

Item	Training Set	Validation Set
Precision (%)	99.08	95.00
Recall (%)	98.18	88.67
F1 Index (%)	98.63	91.73
KC (%)	97.24	82.99
OA (%)	98.62	91.49

401 NT= 300.

402 Fig. 8 presented the success rate curve and predicted rate curve versus TS and VS. The AUC
403 for success rate curve and for prediction rate curve were 0.9936 and 0.9677, respectively. A model
404 can be considered as appropriate for this kind of risk prediction when the AUC had a value above
405 0.5 (Chen et al., 2018b; Achour and Pourghasemi, 2019). The landslide risk map had a better
406 success and prediction rates by RF classification model in this study compared with the results of
407 other scholars (Chung et al., 2003; Nicu, 2018; Depicker et al., 2020). Hence, RF classification
408 algorithm has a good predictive capability, and can be extended elsewhere for application.
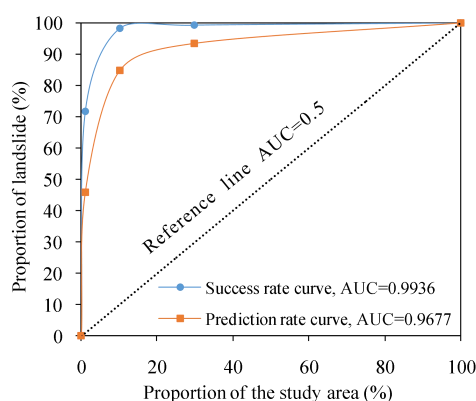

409 **Fig. 8** Success rate curves and prediction rate curves with associated AUC values.

410 *3.5 Case verification*

411   The landslides in the newly-built campus of No. 6 Middle School and the Longzhu Temple
412 in Ruijin took place very recently. The two new landslide events were predicted as Extremely
413 high-risk zones in the risk map (Fig. 9a). During the field investigation in July 2019, the middle
414 school was closed due to this disastrous effect (Fig. 9b); at the Longzhu Temple there were
415 significant ground bulges along the behind and side wall feet because of the extrusion provoked by
416 the downward slide of the upper slope (Fig. 9c). Thus, both sites are in danger as landslides
417 continue gradually.
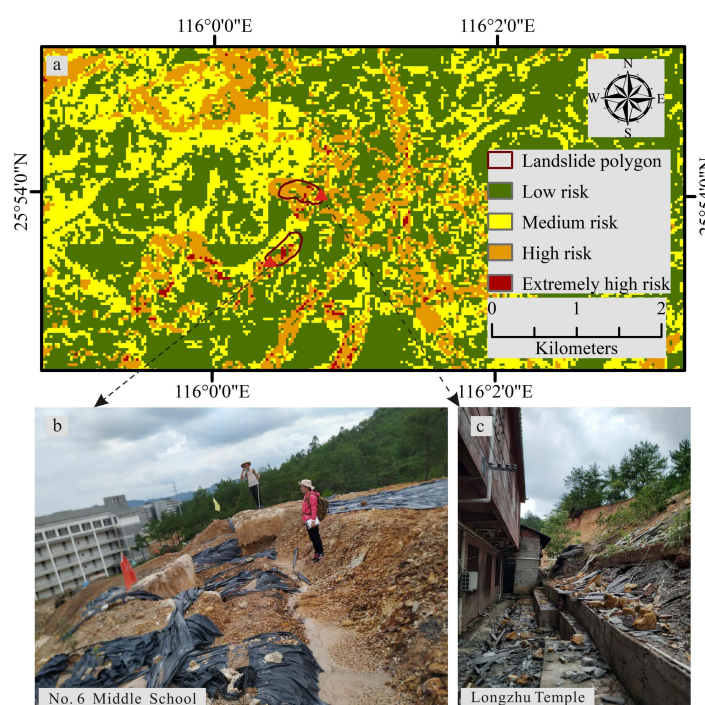
418   **Fig. 9** The landslide risk zoning map and photographs of landslide in field survey.

