RESEARCH ARTICLE

Machine learning for surficial geologic mapping

Sarah E. Johnson ![ORCID] | William C. Haneberg ![ORCID]

Department of Earth and Environmental Sciences, University of Kentucky, Lexington, Kentucky, USA

Correspondence
Sarah E. Johnson, Department of Earth and Environmental Sciences, University of Kentucky, 111 Slone Bldg., Lexington, KY 40506, USA.
Email: sarah.johnson11@uky.edu

Abstract

Surficial geologic maps contribute to decisions regarding natural hazard mitigation, land-use planning and infrastructure development. However, geologic maps may not adequately convey the uncertainty inherent in the information shown. In this study, we use machine learning and lidar elevation data to produce surficial geologic maps for parts of two quadrangles in Kentucky. We measured the performance of eight supervised machine learning methods by comparing the overall accuracy and F1 scores for each geologic unit. Surficial geologic units include residuum, colluvium, alluvial and lacustrine terraces, high-level alluvial deposits and modern alluvium. The importance of 41 moving-window geomorphic variables, including slope, roughness, residual topography, curvature, topographic wetness index, vertical distance to channel network and topographic flatness, was reduced to 12 variables by ranking the importance of each variable. The gradient-boosted trees model produced the classifier with the greatest overall accuracy, producing maps with overall accuracies of 87.4% to 90.7% in areas of simple geology and 80.7% to 81.6% in areas with more complex geology. The model produced high F1 scores of up to 96.2% for colluvium but was not as good at distinguishing between units found in the same geomorphic position, such as high-level alluvium and residuum, both of which are found on ridgelines. Probability values for each geologic unit at each cell are conveyed using gradations of colour and eliminate the need for drawn boundaries between units. Machine learning may be used to create accurate surficial geologic maps in areas of simple geology; in more complex areas, highlight that additional information obtained in the field is necessary to distinguish between units.

KEYWORDS
gradient-boosted trees, landform mapping, lidar, machine learning, mapping probability, surficial geology, uncertainty

1 | INTRODUCTION

The making of geologic maps is a fundamental contribution of geologists to society and industry. However, formatting limitations require geologists and cartographers to make decisions about what information to display and omit, as well as how to show boundaries and areas of geologic uncertainty or transition. Traditionally, there have been few options to communicate the nuance and uncertainty in the mapping process. Perhaps not fully understood by many map users is that geologic maps may convey a level of certainty, which obscures the myriad of decisions and interpretations made by the mapper.

The combination of machine learning with surface, geophysical and remote sensing data provides a new tool for producing geologic maps, which can also convey the underlying uncertainty intrinsic to any geologic map.

The utility of geologic maps in locating natural resources and for engineering projects has been recognized since the first widely published geologic map of England by Smith (1815) and Winchester (2009). In addition to the exploration of mineral, energy and water resources, geologic maps support decision-making processes for natural hazards such as landslides, rockfall, floods, earthquakes and karst; for land-use planning, climate change and environmental impact

evaluations, identifying health hazards, and in locating and building infrastructure such as roads, buildings, dams and utilities (Bernknopf et al., 1993; Derouin, 2021; GSA, 2022). The economic value of geologic maps includes not only the resource development but also the prevention of economic loss due to poor land-use decisions and documented benefit-to-cost ratio for geologic mapping ranges from 4:1 to >100:1 (Berg, MacCormack, & Russell, 2019).

A map of geologic units and boundaries may convey a certainty to the user that the map gives exact and perfect information about the geology at every map location. In reality, mapped geological boundaries are inherently uncertain for several reasons: The boundary may be gradational as opposed to a sharp contact, the width and position of boundaries are generalized and dependent on the scale of the map, and boundaries may be obscured in the field (Lark et al., 2015). As we are unable to directly observe geologic variables at every point, we use models—either mental or computational—to fill in the gaps (Kirkwood, 2022). Individual mappers might use different models and thus produce different maps of the same location. In addition, decisions mappers used to define boundaries might be poorly documented, and the depth of information that went into the creation of a map may not be conveyed in the final product (Jones et al., 2004; Kirkwood, 2022). As there are many potential models to interpolate between direct observations, the necessity of quantifying and conveying uncertainty is an essential part of geologic mapping (Tikoff et al., 2023) and soil mapping (Boettinger et al., 2010; Khaledian & Miller, 2020).

Technological advances throughout history have facilitated changes in geologic mapping such as GPS, GIS, aerial photography, satellite imagery, lidar, geophysical and geochemical data sets. Over the past decade, the expanded deployment of traditional sensors and the development of new data sources such as lidar and satellite data have increased the amount of data available to geoscientists (Bergen et al., 2019). The availability of these large data sets, in combination with advances in computational capacity, has made machine learning (ML) increasingly applicable to modelling a variety of Earth processes. Advantages of ML specific to geologic mapping are the ability to process data sets with numerous variables (high-dimensionality data sets) in order to replicate human mapping performance, to produce maps more quickly than humans can, to produce maps using a consistent well-documented model across large regions, to quantify the probability of each geologic unit for every cell of the map and to convey those probabilities as part of the final map (Bergen et al., 2019; Cracknell & Reading, 2013). Probability is the likelihood of a particular classification; a probability of 0.5 that a cell belongs to a particular geologic unit would indicate considerable uncertainty about that classification, while a probability of 0.9 or above would indicate a high level of certainty about the classification, and a very low probability would indicate a high level of certainty that it does not belong to that particular classification.

There are two main types of ML: supervised classification, which requires labelled examples to learn to distinguish patterns and then make predictions or classify previously unseen data, and unsupervised classification techniques, which learn to distinguish patterns in data without any labels provided. A ML map can document the probability for each class at any point on the map and provide a consistent and well-documented mapping process. Thus the map user can obtain information about the range of geological properties they might encounter at any particular location (Kirkwood, 2022).

While there are examples of geologic maps of bed-rock produced using ML methods, there is also great potential for ML in mapping surficial geology. Maps of bed-rock geology in New South Wales, Australia, were produced from dispersed geologic data in conjunction with geophysical and remote sensing data (Cracknell & Reading, 2014) and in the Eastern Goldfields of Australia using only geophysical data (Kuhn, Cracknell, & Reading, 2018). Geochemical maps in England were generated using sparse geochemical data in combination with geophysical and remote sensing data (Kirkwood et al., 2016, 2022). However, fewer maps have been created of surficial geology using ML, such as regolith mapping in West Africa using airborne geophysics and remote sensing data (Metelka et al., 2018), though there are examples of surficial geologic hazards identified utilizing ML such as landslide-susceptibility maps produced using landslide inventories and lidar data in Kentucky (Crawford et al., 2021) and Japan (Dou et al., 2019), and the identification of sinkholes in Kentucky (Zhu & Pierskalla, 2016). Given the dramatic increase in the number of geophysical and other data sets publicly available, there is great potential for the expansion of ML to geologic mapping applications throughout the world, which may be particularly useful where ground information is sparse or non-existent.

In this study, we examine the utility of ML to the process of mapping the surficial geology of Kentucky using the characteristics of the terrain derived solely from lidar-based digital elevation models (DEMs). Our goal was to evaluate whether the process of mapping surficial geologic units could be automated using currently available off-the-shelf machine learning software to produce maps where classification uncertainty is conveyed, and which are useful within the context of a production-oriented geological survey mapping programme. The Commonwealth of Kentucky, USA, is the first state with bed-rock fully mapped at a scale of 1:24 000. The cost of mapping bed-rock for all 707 quadrangles was initially justified by the economic development of coal, oil, natural gas and minerals; today, the use of maps for management of land, water and the environment has supplanted their use for natural resource development (Bhagwat & Ipe, 2000). To address these needs, geologic mapping of surficial deposits began in 2004 and continues today (KGS, 2021). Mapping the surficial geology of each quadrangle using traditional means requires multiple steps: contacts between surficial units are mapped based on topographic features derived from lidar data, initial maps are used in the field to resolve ambiguous features, contacts and units, and choose sample locations to further inform and update the maps, and final maps are produced. The process takes approximately 1 year for a standard 7.5-min topographic quadrangle published at a scale of 1:24 000 (Hammond et al., 2017). To test the utility of ML to create surficial geologic maps, we first test eight different ML models and 41 terrain variables to determine the best performing model, and the key variables needed to produce maps efficiently. Using the best performing model and most important variables, three areas in each of two quadrangles are then mapped, and the results are quantitatively and qualitatively evaluated.

2 | GEOLOGIC SETTING

The two quadrangles used for this study are both situated within the Outer Bluegrass physiographic region of Kentucky, which is characterized by a dissected terrain that has formed in gently dipping Palaeozoic carbonates and shales (McDowell, 1986). A dendritic drainage system

contributes to the larger river valleys, which are entrenched 60–90 m below the ridges and hold a sequence of lacustrine and alluvial terraces above the modern floodplain. The De Mossville Quadrangle (Massey, 2017) is in Pendleton County, KY, about 34 km south of Cincinnati, OH, and the Ohio River (Figure 1). The Pitts Point quadrangle (Massey et al., 2019) is in Bullitt County, approximately 160 km to the southeast of the De Mossville quadrangle and 48 km south of Louisville, KY. The Licking River meanders north towards the Ohio River though the De Mossville quadrangle, and the Salt River flows to the Ohio River 15 km to the north through the Pitts Point quadrangle.

Bed-rock has weathered to produce ridgetops of residuum and moderate to steep slopes of colluvium. In both quadrangles, the residuum of the ridgetops is interrupted by high-level alluvial and lacustrine deposits or terraces that date to the Pliocene or early Pleistocene (KGS, 2023; Massey, 2017; Massey et al., 2019). These deposits record the presence of an ancestral river system, which predates the down cutting of the Ohio River to its current level (McDowell, 1986; Potter, 2007). As the Laurentide ice sheet advanced and blocked this ancestral river system, the Ohio River valley was formed (Durrell, 1982; Granger & Smith, 1998). The former plateau was dissected in a dendritic pattern as the landscape eroded to meet this new base level. Subsequent glacial advances and their deposits blocked drainages, forming lakes, their presence recorded by extensive lacustrine terraces (middle to late Pleistocene) found along the river valleys including both quadrangles in this study. These lacustrine terraces have been dissected by late-Pleistocene alluvial terraces into which the modern floodplain is entrenched.

Surficial geologic units in both quadrangles form both natural resources and hazards. Outwash and alluvial terraces form resources of sand and gravel for building materials and freshwater aquifers (McDowell, 1986; Potter, 2007). The steep colluvium covered slopes developed from shale-rich bed-rock are prone to landsliding (Baum & Johnson, 1996; Crawford, 2012; Johnson et al., 2023).

3 | METHODS

3.1 | Overview

The utility of ML for mapping surficial geology in Kentucky was assessed by training and testing eight ML methods using publicly available digital elevation models (DEMs) and surficial geologic maps
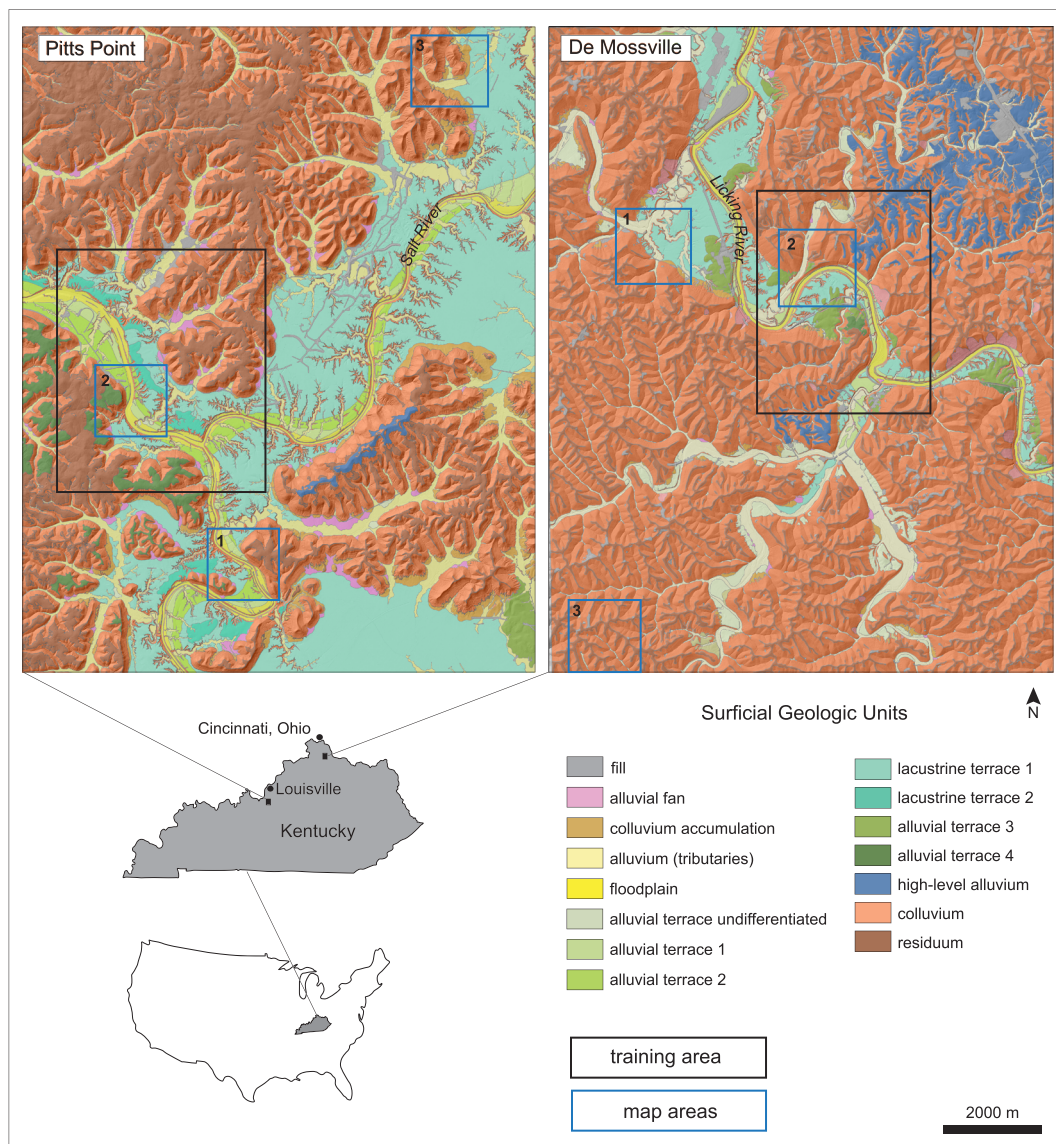

