Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/190217

Prediction of Landslides Using Data Mining Techniques

Raed Alazaidah[1], Mohammad Subhi Al-Batah[2], Ghassan Samara[1], Mo'ath Alluwaici[1], Mohammed Abu Safaqah[3], Mohammad Aljaidi[1], Suleiman Ibrahim Mohammad[4,5], Asokan Vasudevan[6] and Yujie Zhang[7]*

[1]Faculty of Information Technology, Data Science and AI Department, Zarqa University, Zarqa, Jordan
[2]Faculty of Science and Information Technology, Jadara University, Irbid, Jordan
[3]Mechanical Engineering Department, Global College of Engineering and Technology (GCET), Muscat, Sultanate of Oman
[4]Electronic Marketing and Social Media, Economic and Administrative Sciences Zarqa University, 13110 Zarqa, Jordan
[5] INTI International University, 71800 Negeri Sembilan, Malaysia
[6]Faculty of Business and Communications, INTI International University, Persiaran Perdana BBN Putra Nilai, 71800 Nilai, Negeri Sembilan, Malaysia
[7]Faculty of Liberal Arts, Shinawatra University, Thailand

Received: 25 Nov. 2024, Revised: 9 Jan. 2024, Accepted: 13 Jan. 2025
Published online: 1 Mar. 2025

Abstract: Landslide is a phenomenon that can happen suddenly or slowly over a long period of time. It is defined as a mass movement of different materials such as rubbles or stones. Landslides not only may cause huge loss of lives, properties, livestock but also have a bad impact on the environment. Many classification models have been proposed and utilized in aim of prediction of landslides. Hence, this paper aims to conduct an extensive evaluation of large number of classification models that adopt several learning strategies using a primary dataset that has been collected specifically for this purpose. Moreover, another main objective of this paper is to deter-mine the best feature selection method to use among four well-known methods. The evaluation phase considers several related evaluations. The results revealed that RandomForest, had the best performance classifiers, with 87.73% accuracy. Moreover, Correlation Attribute Evaluator method showed the best predictive performance.

Keywords: Classification, data mining, feature selection, landslide, prediction

1 Introduction

Landslides is a geological phenomenon that manifests in several forms such as mass movement, rockfalls, mudflows, and debris flows among other forms [1,2,3]. The main cause for landslides (sometime called landslips) is Gravity [4,5,6]. Nevertheless, other main reasons may cause landslides such as earthquake, intensive or heavy rainfall, or even by human development activities like urban sprawl and mining [7,8,9].

Landslides have very bad impacts such as the loss of lives, infrastructure destruction, lands damages, and huge loss of natural resources [10,11,12]. According to recent statistics, around 18 million people in Syria and Turkey have been impacted by the series of earthquakes and landslides occurred during the first few months of 2023. Moreover, around 59,000 died and nearly 130, 000 people have been injured. Millions of citizens have been relocated or displaces, and more than 10 million in need for urgent aid. According to the World Bank, the immediate damage caused by the recent Turkey landslides is estimated at 34 billion dollar which equals to around 4% of Turkey annual economic output.

Consequently, this paper aims to contribute to the global efforts of reducing the high costs of landslides via two main objectives. The first one is to identify the best classification model among twenty-seven different models that accurately predicts landslide event based on specific features. The second objective is to determine the best feature selection and ranking method that suits this kind of datasets and helps improving the predictive performance of the classification models. For this objective, four well-known feature selection and ranking methods have been evaluated and considered.

Hence, the main task of this research is classification. Classification is one of the most well-known tasks in data mining, data science, and machine learning [13,14,15]. This task aims to highly predict the class label for unseen

* Corresponding author e-mail: dr_sliman@yahoo.com

instance based on the values of the associated features in the training set [16,17,18]. Classification is generally divided into Single Label Classification (SLC) and Multi Label Classification (MLC) [19,20,21] . SLC forces each instance to be associated with only one class label, while instances in MLC may be linked or associated to one or more class labels [22,23,24]. Hence, MLC is much more complicated than SLC. SLC itself could be divided into two main subtypes: binary classification and multi class classification [25,26,27]. The first one consists of only two class labels while the second one consists of at least three class labels [28,29,30]. This paper is more interested in SLC, and specifically, binary classification; since the dataset considered in this paper consists of two class labels only (Slide, NoSlide).

The rest of this paper is organized as follows: Section 2 surveys the most related work to the prediction of landslides task. Section 3 presents the methodology, data, results, and discussion regarding the most significant findings. Section 4 concludes and suggests some of future directions.

2 Related Work

Various machine-learning algorithms have been utilized for landslide prediction. artificial neural networks inspired by biological neural networks have been employed [31, 32,33]. Fuzzy logic algorithms have also been used to evaluate the spatial distribution of landslides [34]. SVM has been widely and effectively applied in landslide prediction [35], outperforming conventional Logistic Regression (LR) in terms of performance [36]. LR has been extensively employed in landslide studies and has shown promise in spatial prediction [37]. In comparative analyses, LR has proven superior to artificial neural networks and likelihood ratio methods for landslide analysis [38]. While Fisher's Linear Discriminant Analysis (FLDA) is commonly used for complex data classification, its application in landslide research remains limited [39]. Bayesian Networks (BN) hold potential for hazard assessment but are rarely used in landslide hazard evaluation [40]. BN has been successfully employed in assessing earthquake-induced landslide susceptibility and debris flow hazard [41]. Additionally, the Naive Bayes (NB) method has shown success in landslide assessment studies, demonstrating its efficacy as a machine learning method for spatial prediction [42].

In [31] authors compared the performance of two Artificial Neural Network (ANN) algorithms, MLP and RBF, in predicting landslide susceptibility in the Vaz Watershed, Iran. Using aerial photographs and field surveys, 136 landslide locations were identified and used to create a landslide inventory map. The map was divided into training and validation datasets. Nine conditioning factors were considered, and MLP with the Broyden-Fletcher-Goldfarb-Shanno learning algorithm showed better performance than RBF in mapping landslide susceptibility. The validation results demonstrated accuracies of 90.85% for RBF and 91.93% for MLP, indicating the effectiveness of the ANN approach in mapping landslide susceptibility and its potential for guiding land use planning in the Vaz Watershed, Iran.