419 **4. Conclusions**

420   The prediction and prevention of landslide disasters have become essential to secure our
421 society. This paper presents a research using multi-source geo-environmental dataset as input
422 variables followed by a RF Classification to model and predict the landslide risk in Ruijin County.
423 Our results were obtained with high reliability of which OA is 98.62% and 91.49% versus TS and
424 VS. We believe that our research will be helpful for local government to take action on prevention
425 and early warning of geohazard to ensure people's safety and property, and to provide theoretical
426 advice for the infrastructure construction and urban planning.

427   Our study revealed the critical role of human activity, in particular, road construction, in
428 landslide events. Most of the observed landslides in Ruijin were actually "man-made". In future
429 road development we should take its negative impacts into account.

430   Another innovation lies in finding out a rational digitization and value assignment approach for
431   non-digital geo-environmental factors such as geological strata, faults, soil, roads and rivers so that
432   quantitative risk modeling and prediction using RF algorithm can be smoothly achieved.
433   Our study also illustrates that combination of remote sensing, geological, geomorphic, climatic
434   and human dimensional data is relevant for such geohazard risk zoning and mapping. RF algorithm
435   is able to satisfactorily achieve such task. This study can hence serve as a prototype for similar
436   research elsewhere.

437

Acknowledgments

445   **Conflicts of Interest:** The authors declare no conflict of interest.

446   **Author Contributions:** Conceptualization, X.Z., W.W. and Z.L.; methodology, W.W. and X.Z.; software,
447   W.W.; validation, X.Z., J.J., M.Z. and Y.L.; formal analysis, X.Z., P.O., W.H. and Y.Z.; investigation, X.Z.,
448   W.W., Y.S., Z.W., P.O., W.H., Y.Z., L.X. and X.H.; resources, G.Z., R.C., Y.S., Z.W. and T.L.; data curation,
449   X.Z., P.O., W.H. and Y.Z.; writing—original draft preparation, X.Z. and W.W; writing—review and editing,
450   W.W.; visualization, F.X. and J.L.; supervision, W.W., Y.Q., S.P., C.S., Y.B., X.Z., X.L., and W.L.; project
451   administration, W.W. and G.Z.; funding acquisition, W.W. All authors have read and agreed to the published
452   version of the manuscript.

453

References

455   Achour, Y. and Pourghasemi, H. R.: How do machine learning techniques help in increasing accuracy of
456   landslide susceptibility maps?, Geoscience Frontiers., doi:10.1016/j.gsf.2019.10.001, 2019.
457   Aleotti, P. and Chowdhury, R.: Landslide hazard assessment: summary review and new perspectives, Bulletin
458   of Engineering Geology & the Environment., 58, 21-44, doi: 10.1007/s100640050066, 1999.
459   Arabameri, A., Pourghasemi, H. R., and Yamani, M.: Applying different scenarios for landslide spatial
460   modeling using computational intelligence methods, Environmental Earth Sciences., 76,
461   doi:10.1007/s12665-017-7177-5, 2017.
462   Arabameri, A., Roy, J., Saha, S., Blaschke, T., Ghorbanzadeh, O., and Tien Bui, D.: Application of
463   Probabilistic and Machine Learning Models for Groundwater Potentiality Mapping in Damghan
464   Sedimentary Plain, Iran. Remote Sensing., 11, doi:10.3390/rs11243015, 2019.
465   Arabameri, A., Saha, S., Roy, J., Chen, W., Blaschke, T., and Tien Bui, D.: Landslide Susceptibility
466   Evaluation and Management Using Different Machine Learning Methods in The Gallicash River
467   Watershed, Iran, Remote Sensing., 12, doi:10.3390/rs12030475, 2020.

