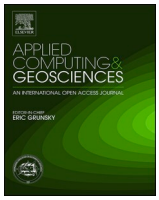
Applied Computing and Geosciences 24 (2024) 100203

Contents lists available at ScienceDirect

Applied Computing and Geosciences

journal homepage: www.sciencedirect.com/journal/applied-computing-and-geosciences



Mapping landforms of a hilly landscape using machine learning and high-resolution LiDAR topographic data

Netra R. Regmi [a,*], Nina D.S. Webb [b], Jacob I. Walter [a], Joonghyeok Heo [c], Nicholas W. Hayman [a]

[a] Oklahoma Geological Survey, The University of Oklahoma, Norman, OK, USA
[b] Department of Physics, Washington University in St. Louis, St. Louis, MO, USA
[c] Department of Geosciences, University of Texas – Permian Basin, Odessa, TX, USA

ARTICLE INFO

Keywords:
Hilly landscapes
Landform mapping
Ozark mountains
Machine learning
Random forest modeling

ABSTRACT

Landform maps are important tools in assessment of soil- and eco-hydrogeomorphic processes and hazards, hydrological modeling, and natural resources and land management. Traditional techniques of mapping landforms based on field surveys or from aerial photographs can be time and labor intensive, highlighting the importance of remote sensing products based automatic or semi-automatic approaches. In addition, the time-intensive manual labeling can also be subjective rather than an objective identification of the landform. Here we implemented such an objective approach applying a random forest machine learning algorithm to a set of observed landform data and 1m horizontal resolution bare-earth digital elevation model (DEM) developed from airborne light detection and ranging (LiDAR) data to rapidly map various landforms of a hilly landscape. The landform classification includes upland plateaus, ridges, convex slopes, planar slopes, concave slopes, stream channels, and valley bottoms, across a 400 km^2 hilly landscape of the Ozark Mountains in northeastern Oklahoma. We used 4200 landform observations (600 per landform) and eight topographic indices derived from 2 m, 5 m and 10 m resolution LiDAR DEM in random forest algorithm to develop 2 m, 5 m and 10 m resolution landform models. We test the effectiveness of DEM resolution in mapping landforms via comparison of observed landforms with modeled landforms. Results showed that the approach mapped ~84% of observed landforms when covariates were at 2 m resolution to ~89% when they were at 10 m resolution. However, predicted maps showed that the 2 m resolution covariates performed better at capturing accurate landform boundaries and details of small-sized landforms such as stream channels and ridges. The approach presented here significantly reduces the time required for mapping landforms compared to traditional techniques using aerial imagery and field observations and allows incorporation of a wide variety of covariates. The landform map developed using this approach has several potential applications. It could be utilized in hydrological modeling, natural resource management, and characterizing soil-geomorphic processes and hazards in a hilly landscape.

1. Introduction

Landform maps are important tools for natural resources and agricultural management, assessment of soil-spatial variation, erosion potential, hydrological modeling, quantifying biophysical and biogeochemical functioning, and mapping soil- and hydro-geomorphic hazards (Garcia-Aguirre et al., 2007; Khan et al., 2014; Regmi and Rasmussen, 2018; Summerell et al., 2005). Traditional approaches of mapping landforms rely on field investigations and manual interpretation of topographic maps and aerial photographs (Dikau et al., 1991; Hammond, 1964). However, such mapping approaches are time consuming, labor intensive, and subjective. Over the last three decades, advances in remote sensing techniques have provided variety of products, including satellite imagery and digital elevation model (DEM), which have been used in mapping landforms in multiple scales, using expert-based, qualitative, and quantitative approaches (Burrough et al., 2000; Irvin et al., 1997; Jasiewicz and Stepinski, 2013; Mashimbye and Loggenberg, 2023; Minár et al., 2023; Pennock and Corre, 2001; Prima et al., 2006; Regmi et al., 2017; Regmi and Rasmussen, 2018; Siervo et al., 2023; Smith et al., 2006; Smith and Clark, 2005; Veronesi and

* Corresponding author.
E-mail address: netraregmi@ou.edu (N.R. Regmi).

https://doi.org/10.1016/j.acags.2024.100203
Received 4 June 2024; Received in revised form 18 October 2024; Accepted 22 October 2024
Available online 23 October 2024

Hurni, 2014; Zhao et al., 2017). Two significant advances in quantitative approaches can broadly be categorized as: (i) unsupervised classification where the mapping is based primarily on univariate and multivariate clustering and pattern recognition of surface topographic geometries derived from remote sensing products (i.e. slope gradient and curvature from DEM) (Burrough et al., 2000; Irvin et al., 1997; Regmi and Rasmussen, 2018) and (ii) supervised classification where models are developed by establishing empirical or statistical relationships between observed dataset and surface topographic geometries (Mithan et al., 2019; Prima et al., 2006). These approaches implement quantitative techniques, such as fuzzy logic and ISODATA classification (MacMillan et al., 2000; Regmi and Rasmussen, 2018), object-based segmentation (Bishop et al., 2001; Drăguţ and Blaschke, 2006), pattern recognition (Jasiewicz and Stepinski, 2013; Libohova et al., 2016), multivariate statistical analysis (Adediran et al., 2004; Friedrich, 1998), and machine learning algorithms including random forest classification (Veronesi and Hurni, 2014), deep learning (Du et al., 2019; Li et al., 2020) and many other (Middleton et al., 2020; Stepinski et al., 2007a, 2007b). However, only limited studies have applied such techniques to high-resolution LiDAR topographic data (Bishop et al., 2012) in order to prepare a high-resolution map of landforms across large hilly landscapes. Here we implement a random forest algorithm, a supervised machine learning approach, over topographic geometries derived from high-resolution bare-earth light detection and ranging (LiDAR) topographic data to test the effectiveness of high-resolution DEM in differentiating landforms of a range of type and size in a hilly landscape. We also test data-resolution sensitivity in mapping such landforms, which may be important for working with future satellite-based time series data.

2. Landforms in the study area

A landform can be defined simply as a physical feature of the Earth's surface, developed by natural processes. It possesses distinguishable shape and geometry that can be recognized relative to adjacent geomorphic features (Evans, 2012). Landscapes in different environments exhibit landforms of a range of types and sizes, however, landforms such as ridges, hillslopes, plateaus, channels, and floodplains are common in any of the hilly and mountainous landscapes (Siervo et al., 2023). Each of these landforms possesses unique surface geometries, cover a large part of such landscapes, and comprise various domains formed by soil-hydrogeomorphic and sediment transport processes (DiBiase et al., 2017; Roering et al., 1999). Consequently, they play a crucial role in soil-hydrogeomorphic processes and hazards, such as mass movement and hillslope and stream erosion (Berhe et al., 2008; Derakhshan-Babaei et al., 2021). Many of these landforms have relatively distinct boundaries and can be delineated from field surveys and remote sensing image-based mapping (Evans, 2012). Hillslopes, however, are considered as a separate mapping problem, and further classified into smaller units, such as concave, planar and convex geometries, which often been termed as land surface form, element or component (Evans, 2012; Pitty, 2020; Romstad and Etzelmüller, 2009). Mapping these units of a large hilly landscape is challenging given their sizes tend to be relatively small and their characteristics vary continuously across the landscape. Given these challenges, to facilitate systematic mapping of landforms we explore below a machine-learning remote sensing method.

The study area is a ~400 km^2 forested hilly region with moderately rugged topography and dendritic drainage patterns in the Ozark Mountains in the US southern mid-continent (Figs. 1 and 2). The