In [34], the utilization of LiDAR data for landslide susceptibility mapping in Ulu Klang, Malaysia was investigated. Nine conditioning factors were derived from the LiDAR data, and an ensemble approach combining Support Vector Machine (SVM) and Evidential Belief Function (EBF) was suggested to enhance prediction accuracy. EBF was employed to evaluate the impact of conditioning factor classes and assign weights, which were then incorporated into SVM modeling. The RBF kernel exhibited superior efficiency, resulting in success rates of 83.04% and 80.04% for the ensemble EBF and RBF-SVM methods, respectively, measured by the Area Under the Curve (AUC). The proposed ensemble technique improved both processing speed and outcomes by capitalizing on the strengths of EBF and SVM.

In [35], a hybrid model is introduced for assessing regional-scale landslide susceptibility in the Zigui-Badong area near the Three Gorges Reservoir, China. The model integrates rough set theory and support vector machine techniques with GIS and remote sensing data. It identifies important environmental factors and predicts landslide susceptibility, generating a map that highlights areas with medium to high susceptibility. The evaluation confirms the model's reliability and superior predictive ability compared to a standard SVM model, underscoring its value for regional-scale landslide susceptibility mapping in the studied area.

In [37], a machine learning approach is introduced for spatial modeling in landslide susceptibility assessment, with a comparison of SVM, Decision Trees (DT), and LR. The study carried out in Fruška Gora Mountain, Serbia, showed that SVM yielded superior results compared to the other algorithms. The SVM classifier was also compared to the Analytical Hierarchy Process (AHP) method, and it outperformed AHP across all evaluation metrics, including the κ index, area under the ROC curve, and false positive rate in the stable ground class.

In [38] different models, such as the likelihood ratio, LR, and ANN were employed and validated to assess the susceptibility of landslides in Youngin, Korea, utilizing a geographic information system. A spatial database containing information on landslide location, topography, soil, forest, geology, and land use was utilized. Fourteen factors associated with landslides were computed or extracted from the database, and landslide susceptibility indexes were calculated using these models. To validate the results, the study area was divided equally into west and east sides, with the west side used for susceptibility assessment and the east side for verification. The evaluation, based on success and prediction rates, demonstrated a satisfactory agreement between the susceptibility map and the available landslide data.

In [41], a hybrid approach utilizing Bayesian Network (BN) is introduced to examine the factors that cause landslides during earthquakes and evaluate their impacts in Beichuan, China. The analysis highlights lithology and Arias intensity as significant factors influencing landslides in the region. By employing the BN model, a high accuracy of 93% is achieved in detecting landslides, showcasing the model's efficacy for assessing and predicting landslide occurrences.

In [42], three data mining approaches (SVM, DT, NB models) are compared for predicting landslide hazards in HoaBinh province, Vietnam. A landslide inventory map is generated from 118 locations, with 70% used for training and 30% for validation. Ten conditioning factors are taken into account, and the models are used to calculate landslide susceptibility indexes. The validation results reveal that SVM models exhibit the strongest prediction capability, while LR models follow closely. DT models perform the least effectively, and NB models also exhibit lower prediction capability when compared to SVM models.

In [43], the authors identified the limitations of Associative Classification (AC) methods in solving multi-label classification (MLC) problems. To overcome this, they proposed a modified version called msCBA, which incorporates multiple class labels and adjusts the rule order procedure for MLC datasets. Comparative evaluations using various MLC algorithms and datasets emphasized the importance of local label dependencies and highlighted the benefits of AC in MLC, such as generating accurate classifiers and revealing hidden information through interpretable rules. The adapted version, ML-CBA, outperformed other MLC algorithms across different learning strategies, as demonstrated by multiple evaluation metrics.

In [44] authors predicted crowdfunding campaign success using Kickstarter data through their research study. They employed various phases, including data scraping, wrangling, exploration, model construction, evaluation, and variable importance ranking. Among the four machine learning algorithms used, Random Forest achieved the highest classification accuracy (94%), followed closely by Deep Learning (93%). The K-Nearest Neighbor model stood out with exceptional performance, reaching a 97.9% accuracy score and an impressive area under the curve (AUC) performance of 98.3%. This research outperformed previous studies' models and provided accurate predictions for crowdfunding campaign success.

In [45] Authors explore the creation of an innovative hybrid functional machine learning algorithm to forecast shallow landslides using remote sensing data. The article highlights that while the proposed method has yielded satisfactory outcomes, no single hybrid model has demonstrated superiority over others.

In [46] authors introduce a hybrid machine learning algorithm, ABSGD, to predict landslides in the Sarkhoon watershed in Iran. It combines Stochastic Gradient Descent (SGD) and AdaBoost (AB) Meta classifier. The authors utilized 20 landslide conditioning factors, ranked using the LSSVM technique. The ABSGD model outperformed other benchmark models, achieving an AUC of 0.86%. Distance to the road was identified as the most significant factor for landslide occurrence. The study emphasizes the importance of reliable susceptibility maps for land management and decision-making. ABSGD, a combination of SGD and AdaBoost, improves the accuracy of predictive landslide susceptibility mapping.

In [50] Authors conducted a study comparing traditional statistical models with newer machine learning models for regional landslide prediction in data-limited areas. Multiple techniques were examined and tested in various regions of Lower Austria. The results indicated that Random Forest and bundling classification methods achieved the highest predictive accuracy, although there were no significant differences among the modeling techniques. Accuracy metrics varied across different areas analyzed. Key predictors varied among the models, but variables like slope angle, surface roughness, and plan curvature consistently ranked as significant. Some models showed spatial irregularities in prediction maps due to predictor splits, while others produced smoother surfaces. The study suggests that this evaluation approach can aid in selecting the most suitable technique for landslide susceptibility modeling.

The authors of [51] proposed a method to address spatial disagreement and uncertainties in landslide prediction. They developed susceptibility maps for Cox's Bazar district in Bangladesh using four machine learning algorithms, achieving high accuracy. To handle uncertainties, they introduced a LR model that integrated the outcomes of the four models. The Random Forest model had the greatest influence in predicting observed landslide locations. By combining the results of the four models, a more refined landslide susceptibility map was created, improving spatial agreement and prediction accuracy compared to individual models. This approach effectively reduces uncertainties and enhances landslide predictions.