468 Assilzadeh, H., Levy, J. K., and Wang, X.: Landslide Catastrophes and Disaster Risk Reduction: A GIS
469   Framework for Landslide Prevention and Management, Remote Sensing., 2, 2259-2273,
470   doi:10.3390/rs2092259, 2010.
471 Ayalew, L. and Yamagishi, H.: The application of GIS-based logistic regression for landslide susceptibility
472   mapping in the Kakuda-Yahiko Mountains, Central Japan, Geomorphology., 65, 15-31,
473   doi:10.1016/j.geomorph.2004.06.010, 2005.
474 Breiman, L.: Random Forests, Machine Learning., 45, 5-32, doi:10.1023/A:1010933404324, 2001.
475 Cao, J., Zhang, Z., Wang, C., Liu, J., and Zhang, L.: Susceptibility assessment of landslides triggered by
476   earthquakes in the Western Sichuan Plateau, Catena., 175, 63-76, doi:10.1016/j.catena.2018.12.013, 2019.
477 Chang, Z., Du, Z., Zhang, F., Huang, F., Chen, J., Li, W., and Guo, Z.: Landslide Susceptibility Prediction
478   Based on Remote Sensing Images and GIS: Comparisons of Supervised and Unsupervised Machine
479   Learning Models, Remote Sensing., 12, doi:10.3390/rs12030502, 2020.
480 Chavez, P. S.: Image-Based Atmospheric Correction-Revisited and Improved, Photogrammetric Engineering
481   & Remote Sensing., 62, 1025-1036, doi:10.1016/0031-0182(96)00019-3, 1996.
482 Chen, W., Xie, X., Peng, J., Shahabi, H., Hong, H., Bui, D. T., Duan, Z., Li, S., and Zhu, A. X.: GIS-based
483   landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based
484   random forest method, Catena., 164, 135-149, doi:10.1016/j.catena.2018.01.012, 2018a.
485 Chen, W., Zhang, S., Li, R., and Shahabi, H.: Performance evaluation of the GIS-based data mining
486   techniques of best-first decision tree, random forest, and naive Bayes tree for landslide susceptibility
487   modeling, Sci Total Environ., 644, 1006-1018, doi:10.1016/j.scitotenv.2018.06.389, 2018b.
488 Chung, C. J. F. and Fabbri, A. G.: Validation of spatial prediction models for landslide hazard mapping,
489   Natural Hazards., 30, 451-472, doi:10.1023/B:NHAZ.0000007172.62651.2b, 2003.
490 Congalton, R. G.: Assessing Landsat classification accuracy using discrete multivariate analysis statistical
491   techniques, Photogramm.eng.remote Sens., 49, 1003-1007, doi:10.1271/bbb1961.44.1003, 1983.
492 Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J. P., Fotopoulou, S., Catani, F., Van Den
493   Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M. G., Pastor, M., Ferlisi, S., Tofani, V.,
494   Hervas, J., and Smith, J. T.: Recommendations for the quantitative analysis of landslide risk, Bulletin Of
495   Engineering Geology And the Environment., 73, 209-263, doi:10.1007/s10064-013-0538-8, 2014.
496 Delvaux, D., Mulumba, J. L., Sebagenzi, M. N. S., Bondo, S. F., Kervyn, F., and Havenith, H. B.: Seismic
497   hazard assessment of the Kivu rift segment based on a new seismotectonic zonation model (western branch,
498   East African Rift system), Journal Of African Earth Sciences., 134, 831-855,
499   doi:10.1016/j.jafrearsci.2016.10.004, 2017.
500 Depicker, A., Jacobs, L., Delvaux, D., Havenith, H. B., Maki Mateso, J. C., Govers, G., and Dewitte, O.: The
501   added value of a regional landslide susceptibility assessment: The western branch of the East African Rift,
502   Geomorphology., 353, doi:10.1016/j.geomorph.2019.106886, 2020.
503 Devkota, K. C., Regmi, A. D., Pourghasemi, H. R., Yoshida, K., Pradhan, B., Ryu, I. C., Dhital, M. R., and
504   Althuwaynee, O. F.: Landslide susceptibility mapping using certainty factor, index of entropy and logistic
505   regression models in GIS and their comparison at Mugling–Narayanghat road section in Nepal Himalaya,
506   Natural Hazards., 65, 135-165, doi:10.1007/s11069-012-0347-6, 2012.
507 Dou, J., Yunus, A. P., Tien Bui, D., Merghadi, A., Sahana, M., Zhu, Z., Chen, C. W., Khosravi, K., Yang, Y.,
508   and Pham, B. T.: Assessment of advanced random forest and decision tree algorithms for modeling

509   rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan, Sci Total Environ., 662,

510   332-346, doi:10.1016/j.scitotenv.2019.01.221, 2019.