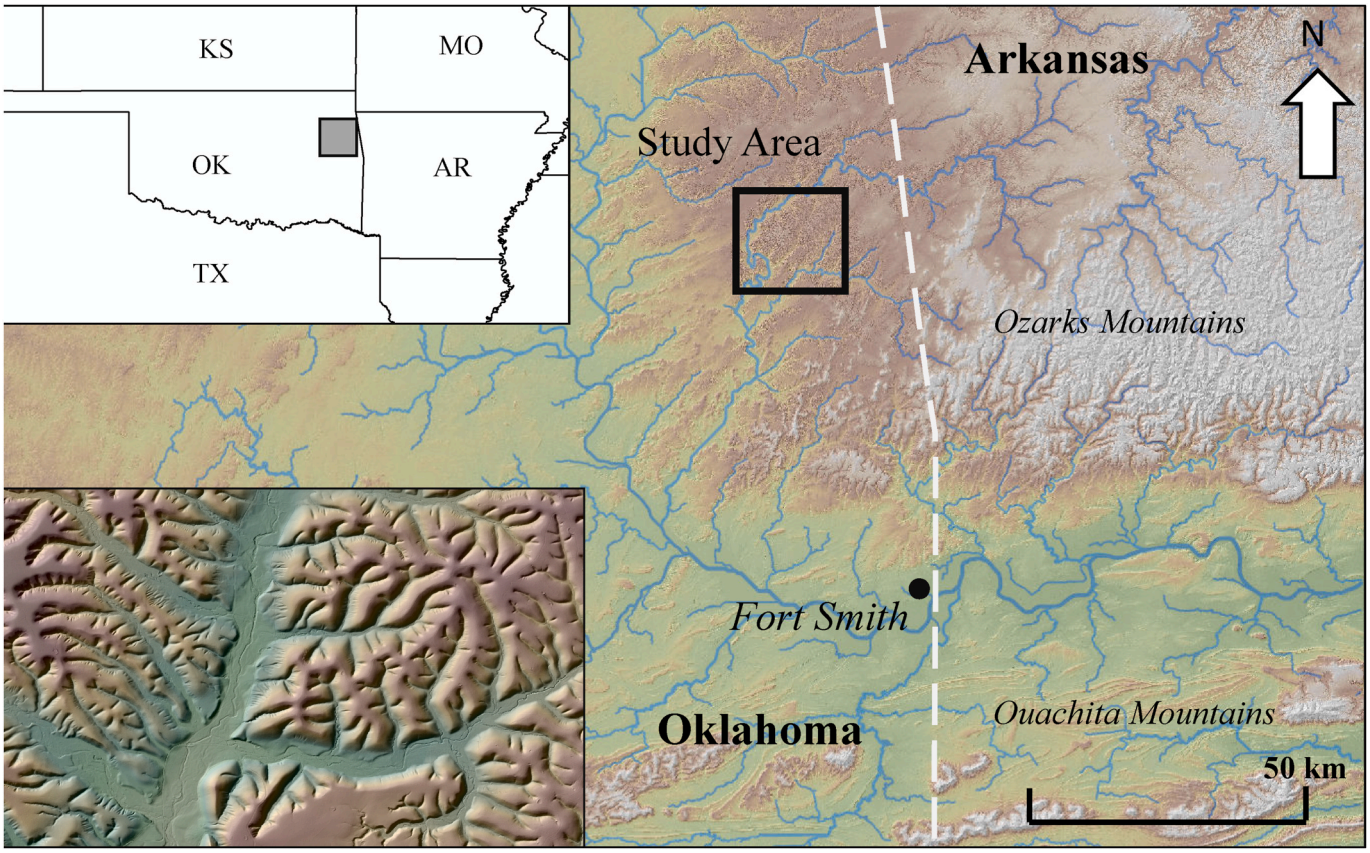


Fig. 1. Location of the study. The study area is a part of the Ozark Mountains in eastern Oklahoma, USA. The insets show the location of the study area within the state of Oklahoma (OK) and a shaded-relief image of a portion of the study area, developed from 1 m LiDAR topographic data. Acronyms: KS-Kansas, MO-Missouri, OK-Oklahoma, AR-Arkansas, TX-Texas.

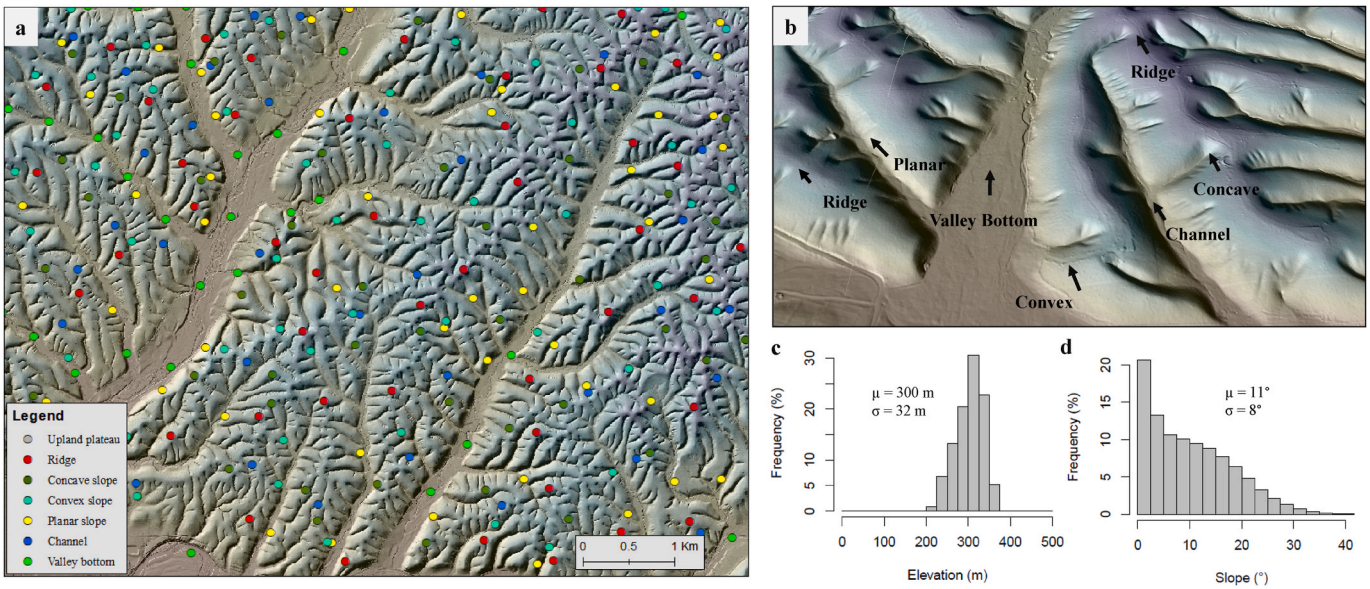


Fig. 2. (a) Locations of observed landforms in a part of the study area. (b) Shaded relief image derived from 1 m bare-earth LiDAR topographic data showing examples of target landforms. (c, d) Histograms showing the distribution of elevation and slope. The mean, μ, and standard deviation, σ, values of slope and elevation are listed.

elevation ranges from 200 m to 400 m with an average of 300 m and a standard deviation of 32 m (Fig. 2). Hillslopes are relatively gentle with an average slope gradient of 11° and a standard deviation of 8°. The geology is dominated by limestone and chert as the primary bedrock lithology and shale and marlstone as secondary lithology (Heran et al., 2003) (See Supplementary Fig. S1). Quaternary-age materials mantle many slopes. Dominant hillslopes consist of a thin veneer of soil with patchy bedrock exposure, and a few colluvial and talus deposits including small-scale colluvial fans and talus cones are also present. Major landforms including ridges, concave, convex and planar hillslopes, channels, and valley bottoms are well developed. These landforms exist in a range of spatial scales ranging from narrow ridges and stream channels to relatively broad valley bottoms and upland plateaus (Figs. 1 and 2). Planar, concave and convex hillslopes are medium in size. Erosional features are also present, such as small-scale rills and gullies.

3. Materials and methods

We consider modeling using high-resolution remote sensing datasets, such as LiDAR topographic data, and a robust method for handling large volumes of remote sensing data (i.e., machine learning) can be a time-

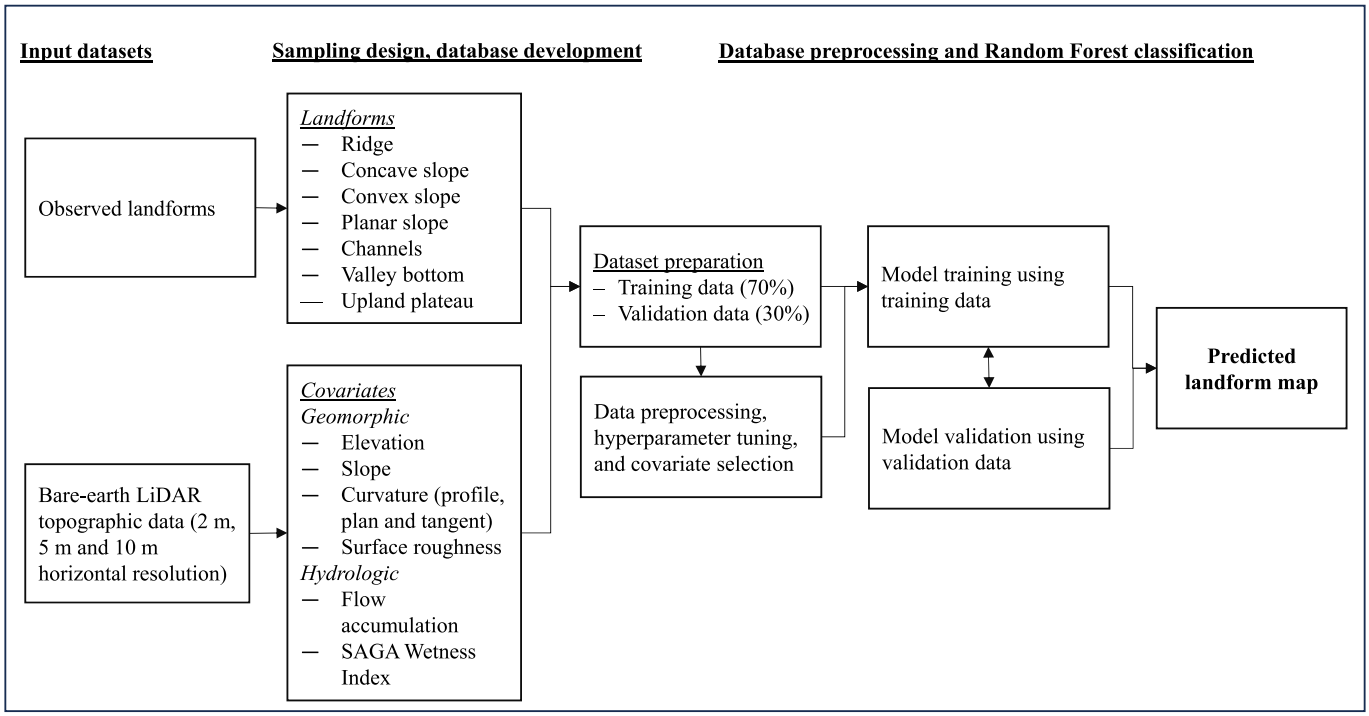


Fig. 3. Flowchart of methodology.