3 Research Methodology and Empirical Analysis

In this section, the research methodology, the dataset, as well as the evaluation results for the considered classifiers and feature selection methods are all presented. First, the methodology is described in Section 3.1. Then, the dataset with its all-relevant details is provided in Section 3.2. Section 3.3 presents the results for the comparative analysis of the considered classifiers and using several well-known metrics. Section 3.4 attempts to identify the most appropriate feature selection method to use with the considered dataset with respect to Accuracy and Time metrics. Finally, Section 3.5 discusses the main results and findings.

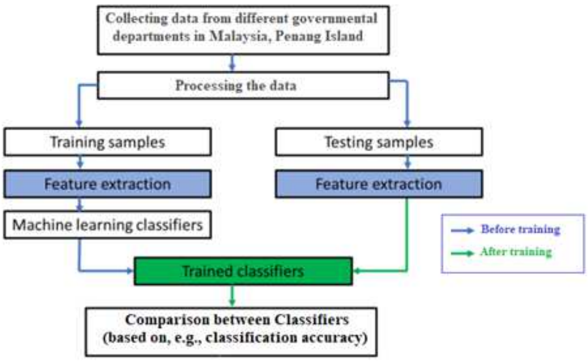

Fig. 1: This Research methodology

3.1 Research Methodology

Figure 1 represents the research methodology followed in this paper. The first step is the data collection step where data has been collected from several governmental departments and centers in Malaysia, Penang Island. Penang island has been chosen due to the high frequency of landslides events it suffers every year. More information regarding the collected data is provided in Section 3.1. The second step is the preprocessing step where data has been cleaned and checked for eliminating duplicate values. Also, this step aims to make the data balanced as much as possible. The third step aims to identify the best feature selection method that highly suits this type of dataset. The fourth step aims to meet the first objective of this research, that is, to identify the best classifier that can highly and accurately predict landslides, while the fifth step discusses the main results and findings.

3.2 Dataset Description

One dataset has been used in this research. The dataset is a primary one and has been collected in Penang island-Malaysia. The dataset consists of 10, 000 instances and twenty one numeric features in addition to the class feature. All instances in the dataset are linked to one of two possible class labels (Slide, No_Slide). The ”Slide” label is associated with (4967) instances, where the ”No_Slide” class label is associated with 5022. Therefore, the dataset does not suffer from the problem of imbalance class distribution. Table 1 depicts the main characteristic of the dataset considered in this research. It is worth mentioning that very few studies have considered the last six features (F16, F17,..., F21).

Twenty-one features have been considered in this dataset. All of these features are of numeric type. Table 2 represents more information regarding these features.

Table 1: Dataset characteristics

Name	Landslide
Type	Binary classification dataset
No. of attributes	21
No. of instances	10,000
Number of class labels	2 (Slide, No-Slide)
Type of attributes	Numeric (continuous)
Missing values	No

Table 2: Features description

No.	Name	Symbol	Equation
1	Slope	F1	$270° + \arctan\left(\frac{f_y}{f_x}\right) - 90° \frac{f_x}{\lvert f_y \rvert}$
2	Curvature	F2	$ProfileCurvature + PlanCurvature$
3	Drainage distance	F3	No equation is needed
4	Fault line distance	F4	No equation is needed
5	Type of land cover	F5	No equation is needed
6	Geology	F6	No equation is needed
7	Level of Plan curvature	F7	$\left(\frac{z_2+z_5}{2} - z_5\right)/2w$
8	Distance from nearest road	F8	No equation is needed
9	Curvature profile	F9	$\left(\frac{z_2+z_5}{2} - z_5\right)/2w$
10	Level of Rain perception	F10	No equation is needed
11	Angle of slope	F11	$\arctan\sqrt{f_x^2 + f_y^2}$
12	Elevation from sea	F12	No equation is needed
13	Type of vegetation	F13	No equation is needed
14	Soil type	F14	No equation is needed
15	Curvature Tangent	F15	$-\frac{q^2r-2pqs+p^2t}{p^2+q^2\sqrt{1+p^2+q^2}}$
16	Area of surface	F16	If adjustment factor value $> 1 \frac{\mathrm{C}^2}{\cos(\text{ stopeangle })}$ If adjustment factor value $= 1, C^2$
17	Roughness	F17	$\sqrt{\frac{1}{\mathrm{N}}\Sigma_{\mathrm{i}}^{\mathrm{N}}(\mathrm{S_i} - \bar{s})^2}$
18	Diagonal length	F18	$\sqrt{f_x^2 + f_y^2}$
19	Curvature longitude	F19	$-2\left\{\frac{p^2r+pqs+q^2t}{p^2+q^2}\right\}$
20	Rigidity level	F20	$\frac{\text{surfaceareaof } 3x3 \text{ neighborhoodwindows}}{\text{plane areaof } 3x3 \text{ neighborhoodwindows}}$
21	Cross curvature	F21	$2\left\{\frac{q^2r-pqs+tp^2}{p^2+q^2}\right\}$

3.3 Evaluation Results for the Considered Classifiers Using All features with Respect to Several Evaluation Metrics and without Using Discretization

This section aims to identify the most appropriate classifier to predict landslide with respect to several significant and well-known metrics such as Accuracy, True Positive (TP) rate, False Positive (FP) rate, and Precision [52]-[54]. The previously mentioned metrics are calculated using the following equations:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{TP rate} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{FP rate} = \frac{FP}{FP + TN} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP'} \quad (4)$$

For this main objective, twenty-seven classifiers have been extensively evaluated. These classifiers belong to six learning strategies.

The Trees' learning strategy has been represented by: DecisionStump, J48, RandomForest, RandomTree, REPTree. The Rule's based strategy has been represented using five classifiers: DecisionTable, JRip, OneR, PART, ZeroR. For Meta learning strategy, the following classifiers have been used: AdaBoostM1, ClassificationViaRegression, FilteredClassifier, LogitBoost, MultiClassClassifier.

The Lazy learning strategy is represented through three different classifiers: IBK, KStar, LWL. Moreover, from the Functions learning strategy, the following classifiers have been considered: Logistic, MultilayerPerceptron, SimpleLogistic, SMO, VotedPerceptron. Finally, four classifiers have been considered to represent the Bayes learning strategy: BayesNet, NaiveBayes, NaiveBayesMultinominal, NaiveBayesUpdateable.

All these classifiers have been used with their default setting using Python. Finally, 10-folds cross-validation has been used in extracting and validating the results.

It is worth mention, for this section, neither feature selection step has been used, nor discretization.