511   Fan, L., Lehmann, P., and Or, D.: Effects of soil spatial variability at the hillslope and catchment scales on

512   characteristics of rainfall-induced landslides, Water Resources Research., 52, 1781-1799,

513   doi:10.1002/2015WR017758, 2016.

514   Frattini, P., Crosta, G., and Carrara, A.: Techniques for evaluating the performance of landslide susceptibility

515   models, Engineering Geology., 111, 62-72, doi:10.1016/j.enggeo.2009.12.004, 2010.

516   Froude, M. J. and Petley, D. N.: Global fatal landslide occurrence from 2004 to 2016, Natural Hazards And

517   Earth System Sciences., 18, 2161-2181, doi:10.5194/nhess-18-2161-2018, 2018.

518   Furlani, S. and Ninfo, A.: Is the present the key to the future?, Earth-Science Reviews., 142, 38-46,

519   doi:10.1016/j.earscirev.2014.12.005, 2015.

520   Gallen, S. F., Clark, M. K., and Godt, J. W.: Coseismic landslides reveal near-surface rock strength in a

521   high-relief, tectonically active setting, Geology., 43, 11-14, doi:10.1130/g36080.1, 2014.

522   Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., and Reichenbach, P.: Comparing landslide inventory

523   maps, Geomorphology., 94, 0-289, doi:10.1016/j.geomorph.2006.09.023, 2008.

524   García-Rodríguez, M. J., Malpica, J. A., Benito, B., and Díaz, M.: Susceptibility assessment of

525   earthquake-triggered landslides in El Salvador using logistic regression, Geomorphology., 95, 172-191,

526   doi:10.1016/j.geomorph.2007.06.001, 2008.

527   Genuer, R., Poggi, J. M., and Tuleau Malot, C.: Variable selection using random forests, Pattern Recognition

528   Letters., 31, 2225-2236, doi:10.1016/j.patrec.2010.03.014, 2010.

529   Goetz, J. N., Brenning, A., Petschko, H., and Leopold, P.: Evaluating machine learning and statistical

530   prediction techniques for landslide susceptibility modeling, Computers & Geosciences., 81, 1-11,

531   doi:10.1016/j.cageo.2015.04.007, 2015.

532   Gritzner, M. L., Marcus, W. A., Aspinall, R., and Custer, S. G.: Assessing landslide potential using GIS, soil

533   wetness modeling and topographic attributes, Payette River, Idaho, Geomorphology., 37, 149-165,

534   doi:10.1016/s0169-555x(00)00068-4, 2001.

535   Guzzetti, F., Carrara, A., Cardinali, M., and Reichenbach, P.: Landslide hazard evaluation: a review of current

536   techniques and their application in a multi-scale study, Central Italy, Geomorphology., 31, 181-216,

537   doi:Doi 10.1016/S0169-555x(99)00078-1, 1999.

538   Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M. and Ardizzone, F.: Probabilistic landslide hazard

539   assessment at the basin scale, Geomorphology., 72, 272-299, doi:10.1016/j.geomorph.2005.06.002, 2005.

540   Huang, Y. and Zhao, L.: Review on landslide susceptibility mapping using support vector machines, Catena.,

541   165, 520-529, doi:10.1016/j.catena.2018.03.003, 2018.

542   Kitutu, M. G., Muwanga, A., Poesen, J. and Deckers, J. A. Influence of soil properties on landslide

543   occurrences in Bududa district, Eastern Uganda, African Journal Of Agricultural Research., 4, doi:611-620,

544   10.1021/jf900470c, 2009.

545   Li L. P., Lan H. X., Guo C. B., Zhang Y. S., Li Q. W., and Wu Y. M.: A modified frequency ratio method for

546   landslide susceptibility assessment, Landslides., doi:10.1007/s10346-016-0771-x, 2017.