and cost-effective way to characterize landforms of a hilly landscape. We here test this approach using a set of observed landform data, 1 m bare-earth LiDAR DEM and a random forest supervised machine learning algorithm across a 400 km^2 hilly landscape. We specifically segment the landscape into areas of relatively homogeneous surface geometries based on the similarity in pattern recognized from DEM-derived surface geometries. We then classify these segments into different landform types. The methodology of this study consists of three main steps (Fig. 3). First, we prepare the input dataset including observed landforms and associated topographic indices (covariates). The step is followed by the preparation and preprocessing of training and validation dataset. The second step involves determining all the required parameters (i.e. hyperparameter) needed to run random forest classification, and validation of the model. Finally, we assess the performance of the model, test the validity of the models, and compare models to determine the best performing model. We then use the best performing models to develop landform maps.

3.1. Input dataset

3.1.1. Observed landforms

We used 1 m horizontal resolution bare-earth LiDAR DEM from USGS 3DEP program (https://www.usgs.gov/core-science-systems/ngp/3dep) and high-resolution imagery available in Google Earth to prepare a set of observed landforms required for supervised learning. Seven landform types including upland plateau (flat or very gentle slopes upland), ridge, concave, convex and planar hillslopes, stream channel, and valley bottom, were identified at randomly distributed locations across the study area based on manual interpretation of 1 m DEM-derived hillshade images and aerial imagery and represented by points (Fig. 2a). Altogether 600 randomly distributed points per landform (total 4200 points) were mapped from hillshade image (Fig. 2a) and verified by three-dimensional visualization of high-resolution aerial imagery and LiDAR DEM in ArcGIS®. While rills and gullies, bedrock exposure, small streams, and braided channels are visible upon close inspection of the DEM, we exclude them from the automatic classification because the resolution of the 1 m LiDAR DEM may limit their automatic identification.

3.1.2. Covariates

Several topographic indices are derived from the LiDAR DEM and used as input independent variables or covariates in the random forest landform classification (Table 1). One limitation of LiDAR topographic data is it commonly exhibits high local variability in surface geometries because of the presence of pits associated with the upheaval or decay of tree roots, or dense vegetation or brush that has been misclassified as bareearth (Lashermes et al., 2007). We smoothed the DEM by averaging elevations within roving windows of 12 m search radius, which is a scale slightly larger than the scale at which most of these features exist (Lashermes et al., 2007; Regmi et al., 2019; Roering et al., 2010). Our study area is devoid of dense forest and significant human disturbances, such as roads and buildings. However, the area has clumps of brush and trees that introduced some uncertainties into the DEM. The smoothed DEM was then resampled into 2 m, 5 m and 10 m DEMs implementing a cubic convolution interpolation technique to test the effectiveness of DEM resolution in mapping landforms. One-meter DEMs were considered unsuitable for this mapping because of the presence of noise associated with larger vegetation pits, human-made structures including roads and trails, and slope materials which could not be ameliorated by smoothing. DEMs coarser than 10 m resolution were not used because of the small spatial scale of some of the landforms (i.e. channels and concave slopes) (Fig. 2).

Table 1
Environmental covariates used and their significance in hillslope processes.

Covariates	Significance
Elevation	Potential energy and micro-climate
Slope	Steepness, overland and sub-surface flow velocity
Curvature (plan, profile and tangent)	3D surface geometry (i.e., concave, convex or planar) along plan and profile forms as well as along the tangent of plan and profile planes. Thus, direction and convergence or divergene of flow.
Surface Roughness	Variation in gradient and curvature along and across the slope as a result of variation in particle size distribution and surface features including rills, gullies, vegetation mounds, etc.
Flow Accumulation	Upstream contributing area, and thus the amount of water derived upslope
SAGA wetness index	Potential of soil water content

From each resampled DEM maps of eight topographic indices including elevation, slope, plan curvature, profile curvature, tangent curvature, surface roughness, upstream flow accumulation and SAGA wetness index (Table 1) were developed. These indices were selected following previous studies that implement DEM-derived topographic indices for mapping landforms (Meles et al., 2020; Minar and Evans, 2008; Regmi and Rasmussen, 2018) and observations of their distributions across different landform types (Supplementary Figs. S2–S4). Slope, curvature (plan, profile and tangent), surface roughness and flow accumulation were developed using inbuilt algorithms in ArcGIS® software. All of these maps were computed from DEM using a 3×3 cell moving window. Slope and curvature together revel the three-dimensional geometry of a landform, and thus can characterize the role of landform in surface hydrological and sediment transport processes including flow velocity, flow direction, and convergence or divergence of flow (Roering et al., 2007). Surface roughness, an index that can characterize surface irregularities contributed by slope materials and erosional and depositional features (Regmi et al., 2019), was computed as a standard deviation of slope within 3×3 cell moving window following Frankel and Dolan (2007). This index can distinguish landforms with smooth surfaces (i.e., planar slopes) from landforms where the geometry changes sharply (i.e., ridges and channels). Flow accumulation computed using the D-inf algorithm (Tarboton, 1997) measures the upstream contributing area which can be used to differentiate upland and lowland landforms. For example, upland landforms, such as ridges and plateaus, tend to have significantly small upstream contributing areas compared to that of lowland landforms including channels and valley bottoms. The SAGA wetness index, defined as the natural logarithmic ratio of upstream contributing area to the tangent of the local slope (Böhner and McCloy, 2006), was developed using inbuilt algorithm in SAGA GIS ® (http://www.saga-gis.org). The index can characterize the potential of soil-water content (Raduła et al., 2018). For example, a slope having a large upstream area and gentle slope tends to have higher soil-water content, and conversely, a steep slope with small upstream contributing area tends to have lower soil-water content. In summary, all these indices can characterize both three-dimensional geometry and position of the landforms across the landscape, and provide quantitative measurements that will aid in classifying the broad surface hydrological and soil-geomorphic process domains (Montgomery, 1999) of a hilly landscape.

3.2. Sampling design and data pre-processing

Altogether 600 randomly distributed observations were made per landform (4200 total). Covariate values were extracted to each data points, and the distribution of covariate values were analyzed to determine if any outlier in the dataset exist, and then outliers and spurious values associated with DEM artifacts and noises were excluded from the further analysis. Next, we randomly divide the entire population into training (70% of the total) and validation datasets (30% of the total) where each of the datasets consists of an equal number of observations for each landform type. The rationale for using 70/30 split is that 70% of the data is expected to be sufficient for producing stable models that can capture the underlying patterns and relationships in the data,

while the remaining 30% is set aside for independent validation of the models. Both datasets were further pre-processed using the Caret package in R® (https://www.r-project.org/) (Kuhn et al., 2020). Covariates that are irrelevant or redundant from the analysis were eliminated using recursive feature elimination (RFE) technique (Kuhn, 2012). The approach iteratively trains the model, ranks covariates, and then removes the lowest-ranking covariate in a recursive fashion. The optimum covariates were then selected based on covariate ranking and the model performance obtained after each RFE. The selected datasets were then implemented for random forest hyperparameter tuning and landform model development.

3.3. Random forest classification

The random forest algorithm, which is a nonparametric supervised approach, has recently been applied to mapping landforms and associated surface processes, such as landslides, erosion and soil (Harris and Grunsky, 2015; Shruthi et al., 2014; Taalab et al., 2018; Youssef et al., 2016; Zhao et al., 2017). The algorithm performs classification and regression by building an ensemble of decision trees. Each tree is trained on a random subset of the observed dataset containing dependent (i.e., landform type) and independent variables or covariates (i.e., slope, curvature) through a process called bootstrap sampling or bagging, and provides prediction of dependent variable (Breiman and Cutler, 2011; Breiman et al., 1984). The underlying idea of ensemble approach is to generate a set of weak classifiers (individual decision trees) that, when combined, form a stronger model. By using different subsets of the training samples for each decision tree, the correlation between the trees is reduced, which improves the overall performance of the model. The approach makes overall predictions by taking the average of the decisions made by each of the trees in the case of regression, and the majority of such decisions in the case of classification.