According to Table 3, we can conclude that RandomForest, which belongs to the Trees learning strategy, is the best classifier with respect to the considered evaluation metric in the considered dataset. It has a high Accuracy, a high TP rate, the lowest FP rate, and a high Precision, even if it does not have the lowest time.

In our case, the TP rate, also known as Recall, is the most crucial parameter since it indicates how effectively our model can identify relevant information. Where Recall indicates the number of times we correctly recognized a nation as having a landslide, we must avoid any cases where the model assumed that the country did not have a landslide.

Table 3: Evaluation results for the considered classifiers using Accuracy, TP, and FP metrics

Strategy	Classifier	Accuracy	TP Rate	FP Rate	Precision
Trees	DecisionStump	71.140	0.711	0.285	0.801
	J48	81.130	0.811	0.188	0.814
	RandomForest	87.730	0.877	0.122	0.888
	RandomForest	87.730	0.877	0.122	0.888
	RandomTree	80.430	0.804	0.195	0.805
	REPTree	81.690	0.817	0.182	0.821
Rules	DecisionTable	83.760	0.838	0.162	0.838
	JRip	80.870	0.809	0.191	0.811
	OneR	69.140	0.691	0.307	0.698
	PART	81.590	0.816	0.184	0.817
	ZeroR	50.330	0.503	0.503	0.503
Meta	AdaBoostM1	72.640	0.726	0.271	0.760
	ClassificationViaRegression	78.010	0.780	0.218	0.798
	FilteredClassifier	81.740	0.817	0.181	0.827
	LogitBoost	74.690	0.747	0.251	0.778
	MultiClassClassifier	72.730	0.727	0.271	0.745
Lazy	IBK	77.650	0.777	0.223	0.780
	KStar	85.540	0.855	0.143	0.872
	LWL	71.540	0.715	0.281	0.802
Functions	Logistic	72.730	0.727	0.271	0.745
	MultilayerPerceptron	77.210	0.772	0.227	0.785
	SimpleLogistic	72.850	0.729	0.270	0.752
	SMO	72.710	0.727	0.270	0.781
	VotedPerceptron	72.740	0.727	0.271	0.751
Bayes	BayesNet	73.350	0.734	0.264	0.781
	NaiveBayes	72.890	0.729	0.269	0.770
	NaiveBayesMultinomial	70.720	0.707	0.291	0.733
	NaiveBayesUpdateable	72.890	0.729	0.269	0.770

3.4 Identifying the Best Feature Selection Method

This section aims to identify the best feature selection method that suits the considered dataset in a best way. To achieve this significant goal, four well-known feature selection and ranking methods have been considered. These feature selection methods are: ClassifierAttributeEval (ClAE), CorrelationAttributeEval (CoAE), GainRatioAttributeEval (GRAE), InfoGainAttributeEval (IGAE). These methods have been chosen due to their popularity in the domain. All feature selection methods have been used with their default settings as implemented in python.

Table 4 depicts the ranks for the best 0.75% of the features (best 15 features) of the considered dataset, using the four previously mentioned feature selection methods.

The evaluation process of the four feature selection methods considers one metric, that is, Accuracy. Also, the evaluation process considers the classifiers results when considering 0.50% (best 10 features) and 0.75% of the features in the considered dataset.

Table 4: Best 0.75% features of the considered dataset using four feature selection methods

Feature No.	ClAE	CoAE	GRAE	IGAE
1	21	16	6	15
2	7	6	2	14
3	8	5	4	16
4	10	4	21	3
5	6	3	15	6
6	5	2	5	2
7	4	7	8	21
8	3	1	9	5
9	2	8	7	9
10	9	15	16	8
11	11	21	3	13
12	20	19	1	1
13	18	18	13	12
14	19	12	12	18
15	12	13	17	17

Table 5 depicts the evaluation results of the twenty-seven classifiers when using ClAE method and considering 0.50% and 0.75% of the features in the considered dataset with respect to Accuracy metric.

Table 5: Evaluation results of the considered classifiers using ClAE

Strategy	Classifier	50% Features Accuracy	75% Features Accuracy
Trees	DecisionStump	70.120	70.120
	J48	74.520	78.770
	RandomForest	74.260	83.660
	RandomTree	70.280	77.230
	REPTree	74.380	78.030
Rules	DecisionTable	73.710	80.030
	JRip	73.480	76.010
	OneR	69.030	69.030
	PART	73.750	77.520
	ZeroR	50.330	50.330
Meta	AdaBoostM1	72.580	72.630
	ClassificationViaRegression	73.860	76.360
	FilteredClassifier	73.930	77.970
	LogitBoost	72.750	73.380
	MultiClassClassifier	71.270	72.370
Lazy	IBK	71.390	75.500
	KStar	75.290	81.300
	LWL	70.030	70.100
Functions	Logistic	71.270	72.370
	MultilayerPerceptron	73.000	74.630
	SimpleLogistic	72.040	72.430
	SMO	71.100	72.520
	VotedPerceptron	71.370	72.320
Bayes	BayesNet	72.740	73.320
	NaiveBayes	72.470	72.540
	NaiveBayesMultinomial	70.460	71.900
	NaiveBayesUpdateable	72.470	72.540

According to the ClAE results, we can conclude that KStar is the best classifier when we choose 50% of the feature, but if we compare the results when we don't apply the CIAE methods, we can see that it reduced the Accuracy by 10.25%, which dropped the Accuracy significantly. Whereas RandomForest is the best classifier when 75% of the feature are chosen, they also reduce the Accuracy by 4.07% when ClAE methods are applied to the datasets, but they also have the best Accuracy among other learning strategies.

Table 6 depicts the evaluation results of the twenty seven classifiers when using CoAE method and considering 0.50% and 0.75% of the features in the considered dataset with respect to Accuracy metric.

According to the CoAE results in Table 6, we can conclude that RandomForest is the best classifier when we choose 50% or 75% of the feature. However, we can see that when we applied the CoAE methods, it reduced the accuracy by 7.78% when we chose 50% and 7.2% for 75% of the feature in the dataset.

Table 7 depicts the evaluation results of the twenty seven classifiers when using GRAE method and considering 0.50% and 0.75% of the features in the considered dataset with respect to Accuracy metric.