FIGURE 1 Location and surficial geology of the 7.5-min Pitts Point and De Mossville quadrangles in Kentucky, USA.

for two quadrangles in Kentucky. A summary workflow is shown in Figure 2, and each step is described in detail in the following paragraphs.

3.2 | Surficial geologic quadrangles

Machine learning methods were applied to two surficial geologic maps published as 7.5-min quadrangles by the Kentucky Geological Survey (KGS). These were created using traditional field-based and digital mapping methods. One quadrangle from each region, Pitts Point (SE corner 85° 45′, 37° 52′ 30″) (Massey et al., 2019) and De

Mossville (SE corner 84° 22′ 30″, 38° 45′) (Massey, 2017) were each chosen to include a diverse set of surficial geologic units including residuum and colluvium derived from bed-rock units, high-level pre-glacial alluvial deposits, a sequence of alluvial and lacustrine terraces, modern alluvial deposits and fill.

3.3 | Digital elevation maps and derivative maps

The DEMs for the Pitts Point and De Mossville Quadrangles are publicly available through KyFromAbove, the state's elevation data and aerial photography programme, which is accessible digitally at
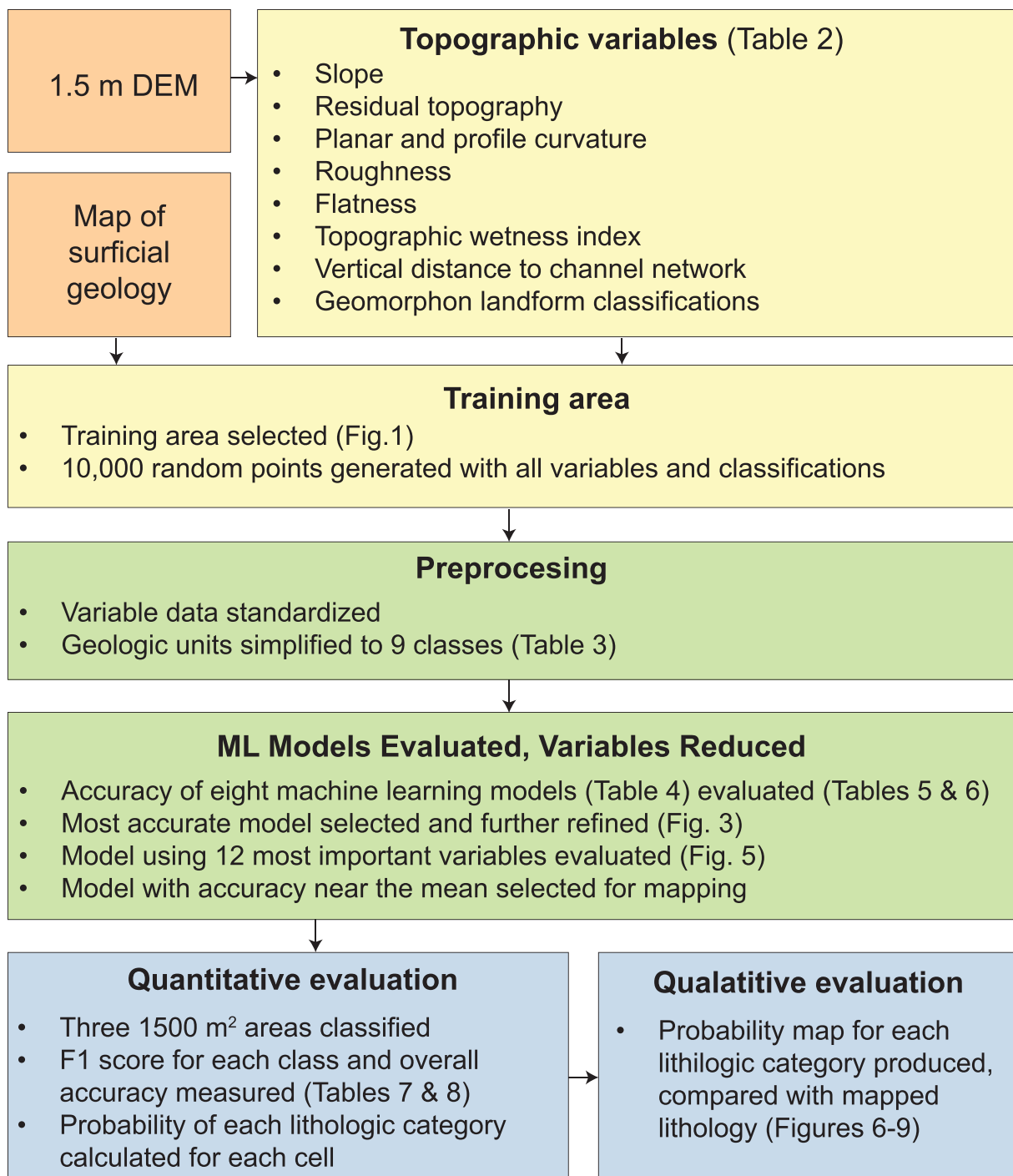

FIGURE 2 Overview of workflow.

kyfromabove.ky.gov. The maps were derived from lidar data and have a 5 ft (1.524 m) cell size. To test which geomorphic variables are the most important for the ML models, a series of 41 derivative maps were made from the DEMs using ESRI ArcPro GIS software (Table 2). The derivative maps represent geomorphic attributes independent of their specific location (i.e., latitude and longitude). Moving windows with radii ranging from three cells to 500 cells were used to measure qualities of the geomorphology at a variety of scales such that each cell has information about its local vicinity as well as a comparison to or measure of the surrounding terrain (Table 1). A multiscale digital terrain analysis approach is important, as the spatial patterns of landforms may be more visible at one scale than another or more related to some terrain attributes than others (e.g., Behrens et al., 2010). The topographic features that we are capturing include small-scale features such as gullies and small depressions in the residuum and alluvial terrace 4, moderate-scale features such as ridges of residuum, ridges of high-level alluvium, alluvial terraces and small tributary valleys; and large-scale features such as the trunk valleys, and lacustrine terraces (Table 1). All variables were standardized in ArcPro so that the mean value was 0, and the standard deviation was 1 in order to make the units of each variable comparable (Shanker, Hu, & Hung, 1996).

A series of residual topographic maps and slope maps were made using the original unsmoothed DEM and smoothed DEMs. Smoothed DEMs were produced using a focal statistics tool in GIS, which calculates a statistic for a defined region around a particular cell, in this case the mean elevation was calculated within a circular radius around each cell ranging from 16 cells to 1500 cells (49 to 2740 m). Residual topographic maps were produced by subtracting each smoothed DEM from the elevation for each cell, thus allowing the position of a cell relative to its surroundings to be represented, with positive values representing ridges or peaks, negative values representing depressions or valleys and values close to zero representing an even slope or flat area (Weiss, 2001; De Reu et al., 2013; Haneberg et al., 2005; Muñoz & Valeriano, 2014). Slope maps were produced from the original DEM and smoothed DEMs with moving windows of 4, 8, 16, 24 and 32 cells (Table 2).

Profile and planar curvature are the second derivative of the surface or the slope of the slope. Profile curvature is calculated parallel to the slope, and positive values indicate the surface is upwardly concave at that cell, negative values indicate the slope is upwardly convex at that cell, and a value of zero indicates that the slope is linear. Planar curvature is perpendicular to the orientation of the maximum slope. A positive value indicates that the surface is laterally convex, a negative value indicates that the surface is laterally concave, and a zero value that the surface is linear at that cell. Profile and planar curvature were calculated in ArcPro using a 3 by 3 window of cells, using the original DEM and smoothed DEMs with a radius of 50, 100 and 500 cells (Table 2).

TABLE 1 Scale of topographic features in the Pitts Point (PP) and De Mossville (DM) quadrangles.

Topographic features	Feature width
Trunk valleys	1500 to 2700 m
Tributary valleys with alluvium	300 to 1000 m
Tributary valleys without alluvium	130 to 260 m
Lacustrine terraces	2300 to 2700 m PP; 800 to 1500 m DM
Alluvial terraces	150 to 600 m PP; 20 to 400 m DM
Width of floodplains	100 to 275 m PP; 50 to 175 m DM
Tributaries within terraces	30 to 150 m
Colluvial slopes on valley walls	150 to 300 m PP; up to 900 m DM
Colluvial slopes of tributaries within terraces	3 to 30 m
Ridges of residuum	15 to 300 m
Ridges of alluvial terrace 4	10 to 250 m PP
Ridges of high-level alluvium	30 to 250 m DM
Gullies and small depressions in residuum and alluvial terrace 4	1.5 to 8 m
Width of roads and other fill	12 to 100 m

TABLE 2 Definition of variables.

Variable	Description
Digital elevation model (DEM)	Mean height of cell measured from mean sea level, cell dimensions 1.524 m
Smoothed DEM	Mean elevation within a radius ranging from 4 to 1500 cells. Used to calculate residual topography, slope and curvature maps
Slope	Slope original DEM, using a 3 by 3 cell moving window
Slope of smoothed DEM radius 4, 8, 16, 24, 32	Slope of the smoothed DEM
Residual Topography (RT), radius 16, 24, 32, 50, 100, 150, 200, 500, 1000, 1500	The elevation of the original DEM minus the smoothed DEM
Vertical distance to channel network	Vertical distance to a channel network, which accumulates from an area of least 450,000 m^2
Topographic Wetness Index (TWI)	Areas that are likely to be wetter due to their position in the landscape. ln (upslope area/tan [local slope])
Profile curvature of DEM, smoothed DEM radius 5, 50, 100, 500	Profile curvature of the original and smoothed DEMs. The second derivative of the profile slope, or the slope of the slope
Plan curvature of DEM, smoothed DEM radius 5, 50, 100, 500	Plan curvature of the original and smoothed DEM. The second derivative of the plan view, or the slope of the plan view slope
Roughness radius 4, 8, 16, 24, 32, 50, 100	Standard deviation of the slope angles derived from the original DEM within a radius of 16, 24, 32, 50 and 100 cells
Flatness, radius 50, 100, 150, 200, 250, 500	Standard deviation of the elevation values from the original DEM, within a radius of 50, 100, 150, 200, 250 and 500 cells
Geomorphons, search distance 15.2, 30.5, 45.7 m	Landforms classified into common landform types: flat, peak, ridge, shoulder, spur, slope, hollow, footslope, valley and pit

Surface roughness maps are a measure of the distribution of slope angles within a particular area. They were produced using the focal statistics tool to calculate the standard deviation of the slope angles derived from the original DEM within a circular radius ranging from 16 to 150 cells (Grohmann, Smith, & Riccomini, 2011). A measure of topographic flatness was calculated from the standard deviation of elevations from the unsmoothed original DEM using the focal statistics tool and a radius ranging from 100 to 500 cells (Grohmann, Smith, & Riccomini, 2011).