A random forest algorithm generally utilizes various parameters in making decisions including the number of trees, the minimum number of datapoints in each node in a tree, and the number of covariates tried at each node. The split at a node is based on a random subsample of covariates in such a way that the process minimizes the regression or classification error by minimizing covariate noise, bias and variance. Nodes continue to be split until no further improvement in error is achieved. Omitted observations, which are sometimes called “out-of-bag” sample, are used to compute the errors. This is commonly referred to as the out-of-bag (OOB) error. The algorithm quantifies the importance of each covariate in model prediction by computing the mean decrease in the Gini (MDG) coefficient, which is the total decrease in node impurities from splitting on the covariates, averaged over all trees (Breiman et al., 1984; Díaz-Uriarte and De Andres, 2006; Veronesi and Hurni, 2014). A higher MDG value generally indicates higher covariate importance. However, it is often called a black box algorithm because of its limited feasibility to gain a full understanding of computation and decision process of each individual tree.

3.3.1. Hyperparameter tuning and model development

We implement the grid search-based algorithm available in the Caret package (Kuhn, 2012) to determine two hyperparameters including number of trees and the maximum number of covariates to split at nodes. In this process models were developed using combinations of several trees (50, 100, 200, 300, 400, and 500) and different numbers of RFE-selected covariates (ranging from 1 to 8). The landform models were developed using RFE-selected training dataset and the optimum hyperparameters identified. Gini index was used as a measure of the best covariate split selection, and trees were grown up to their maximum depth (Supplementary Table S1). Models were also validated implementing five-fold cross validation technique. The approach divides the dataset into five equal subsets, and during each run it uses four sets to develop a training model and tests the performance of the model using the fifth set. In such a way, the algorithm develops five models and independently tests the validity of each model. The final model, which was the average of all the models, was then applied over the RFE-selected covariate maps to develop a landform map.

3.3.2. Evaluation of model performance

Random forest classification makes overall prediction for each landform class based on the majority of the decisions made by all the trees in the forest. In addition, the approach provides a prediction probability for each class based on the proportions of trees that voted for a particular class. We assess the performance of all the models using these results. First, we develop a confusion matrix, which shows metrics of prediction including truly and falsely predicted observations. False predictions include cases where observations are mistakenly classified as other landforms (omission error) and cases where other landforms are incorrectly classified as positive (commission error). Then, the performances of individual classes and all models were determined based on ‘overall accuracy’ computed as a ratio of the number of truly predicted landform observations to the total number of the assessments. High overall accuracy suggests that the model is making accurate predictions. The approach has widely been used to distinguish multiclass predictive performance of the model (Egan and Egan, 1975; Gorsevski et al., 2006; Søreide, 2009). Second, we assess the model performance by evaluating the frequency distribution of truly and falsely predicted assessments as well as by comparing their prediction probabilities.

4. Results

4.1. Covariate selection and model development

The MDG values of all covariates included in three training models suggested that slope, SAGA wetness index, elevation, and surface roughness are some of the major covariates for mapping landforms (Fig. 4). The RFE approach identified the best model performance with 7 covariates in case of 2 m resolution, and all 8 covariates in case of 5 m and 10 m resolutions (Fig. 4). Altogether, the model performance ranges from 70% overall accuracy when only two covariates are used to 89% overall accuracy when all RFE-selected covariates are used (Fig. 4). Models developed using RFE-selected covariates and optimum hyperparameters identified (Supplementary Table S1) mapped observed landforms with overall accuracy ranging from 84% when using 2 m resolution covariates to 89% when using 10 m resolution covariates (Table 2). In each case of training model development, results from five-fold cross validations (Supplementary Table S2) show the performance of five training and five testing models are not significantly different. This suggests that the number of observed landforms and input covariates included in the training data is sufficient to generate stable models.

4.2. Model performance and validation

All three final models produced qualitatively similar landform maps (Figs. 5 and 6). Additionally, the validity assessment of training models by the validation data (30% of the total population or 180 observations per landform) showed that the predictions are not significantly different which indicates that the models are not severely overfitted (Tables 2 and 3). For example, with 2 m covariates the training model truly predicted ~84% of the training dataset and ~85% of the validation dataset which were not included in training model development. Furthermore, results show similar frequency distribution of prediction probabilities of truly predicted (matched) and falsely predicted (missed) landforms across all three models (Fig. 7 and Supplementary Fig. S5). In each case, most of the observations were truly predicted, and most of them have prediction probability close to 1. A very small number of landforms (16% by 2 m model, 11% by 5 m model, and 11% by 10 m model) were falsely predicted (Table 2). Additionally, the distribution of major covariate values in observed and predicted landforms are similar (Fig. 8 and

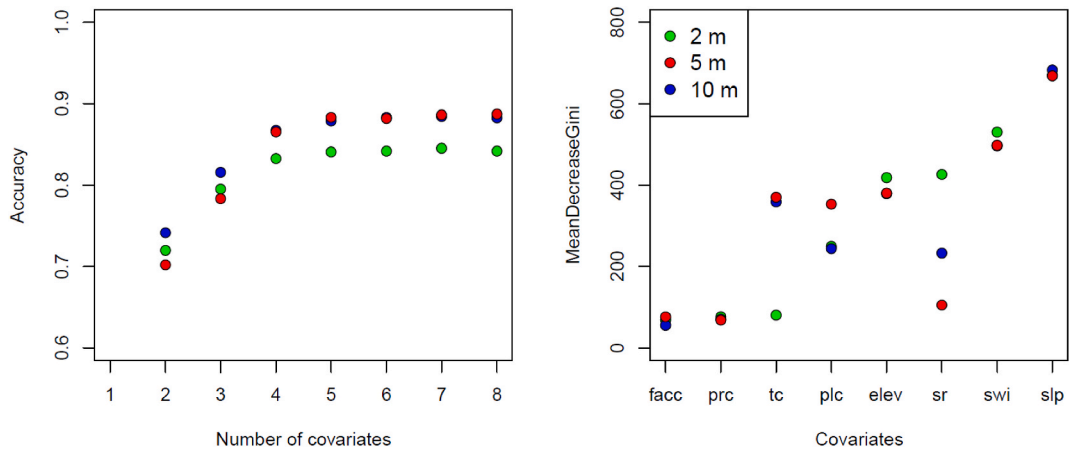


Fig. 4. (a) A plot showing the relationship between the number of covariates and overall accuracies of 2 m, 5 m and 10 m models resulted from recursive feature elimination. (b) Covariate rankings of the best-performing models quantified in terms of MeanDecrease Gini (MDG) values. A higher MDG value indicates higher importance. Acronyms: elev: elevation, slp: slope, sr: surface roughness, swi: SAGA wetness index, prc: profile curvature, tc: tangent curvature, plc: plan curvature, and facc: flow accumulation.

Table 2
Confusion matrix and performance of training models evaluated using training datasets comprising of covariates developed from 2 m, 5 m and 10 m DEMs. Both class accuracy and overall model accuracy were computed as a ratio of the number of truly predicted landform observations to the total number of assessments. UP: Upland Plateau; R: Ridge; CvS: Concave Slope; CsX: Convex Slope; PS: Planar Slope; C: Channel; VB: Valley Bottom.

Input data	Landforms	UP	R	CvS	CxS	PS	C	VB	Class accuracy (%)	Overall accuracy (%)
2 m covariates	UP	354	38	1	8	0	1	18	84.29	84.39
	R	53	302	0	53	0	4	8	71.90	
	CvS	0	0	357	2	29	31	1	85.00	
	CxS	3	36	1	345	30	2	3	82.14	
	PS	1	0	12	22	376	9	0	89.52	
	C	1	1	23	5	16	358	16	85.24	
	VB	11	8	1	1	0	10	389	92.62	
5 m covariates	UP	376	24	1	6	2	0	14	88.89	88.61
	R	34	346	0	35	0	1	3	82.58	
	CvS	0	0	380	1	14	25	0	90.48	
	CxS	1	29	0	372	17	0	1	88.57	
	PS	0	1	5	17	385	12	0	91.67	
	C	0	0	25	5	16	350	23	83.53	
	VB	8	0	0	0	0	15	398	94.54	
10 m covariates	UP	374	29	0	6	5	0	6	89.05	88.71
	R	23	373	0	22	2	0	0	88.81	
	CvS	1	1	391	1	2	23	1	93.10	
	CxS	1	34	7	360	16	2	0	85.71	
	PS	0	2	4	16	368	30	0	87.62	
	C	1	0	36	2	16	348	17	82.86	
	VB	7	0	0	1	0	18	394	93.81	