According to the GRAE results, we can conclude that RandomForest is the best classifier when we choose 50% or 75% of the feature, but if we compare the results when we don't apply the GRAE methods, we can see that it reduced the Accuracy by 11.02% when we choose 50% feature, which dropped the Ac- curacy significantly. Whereas when 75% of the feature are chosen, they also reduce the Accuracy by 5.5%, but they also have the best Accuracy among other learning strategies.

Table 6: Evaluation results of the considered classifiers using CoAE

Strategy	Classifier	Accuracy	TP Rate	FP Rate	Precision
Trees	DecisionStump	71.140	0.711	0.285	0.801
	J48	77.480	0.775	0.225	0.812
	RandomForest	79.950	0.800	0.200	0.826
	RandomTree	75.040	0.750	0.250	0.808
	REPTree	76.780	0.768	0.232	0.812
Rules	DecisionTable	75.630	0.756	0.244	0.803
	JRip	76.070	0.761	0.239	0.809
	OneR	69.140	0.691	0.309	0.698
	PART	75.310	0.753	0.247	0.803
	ZeroR	50.330	0.503	0.503	0.503
Meta	AdaBoostM1	72.540	0.725	0.275	0.757
	ClassificationViaRegression	75.880	0.759	0.241	0.804
	FilteredClassifier	75.120	0.751	0.249	0.803
	LogitBoost	72.980	0.729	0.271	0.766
	MultiClassClassifier	71.880	0.719	0.281	0.756
Lazy	IBK	77.870	0.779	0.221	0.810
	KStar	78.330	0.783	0.217	0.814
	LWL	71.520	0.715	0.285	0.802
Functions	Logistic	71.880	0.718	0.282	0.756
	MultilayerPerceptron	74.440	0.744	0.256	0.786
	SimpleLogistic	72.190	0.722	0.278	0.762
	SMO	71.450	0.714	0.286	0.766
	VotedPerceptron	72.140	0.721	0.279	0.763
Bayes	BayesNet	73.150	0.731	0.269	0.778
	NaiveBayes	72.100	0.721	0.279	0.765
	NaiveBayesMultinomial	70.510	0.705	0.295	0.738
	NaiveBayesUpdateable	72.100	0.721	0.279	0.765

Table 7: Evaluation results of the considered classifiers using GRAE

Strategy	Classifier	50% Features Accuracy	75% Features Accuracy
Trees	DecisionStump	71.140	71.140
	J48	73.850	77.620
	RandomForest	76.710	82.180
	RandomTree	71.370	76.460
	REPTree	74.830	78.170
Rules	DecisionTable	74.170	77.070
	JRip	73.490	77.060
	OneR	69.140	69.140
	PART	73.240	75.290
	ZeroR	50.330	50.330
Meta	AdaBoostM1	72.580	72.530
	ClassificationViaRegression	73.220	74.500
	FilteredClassifier	74.000	75.750
	LogitBoost	72.820	72.900
	MultiClassClassifier	72.310	72.740
Lazy	IBK	73.170	77.110
	KStar	75.620	81.170
	LWL	71.670	71.490
Functions	Logistic	72.310	72.740
	MultilayerPerceptron	72.470	75.050
	SimpleLogistic	72.330	72.460
	SMO	71.780	72.370
	VotedPerceptron	71.400	72.340
Bayes	BayesNet	72.820	73.180
	NaiveBayes	72.620	72.290
	NaiveBayesMultinomial	71.290	70.610
	NaiveBayesUpdateable	72.620	72.290

Table 8 depicts the evaluation results of the twenty-seven classifiers when using IGAE method and considering 0.50% and 0.75% of the features in the considered dataset with respect to Accuracy metric.

The IGAE results show that RandomForest is the best classifier when we select 50% or 75% of the features, but when we compare the results without using the IGAE methods, we find that it significantly decreased Accuracy when we selected 50% of the features, dropping it by

10.84%. When IGAE technique is applied to the dataset, it also reduces the Accuracy by 1.2% when 75% of the features employed, but it does not drop the accuracy extremely low and still has some stability.

Table 8: Evaluation results of the considered classifiers using IGAE

Strategy	Classifier	50% Features Accuracy	75% Features Accuracy
Trees	DecisionStump	71.140	71.140
	J48	74.150	80.540
	RandomForest	76.890	86.530
	RandomTree	72.020	80.240
	REPTree	74.660	80.730
Rules	DecisionTable	74.250	79.930
	JRip	73.680	80.470
	OneR	69.140	69.140
	PART	73.870	77.520
	ZeroR	50.330	50.330
Meta	AdaBoostM1	72.510	72.530
	ClassificationViaRegression	73.190	78.360
	FilteredClassifier	74.030	78.640
	LogitBoost	72.650	73.070
	MultiClassClassifier	72.760	72.640
Lazy	IBK	73.370	78.840
	KStar	75.800	84.860
	LWL	71.440	71.480
Functions	Logistic	72.760	72.640
	MultilayerPerceptron	73.090	76.990
	SimpleLogistic	72.570	72.420
	SMO	72.210	72.370
	VotedPerceptron	71.550	71.950
Bayes	BayesNet	72.780	73.370
	NaiveBayes	72.540	72.310
	NaiveBayesMultinomial	71.560	70.410
	NaiveBayesUpdateable	72.540	72.310

Table 9 summarizes the results obtained from Tables 5 to 8 in order to identify the best feature selection and ranking method based on Accuracy for the twenty-seven classifiers.v

Table 9: Summarization of the results

Method	50% Features		75% Features	
	Best Accuracy	Best classifier	Best Accuracy	Best classifier
ClAE	74.520	RandomForest	78.770	RandomForest
CoAE	79.950	RandomForest	80.530	RandomForest
GRAE	76.710	RandomForest	82.180	RandomForest
IGAE	76.890	J48	86.530	RandomForest

We can also conclude that RandomForest is the dominant classifier that achieves the best accuracy results in most cases, regardless of the feature selection method being used or the number of features in the dataset.

3.5 Discussion of the Main Results and Findings

The main purpose of the research is to determine the best classifier that could highly predict landslide events among twenty-seven classifiers that belong to six learning strategies, and using several evaluation metrics such as Accuracy, Precision, TP ratio and FP ratio. Moreover, one of the main metrics to determine the best classifier is the stability of the classifier for the landslide dataset when we reduce the number of features. As we know, in many cases there are a lot of features that are not important to the problem statement, which makes it necessary to use some feature selection methods to remove these features and make the classifier more powerful. In many datasets, there are also some features that are missing that may affect the performance of the classifier, and for that reason, in this research, we focus on the stability of the classifier even if we reduce the number of features.