The topographic wetness index (TWI) is a measure of areas likely to be wetter due to their topographic positions and was calculated by taking the natural log of the upslope area, divided by the tangent of the local slope in radians (Beven & Kirkby, 1979). The vertical distance to the channel network was calculated for a network, which accumulates from a heuristically chosen value of at least 450 000 m^2.

In addition to the topographic variables described above, we used geomorphons, which use the DEM and search distances from each cell in order to classify landforms into 10 common landform types: flat, peak, ridge, shoulder, spur, slope, hollow, footslope, valley and pit (Jasiewicz & Stepinski, 2013). For this study, geomorphons were calculated in ArcPro using search distances of 15.2, 30.5 and 45.7 m.

3.4 | Geologic classes

The geologic units were simplified to avoid undersampling smaller scale features such as alluvial fans, colluvium accumulation zones, landslides and different varieties of fill (Table 3). Colluvium is defined to include landslides and colluvium accumulation zones found at the base of some slopes. Alluvium is defined to include alluvial fans, alluvium in tributaries and old alluvium. The lacustrine and alluvial terraces were divided into three terrace units: alluvial terraces 3&4 are at the highest level; lacustrine terraces 1&2 lie below this; and alluvial terraces 1&2 are at a lower level nearest the modern floodplain. Fill includes engineered and nonengineered artificial fill.

3.5 | Machine learning methods

Supervised ML methods use training data that has a known classification, in this case the mapped surficial geologic unit, to train a classifier

TABLE 3 Simplified lithological categories.

Surficial geologic units for the Pitts Point and De Mossville Quadrangles
Residuum
Colluvium (includes landslides and colluvial accumulation at the base of slopes)
High-level alluvium (high-elevation preglacial river system deposits)
Alluvial terrace 3&4
Lacustrine terraces 1&2
Alluvial terraces 1&2
Floodplain (main river valley)
Alluvium (includes alluvial fans, alluvium in tributaries, old alluvium)
Artificial fill (engineered and other fill)

which is then used to classify previously unobserved data. To build the classifier and measure how well it performs, the data set is randomly divided into a training set, which is used to train the model, and a testing set, which is used to measure the performance of the model. We used a data set of 10 000 random points within a training area for each quadrangle (Figure 1). As the geologic classes are imbalanced (i.e., there are far more examples of colluvium than the other units in the data set), we used stratified sampling with proportional allocation. We then used fivefold cross-validation to test each model, which was repeated for a total of ten folds.

We measured the performance of eight supervised ML methods in Mathematica. The eight ML methods tested represent the main types of supervised learning algorithms and are available in the software used for the analysis ('Mathematica', 2022; Bergen et al., 2019): logistic regression, support vector machine, nearest neighbour, decision tree, random forest, gradient-boosted trees, naïve Bayes, and neural network. These methods are summarized in Table 4 and further described below. We used the automated function 'Classify' in Mathematica, which trains a number of hyperparameters automatically on a subset of the training set and then chooses the best configuration for the full training set (Bernard, 2021). The method that produced the greatest overall accuracy and F1 scores out of the eight tested models was then refined by further tuning model hyperparameters. The values used for all models are included in the supplementary materials.

TABLE 4 Summary of the ML models used in this study.

Machine learning method	Algorithm type	Description
Logistic regression	Binary classifier	Uses a logistic function to estimate the probability that a data point belongs to one of two classes
Support vector machine	Binary classifier	Finds a hyperplane that separates training data into two classes. A multiclass classification is reduced to a set of binary classification problems
Nearest neighbour	Instance-based	Infers the class of each example by comparing to the nearest neighbour in the feature space and picking the commonest class or average value
Decision tree	Decision Tree	A branching treelike structure where each node represents a test on each feature, each branch represents the outcome of a test, each leaf represents a prediction or class probability
Random forest	Ensemble	An ensemble of weaker decision trees independently trained on a subset of the training data. Predictions are combined
Gradient-boosted trees	Ensemble	An ensemble of weaker decision trees is trained sequentially, compensating for the weaknesses of previous trees
Naïve Bayes	Bayesian	Applies Bayes's theorem, which assumes that features are independent given the class, regardless of possible correlations between features
Neural network	Neural Network	Consists of stacked layers each of which performs a single computation and passes that value to the next layer

Logistic regression (LR) uses a logistic function to estimate the probability that a data point belongs to one of two classes (Cox, 1958). This method has been used to map soil type based on terrain parameters (e.g., Giasson et al., 2008) and landslide susceptibility mapping (Crawford et al., 2021). The optimization method used in this study is the limited memory Broyden-Fletcher-Goldfar-Shanno (LBFGS) algorithm (Liu & Nocedal, 1989). Overfitting of data is prevented by utilizing L1 and L2 regularization values, which add a penalty as the model complexity increases (Ng, 2004). Our models used L1 regularization values of 0 and L2 regularization values ranging from 0.000001 to 1.

The support vector machine (SVM) method is designed to construct an optimal decision surface, or hyperplane, in the feature space that separates classes (Cortes & Vapnik, 1995; Heung et al., 2016). To create nonlinear class boundaries for the classification of complex data sets, the training data set is transformed to a high-dimension feature space using a polynomial or radial basis kernel, which may be divided by a linear decision surface; in the original non-transformed space, this surface becomes a nonlinear hyperplane (Cortes & Vapnik, 1995; Heung et al., 2016). The hyperplane is constructed to have a maximum margin, or distance, from individual observations, making the results more applicable to the unseen data (Khaledian & Miller, 2020). While the models are fast to run, training time can be slow as the similarity of every pair of training examples is computed (Bernard, 2021). SVM has been used for digital soil mapping (Kovačević, Bajat, & Gajić, 2010) and creating landslide susceptibility maps (Huang & Zhao, 2018; Kavzoglu, Sahin, & Colkesen, 2014). Our model uses an exponential radial basis function as the kernel. The gamma scaling parameter that controls the influence of the support vectors is automatically chosen in Mathematica, with values ranging from 0.0121833 to 0.0319713. The multiclass strategy we used is one versus one, which tests each class against each other.

The nearest neighbour, also known as the k-nearest neighbour (KNN) method, assumes that examples that have similar features likely have similar labels (Bernard, 2021). This method picks the most common class among the KNN, where k is the number of neighbours considered. While training is fast, this method does not have the capability to learn that some variables are more important than others; thus, the resulting model may not produce the best predictions (Bernard, 2021). While this model is one of the simplest machine learning methods, it can be useful and has been used for digital mapping of soils (Khaledian & Miller, 2020; Mansuy et al., 2014). Our models used k-values of 10 and 20, a regularization parameter of 0.5, and scanned the entire data set to compute the k-nearest examples.

The decision tree (DT) model has a treelike structure, where each node represents a test on a feature, and each branch represents the test outcome. The nodes and branches continue until a leaf is reached, with the leaf representing the class prediction. The DT model may not provide great models in terms of predictions, but it is fast (Bernard, 2021). Perhaps more importantly, DT is the basis for two more complex but better performing ML methods, random forest and gradient-boosted trees. The DT is constructed using the classification and regression tree (CART) algorithm in Mathematica. DT has been used for soil mapping (Hatefard et al., 2019), landslide susceptibility mapping (Yeon, Han, & Ryu, 2010) and subsidence maps (Lee & Park, 2013).

The random forest (RF) method is an ensemble learning algorithm that works by constructing a collection of DT classifiers, and each tree votes for the most popular class (Breiman, 2001; Khaledian & Miller, 2020). Each DT is trained independently on a random subset of the original training data set (bootstrap samples). The RF method is generally fast and makes good predictions, making it a great out of the box method (Bernard, 2021). It has been used for soil mapping (Dharumarajan & Hegde, 2022; Gruber et al., 2019; Heung et al., 2016), geologic mapping (Cracknell & Reading, 2013), landslide susceptibility maps (Sun et al., 2020) and sinkhole mapping (Zhu & Pierskalla, 2016). For our models, we used the default values in Mathematica, so each tree was constructed using 15% of the training set, 100 trees were trained to produce the forest, and the maximum number of samples in each leaf was 2.

The gradient boosted trees (GBT) method is an ensemble learning algorithm where models are trained in a sequence, and each new model corrects the errors from the previous model. In this way, weak DT learners are combined into a stronger 'boosted' learner. GBT generally produces predictions that are as good or even better than RF (Bernard, 2021). The GBT model is slower to train than RF, and there are more hyperparameters to tune including the learning rate, the size or depth of the trees and the number of trees. A slower learning rate produces a model that generalizes better than a faster rate. Deeper trees with multiple nodes can capture more complex relationships between features. The number of trees is also important, as too many trees can lead to overfitting. GBT has been used for digital soil mapping and landslide susceptibility mapping using terrain features as inputs (Dyer et al., 2024; Hitziger & Ließ, 2014; Sahin, 2020). Mathematica trains a number of hyperparameters automatically on a subset of the training set and then chooses the best configuration for the full training set (Bernard, 2021). For our models, the number of training rounds was the default value of 50, the number of leaves ranged from 13 to 1023, the learning rate ranged from 0.1 to 0.2, the maximum tree depth was the default value of 6, and the leaf size ranged from 7 to 90.

The naïve Bayes (NB) method is based on Bayes' theorem, which assumes that all variables in a data set are independent from each other. While it is a simpler model than the ensemble methods described above, it has been used for mapping and landslide susceptibility studies, though in comparison with other ML methods, it sometimes fall short (Harvey & Fotopoulos, 2016; Pham et al., 2017).

Neural network methods, also known as artificial neural networks (ANN), mimic biological neural networks. The structure consists of interconnected nodes, or artificial neurons, that are organized into layers, and information is passed from one layer of the network to another (Khaledian & Miller, 2020). While there are several types of ANN architectures, the one used for this study is a multilayer perceptron with self-normalizing architecture. It uses dropout for regularization and a scaled exponential linear unit activation function (Bernard, 2021). Variables are introduced to the input neurons, which are then connected to multiple layers of hidden neurons where computations are made using the activation function and are then linked to output layer where the predictions are made (Heung et al., 2016; Jaiswal, 2024). During the training period, the output is compared to the input and the residual calculated, and the network is repeatedly adjusted and the residual calculated again until it is minimized (Khaledian & Miller, 2020).

Neural networks of various architectures have been used for digital soil mapping, geologic mapping and landslide susceptibility studies (Kawabata & Bandibas, 2009; Kalambukattu, Kumar, & Arya Raj, 2018; Bodaghabadi et al., 2015). For this study, Mathematica trains a subset of the training set and then chooses the best configuration for the full training set. Values for the network depth ranged from 2 to 8, and the maximum number of training rounds ranged from 10 to 300.

3.6 | Variables and dimensionality reduction

A total of 41 variables were initially used to train the eight machine learning models. The dimensionality of the data set was then reduced to the 12 most important variables to lessen the computing power necessary to perform further classifications of larger map areas. We evaluated the variable importance by using the Cumulative Feature Impact Plot in Mathematica, which ranks the importance of each variable in the model. We ranked the importance of each variable for each model produced from the tenfold and chose the 12 top-ranked variables. A distribution of classifier accuracy with all 41 variables versus a model trained with just the 12 selected variables was performed to measure the loss of accuracy.

3.7 | Quantifying model performance

The performance of eight machine learning methods was evaluated by comparing their overall accuracy and the F1 scores for each geologic unit. Accuracy is defined as the fraction of correctly classified examples:

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

where TP = the number of true positive predictions, TN = the number of true negative predictions, FP = the number of false positive predictions (Type I errors), and FN = the number of false negative predictions (Type II errors). Accuracy is a useful metric for comparing the models which all use the same data. However, as the distribution of classes is not even (e.g., there are many more samples of colluvium than high-level alluvium or fill), accuracy is not a good way to measure model performance for each geologic unit (Forman & Scholz, 2010). For example, a good performance of the model to predict colluvium might overshadow its poor performance in predicting fill. The F1 score, on the other hand, accounts for class imbalance through the metrics of Precision and Recall for each class:

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

3.8 | Lithologic probability maps

Qualitative evaluations of the best and worst models were produced by creating maps for three 1.5 × 1.5 km areas for each quadrangle. Map 1 for both the Pitts Point and De Mossville quadrangles is just outside of the training area. Map 2 is within the training area, and Map 3 is 4.5 to 6 km away from the training area boundary for the Pitts Point and De Mossville quadrangles, respectively (Figure 1). The probability of every cell belonging to each geological unit was calculated using the ML model. These probabilities were then brought into GIS software, and probability map layers were produced for each unit. These maps were symbolized so that the darkest shade represents a probability of ≥80%, the middle tone represents probabilities of 60% to 80%, the lightest tone represents probabilities of 40% to 60%, and probabilities of <40% are completely transparent. In this way, each map layer preserves the underlying probability values, and different map users could change the symbology according to their purpose. The probabilities can also be shown by querying specific points if the maps are being used in GIS software. The resulting ML maps can then be compared with the hand-digitized maps.