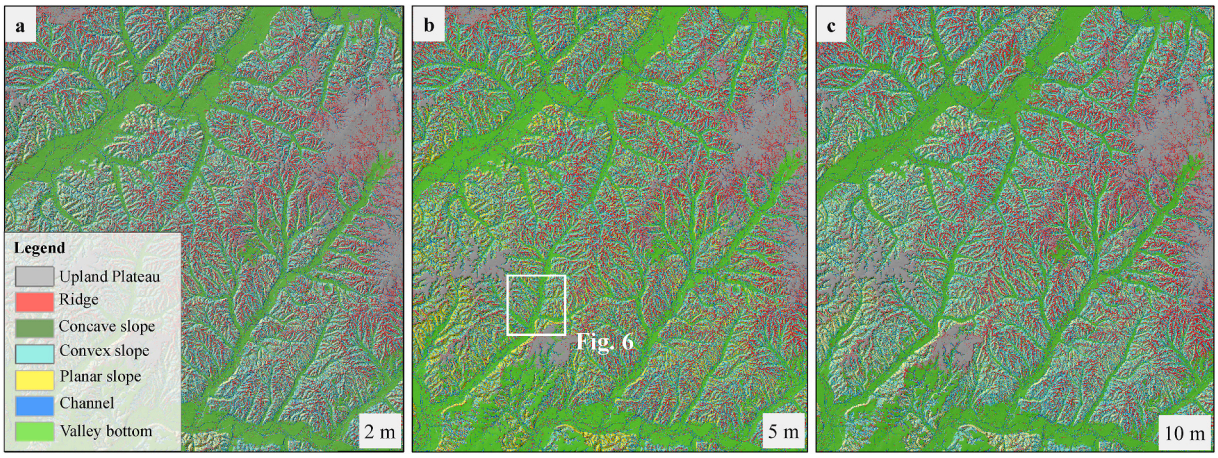


Fig. 5. Figures showing landform maps developed using training datasets comprising of: (a) 2 m, (b) 5 m, and (c)10 m covariates.

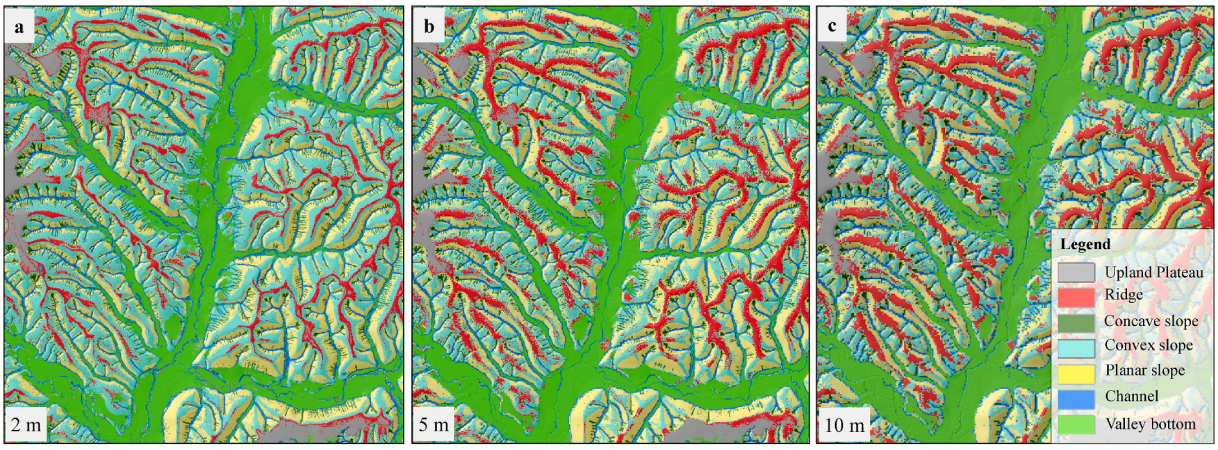


Fig. 6. Figures showing landform maps of a selected area shown in Fig. 5b. Landform maps are developed using: (a) 2 m, (b) 5 m and (c)10 m covariates. See Supplementary Figs. S8 and S9 for comparison among these landform maps.

Table 3
Confusion matrix and performance of training models evaluated using validation datasets comprising of covariates developed from 2 m, 5 m and 10 m DEMs. See Table 2 caption for acronyms.

Input data	Landforms	UP	R	CvS	CxS	PS	C	VB	Class accuracy (%)	Overall accuracy (%)
2 m covariates	UP	158	36	0	1	0	1	8	77.45	85.08
	R	7	123	0	15	0	0	0	84.83	
	CvS	0	0	158	1	5	8	0	91.86	
	CxS	6	18	0	151	6	1	1	82.51	
	PS	0	0	11	11	163	7	0	84.90	
	C	0	1	11	1	6	152	7	85.39	
	VB	8	2	0	0	0	11	164	88.65	
5 m covariates	UP	159	14	0	0	0	0	3	90.34	89.02
	R	11	152	0	14	0	0	0	85.88	
	CvS	0	0	164	1	0	13	0	92.13	
	CxS	3	13	0	156	3	2	0	88.14	
	PS	0	0	4	8	173	6	0	90.58	
	C	0	0	11	1	4	147	6	86.98	
	VB	7	1	1	0	0	12	171	89.06	
10 m covariates	UP	165	8	0	2	0	0	7	90.66	88.46
	R	8	155	0	11	0	0	2	88.07	
	CvS	0	0	159	0	0	13	0	92.44	
	CxS	1	16	1	160	10	1	0	84.66	
	PS	0	1	5	5	157	4	0	91.28	
	C	0	0	15	0	13	154	8	81.05	
	VB	6	0	0	2	0	8	163	91.06	

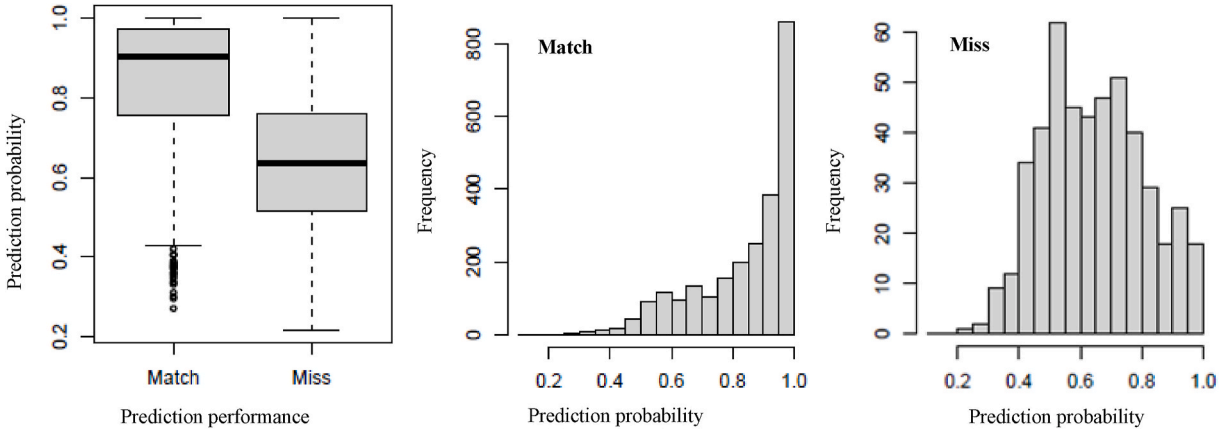


Fig. 7. Prediction probability of truly predicted (match) and falsely predicted (missed) landforms of the validation datasets by final models developed using 2 m covariates. Note the median prediction probabilities for matched and missed landforms are 0.87 and 0.62. Similar distributions for 5 m and 10 m models are shown in Supplementary Fig. S5.