In this research, we got the results of 27 classifiers that belong to different learning strategies without any feature selection methods, and after that, we applied four well-mown feature selection methods to test the stability of the classifier, after some features were dropped.

When we applied the classifiers without feature selection methods, we concluded that RandomForest, which belongs to Tree learning strategies, had the best performance among them, with 87.73% Accuracy rate. Then, we applied the four feature selection methods to all classifiers to test if the RandomForest would have the best performance, and for that, we chose 50% and 75% of the features of the considered dataset.

According to the results of the feature selection methods, we can conclude that RandomForest is the best classifier when we choose 50% or 75% of the features in many cases. However, regarding the performance of the considered feature selection methods, we can see that the CoAE method is better than the ClAE, GRAE, and IGAE methods when we choose 50% of the features. Moreover, when we choose 75% of the features, IGAE has the best Accuracy.

From the results of the best feature selection method, which indicate that CoAE is the best one when we choose 50% of the features, we can see that the Accuracy drops significantly to 7.78%, where it drops by 1.2% when we apply IGAE when it is used with 75% of the features, which indicates that the Random Forest does not drop the Accuracy extremely low and still has some stability.

4 Conclusion and Future Work

In this paper, an investigation regarding the best classifier that can highly and accurately predict landslides has been conducted. The investigation considered twenty-seven different classifiers that belong to six well-known learning strategies. Also, four popular feature selection and ranking methods have been considered in the investigation. Moreover, a primary dataset has been collected specifically from an area that suffers continuously from landslides with several unique features that have not been investigated enough in the previous studies. The results showed that Random Forest is the best classifier to predict landslide, and both CoAE and IGAE are the best choice to handle the task of feature selection in such datasets.

Acknowledgement

This research is partially funded by Zarqa University

References

[1] Pham, B. T., Prakash, I., & Bui, D.T. (2018). Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees.*Geomorphology*, **303**, 256-270.

[2] A.M. Alrabei, Green electronic auditing and accounting information reliability in the Jordanian social security corporation: the mediating role of cloud computing, *International Journal of Financial Studies*,**11**, 114 (2024).

[3] A.M. Alrabei, The mediating effect of COVID 19—pandemic on the Nexus between accounting information systems reliability and e-commerce: from the perception of documentary credit employees, *Inf. Sci. Lett.*,**12**, 2867-2876 (2023).

[4] Zhang, Y., Tang, J., Cheng, Y., Huang,L., Guo, F., Yin, X., & Li, N. (2022). Prediction of landslide displacement with dynamic features using intelligent approaches. *International Journal of Mining Science and Technology*, **32**(3), 539-549.

[5] A.M. Alrabei, O. Jawabreh, A.M.M. Saleh, Accounting Information and Role It on Financial Reports Quality in Jordanian Hotels, and Social Performance as a Mediating Effect, *International Journal of Sustainable Development & Planning*,**18**, 2271-2279 (2023).

[6] A.M.A. Alrabei, Perception of Jordanian Banks Employees on the Relationship between Accounting Information Quality (AIQ) and Documentary Credits, *International Journal of Applied Business and Economic Research*,**15**, 409-419 (2017).

[7] Han, H., Shi, B., & Zhang, L. (2021). Prediction of landslide sharp increase displacement by SVM with considering hysteresis of groundwater change. *Engineering Geology*, **280**, 105876.

[8] A.M. Alrabei, The influence of accounting information systems in enhancing the efficiency of internal control at Jordanian commercial banks, *Journal of Management Information and Decision Sciences*,**24**, 1-9 (2021).

[9] A.M. Alrabei, D.S. Ababnehi, The Moderating Effect of Information Technology on the Relationship between Audit Quality and the Quality of Accounting Information: Jordanian Auditors' Perception, *Journal of Theoretical and Applied Information Technology*,**99**, 3365-3378 (2021).

[10] Shan, Y., Chen, S., & Zhong, Q. (2020). Rapid prediction of landslide dam stability using the logistic regression method. *Landslides*, **17**, 2931-2956.

[11] A.M. Alrabei, L.N. Al-Othman, T.A. Abutaber, M.S. Alathamneh, T.M. Almomani, M.H. Qeshta, S.A.M. Amareen, Nexus between Intellectual Capital and Financial Performance Sustainability: Evidence from Listed Jordanian Firms. *Appl. Math.*,**17**, 881-888 (2023).

[12] A.M. Alrabei, L.N. Al-Othman, F.A. Al-Dalabih, T. Taber, B.J. Ali, S.A.M. Amareen, The impact of mobile payment on the financial inclusion rates, *Information Sciences Letters*,**11**, 1033-1044 (2022).

[13] Alazaidah, R., Ahmad, F. K., & Mohsin, M. (2020). Multi label ranking based on positive pairwise correlations among labels. *The International Arab Journal of Information Technology*, **17**(4), 440-449.

[14] A. Jahmani, O. Jawabreh, R. Abokhoza, A.M. Alrabei, The impact of marketing mix elements on tourist's satisfaction towards Five Stars Hotel Services in Dubai during COVID-19, *Journal of Environmental Management & Tourism*,**14**, 335-346 (2023).

[15] T. Almomani, M. Almomani, M. Obeidat, M. Alathamneh, A. Alrabei, M. Al-Tahrawi, D. Almajali, Audit committee characteristics and firm performance in Jordan: The moderating effect of board of directors' ownership, *Uncertain Supply Chain Management*,**11**, 1897-1904 (2023).

[16] Junoh, A. K., Ahmad, F. K., Mohsen, M.F. M., & Alazaidah, R. (2018). Open research directions for multi label learning. In *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)* (pp. 125-128). IEEE.

[17] H.A. Owida, N.M. Turab, J. Al-Nabulsi, Carbon nanomaterials advancements for biomedical applications, *Bulletin of Electrical Engineering and Informatics*,**12**, 891-901 (2023).

[18] H. Abu Owida, Recent biomimetic approaches for articular cartilage tissue engineering and their clinical applications: narrative review of the literature, *Advances in Orthopedics*,**2022**, 8670174 (2022).

[19] Alazaidah, R., Ahmad, F. K., & Mohsen, M. F. M. (2017). A comparative analysis between the three main approaches that are being used to solve the problem of multi label classification. *International Journal of Soft Computing*,**12**(4), 218-223.