4 | RESULTS

4.1 | Model accuracy and F1 scores

Of the eight models tested, the GBT model produced the classifier with the greatest overall accuracy for both the Pitts Point and De Mossville quadrangles, with an overall accuracy of 0.926 ± 0.006 and 0.901 ± 0.006, respectively (Figure 3, Figure 4, Tables 5 and 6), and the NB model produced the least accurate classifier in both cases, of 0.811 ± 0.007 and 0.787 ± 0.011, respectively. The GBT model also produced the highest F1 scores for all geologic units for both quadrangles. In the Pitts Point quadrangle, the colluvium, residuum, alluvial terraces 1&2, lacustrine terraces 1&2 and floodplain have F1 scores of 90% or greater, while the fill, alluvium and alluvial terraces 3&4 have lower F1 scores (Table 5). In the De Mossville quadrangle, the colluvium and floodplain have F1 scores of 90% or greater, while the fill, alluvium, high-level alluvium, terraces and residuum have lower F1 scores (Table 6). Confusion matrices for one of the models for the Pitts Point quadrangles give a more detailed view of how the model performed (Figure 4).

4.2 | Hyperparameter tuning of GBT model

The GBT model was further refined by tuning the learning rate and the leaf size; the maximum training rounds and maximum depth were left at the automatically chosen values of 50 and 6, respectively. A training and testing fold was selected, and learning rates of 0.05, 0.1, 0.15, 0.2 and 0.25 were tested, with the rate of 0.15 producing the model with the greatest accuracy. The leaf size specifies the minimum number of samples required in a leaf, or terminal node. As the classes are imbalanced, lower values for leaf size were tested. Leaf sizes of 15, 30 and 45 were tested, and a leaf size of 30 produced the model with the greatest accuracy.
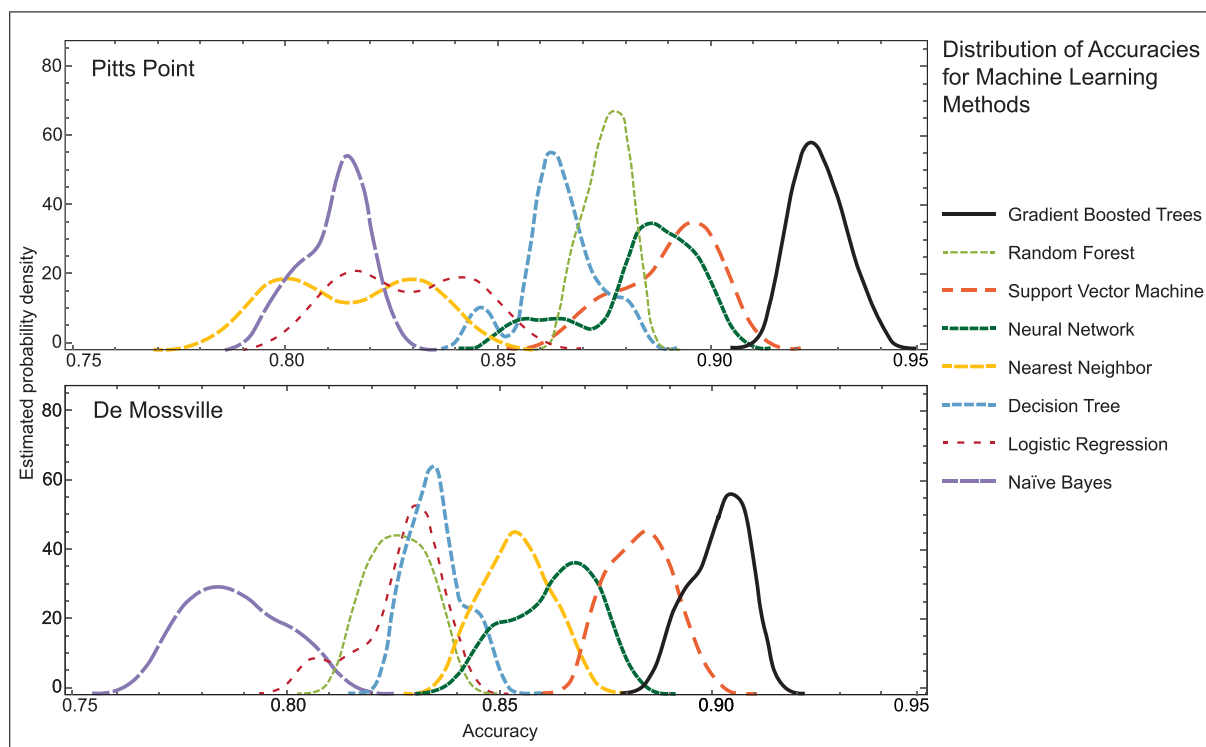
FIGURE 3 Distribution of accuracies for each model tested for the Pitts Point and De Mossville Quadrangles. In both cases, the gradient-boosted trees model produced the greatest accuracy, and the naïve Bayes model produced the least accurate classifier.

4.3 | Variable importance

The 41 variables tried were reduced to the 12 most important variables for each quadrangle. The variables that have the greatest contribution to model accuracy for the Pitts Point quadrangle were the elevation, slope, slope radius 4 and 8, roughness radius 8, residual topography radius 150, 500 and 1500 and the vertical distance to channel network; TWI, and flatness radius 100 and 250. The variables that have the greatest contribution to model accuracy for the De Mossville quadrangle were the elevation, slope, slope radius 4, 8 and 16, roughness radius 8 and 16, residual topography radius 50, 500 and 1500, the vertical distance to channel network and flatness radius 50. A distribution of the resulting accuracies using just the important variables is compared with that from the original classifier in Figure 5. The mean accuracy using all variables for the Pitts Point and De Mossville Quadrangles is 0.926 and 0.901, respectively, and the mean accuracy by using just the most important variables is 0.909 and 0.892. The loss of accuracy then is 0.017 and 0.009, respectively.

4.4 | Probability maps and F1 scores

The most accurate classifier (GBT) with the 12 most important variables was used to classify all cells in three 1.5 × 1.5 km map areas for each quadrangle (Figure 1), and a surficial geologic map was produced using the results. The overall accuracy and F1 scores for each geologic unit are presented in Tables 7 and 8, and the resulting probability maps are presented in Figures 6–9, along with the hand-digitized maps. For comparison, the worst classifier (NB) is also used to classify Map 1 for each quadrangle (Figures 6 and 8).

4.4.1 | Pitts Point

For the Pitts Point quadrangle, the overall accuracy for Map 1 (just outside the training area) is 0.816 using the GBT classifier and to 0.684 using the NB classifier (Table 7). The overall accuracy for Map 2 (within the training area) is 0.912, and for Map 3 approximately 4.5 km away from the training area, it is 0.874. F1 scores are above 0.91 in all map areas for the colluvium and are lowest for the fill and alluvium. The F1 scores for most lithologies are higher in Map 2 than the other two map areas.

The best classifier (GBT) does an excellent job of mapping the colluvium in Map 1 (Figure 6). The boundaries between the terraces and the colluvium are well-defined and in the correct locations. The boundary between the lacustrine and alluvial terraces does not exactly match the boundary in the original map but is reasonably close. The classifier's accuracy is weakest where it misclassifies some of the residuum in the NE quadrant as alluvial terrace 4 and some of the alluvial terrace 4 as residuum in the SW quadrant. Some of these areas have lower probability values (shown as lighter colours). In addition, some areas of the floodplain in the SE quadrant are misclassified as alluvium, alluvial terrace and fill. One feature not mapped in the original map, a roadway that trends NW/SE on the east side of the floodplain, was recognized by the classifier as fill.

The worst performing classifier (NB) also misclassifies the residuum and alluvial terrace. There are extensive areas of the floodplain that are misclassified as fill and alluvium. Much of the alluvial terrace 1&2 is mapped with low confidence or misclassified as lacustrine terrace, and the floodplain is much less extensive than it should be.

The classifications in Map 2 from Pitts Point (Figure 7) are much more accurate across the board, as would be expected as this map is
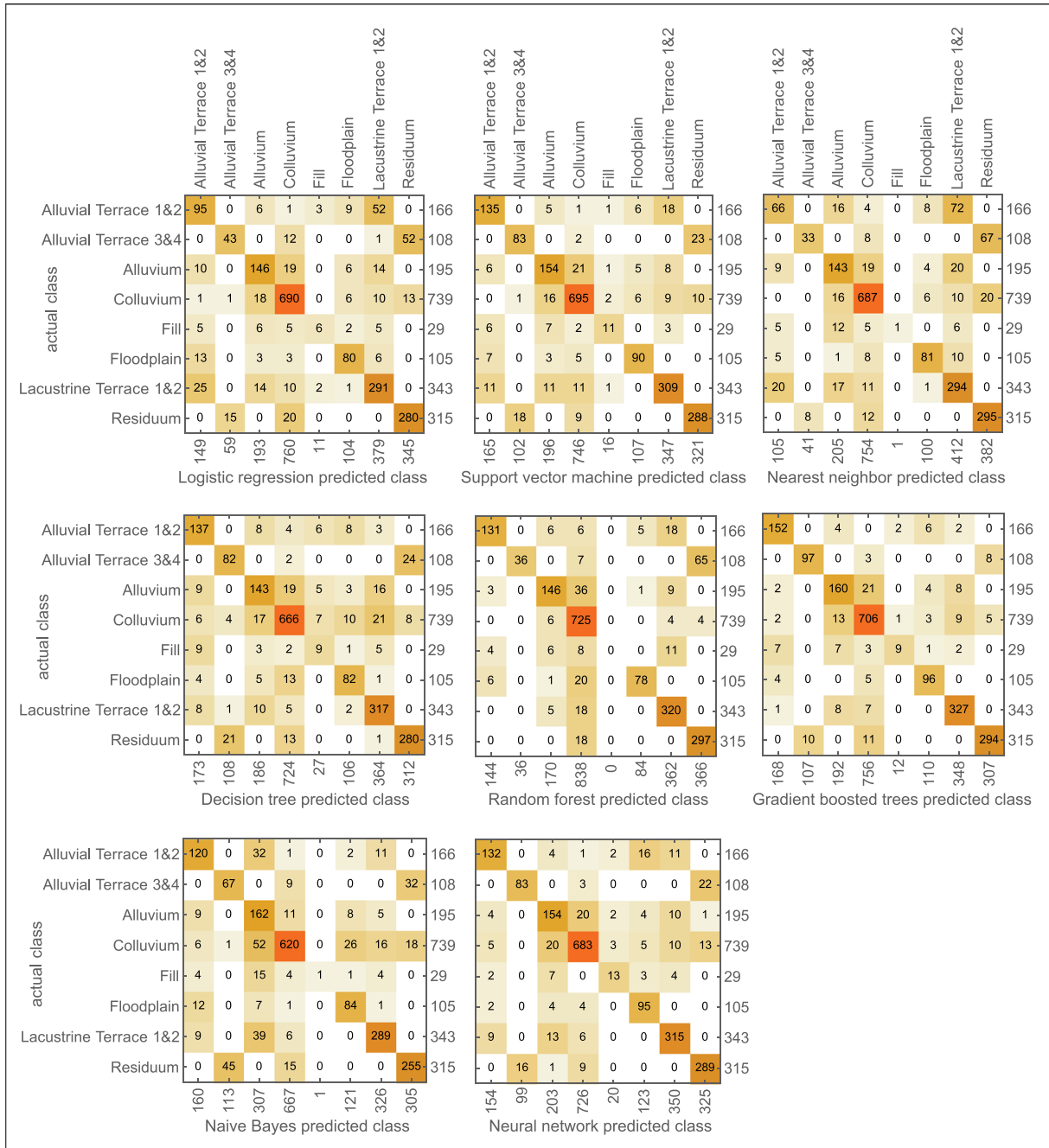
FIGURE 4 Confusion matrices for one training set for the Pitts Point quadrangle using each of the eight ML methods.

within the training sample area. The delineation of the colluvium is excellent, and the boundaries between the lacustrine terraces, alluvial terraces and floodplain are sharp and in the correct location. The boundary between alluvial terrace 4 and the colluvium is mostly correct, though there are small areas where residuum is incorrectly classified as alluvial terrace 4. While most of the floodplain and alluvium is accurately mapped, some of the alluvium is misclassified as alluvial terrace 1&2. One of the alluvial fans (which are grouped together with alluvium for this study) in the SW quadrant is mapped as alluvium, and the other is classified as a combination of colluvium and lacustrine terrace. The areas of fill in the NW quadrant are roughly identified. Two isolated remnants of lacustrine terrace preserved in the central area of the map are partially misclassified as fill.