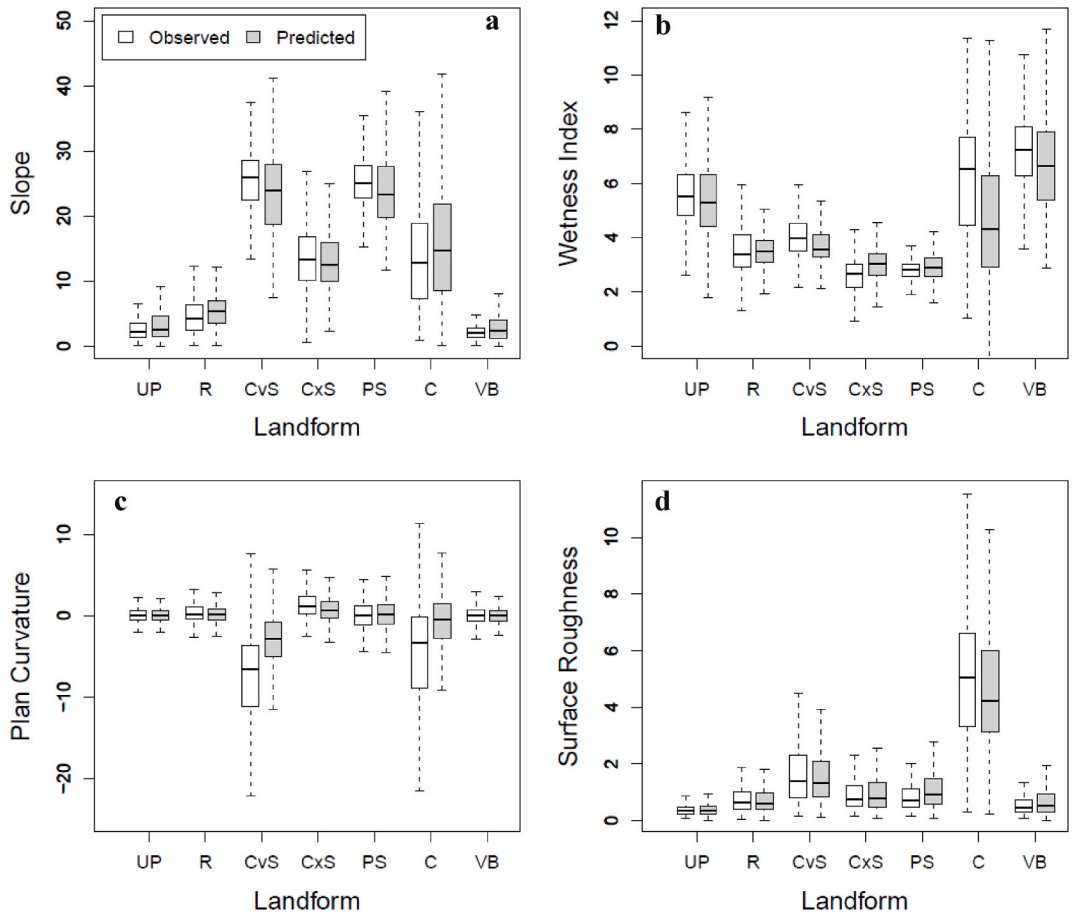


Fig. 8. Box plots showing the distribution of four geometries in observed landforms and in landforms predicted by 2 m covariates. The horizontal line inside each box represents the median value. The lower and upper limits of the box represent the 25th and 75th percentiles. Whiskers show the 10th and 90th percentiles. UP: Upland Plateau; R: Ridge; CvS: Concave Slope; CsX: Convex Slope; PS: Planar Slope; C: Channel; VB: Valley Bottom.

Supplementary Figs. S6 and S7). Similarities in the prediction of training and validation dataset and covariate characteristics of observed and predicted landforms indicate that the models are valid, and the overall approach can effectively predict landforms at locations other than the locations used in the training step.

4.3. Landform predictability

The outputs of the models show a range of prediction accuracies for each landform (Tables 2 and 3). The model performances for each landform were: (i) All models performed the best for valley bottoms, (ii) 2 m and 5 m models performed the worst for ridges, while (iii) 10 m model performed the worst for channels. For example, the 2 m model achieved the best accuracy for valley bottoms (~93%) and the worst accuracy for ridges (~72%). Similarly, the 10 m model mapped valley bottoms and ridges with 94% and 89% accuracies, respectively, and channels with the lowest accuracy (~83%). The prediction of other landforms, including upland plateaus, concave slopes, convex slopes,

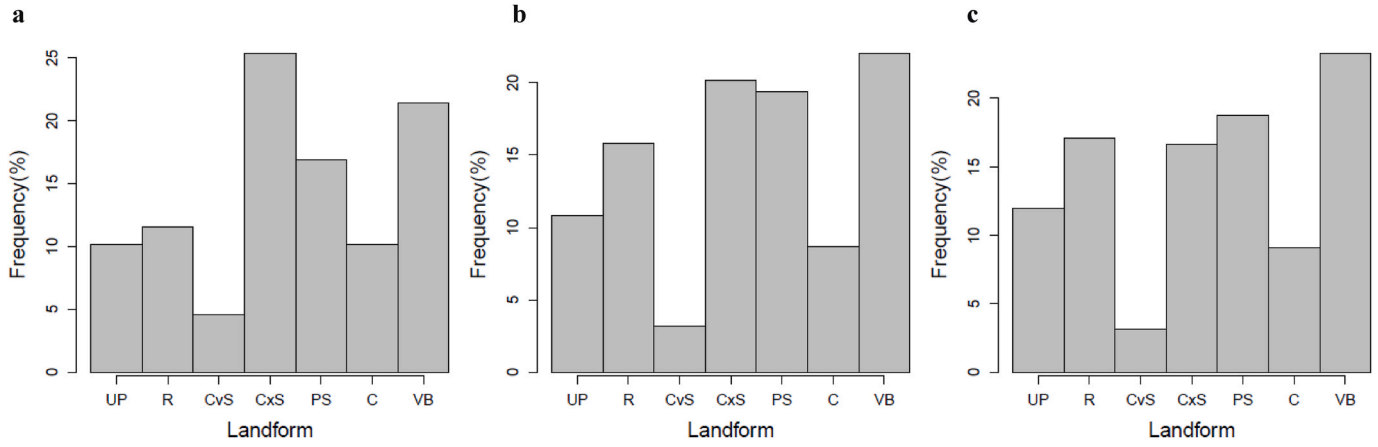


Fig. 9. Frequency distribution of landform cells predicted by the final models developed using: (a) 2 m covariates, (b) 5 m covariates, and (c) 10 m covariates. See Fig. 8 caption for acronyms.

planar slopes, and channels also resulted in a range of accuracies across models, though not significantly different from each other (Tables 2 and 3). Confusion matrices of the predictions (Tables 2 and 3) clearly showed that the models falsely predicted many of the observed landforms. For example, many upland plateaus are misclassified as ridges and valley bottoms, and ridges are misclassified as convex slopes and upland plateaus. Similarly, a significant number of concave slopes are misclassified as channels and planar slopes (Supplementary Figs. S8 and S9).

The frequency distribution of landform cells predicted by 2 m, 5 m and, 10 m models show a similar pattern with slight variation in percent distribution across landform types (Fig. 9). All models predicted convex slopes and valley bottoms with larger frequency and concave slopes and channels with smaller frequency. Fig. 9 also shows coarse resolution models (5 m and 10 m) resulted in higher frequencies of ridge, valley bottom, upland plateau and planar slope and smaller frequencies of concave and convex slopes compared to the 2 m model results. This is well-illustrated through a comparison between predicted landform maps (Fig. 6, and Supplementary Figs. S8 and S9), which show 5 m and 10 m resolution maps clearly illustrating wider ridges, likely the result of these models mischaracterizing some convex slopes as ridges. Additionally, coarser models falsely predicted many of the concave hillslopes as channels and small-sized channels in lowlands as valley bottoms (Figs. 5 and 6, and Supplementary Figs. S8 and S9).

5. Discussion

5.1. Model results and landform predictability

Final models using 2 m, 5 m, and 10 m resolution covariates mapped landforms with overall accuracy of 84%, 89%, and 89%, respectively. These results indicate that DEMs finer than 10 m in resolution, and limited observed data (i.e., 600 observations per landform across 400 km^2 area in this study), can produce acceptable maps of landforms across a large hilly landscape. The landform maps based on 10 m resolution mischaracterize several features. For example, many lowland channels are mischaracterized as valley bottoms, convex slopes next to ridges are mischaracterized as ridges, and the boundaries between landforms are not as smooth and differ in extent when compared to the 2 m and 5 m maps (Fig. 6, and Supplementary Figs. S8 and S9). On the other hand, we observe slightly lower predictive performance of 2 m covariate-based models that is attributable to noise on the DEM (Supplementary Fig. S7) contributed by vegetation pits and clusters, DEM artifacts (i.e., striping artifacts), roads, trails, and small-scale geomorphic features that exhibit significantly different geometries compared to their surroundings (i.e., steep channel walls and patchy bedrock cliffs). Nevertheless, the overall results indicate that higher resolution covariates were more effective in mapping most landforms, especially narrow channels and ridges, and in distinguishing between concave, convex and planar hillslopes. Results also show that large-sized landforms, such as valley bottom and upland plateau, can accurately be mapped with low resolution covariates. Given those general observations on the quality of the landform predictions, careful selection of the input dataset, such as the resolution of covariates, may depend on the size of the study area as well as the types and size of landforms to be mapped and the degree of spatial generalization (or scale) of a landform map one seeks to produce. In addition, the quality of the models may also depend on the number of observations per landform that are used to train models, the proportions of different landforms, the degree of anthropogenic landscape alterations, and the specific landforms as well as the general geomorphic characteristics of the landscape being mapped, which were not tested here. This study shows limited observations of landforms (i.e., 600 observations per landform across a 400 km^2 area) and finer-than-10 m resolution covariates are enough for mapping types and spatial scale of major landforms across a hilly landscape.