[20] H.A. Owida, J.I. Al-Nabulsi, N.M. Turab, F. Alnaimat, H. Rababah, M. Shakour, Autocharging techniques for implantable medical applications, *International Journal of Biomaterials*,**2021**, 6074657 (2021).

[21] H.A. Owida, B.A.H. Moh'd, N. Turab, J. Al-Nabulsi, S. Abuowaida, The Evolution and Reliability of Machine Learning Techniques for Oncology, *International Journal of Online & Biomedical Engineering*,**19**, 110 (2023).

[22] Alluwaici, M. A., Junoh, A. K., & Alazaidah, R. (2020). New problem transformation method based on the local positive pairwise dependencies among labels. *Journal of Information & Knowledge Management*, **19** (01), 2040017.

[23] H. Abu Owida, Developments and clinical applications of biomimetic tissue regeneration using 3D bioprinting technique, *Applied Bionics and Biomechanics*,**2022**, 2260216 (2022).

[24] H.A. Owida, O.S.M. Hemied, R.S. Alkhawaldeh, N.F.F. Alshdaifat, S.F.A. Abuowaida, Improved deep learning approaches for covid-19 recognition in ct images, *Journal of Theoretical and Applied Information Technology*,**100**, 4925-4931 (2022).

[25] Junoh, A. K., AlZoubi, W. A., Alazaidah, R., & Al-luwaici, W. (2020). New features selection method for multi-label classification based on the positive dependencies among labels. *Solid State Technology*, **63**(2s).

[26] H. Abu Owida, G. AlMahadin, J.I. Al-Nabulsi, N. Turab, S. Abuowaida, N. Alshdaifat, Automated classification of brain tumor-based magnetic resonance imaging using deep learning approach, *International Journal of Electrical & Computer Engineering*,**14**, 3150 (2024).

[27] H.A. Owida, M.R. Hassan, A.M. Ali, F. Alnaimat, A. Al Sharah, S. Abuowaida, N. Alshdaifat, The performance of artificial intelligence in prostate magnetic resonance imaging screening, *International Journal of Electrical and Computer Engineering*,**14**, 2234-2241 (2024).

[28] Al-Batah, M. S., Alzyoud, M., Alazaidah, R., Toubat, M., Alzoubi, H., & Olaiyat, A. (2022). Early Prediction of Cervical Cancer Using Machine Learning Techniques. *Jordanian Journal of Computers and Information Technology (JJCIT)*, **8**(04).

[29] O. Alidmat, H.A. Owida, U.K. Yusof, A. Almaghthawi, A. Altalidi, R.S. Alkhawaldeh, S. Abuowaida, N. Alshdaifat, J. AlShaqsi, Simulation of crowd evacuation in asymmetrical exit layout based on improved dynamic parameters model, *IEEE Access*,**11**, 1-15 (2024).

[30] R. Alazaidah, H.A. Owida, N. Alshdaifat, A. Issa, S. Abuowaida, N. Yousef, A comprehensive analysis of eye diseases and medical data classification. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*,**22**, 1422-1430 (2024).

[31] Zare, M., Pourghasemi, H.R., Vafakhah, M. & Pradhan, B. (2013). Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. *Arabian Journal of Geosciences*, **6**, 2873-2888.

[32] H.A. Owida, N. Alshdaifat, A. Almaghthawi, S. Abuowaida, A. Aburomman, A. Al-Momani, M. Arabiat, H.Y. Chan, Improved deep learning architecture for skin cancer classification, *Indonesian Journal of Electrical Engineering and Computer Science*,**36**, 501-501 (2024).

[33] N. Alshdaifat, H.A. Owida, Z. Mustafa, A. Aburomman, S. Abuowaida, A. Ibrahim, W. Alsharafat, Automated blood cancer detection models based on EfficientNet-B3 architecture and transfer learning, *Indonesian Journal of Electrical Engineering and Computer Science*,**36**, 1731-1738 (2024).

[34] Jebur, M.N., Pradhan, B. & Tehrany, M.S. (2014). Manifestation of LiDAR-derived parameters in the spatial prediction of landslides using novel ensemble evidential belief functions and support vector machine models in GIS. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **8** (2), 674-690.

[35] Peng, L., Niu, R., Huang, B., Wu, X., Zhao, Y. & Ye, R. (2014). Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China. *Geomorphology*, **204**, 287-301.

[36] Marjanović, M., Kovačević, M., Bajat, B. & Voženílek, V. (2011). Landslide susceptibility assessment using SVM machine learning algorithm. *Engineering Geology*, **123**(3), 225-234.

[37] Kavzoglu, T., Sahin, E.K. & Colkesen, I. (2014). Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides*, **11**, 425-439.

[38] Lee, S., Ryu, J.H. & Kim, I.S. (2007) . Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: case study of Youngin, Korea. *Landslides*, **4**, 327-338.

[39] Durrant, R.J. & Kabán, A. (2010). Compressed Fisher linear discriminant analysis: Classification of randomly projected data. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1119-1128).

[40] Murillo-García, F.G. & Alcántara-Ayala, I. (2015). Landslide susceptibility analysis and mapping using statistical multivariate techniques: Pahuatlán, Puebla, Mexico. *Recent advances in modeling landslides and Debris flows*, 179-194.

[41] Song, Y., Gong, J., Gao, S., Wang, D., Cui, T., Li, Y. & Wei, B. (2012). Susceptibility assessment of earthquake-induced landslides using Bayesian network: A case study in Beichuan, China. *Computers & Geosciences*, **42**, 189-199.

[42] Tien Bui, D., Pradhan, B., Lofman, O. & Revhaug, I. (2012). Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and Naive Bayes Models. *Mathematical problems in Engineering*, 2012.

[43] Alazaidah, R., Samara, G., Almatarneh, S., Hassan, M., Aljaidi, M. & Mansur, H. (2023). Multi-Label Classification Based on Associations. *Applied Sciences*, **13**(8), 5081.

[44] Banimustafa, A., Almatarneh, S., Bulkrock, O., Samara, G. & Aljaidi, M. (2022). A Data Science Approach for Predicting Crowdfunding Success. In *2022 International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI)* (pp. 1-6). IEEE.

[45] Tien Bui, D., et al. (2019). Shallow landslide prediction using a novel hybrid functional machine learning algorithm. *Remote Sensing*, **11**(8), 931.