In Map 3 from Pitts Point (Figure 7), the boundary between the colluvium and residuum, and the residuum itself, is accurately delineated. The boundary of the lacustrine terrace and alluvium is roughly correct in the central part of the map but is misclassified as alluvium in the northern and southern portions. The areas of colluvium accumulation (which are grouped together as colluvium for the classification) are mapped as a splotchy combination of colluvium and alluvium. Fill is misclassified as alluvium. The heads of tributaries consisting of alluvium are identified along the eastern edge of the map.

4.4.2 | De Mossville

The overall accuracy for Map 1 in the De Mossville quadrangle (just outside the training area) is 0.807 using the GBT classifier and 0.708 using the NB classifier (Table 8). The overall accuracy for Map 2 (within the training area) is 0.888; and for Map 3, it is 0.907. F1

TABLE 5 F1 scores for each lithologic class for the Pitts Point quadrangle and the overall accuracy for each machine learning model.

Pitts Point geologic unit	# samples	Gradient-boosted trees	Random Forest	Support vector machine	Neural network	Nearest neighbours	Decision tree	Logistic regression	Naïve Bayes
Residuum	1576	0.950 ± 0.008	0.884 ± 0.009	0.914 ± 0.014	0.898 ± 0.020	0.845 ± 0.011	0.898 ± 0.012	0.850 ± 0.011	0.823 ± 0.012
Colluvium	3696	0.950 ± 0.004	0.921 ± 0.004	0.943 ± 0.007	0.939 ± 0.005	0.920 ± 0.006	0.915 ± 0.007	0.926 ± 0.004	0.894 ± 0.008
Alluvial terrace 3&4	539	0.894 ± 0.025	0.529 ± 0.042	0.782 ± 0.063	0.737 ± 0.113	0.485 ± 0.055	0.777 ± 0.022	0.533 ± 0.050	0.576 ± 0.040
Lacustrine terraces 1&2	1713	0.941 ± 0.012	0.920 ± 0.032	0.904 ± 0.011	0.904 ± 0.016	0.804 ± 0.033	0.900 ± 0.010	0.832 ± 0.028	0.864 ± 0.009
Alluvial terraces 1&2	832	0.891 ± 0.036	0.819 ± 0.051	0.765 ± 0.043	0.747 ± 0.076	0.380 ± 0.152	0.782 ± 0.047	0.512 ± 0.086	0.732 ± 0.053
Floodplain	526	0.907 ± 0.020	0.836 ± 0.032	0.818 ± 0.034	0.830 ± 0.039	0.730 ± 0.034	0.820 ± 0.028	0.735 ± 0.036	0.730 ± 0.031
Alluvium	974	0.856 ± 0.014	0.820 ± 0.015	0.821 ± 0.018	0.824 ± 0.020	0.745 ± 0.020	0.751 ± 0.032	0.767 ± 0.021	0.687 ± 0.028
Fill	144	0.473 ± 0.103	0.000 ± 0.000	0.427 ± 0.101	0.454 ± 0.065	0.111 ± 0.069	0.266 ± 0.086	0.324 ± 0.084	0.208 ± 0.081
Overall Accuracy	10 000	0.926 ± 0.006	0.875 ± 0.005	0.890 ± 0.011	0.883 ± 0.016	0.814 ± 0.016	0.864 ± 0.010	0.828 ± 0.014	0.811 ± 0.007

TABLE 6 F1 scores for each lithologic class for the De Mossville quadrangle, and the overall accuracy for each machine learning model.

De Mossville geologic unit	# samples	Gradient-boosted trees	Random Forest	Support vector machine	Neural network	Nearest neighbours	Decision tree	Logistic regression	Naïve Bayes
Residuum	702	0.805 ± 0.024	0.613 ± 0.030	0.767 ± 0.025	0.744 ± 0.020	0.743 ± 0.031	0.659 ± 0.032	0.682 ± 0.026	0.614 ± 0.037
Colluvium	5943	0.942 ± 0.006	0.886 ± 0.003	0.938 ± 0.003	0.935 ± 0.003	0.931 ± 0.004	0.914 ± 0.006	0.923 ± 0.006	0.891 ± 0.009
High-level alluvium	610	0.854 ± 0.017	0.681 ± 0.041	0.826 ± 0.017	0.797 ± 0.021	0.796 ± 0.030	0.747 ± 0.028	0.705 ± 0.030	0.694 ± 0.034
Alluvial Terraces 3&4	467	0.883 ± 0.021	0.790 ± 0.036	0.857 ± 0.029	0.826 ± 0.026	0.790 ± 0.030	0.790 ± 0.028	0.738 ± 0.027	0.783 ± 0.028
Lacustrine terraces 1&2	597	0.874 ± 0.022	0.816 ± 0.026	0.821 ± 0.033	0.762 ± 0.029	0.744 ± 0.029	0.784 ± 0.036	0.688 ± 0.022	0.742 ± 0.038
Alluvial terraces 1&2	356	0.868 ± 0.033	0.750 ± 0.041	0.770 ± 0.043	0.682 ± 0.049	0.644 ± 0.024	0.714 ± 0.048	0.592 ± 0.046	0.677 ± 0.043
Floodplain	382	0.938 ± 0.021	0.875 ± 0.016	0.917 ± 0.026	0.875 ± 0.045	0.838 ± 0.040	0.876 ± 0.025	0.803 ± 0.027	0.792 ± 0.038
Alluvium	645	0.862 ± 0.023	0.717 ± 0.049	0.826 ± 0.021	0.783 ± 0.035	0.732 ± 0.031	0.741 ± 0.030	0.707 ± 0.023	0.644 ± 0.028
Fill	298	0.405 ± 0.072	0.000 ± 0.000	0.348 ± 0.073	0.293 ± 0.071	0.151 ± 0.049	0.226 ± 0.061	0.217 ± 0.063	0.203 ± 0.042
Overall Accuracy	10 000	0.901 ± 0.006	0.826 ± 0.007	0.883 ± 0.007	0.862 ± 0.010	0.853 ± 0.008	0.834 ± 0.006	0.826 ± 0.010	0.787 ± 0.011

scores are above 0.919 in all map areas for the colluvium using the GBT model and lowest for the fill. The F1 scores for the residuum and colluvium are higher in Map 3 than the other two map areas. The F1 scores for the residuum and high-level alluvium are low for Map 2.

The probability maps for Map 1 show that the best classifier (GBT) does an excellent job of mapping the colluvium (Figure 8). The lacustrine terrace and all are accurately mapped in much of the map area, though there is a misclassification as alluvial terrace 1&2 near the areas of alluvium, and some of the lacustrine terrace is mis-classified as alluvium. The classifier misclassifies some of the residuum

as high-level alluvium on the ridgetops. Some of the areas of fill in the SW quadrant have been correctly classified, but larger areas of fill in the NW quadrant have not been classified correctly, and some areas of alluvium have been incorrectly classified as fill. Areas of colluvium accumulation in the SW quadrant have been incorrectly classified as a combination of lacustrine terrace and alluvium.

The worst performing classifier (NB) produced probability maps with large areas of such low confidence in the lacustrine terrace area that they mapped as transparent, though the areas of residuum, colluvium and alluvium are roughly correct.
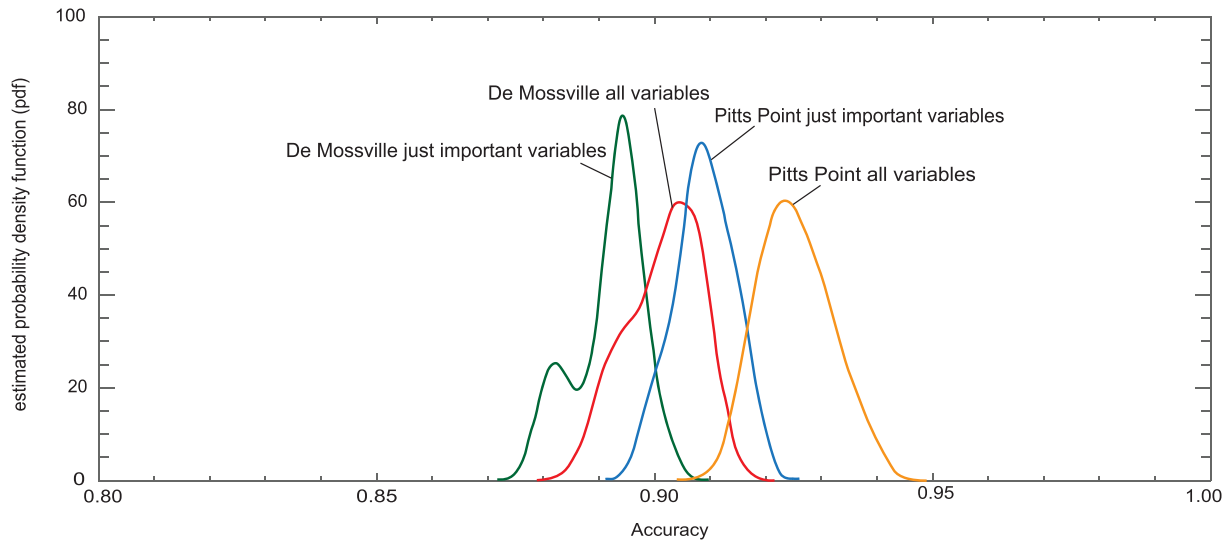
FIGURE 5 Accuracy distribution of the gradient-boosted trees model with the original 41 variables and with the 12 most important variables for the Pitts Point and De Mossville quadrangles.

TABLE 7 F1 scores for each class and overall accuracy results for the maps in the Pitts Point quadrangle.

Pitts Point Geologic unit	F1 scores tenfold cross-validated models	Map 1 (just outside training area)			Map 2 (within training area)		Map 3 (4.5 km distant)	
		# of samples	F1 score GBT	F1 score Naïve Bayes	# of samples	F1 score GBT	# of samples	F1 score GBT
Residuum	0.933 ± 0.012	101 846	0.851	0.880	38 197	0.880	18 810	0.874
Colluvium	0.942 ± 0.005	426 752	0.933	0.906	265 195	0.917	656 125	0.936
Alluvial terraces 3&4	0.850 ± 0.043	8166	0.337	0.248	111 024	0.953	0	-
Lacustrine terraces 1&2	0.927 ± 0.007	152 551	0.806	0.672	256 506	0.957	213 565	0.880
Alluvial terraces 1&2	0.889 ± 0.017	149 391	0.710	0.319	133 296	0.885	0	-
Floodplain	0.894 ± 0.017	129 278	0.706	0.433	109 501	0.904	0	-
Alluvium	0.824 ± 0.017	30 504	0.415	0.209	70 806	0.775	100 549	0.615
Fill	0.400 ± 0.094	1470	0.000	0.022	15 439	0.703	10 933	0.000
Overall accuracy	0.909 ± 0.005		0.816	0.684		0.912		0.874

In Map 2 from the De Mossville quadrangle (Figure 9), which is within the training area, the classifications are much more accurate overall than for Map 1. The delineation of the colluvium and terraces are well-defined and in the correct location. Some of the high-level alluvium along the eastern edge of the map has been correctly classified, but some has been misclassified as residuum. Residuum along the ridgetops has a gradational contact with the colluvium. The alluvial terrace 3&4 is roughly correct, though the boundaries may be misclassified as lacustrine terrace. Small regions of alluvial terrace 1&2 are correctly mapped. Some of the fill has been correctly classified, particularly in the NE quadrant along the ridgetops and colluvial valleys and the roadway in the SE quadrant but was not recognized within the terrace area.

In Map 3 from De Mossville (Figure 9), the boundary between the colluvium and residuum, and the residuum itself, is quite close to the original map, though the ends of the ridges have a lower probability of being residuum, and grade into the colluvium. In one valley in the SW quadrant, the alluvium in the valley bottom is not recognized by the classifier. Some areas of fill within the colluvial valleys are correctly classified but fill but along the ridgetops; there are some areas that are classified as fill but that were not included on the original map. There are some areas along the ridgetops that are incorrectly classified as high-level alluvium.

5 | DISCUSSION

5.1 | Model performance

The GBT model produced the classifier with the highest overall accuracy for each map area and the highest F1 score for each class. While GBT and the RF methods both utilize independent decision trees,

TABLE 8 F1 scores for each class and overall accuracy results for the maps in the De Mossville quadrangle.