The models show subtle differences in their prediction accuracy for different landforms (Tables 2 and 3). The higher accuracy is likely the result of covariates being able to effectively characterize the distinguishing characteristics of landforms. Conversely, lower accuracy could stem from several reasons: covariates failing to fully characterize landform characteristics, errors introduced by DEM noises, and the resolution of the covariates being insufficient to extract signatures of small-sized landforms. For example, models predict valley bottoms and upland plateaus with higher accuracy compared to ridges and channels. Valley bottoms and upland plateaus share similar surface morphologies, such as surface roughness, slope, and curvature. However, they differ primarily in terms of elevation and upstream contributing area, which were well-characterized by the elevation and flow accumulation covariates. In contrast, the worst prediction for ridges and channels could be attributed to their spatial scale and local-scale DEM noises. These landforms are relatively small in size compared to others, and the low-resolution covariates may have failed to adequately characterize them. Moreover, the local-scale noise present in the higher-resolution covariates could have adversely influenced their prediction. Nevertheless, we believe additional covariates such as landscape position index, local relief, and divergence and convergence indices (Deumlich et al., 2010; Dowling et al., 2003; Evans et al., 2016; Meles et al., 2020) could increase the model accuracy and the quality of landform maps. Additionally, to reduce the effect of noise, different degrees of smoothing can be implemented for the DEMs with the same cell size, and differently smoothed DEMs can be used for mapping different landforms. For example, a high-resolution DEM with a high degree of smoothing applied may perform better for the large-sized landforms. However, optimal smoothing parameters need to be determined in each case, because a subtle difference in geometry exists among landforms and smoothing can significantly minimize these differences. Furthermore, given the landforms are different sized, we believe a multi-scale mapping by involving multiscale covariates in model development or by segmenting the higher resolution map (i.e., 2 m one) using multi-scale segmentation approach or object-based multi-scale segmentation of covariate in multidimensional space (Deumlich et al., 2010; Drăguţ and Blaschke, 2006) can also increase the model performance and the quality of predicted map.

5.2. Model applications

The approach presented here offers a robust way to map fluvial basin landforms with satisfactory performance (Figs. 5, 6 and 8, Supplementary Figs. S6–S9, and Tables 2 and 3). Our approach is data-driven supervised learning, where users can integrate a wide variety of predictors, including both continuous and categorical covariates, and train models to differentiate landforms. This flexibility makes the approach effective for mapping both large- and small-scale discrete and spatially continuous landforms, and highlights its potential for mapping other landform types beyond those presented here. We consider this flexibility a key advantage of our approach over existing models that classify a predefined set of landforms using DEM-derived landscape geometries, such as slope, local relief, curvature, and topographic position (Drăguţ and Blaschke, 2006; Minár et al., 2023) or by applying pattern recognition algorithms to DEM, such as geomorphon tool (Jasiewicz and Stepinski, 2013; Stepinski and Jasiewicz, 2011). However, the performance of our model is solely dependent upon the quantity, quality and spatial distribution of observed landforms, as well as the selection of the appropriate covariates, which can be time-consuming. In addition, the random forest model offers limited insight into the decision process of each individual tree and therefore the relationship between landform and covariates cannot be examined individually for every tree in the forest. The workflow, as presented, successfully mapped landform types that characterize major process domains of hilly landscapes (Montgomery, 1999; Scott and Wohl, 2019). This approach, thus, has further potential applications in generating input landform maps needed for surface hydrological modeling (Vannametee et al., 2013),

soil-landscape characterization (Regmi and Rasmussen, 2018), ecosystem and habitat assessment (Burnett and Blaschke, 2003), and hazard assessment, such as landslide and arid-region dust hazards (Iwahashi et al., 2001; Minar and Evans, 2008; Tian et al., 2010). Results clearly show that the overall quality of landform mapping depends on DEM resolution, or the spatial scale of the landform map one seeks to produce. Lower resolution topographic data can equally be applicable to larger-scale landforms; thus, this approach can be applied anywhere across the US and worldwide given the fact that the entire US is covered by 10 m horizontal resolution National Elevation Datasets (NED). In addition, many of the US regions are covered by higher resolution IfSAR and LiDAR datasets (https://www.usgs.gov/core-science-systems/ngp/3dep), and the entire world is almost covered by 15 m Aster Global DEM (https://lpdaac.usgs.gov/products/astgtmv003/) and 30 m SRTM DEM (https://earthexplorer.usgs.gov/).

6. Conclusions

The approach highlights the application of the random forest machine learning algorithm and high-resolution LiDAR topographic data in mapping spatially homogenous landscape components or landforms with limited observed dataset. The approach identified and mapped seven landforms including upland plateau, ridge, concave, convex and planar hillslopes, channel and valley bottom across a large landscape with satisfactory performance; and therefore, is a technique for producing multi-scale landform maps needed for soil and eco-hydrogeomorphic processes and hazard assessment. Overall, the mapping performance was best for valley bottoms and worst for ridges and channels. The approach can incorporate a wide variety of numerically continuous and categorical covariates to improve mapping efficacy. In addition, the approach can reduce the resources needed to provide detailed landform characteristics (i.e. geometry and size) and associated errors to landscape models relevant to hydrology, agriculture, and geosciences.

CRediT authorship contribution statement

Netra R. Regmi: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Nina D.S. Webb:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation. **Jacob I. Walter:** Writing – review & editing, Visualization, Validation, Funding acquisition. **Joonghyeok Heo:** Writing – review & editing, Software. **Nicholas W. Hayman:** Writing – review & editing, Visualization, Validation, Resources.

Code availability

Program language: R 4.3.1.
Software required: R, RStudio.
The source codes are available for downloading at the link: https://github.com/netraregmi/landform.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The research was funded by FEMA Hazard Mitigation Grant Program under DR-4315-0030-OK via the Oklahoma Dept. of Emergency Management and partly contributed by NASA Earth Surface and Interior grant 80NSSC22K1723. The views and conclusions contained in this paper are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of the U.S. government.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.acags.2024.100203.

References

Adediran, A.O., Parcharidis, I., Poscolieri, M., Pavlopoulos, K., 2004. Computer-assisted discrimination of morphological units on north-central Crete (Greece) by applying multivariate statistics to local relief gradients. Geomorphology 58, 357–370.
Berhe, A.A., Harden, J.W., Torn, M.S., Harte, J., 2008. Linking soil organic matter dynamics and erosion-induced terrestrial carbon sequestration at different landform positions. J. Geophys. Res.: Biogeosciences 113.
Bishop, M.P., Bonk, R., Kamp Jr, U., Shroder Jr, J.F., 2001. Terrain analysis and data modeling for alpine glacier mapping. Polar Geogr. 25, 182–201.
Bishop, M.P., James, L.A., Shroder Jr, J.F., Walsh, S.J., 2012. Geospatial technologies and digital geomorphological mapping: concepts, issues and research. Geomorphology 137, 5–26.
Böhner, J., McCloy, K.R., 2006. SAGA-analysis and modelling applications. Collection Göttinger geographische Abhandlungen 115.
Breiman, L., Cutler, A., 2011. Manual–setting up, using, and understanding random forests V4. 0. 2003. URL. https://www.stat.berkeley.edu/~breiman/Using_random_forests_v4.0.pdf.
Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. Taylor & Francis, p. 368.
Burnett, C., Blaschke, T., 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. Ecol. Model. 168, 233–249.
Burrough, P.A., van Gaans, P.F., MacMillan, R., 2000. High-resolution landform classification using fuzzy k-means. Fuzzy Set Syst. 113, 37–52.
Derakhshan-Babaei, F., Nosrati, K., Mirghaed, F.A., Egli, M., 2021. The interrelation between landform, land-use, erosion and soil quality in the Kan catchment of the Tehran province, central Iran. Catena 204, 105412.
Deumlich, D., Schmidt, R., Sommer, M., 2010. A multiscale soil–landform relationship in the glacial-drift area based on digital terrain analysis and soil attributes. J. Plant Nutr. Soil Sci. 173, 843–851.
Díaz-Uriarte, R., De Andres, S.A., 2006. Gene selection and classification of microarray data using random forest. BMC Bioinf. 7, 3.
DiBiase, R.A., Lamb, M.P., Ganti, V., Booth, A.M., 2017. Slope, grain size, and roughness controls on dry sediment transport and storage on steep hillslopes. J. Geophys. Res.: Earth Surf. 122, 941–960.
Dikau, R., Brabb, E.E., Mark, R.M., 1991. Landform Classification of New Mexico by Computer. US Dept. Of the Interior. US Geological Survey.
Dowling, T., Summerell, G.K., Walker, J., 2003. Soil wetness as an indicator of stream salinity: a landscape position index approach. Environ. Model. Software 18, 587–593.
Drăguţ, L., Blaschke, T., 2006. Automated classification of landform elements using object-based image analysis. Geomorphology 81, 330–344.
Du, L., You, X., Li, K., Meng, L., Cheng, G., Xiong, L., Wang, G., 2019. Multi-modal deep learning for landform recognition. ISPRS J. Photogrammetry Remote Sens. 158, 63–75.
Egan, J.P., Egan, J.P., 1975. Signal Detection Theory and ROC-Analysis. Academic Press, New York, p. 277.
Evans, D.A., Williard, K.W., Schoonover, J.E., 2016. Comparison of terrain indices and landform classification procedures in low-relief agricultural fields. Journal of Geospatial Applications in Natural Resources 1, 1.
Evans, I.S., 2012. Geomorphometry and landform mapping: what is a landform? Geomorphology 137, 94–106.
Frankel, K.L., Dolan, J.F., 2007. Characterizing arid region alluvial fan surface roughness with airborne laser swath mapping digital topographic data. J. Geophys. Res.: Earth Surf. 112.
Friedrich, K., 1998. Multivariate distance methods for geomorphographic relief classification. ESB Research Report 259–266.
Garcia-Aguirre, M.C., Ortiz, M.A., Zamorano, J.J., Reyes, Y., 2007. Vegetation and landform relationships at Ajusco volcano Mexico, using a geographic information system (GIS). For. Ecol. Manag. 239, 1–12.
Gorsevski, P.V., Gessler, P.E., Foltz, R.B., Elliot, W.J., 2006. Spatial prediction of landslide hazard using logistic regression and ROC analysis. Trans. GIS 10, 395–415.
Hammond, E.H., 1964. Analysis of properties in land form geography: an application to broad-scale land form mapping. Ann. Assoc. Am. Geogr. 54, 11–19.
Harris, J., Grunsky, E.C., 2015. Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. Comput. Geosci. 80, 9–25.
Heran, W.D., Green, G.N., Stoeser, D.B., 2003. A Digital Geologic Map Database for the State of Oklahoma. USGS. Open-File Report 2003-247.
Irvin, B.J., Ventura, S.J., Slater, B.K., 1997. Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin. Geoderma 77, 137–154.
Iwahashi, J., Watanabe, S., Furuya, T., 2001. Landform analysis of slope movements using DEM in Higashikubiki area, Japan. Comput. Geosci. 27, 851–865.