547   Monsieurs, E., Jacobs, L., Michellier, C., Basimike Tchangaboba, J., Ganza, G. B., Kervyn, F., Maki Mateso,

548   J. C., Mugaruka Bibentyo, T., Kalikone Buzera, C., Nahimana, L., Ndayisenga, Aloys., Nkurunziza, P.,

549   Thiery, W., Demoulin, A., Kervyn, M., and Dewitte O.: Landslide inventory for hazard assessment in a

Natural Hazards
and Earth System
Sciences
Discussions

Open Access

EGU

550   data poor context: a regional-scale approach in a tropical African environment, Landslides., 15, 2195-2209,
551   doi:10.1007/s10346-018-1008-y, 2018.

552   Montgomery, D. R. and Dietrich, W. E.: A physically based model for the topographic control on shallow
553   landsliding, Water Resources Research., 30, doi: 10.1029/93wr02979, 1994.

554   Montgomery, D. R., Schmidt, K. M., Greenberg, H. M., and Dietrich, W. E.: Forest clearing and regional
555   landsliding, Geology., 28, 311, doi:10.1130/0091-7613(2000)28<311:FCARL>2.0.CO;2, 2000.

556   Nadim, F., Kjekstad, O., Peduzzi, P., Herold, C., and Jaedicke, C.: Global landslide and avalanche hotspots,
557   Landslides., 3, 159-173, doi:10.1007/s10346-006-0036-1, 2006.

558   Nicu, I. C. and Asăndulesei, A.: GIS-based evaluation of diagnostic areas in landslide susceptibility analysis
559   of Bahluieț River Basin (Moldavian Plateau, NE Romania). Are Neolithic sites in danger?,
560   Geomorphology., 314, 27-41, doi:10.1016/j.geomorph.2018.04.010, 2018.

561   Petley, D.: Global patterns of loss of life from landslides, Geology., 40, 927-930, doi:10.1130/g33217.1,
562   2012.

563   Pourghasemi, H. R. and Kerle, N.: Random forests and evidential belief function-based landslide
564   susceptibility assessment in Western Mazandaran Province, Iran, Environmental Earth Sciences., 75,
565   185.181-185.117, doi:10.1007/s12665-015-4950-1, 2016.

566   Pradhan, B.: Remote sensing and GIS-based landslide hazard analysis and cross-validation using multivariate
567   logistic regression model on three test areas in Malaysia, Advances in Space Research., 45, 1244-1256,
568   doi:10.1016/j.asr.2010.01.006, 2010.

569   Qin, Y. and Liu, L.: Quantitative 3D Association of Geological Factors and Geophysical Fields with
570   Mineralization and Its Significance for Ore Prediction: An Example from Anqing Orefield, China,
571   Minerals., 8, 300, doi:10.3390/min8070300, 2018.

572   Reichenbach, P., Busca, C., Mondini, A. C., and Rossi, M.: The Influence of Land Use Change on Landslide
573   Susceptibility Zonation: The Briga Catchment Test Site (Messina, Italy), Environmental Management., 54,
574   1372-1384, doi:10.1007/s00267-014-0357-0, 2014.

575   Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F.: A review of statistically-based
576   landslide susceptibility models, Earth-Science Reviews., 180, 60-91, doi:10.1016/j.earscirev.2018.03.001,
577   2018.

578   Remondo, J., González, A., De Terán, J. R. D., Cendrero, A., Fabbri, A., and Chung, C. J. F.: Validation of
579   landslide susceptibility maps; Examples and applications from a case study in northern Spain, Natural
580   Hazards., 30, 437-449, doi:DOI 10.1023/B:NHAZ.0000007201.80743.fc, 2003.

581   Ruff, M. and Czurda, K.: Landslide susceptibility analysis with a heuristic approach in the Eastern Alps
582   (Vorarlberg, Austria), Geomorphology., 94, 314-324, doi:10.1016/j.geomorph.2006.10.032, 2008.

583   Sameen, M. I., Pradhan, B., and Lee, S.: Application of convolutional neural networks featuring Bayesian
584   optimization for landslide susceptibility assessment, Catena., 186, doi:10.1016/j.catena.2019.104249, 2020.

585   Smets, B., Delvaux, D., Ross, K. A., Poppe, S., Kervyn, M., d'Oreye, N., and Kervyn, F.: The role of inherited
586   crustal structures and magmatism in the development of rift segments: Insights from the Kivu basin,
587   western branch of the East African Rift, Tectonophysics., 683, 62-76, doi:10.1016/j.tecto.2016.06.022,
588   2016.