De Mossville Geologic unit	F1 scores tenfold cross-validated models	Map 1 (just outside training area)			Map 2 (within training area)		Map 3 (6.0 km distant)	
		# of samples	F1 score GBT	F1 score Naïve Bayes	# of samples	F1 score GBT	# of samples	F1 score GBT
Residuum	0.806 ± 0.027	18 469	0.725	0.789	39 172	0.645	193 432	0.810
Colluvium	0.940 ± 0.005	451 110	0.919	0.886	492 653	0.936	778 603	0.962
High-level alluvium	0.843 ± 0.040	0	-	-	27 939	0.626	0	-
Alluvial terraces 3&4	0.888 ± 0.020	0	-	-	57 328	0.800	0	-
Lacustrine terraces 1&2	0.876 ± 0.022	294 286	0.790	0.662	165 953	0.911	0	-
Alluvial terraces 1&2	0.858 ± 0.017	233	0.000	0.000	25 450	0.709	0	-
Floodplain	0.936 ± 0.011	0	-	-	78 298	0.944	0	-
Alluvium	0.846 ± 0.020	212 890	0.790	0.733	95 521	0.903	14 400	0.418
Fill	0.398 ± 0.071	23 011	0.151	0.117	17 626	0.304	13 540	0.217
Overall Accuracy	0.892 ± 0.006		0.807	0.708		0.888		0.907

GBT uses boosting to combine, or create ensembles of, individual trees in a sequence, while RF uses individual trees and combines them in parallel (Dhingra, 2020). Ensemble methods have been shown to work well with sparse ground-truth data, and it gives robust estimates of the probability of predicted lithology (Bergen et al., 2019; Kuhn, Cracknell, & Reading, 2018). In GBT each decision tree is evaluated in terms of its loss, which is high when the classification and prediction do not agree, and each new tree corrects the errors of the previous tree. GBT can thus model more complex relationships and decision boundaries than a RF model (Friedman, 2001).

The NB model produced the classifier with the lowest overall accuracy for each map area, and with some exceptions, the lowest F1 scores for each class. The NB model is a probabilistic method based on Bayes' theorem, which assumes that all the variables in the data set are independent of each other and do not affect each other (i.e., they are naïve). It is a simpler model than either of the ensemble methods described above. This type of model is often used for things like spam filtering and recommendation systems (Gandhi, 2018). In this study, because the variables are not independent but correlate with each other (e.g., the terraces are at a lower elevation and have lower roughness than the residuum), this method did not perform as well as others.

5.2 | Strengths and limitations of machine learning

5.2.1 | Class imbalance

The selection of the training area from which the 10 000 points were randomly chosen influences the depiction or representation of each geologic unit and affects the imbalance between the classes. While an effort was made to include archetypical areas of all map units in the training area, there is a preponderance of certain units. For example, in the Pitts Point and De Mossville quadrangles, 37% and 59% of the sample points, respectively, were colluvium. It is not surprising then that the colluvium had, with two exceptions, the highest F1 scores in

all maps in both quadrangles (up to 0.936 and 0.962, respectively), with the lowest F1 score of 0.886 using the worst performing classifier. This study did not attempt to map geologic units with only a small arial extent, such as the alluvial fans or colluvial accumulation areas, as these regions were not adequately represented in the training area data used to build the classifiers. While we addressed class imbalance through stratified sampling for this study, a future study might include a different sampling strategy, whereby the number of samples for each class are balanced. This might take the form of a series of key map areas with exemplars of each class, instead of a single contiguous rectangular training area as used in the current study. Another approach would be to undersample the majority class (i.e., the colluvium), and duplicate examples, or oversample, from the minority classes (Brownlee, 2021).

5.2.2 | Ease of classification

In addition to class imbalance having an impact on the ability of ML to correctly classify units, there may be some units that are easier to classify than others based on how unique their characteristics are in the topographic data. For example, colluvium is well recognized by ML, perhaps not only because it has the greatest representation in the training data but also because it is found exclusively on slopes, as opposed to the units found on ridge tops or in the valleys. There is a persistent confusion between residuum and high-level alluvium and terrace deposits, which occupy a similar geomorphic position. It may be that distinguishing between these two deposits is beyond the capacity allowed by using just topographic information and may need additional data such as a description of the soil itself, or, as the high-level alluvium is related to the ancient Teays river system, information about the location of that system. That fill does not have high F1 scores in any map is not too surprising. While the naturally occurring surficial geologic units have a particular genesis that defines them, fill may be placed in a variety of locations that do not conform to a general origin, such as roads along ridgetops, terraces or slopes, dams in
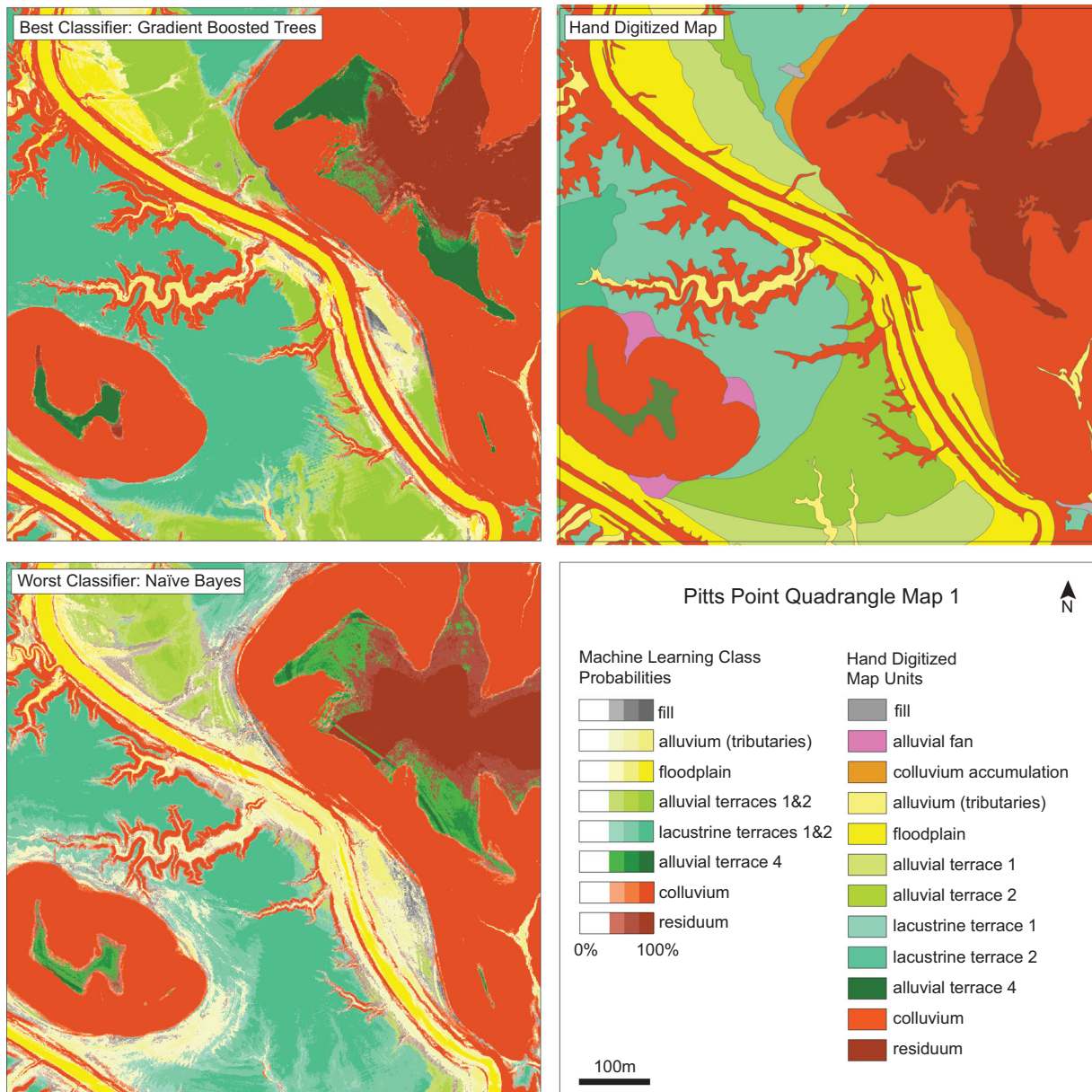
FIGURE 6 Map 1 in the Pitts Point quadrangle, which lies just outside of the training area (Figure 1). The best performing classifier (GBT) and worst performing classifier (naïve Bayes) are compared with the hand-digitized map.

small valleys in the uplands or lowlands, and as the foundations for buildings and other infrastructure. However, there are still examples of locations where fill has been correctly identified, though perhaps with not a great degree of confidence, but a number of dams and roads are at least sketchily mapped in the correct locations (e.g., Figures 8 & 9).

5.2.3 | Topographic control

There is a strong correlation between elevation and the landforms, with residuum residing solely on ridgetops, lacustrine terraces in the valley and so on. In addition, the esp ridges, as they are the remnants of a plateau, are confined to a small range of elevations. It would be informative to use this method in an area of more complex and varied terrain where the elevation does not have as clear a relationship with the surficial deposits.

5.3 | ML performance in areas of complex and simple geology

Map areas within the training areas (Map 2 for both quadrangles) have the highest overall accuracy and F1 scores, as might be expected. Map 3 in both quadrangles lies 4.5–6.0 km away from the training area and have a relatively simple geology. The classifier does do an excellent job mapping the colluvium, residuum and terraces in these more distant areas, though a poorer job mapping alluvium. Map 3 from the De Mossville quadrangle has a very high F1 scores for colluvium of 0.962, and residuum of 0.810, and a much lower F1 score of 0.418 for alluvium. Map 3 from the Pitts Point quadrangle also very high scores for colluvium (0.936), residuum (0.874) and lacustrine terraces (0.880), though a much lower F1 score for alluvium (0.615).

Map 1 in both quadrangles lies just outside of the training area and has a more complex geology than Map 3. The classifier however does perform well, with overall accuracies of 0.816 and 0.807,
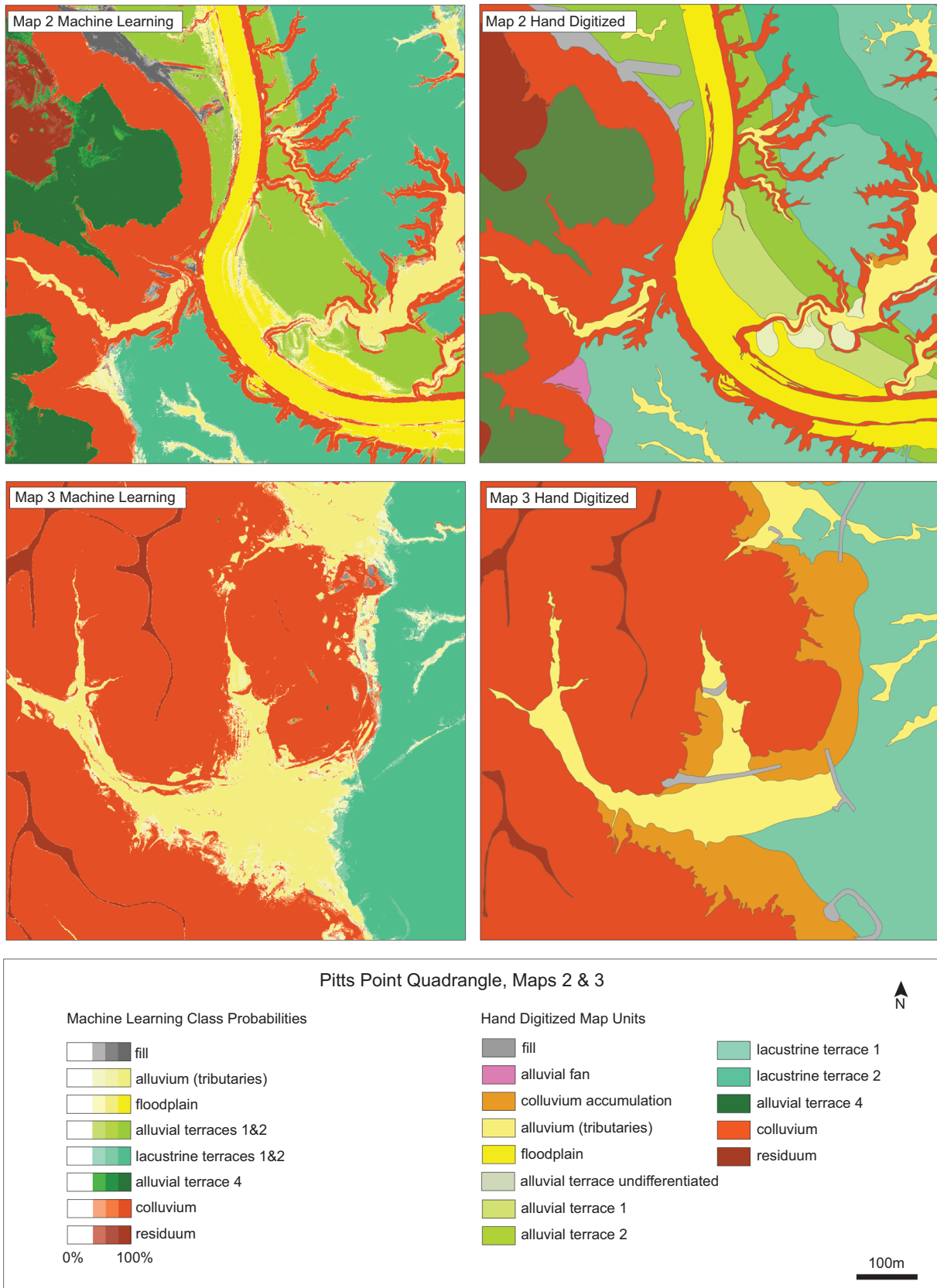
FIGURE 7 Maps 2 and 3 in the Pitts Point quadrangle, which lies within the training area, and 4.5 km away from the training area, respectively (Figure 1).