[46] Chae, B.G., Park, H.J., Catani, F., Simoni, A. & Berti, M. (2017). Landslide prediction, monitoring and early warning: a concise review of state-of-the-art. *Geosciences Journal*, **21**, 1033-1070.

[47] Al-Batah, M.S., et al. (2022). Early prediction of cervical cancer using machine learning techniques. *Jordanian Journal of Computers and Information Technology (JJCIT)*, **8**(04).

[48] Qasem, M.H., et al. (2022). Multi-agent systems for distributed data mining techniques: an overview. *Big Data Intelligence for Smart Applications*, 57-92.

[49] Alhusenat, A. Y., et al. (2023). A secured multi-stages authentication protocol for IoT devices. *Mathematical Modelling of Engineering Problems*, **10**(4).

[50] Goetz, J.N., Brenning, A., Petschko, H. & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Computers & geosciences*, **81**, 1-11.

[51] Ahmed, B., Rahman, M.S., Sammonds, P., Islam, R. & Uddin, K. (2020). Application of geospatial technologies in developing a dynamic landslide early warning system in a humanitarian context: the Rohingya refugee crisis in Cox's Bazar, Bangladesh. *Geomatics, Natural Hazards and Risk*, **11**(1), 446-468.

[52] Owida, H. A., et al. (2023). The evolution and reliability of machine learning techniques for oncology. *International Journal of Online & Biomedical Engineering*, **19**(8).

[53] Abu Owida, H., et al. (2024). Automated classification of brain tumor-based magnetic resonance imaging using deep learning approach. *International Journal of Electrical & Computer Engineering*, **14**(3) .

[54] Owida, H. A., et al. (2021). Autocharging techniques for implantable medical applications. *International Journal of Biomaterials*, **2021**(1), 6074657.

Raed Alazaidah received the B.Sc. degrees in computer information system, M.Sc. degrees in computer science from Al al-Bayt University, Jordan, in 2012 and 2015, respectively, and the Ph.D. degree in computer science from Universiti Sains Malaysia, Malaysia, in 2023. Dr. Abuowaida is currently working as an assistant professor in the Computer Science Department at Zarqa University, Jordan. Her research interests include deep learning, depth estimation, point cloud processing, and computer vision. His ORCID ID is orcid.org/0000-0002-1818-4288.

Mohammad Subhi Al-Batah is a Full Professor at Jadara University, specializing in Artificial Intelligence. He earned his Bachelor's Degree in Computer Engineering from Philadelphia University, Jordan, in 2001, followed by a Master's Degree in Computer and Communication Engineering from Universiti Kebangsaan Malaysia (UKM) in 2004, where he conducted research on IEEE 802.11 Wireless LAN performance. In 2009, he completed his Ph.D. in Artificial Intelligence at Universiti Sains Malaysia (USM), focusing on developing an intelligent classification system using image processing and neural networks. With a strong foundation in AI and over two decades of academic and research experience, Dr. Al-Batah is recognized for his contributions to intelligent systems and advanced computational methods.

Ghassan Samara is an academic and researcher, currently serving as an Associate Professor and Vice Dean of the Faculty of Information Technology at Zarqa University, Jordan. He earned his Ph.D. in Computer Networks from Universiti Sains Malaysia in 2012 and has extensive expertise in areas such as the Internet of Things (IoT), cryptography, vehicular ad hoc networks (VANETs), wireless sensor networks (WSNs), cloud computing, network security, intelligent systems, energy management, and digital forensics. His ORCID ID is orcid.org/0000-0001-8415-0572.

Mo'ath Alluwaici is an assistant professor in computer science department at Zarqa university. Alluwaici got his PhD and MSc degrees from UniMAP University in Malaysia in Applied Mathematics-AI. Alluwaici got his BSc degree in Mathematics and Statistics from JUST university-Jordan. His main interests are machine learning, Fuzzy logic, and data science.

Mohammed Abu Safaqah is an assistant professor in Mechanical & Building Service Engineering Department in the Global College of Science & Technology. Mr. Mohammed did both the bachelor and the master in Structural Mechanics at Jordan University of Science & Technology in Jordan, from which he graduated in 2013. During his master, he focused on an interesting field which was the Functionally Graded materials (FGM). I am also open to exploring other interesting topics in the field of Building Services Engineering and mechanics including structural insulation, smart homes, and structural reliability. Also, I'm very interested in Nano technology and its applications in Civil and Environmental Engineering.

Mohammad Aljaidi received his B.Sc. (with distinction) in computer science from Zarqa University, Jordan, in 2014; the M.Sc. (with distinction) in computer science from Zarqa University, Jordan, in 2017; and the Ph.D. degree in computer science and artificial intelligence from the University of Northumbria at Newcastle, UK, in 2022. Dr. Aljaidi is currently working as an assistant professor with the Computer Science Department, Zarqa University, Jordan. His research interests include but are not limited to, EVs charging management and development, wireless sensor networks (WSNs), AI, cybersecurity, optimization, reinforcement learning. His ORCID ID is orcid.org/0000-0001-9486-3533.

Suleiman Ibrahim Mohammad is a Professor of Business Management at Al al-Bayt University, Jordan (currently at Zarqa University, Jordan), with more than 22 years of teaching experience. He has published over 400 research papers in prestigious journals. He holds a PhD in Financial Management and an MCom from Rajasthan University, India, and a Bachelor's in Commerce from Yarmouk University, Jordan. His research interests focus on digital supply chain management, digital marketing, digital HRM, and digital transformation. His ORCID ID is orcid.org/0000-0001-6156-9063.

Asokan Vasudevan is a distinguished academic at INTI International University, Malaysia. He holds multiple degrees, including a PhD in Management from UNITEN, Malaysia, and has held key roles such as Lecturer, Department Chair, and Program Director. His research, published in esteemed journals, focuses on business management, ethics, and leadership. Dr. Vasudevan has received several awards, including the Best Lecturer Award from Infrastructure University Kuala Lumpur and the Teaching Excellence Award from INTI International University. His ORCID ID is orcid.org/0000-0002-9866-4045.

Yujie Zhang is a faculty member at the Faculty of Liberal Arts, Shinawatra University, Thailand. Zhang actively contributes to academic and research activities within the discipline, fostering interdisciplinary learning and innovation. With expertise in liberal arts education, Yujie Zhang plays a vital role in supporting the university's mission to promote academic excellence and global engagement.