Jasiewicz, J., Stepinski, T.F., 2013. Geomorphons—a pattern recognition approach to classification and mapping of landforms. Geomorphology 182, 147–156.
Khan, U., Tuteja, N.K., Ajami, H., Sharma, A., 2014. An equivalent cross-sectional basis for semidistributed hydrological modeling. Water Resour. Res. 50, 4395–4415.
Kuhn, M., 2012. Variable selection using the caret package. URL. http://cran.cermin.lipi.go.id/web/packages/caret/vignettes/caretSelection.pdf.
Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., 2020. Caret: classification and regression training. R package version 6, 0-86. https://cran.r-project.org/web/packages/caret/caret.pdf.
Lashermes, B., Foufoula-Georgiou, E., Dietrich, W.E., 2007. Channel network extraction from high resolution topography using wavelets. Geophys. Res. Lett. 34.
Li, S., Xiong, L., Tang, G., Strobl, J., 2020. Deep learning-based approach for landform classification from integrated data sources of digital elevation model and imagery. Geomorphology 354, 107045.
Libohova, Z., Winzeler, H.E., Lee, B., Schoeneberger, P.J., Datta, J., Owens, P.R., 2016. Geomorphons: landform and property predictions in a glacial moraine in Indiana landscapes. Catena 142, 66–76.
MacMillan, R., Pettapiece, W., Nolan, S., Goddard, T., 2000. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. Fuzzy Set Syst. 113, 81–109.
Mashimbye, Z.E., Loggenberg, K., 2023. A scoping review of landform classification using geospatial methods. Geomatics 3, 93–114.
Meles, M.B., Younger, S.E., Jackson, C.R., Du, E., Drover, D., 2020. Wetness index based on landscape position and topography (WILT): modifying TWI to reflect landscape position. J. Environ. Manag. 255, 109863.
Middleton, M., Heikkonen, J., Nevalainen, P., Hyvönen, E., Sutinen, R., 2020. Machine learning-based mapping of micro-topographic earthquake-induced paleo-Pulju moraines and liquefaction spreads from a digital elevation model acquired through laser scanning. Geomorphology 358, 107099.
Minár, J., Drăguţ, L., Evans, I.S., Feciskanin, R., Gallay, M., Jenčo, M., Popov, A., 2023. Physical geomorphometry for elementary land surface segmentation and digital geomorphological mapping. Earth Sci. Rev. 104631.
Minar, J., Evans, I.S., 2008. Elementary forms for land surface segmentation: the theoretical basis of terrain analysis and geomorphological mapping. Geomorphology 95, 236–259.
Mithan, H.T., Hales, T.C., Cleall, P.J., 2019. Supervised classification of landforms in Arctic mountains. Permafr. Periglac. Process. 30, 131–145.
Montgomery, D.R., 1999. Process domains and the river continuum 1. JAWRA Journal of the American Water Resources Association 35, 397–410.
Pennock, D., Corre, M., 2001. Development and application of landform segmentation procedures. Soil Tillage Res. 58, 151–162.
Pitty, A.F., 2020. Introduction to Geomorphology. Routledge.
Prima, O.D.A., Echigo, A., Yokoyama, R., Yoshida, T., 2006. Supervised landform classification of Northeast Honshu from DEM-derived thematic maps. Geomorphology 78, 373–386.
Raduła, M.W., Szymura, T.H., Szymura, M., 2018. Topographic wetness index explains soil moisture better than bioindication with Ellenberg's indicator values. Ecol. Indicat. 85, 172–179.
Regmi, N., Dieu, J., Stewart, G., Turner, T., Miller, D., Johnson, A., Haemmerle, H., 2017. Characterizing landforms and associated mass movement processes using an object-based mapping approach. GSA Annual Meeting in Seattle. GSA, Washington, USA-2017.
Regmi, N.R., McDonald, E.V., Rasmussen, C., 2019. Hillslope response under variable microclimate. Earth Surf. Process. Landforms 44, 2615–2627.
Regmi, N.R., Rasmussen, C., 2018. Predictive mapping of soil-landscape relationships in the arid Southwest United States. Catena 165, 473–486.
Roering, J.J., Kirchner, J.W., Dietrich, W.E., 1999. Evidence for nonlinear, diffusive sediment transport on hillslopes and implications for landscape morphology. Water Resour. Res. 35, 853–870.
Roering, J.J., Marshall, J., Booth, A.M., Mort, M., Jin, Q., 2010. Evidence for biotic controls on topography and soil production. Earth Planet Sci. Lett. 298, 183–190.
Roering, J.J., Perron, J.T., Kirchner, J.W., 2007. Functional relationships between denudation and hillslope form and relief. Earth Planet Sci. Lett. 264, 245–258.
Romstad, B., Etzelmüller, B., 2009. Structuring the digital elevation model into landform elements through watershed segmentation of curvature. Proceedings of Geomorphometry 31, 55.
Scott, D.N., Wohl, E.E., 2019. Bedrock fracture influences on geomorphic process and form across process domains and scales. Earth Surf. Process. Landforms 44, 27–45.
Shruthi, R.B., Kerle, N., Jetten, V., Stein, A., 2014. Object-based gully system prediction from medium resolution imagery using Random Forests. Geomorphology 216, 283–294.
Siervo, V., Pescatore, E., Giano, S.I., 2023. Geomorphic analysis and semi-automated landforms extraction in different natural landscapes. Environ. Earth Sci. 82, 128.
Smith, M., Rose, J., Booth, S., 2006. Geomorphological mapping of glacial landforms from remotely sensed data: an evaluation of the principal data sources and an assessment of their quality. Geomorphology 76, 148–165.
Smith, M.J., Clark, C.D., 2005. Methods for the visualization of digital elevation models for landform mapping. Earth Surf. Process. Landforms 30, 885–900.
Søreide, K., 2009. Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research. Journal of clinical pathology 62, 1–5.
Stepinski, T., Vilalta, R., Ghosh, S., 2007a. Machine learning tools for automatic mapping of Martian landforms. IEEE Intell. Syst. 22, 100–106.
Stepinski, T.F., Ghosh, S., Vilalta, R., 2007b. Machine learning for automatic mapping of planetary surfaces. Proceedings of the National Conference on Artificial Intelligence 1807–1812. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.
Stepinski, T.F., Jasiewicz, J., 2011. Geomorphons-a new approach to classification of landforms. Proceedings of Geomorphometry 2011, pp. 109–112.
Summerell, G., Vaze, J., Tuteja, N., Grayson, R., Beale, G., Dowling, T., 2005. Delineating the major landforms of catchments using an objective hydrological terrain analysis method. Water Resour. Res. 41.
Taalab, K., Cheng, T., Zhang, Y., 2018. Mapping landslide susceptibility and types using Random Forest. Big Earth Data 2, 159–178.
Tarboton, D.G., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resour. Res. 33, 309–319.
Tian, Y., Xiao, C., Wu, L., 2010. Slope unit-based landslide susceptibility zonation. 2010 18th International Conference on Geoinformatics. IEEE, pp. 1–5.
Vannametee, E., Karssenberg, D., Hendriks, M., Bierkens, M., 2013. Hortonian runoff closure relations for geomorphologic response units: evaluation against field data. Hydrol. Earth Syst. Sci. 17, 2981–3004.
Veronesi, F., Hurni, L., 2014. Random Forest with semantic tie points for classifying landforms and creating rigorous shaded relief representations. Geomorphology 224, 152–160.
Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., Al-Katheeri, M.M., 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides 13, 839–856.
Zhao, W.-f., Xiong, L.-y., Ding, H., Tang, G.-a., 2017. Automatic recognition of loess landforms using Random Forest method. J. Mt. Sci. 14, 885–897.