respectively. While the classifiers do an excellent job on mapping the colluvium, with F1 scores of 0.933 and 0.919, respectively, F1 scores drop to 0.851 and 0.725 for residuum, primarily because of misclassification as high-level alluvium. Lacustrine terraces 1&2 and

alluvial terraces 1&2 have moderately good F1 scores, of 0.806, 0.710 and 0.790, while the alluvial terraces 3&4 have low F1 scores, of 0.337, likely because they are less well-represented in the data used to train the model. While some of these lower F1 scores may reflect
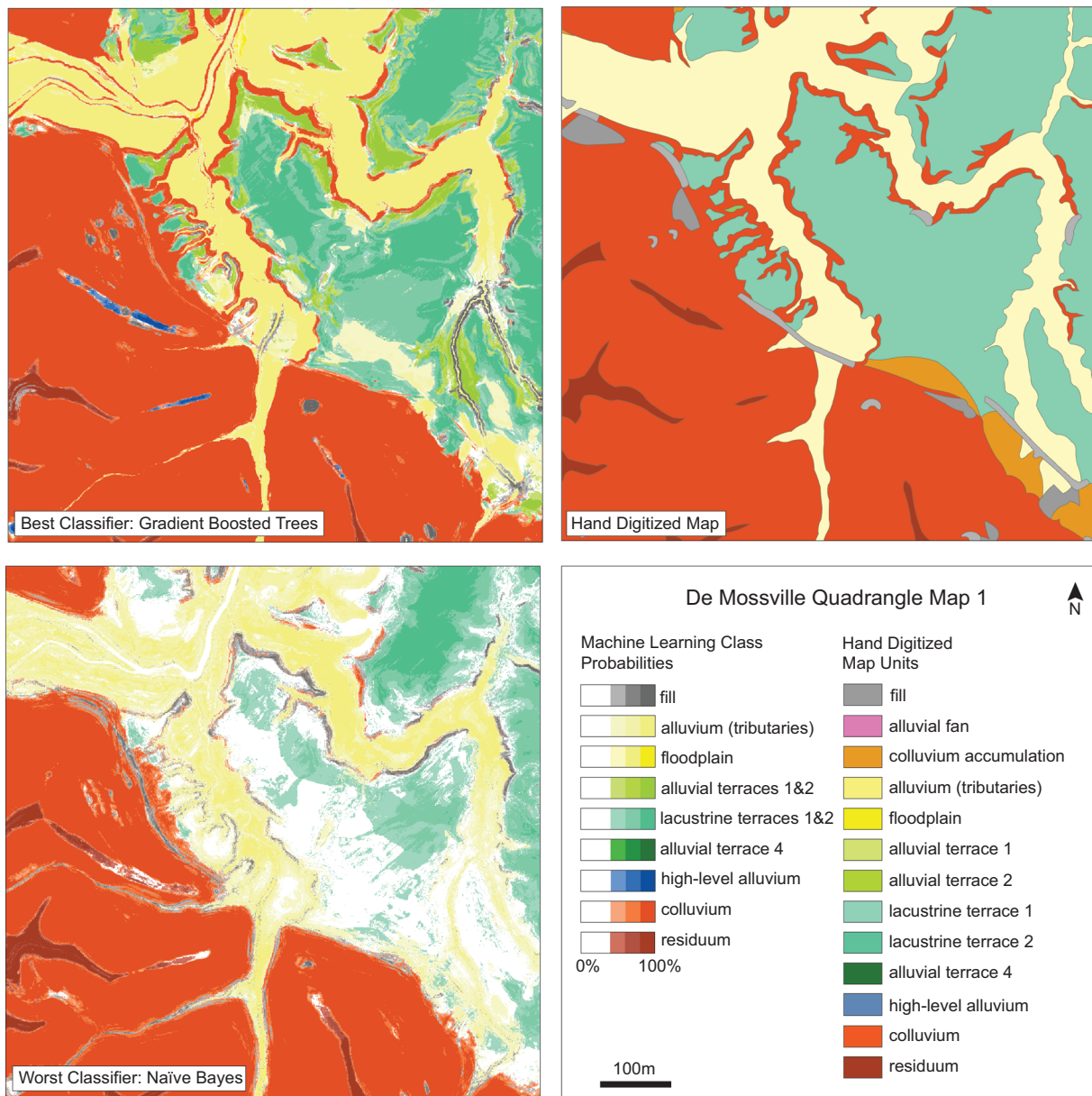
FIGURE 8 Map 1 in the De Mossville quadrangle, which lies just outside of the training area. The best performing classifier (GBT) and worst performing classifier (naïve Bayes) are compared with the hand-digitized map.

that this classifier is not performing well, it is worth keeping in mind that the ML maps display this uncertainty and highlight areas where there is more than one potential geologic interpretation. This may be very useful for pointing out areas where further investigation is needed for accurate mapping. Furthermore, it is possible that the ML maps are picking up on details not noticed during the hand-digitized mapping process.

5.4 | Mapping uncertainty

An advantage of using ML for mapping is that it communicates the probability of each geologic unit at every cell and thus communicates the level of certainty of that classification at every cell. Areas of uncertainty may coincide with lithologic transitions of key geological importance (Cracknell & Reading, 2013). For example, most of the

maps in this study show a gradational boundary between the residuum and colluvium, and this might be useful for the user to know so that this uncertainty could be incorporated into their plan to evaluate a site for potential slope mitigation. Areas of uncertainty may also point to areas of real geologic complexity. For example, in Map 3 for the Pitts Point quad (Figure 7), there are significant areas of colluvium accumulation, which in this study, we have defined as colluvium. In these areas, the ML model has classified this as a patchy combination of colluvium, alluvium and lacustrine terrace. According to the Kentucky Geologic Map Service, colluvium accumulation zones consist of a combination of colluvium, alluvium, lacustrine deposits, alluvial terraces and fans. Thus, the classifier has pointed to the multigenetic origins of these deposits. Using ML for mapping also eliminates the need for boundary lines, which are known to have their own uncertainty issues (Lark et al., 2015). If a boundary between units is desired, the user could apply a threshold for which a boundary can be applied.
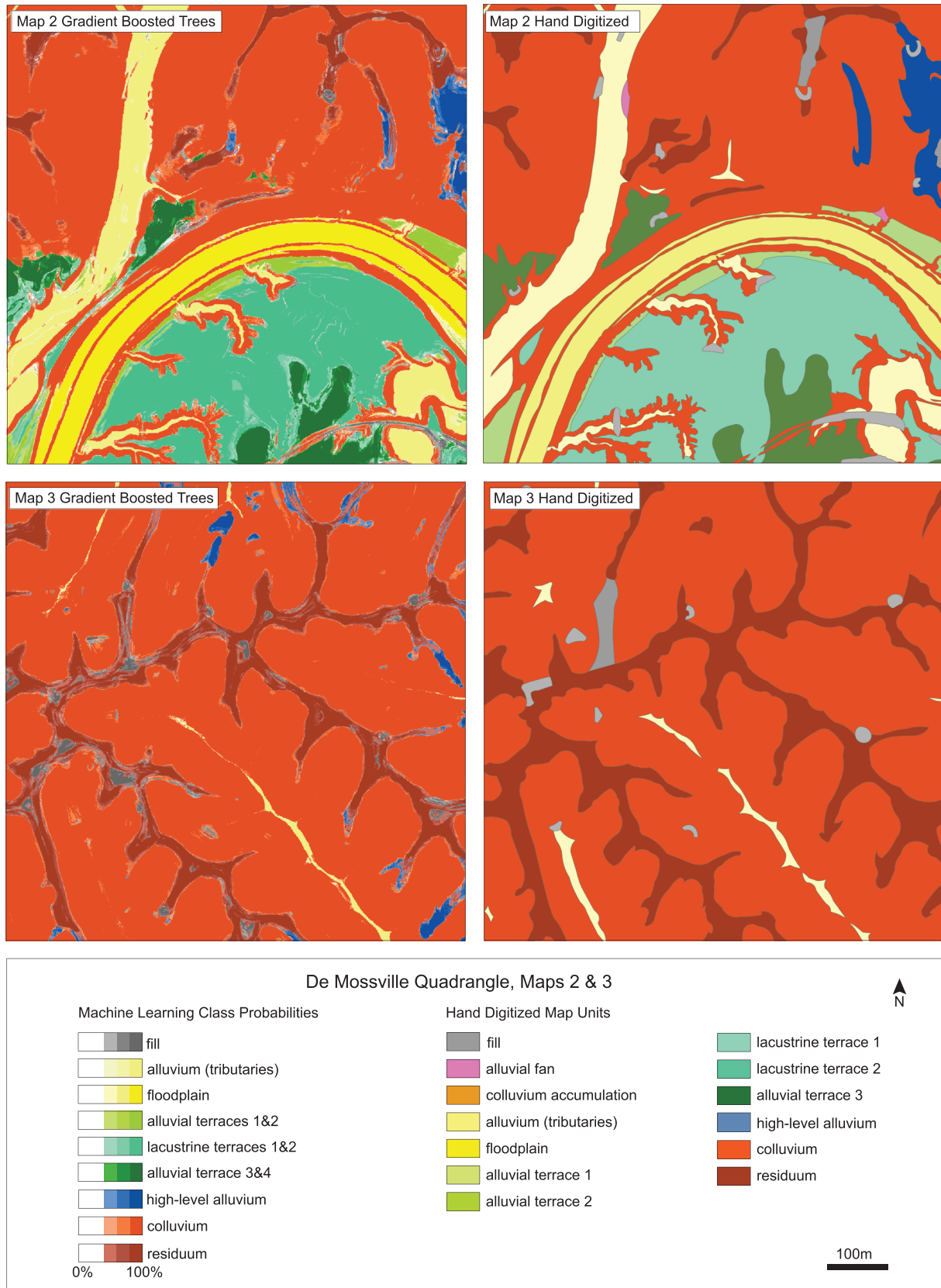
FIGURE 9 Maps 2 and 3 in the De Mossville quadrangle, which lies within the training area, and 6.0 km away from the training area, respectively (Figure 1).

5.5 | Can ML replace human mapping?

For some geologic units, such as the colluvium, ML appears to do an excellent job of classifying. There are extensive areas outside of the

two quadrangles used in this study, where the geology is dominated by colluvium and residuum with some valley alluvium and fill. In these areas of simple geology (i.e., De Mossville Map 3), ML produced maps with overall accuracies of up to 90.7%. This work

demonstrates the strong potential for ML to quickly create accurate surficial geologic maps in geologically simple areas that can closely match human mapping. In more complex areas, particularly where some units, such as terraces, are not well preserved, or there are high-level alluvial deposits on the ridgetops, additional information obtained in the field is necessary. In such areas, ML may not create an accurate map without additional human mapping but could instead provide a base map highlighting where areas of geologic uncertainty lie. There is also the potential for ML performance to be enhanced by including other data sets such as soil samples, well log data or remote sensing data, which could help in distinguishing these geologic units from one another.

5.6 | Future directions

This study used supervised machine learning algorithms, which use human classifications of geologic units as their basis for measuring model performance. A future line of inquiry is to use unsupervised machine learning, in which no classifications are provided, and the algorithm defines its own classification model. Another line of inquiry is to train a classifier using samples, which equally sample all geologic units, thereby eliminating class imbalance, or using an oversampling/undersampling routine. This may allow smaller features like alluvial fans and colluvial accumulation zones to be mapped. However, it may be that some of these features, as they are an amalgamation of multiple types of deposits, may not be classified very well in any case. This study relied solely on the digital elevation model and derivative maps. Other data derived from the original lidar data, such as lidar intensity, could potentially be used as variables, along with existing county soils maps, imagery and other remote sensing data such as synthetic aperture radar (SAR) data. Additional neural network architectures could be explored, such as convoluted neural networks (CNN) or recurrent neural networks (RNN). Furthermore, there is the potential for automatic feature selection from topographic data using a deep neural network (Kirkwood, 2022). It would also be useful to examine how well the models developed in this study might work in areas outside of the two quadrangles used here.

6 | CONCLUSIONS

This study demonstrates that machine learning can be a valuable tool for surficial geologic mapping. It can produce a near-final map in areas of simple geology and a base map where the geology is more complex. A great benefit in using ML is that the probability for each geologic unit at every cell in a map can be conveyed. This can be used to highlight areas where additional fieldwork is required and also communicate to the end user the real geologic uncertainty between units. The GBT method produced the best performing classifier for both map areas tested. It produced very accurate maps of colluvium but had difficulty distinguishing between some units in the same topographic position such as residuum and high-level alluvial deposits. This method has the potential to be useful for mapping the surficial geology in other regions where lidar data is available.