Natural Hazards
and Earth System
Sciences
Open Access
Discussions
EGU

589 Tien Bui, D., Pradhan, B., Lofman, O., and Revhaug, I.: Landslide Susceptibility Assessment in Vietnam
590   Using Support Vector Machines, Decision Tree, and Naïve Bayes Models, Mathematical Problems in
591   Engineering., 2012, 1-26, doi:10.1155/2012/974638, 2012.

592 Wang, H.B., Liu, G. J., Xu, W. Y., and Wang, G. H.: GIS-based landslide hazard assessment: an overview,
593   Prog Phys Geog., 29, 548-567, doi:10.1191/0309133305pp462ra, 2005.

594 Waske, B., van der Linden, S., Oldenburg, C., Jakimow, B., Rabe, A., and Hostert, P.: imageRF – A
595   user-oriented implementation for remote sensing image analysis with Random Forests, Environmental
596   Modelling & Software., 35, 192-193, doi:10.1016/j.envsoft.2012.01.014, 2012.

597 Wasowski, J., Bovenga, F., Nutricato, R., Nitti, D. O., and Chiaradia, M. T.: Applications Of Medium C-Band
598   And High Resolution X-Band Multitemporal Interferometry In Landslide Investigations, ISPRS -
599   International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.,
600   XL-1-W5, 737-743, doi:10.5194/isprsarchives-XL-1-W5-737-2015, 2015.

601 Wu, W. C. and Ai, G.: Risk assessment of natural disasters in the course of selection of nuclear waste disposal
602   (in Chinese with English abstract), Journal of East China Geological Institute., 18, 260-265, 1995.

603 Wu, W. C., De Pauw, E., and Hellden, U.: Assessing woody biomass in African tropical savannahs by
604   multiscale remote sensing, International Journal of Remote Sensing., 34, 4525-4549,
605   doi:10.1080/01431161.2013.777487, 2013.

606 Wu, W. C., Zucca, C., Karam, F., and Liu, G.: Enhancing the performance of regional land cover mapping,
607   International Journal of Applied Earth Observation and Geoinformation., 52, 422-432,
608   doi:10.1016/j.jag.2016.07.014, 2016.

609 Wu, W. C., Zucca, C., Muhaimeed, A. S., Al-Shafie, W. M., Fadhil Al-Quraishi, A. M., Nangia, V., Zhu, M.,
610   and Liu, G.: Soil salinity prediction and mapping by machine learning regression in Central Mesopotamia,
611   Iraq, Land Degradation & Development., 29, 4005-4014, doi:10.1002/ldr.3148, 2018.

612 Wu, W. C.: Application de la géomatique au suivi de la dynamique environnementale en zones arides, Ph.D.
613   Université de Paris 1-Panthéon-Sorbonne, Paris, France, 2003.

614 Yilmaz, I.: Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural
615   networks and their comparison: A case study from Kat landslides (Tokat-Turkey), Computers &
616   Geosciences., 35, 1125-1138, doi:10.1016/j.cageo.2008.08.007, 2009.

617 Youssef, A. M., Maerz, N. H., and Hassan, A. M.: Remote sensing applications to geological problems in
618   Egypt: case study, slope instability investigation, Sharm El-Sheikh/Ras-Nasrani Area, Southern Sinai,
619   Landslides., 6, 353-360, doi:10.1007/s10346-009-0158-3, 2009.

620 Zhang, K., Wu, X., Niu, R., Yang, K., and Zhao, L.: The assessment of landslide susceptibility mapping using
621   random forest and decision tree methods in the Three Gorges Reservoir area, China, Environmental Earth
622   Sciences., 76, doi:10.1007/s12665-017-6731-5, 2017.

623 Zhu, A. X., Miao, Y., Liu, J., Bai, S., Zeng, C., Ma, T., and Hong, H.: A similarity-based approach to
624   sampling absence data for landslide susceptibility mapping using data-driven methods, Catena., 183,
625   doi:10.1016/j.catena.2019.104188, 2019.