ORCID

Sarah E. Johnson https://orcid.org/0000-0001-7335-4959
William C. Haneberg https://orcid.org/0000-0002-1254-2507

REFERENCES

Baum, R.L. & Johnson, A.M. (1996) Overview of landslide problems, research, and mitigation, Cincinnati, Ohio, Area. Report 2059A. U.S. Geological Survey Bulletin. Washington, DC. USGS Publications Warehouse. Available from: https://doi.org/10.3133/b2059A

Behrens, T., Zhu, A.-X., Schmidt, K. & Scholten, T. (2010) Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma*, 155(3–4), 175–185. Available from: https://doi.org/10.1016/j.geoderma.2009.07.010

Berg, R.C., MacCormack, K.E. & Russell, H.A.J. (2019) Benefit-cost analysis for building 3D maps and models. In: *Alberta Energy Regulator/Alberta Geological Survey Special Report*, Vol. 112, p. 8104. Alberta, Canada: Alberta Geological Survey.

Bergen, K.J., Johnson, P.A., de Hoop, M.V. & Beroza, G.C. (2019) Machine learning for data-driven discovery in solid earth geoscience. *Science*, 363(6433), eaau0323. Available from: https://doi.org/10.1126/science.aau0323

Bernard, E. (2021) *Introduction to machine learning*. Champaign: Wolfram Media.

Bernknopf, R.L., Brookshire, D.S., Soller, D.R., McKee, M.J., Sutter, J.F., Matti, J.C., et al. (1993) *Societal value of geologic maps*. Geological Survey Circular: Circular. U.S, p. 1111.

Beven, K.J. & Kirkby, M.J. (1979) A physically based, variable contributing area model of basin hydrology/un Modèle à base physique de zone d'appel variable de l'hydrologie Du Bassin versant. *Hydrological Sciences Bulletin*, 24(1), 43–69. Available from: https://doi.org/10.1080/02626667909491834

Bhagwat, S.B. & Ipe, V.C. (2000) Economic Benefits of Detailed Geologic Mapping to Kentucky. In: *Special Report Special Report 3*. Champaign, IL: Illinois State Geological Survey.

Bodaghabadi, B., Mohsen, J.A.M.-C., Salehi, M.H., Mohammadi, J., Borujeni, I.E., Toomanian, N., et al. (2015) Digital soil mapping using artificial neural networks and terrain-related attributes. *Pedosphere*, 25(4), 580–591. Available from: https://doi.org/10.1016/S1002-0160(15)30038-2

Boettinger, Janis L., David W. Howell, Amanda C. Moore, Alfred E. Hartemink, and Suzann Kienast-Brown. 2010. *Digital soil mapping: bridging research, environmental application, and operation*. Berlin, Germany: Springer Science & Business Media, https://doi.org/10.1007/978-90-481-8863-5

Breiman, L. (2001) Random forests. *Machine Learning*, 45(1), 5–32. Available from: https://doi.org/10.1023/A:1010933404324

Brownlee, J. (2021) Random oversampling and Undersampling for imbalanced classification. *Machine Learning Mastery (blog)*. January 2021, Available from: https://machinelearningmastery.com/random-oversampling-and-undersampling-for-imbalanced-classification/

Cortes, C. & Vapnik, V. (1995) Support-vector networks. *Machine Learning*, 20(3), 273–297. Available from: https://doi.org/10.1007/BF00994018

Cox, D.R. (1958) The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(2), 215–242. Available from: https://doi.org/10.1111/j.2517-6161.1958.tb00292.x

Cracknell, M.J. & Reading, A.M. (2013) The upside of uncertainty: identification of lithology contact zones from airborne geophysics and satellite data using random forests and support vector machines. *Geophysics*, 78(3), WB113–WB126. Available from: https://doi.org/10.1190/geo2012-0411.1

Cracknell, M.J. & Reading, A.M. (2014) Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers & Geosciences*, 63, 22–33. Available from: https://doi.org/10.1016/j.cageo.2013.10.008

Crawford, M.M. (2012) Using LiDAR to map landslides in Kenton and Campbell counties, Kentucky. *Kentucky Geological Survey Report of Investigations*, 24 Series XII, 18. Available from: https://doi.org/10.13023/kgs.ri24.12

Crawford, M.M., Dortch, J.M., Koch, H.J., Killen, A.A., Zhu, J., Zhu, Y., et al. (2021) Using landslide-inventory mapping for a combined bagged-trees and logistic-regression approach to determining landslide susceptibility in eastern Kentucky, USA. *Quarterly Journal of Engineering Geology and Hydrogeology*, 54(4), qjegh2020-177. Available from: https://doi.org/10.1144/qjegh2020-177

Derouin, S. (2021) Detailed geologic mapping helps identify health hazards. *Eos* 102 (July). https://eos.org/articles/detailed-geologic-mapping-helps-identify-health-hazards

Dharumarajan, S. & Hegde, R. (2022) Digital mapping of soil texture classes using random Forest classification algorithm. *Soil Use and Management*, 38(1), 135–149. Available from: https://doi.org/10.1111/sum.12668

Dhingra, C. (2020) A visual guide to gradient boosted trees. *Towards Data Science. 2020*. Available from: https://towardsdatascience.com/a-visual-guide-to-gradient-boosted-trees-8d9ed578b33

Dou, J., Yunus, A.P., Bui, D.T., Merghadi, A., Sahana, M., Zhu, Z., et al. (2019) Assessment of advanced random Forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Science of the Total Environment*, 662, 332–346. Available from: https://doi.org/10.1016/j.scitotenv.2019.01.221

Durrell, R.H. (1982) A recycled landscape. In: *Cincinnati museum of natural history popular publication series*. Cincinnati, Ohio: Cincinnati Museum of Natural History. Available from:. https://books.google.com/books?id=Xc6btgAACAAJ

Dyer, A.S., Mark-Moser, M.K., Duran, R. & Bauer, J.R. (2024) Offshore application of landslide susceptibility mapping using gradient-boosted decision trees: a Gulf of Mexico case study. *Natural Hazards*, 120(7), 6223–6244. Available from: https://doi.org/10.1007/s11069-024-06492-6

Forman, G. & Scholz, M. (2010) Apples-to-apples in cross-validation studies: Pitfalls in classifier performance measurement. *SIGKDD Explorations*, 12(1), 49–57. Available from: https://doi.org/10.1145/1882471.1882479

Friedman, J.H. (2001) Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. Available from: https://doi.org/10.1214/aos/1013203450

Gandhi, R. (2018) Naive Bayes classifier. *Towards Data Science*, 2018, Available from: https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c

Giasson, E., Figueiredo, S.R., Tornquist, C.G. & Clarke, R.T. (2008) Digital soil mapping using logistic regression on terrain parameters for several ecological regions in Southern Brazil. In: Hartemink, A.E., McBratney, A. & Mendonça-Santos, M.D.L. (Eds.) *Digital soil mapping with limited data*. Dordrecht: Springer Netherlands, pp. 225–232 https://doi.org/10.1007/978-1-4020-8592-5_19

Granger, D. E. & Smith, A.L. (1998) Early Laurentide glaciation and creation of Ohio river dated by radioactive decay of cosmogenic Al-26 and Be-10 in proglacial sediments. In *Geological Society of America abstracts with programs*. Boulder, CO: Geological Society of America, Vol. 30, no. 7, p. 298.

Grohmann, C.H., Smith, M.J. & Riccomini, C. (2011) Multiscale analysis of topographic surface roughness in the Midland Valley, Scotland. *IEEE Transactions on Geoscience and Remote Sensing*, 49(4), 1200–1213. Available from: https://doi.org/10.1109/TGRS.2010.2053546

Gruber, F.E., Baruck, J., Mair, V. & Geitner, C. (2019) From geological to soil parent material maps—a random forest-supported analysis of geological map units and topography to support soil survey in South Tyrol. *Geoderma*, 354 (November, 113884. Available from: https://doi.org/10.1016/j.geoderma.2019.113884

GSA. (2022) *GSA position statement: the value of geologic mapping*. Boulder, CO: The Geological Society of America. Available from: https://www.geosociety.org/GSA/Science_Policy/Position_Statements/Current_Statements/gsa/positions/position3.aspx

Hammond, M., Massey, M.A., Andrews, W., Martin, S.L. & Bottoms, A. (2017) Annual workflow of the STATEMAP-funded digital surficial geologic mapping program at the Kentucky geological survey. 291134, Available from: https://doi.org/10.1130/abs/2017NE-291134

Haneberg, W.C., Creighton, A.L., Medely, E.W. & Jonas, D.A. (2005) Use of LiDAR to assess slope hazards at the Lihir gold mine, Papua New Guinea. In: Hungr, O., Fell, R., Couture, R. & Eberhardt, E. (Eds.) *Landslide risk management: Proceedings of international conference on landslide risk management. Vol. supplementary CD. Vancouver, Canada*. London, England: Taylor & Francis.

Harvey, A.S. & Fotopoulos, G. (2016) Geological mapping using machine learning algorithms. In: *ISPRS—International archives of the photogrammetry, remote sensing and spatial information sciences XLI-B8*. Hannover, Germany: International Society for Photogrammetry and Remote Sensing, pp. 423–430. Available from: https://doi.org/10.5194/isprsarchives-XLI-B8-423-2016

Hateffard, F., Dolati, P., Heidari, A. & Zolfaghari, A.A. (2019) Assessing the performance of decision tree and neural network models in mapping soil properties. *Journal of Mountain Science*, 16(8), 1833–1847. Available from: https://doi.org/10.1007/s11629-019-5409-8

Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E. & Schmidt, M.G. (2016) An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, 265, 62–77. Available from: https://doi.org/10.1016/j.geoderma.2015.11.014

Hitziger, M. & Ließ, M. (2014) Comparison of three supervised learning methods for digital soil mapping: application to a complex terrain in the Ecuadorian Andes. *Applied and Environmental Soil Science*, 2014, 1–12. Available from: https://doi.org/10.1155/2014/809495

Huang, Y. & Zhao, L. (2018) Review on landslide susceptibility mapping using support vector machines. *Catena*, 165, 520–529. Available from: https://doi.org/10.1016/j.catena.2018.03.003

Jaiswal, S. (2024) *Multilayer perceptrons in machine learning: a comprehensive guide*. February 2024. New York, NY: Datacamp. Available from: https://www.datacamp.com/tutorial/multilayer-perceptrons-in-machine-learning

Jasiewicz, J. & Stepinski, T.F. (2013) Geomorphons—a pattern recognition approach to classification and mapping of landforms. *Geomorphology*, 182, 147–156. Available from: https://doi.org/10.1016/j.geomorph.2012.11.005

Johnson, S.E., Haneberg, W.C., Bryson, L.S. & Crawford, M.M. (2023) Measuring ground surface elevation changes in a slow-moving Colluvial landslide using combinations of regional airborne Lidar, UAV Lidar, and UAV photogrammetric surveys. *Quarterly Journal of Engineering Geology and Hydrogeology*, 56(2), qjegh2022-078. Available from: https://doi.org/10.1144/qjegh2022-078

Jones, R.R., McCaffrey, K.J.W., Wilson, R.W. & Holdsworth, R.E. (2004) Digital field data acquisition: towards increased quantification of uncertainty during geological mapping. *Geological Society, London, Special Publications*, 239(1), 43–56. Available from: https://doi.org/10.1144/GSL.SP.2004.239.01.04

Kalambukattu, J.G., Kumar, S. & Arya Raj, R. (2018) Digital soil mapping in a Himalayan watershed using remote sensing and terrain parameters employing artificial neural network model. *Environmental Earth Sciences*, 77(5), 203. Available from: https://doi.org/10.1007/s12665-018-7367-9

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(3), 425–439. Available from: https://doi.org/10.1007/s10346-013-0391-7

Kawabata, D. & Bandibas, J. (2009) Landslide susceptibility mapping using geological data, a DEM from ASTER images and an artificial neural network (ANN). *Geomorphology*, 113(1–2), 97–109. Available from: https://doi.org/10.1016/j.geomorph.2009.06.006

KGS. (2021) Kentucky Geological Survey 2020–2021 Annual Report. In: *Annual Report*. Lexington, Kentucky: University of Kentucky. https://storymaps.arcgis.com/stories/5adadf024bef4e779dab2fcc1e0251eb

KGS. (2023) *Kentucky geologic map service*. Lexington, KY: Kentucky Geological Survey. https://kgs.uky.edu/kygeode/geomap/

Khaledian, Y. & Miller, B.A. (2020) Selecting appropriate machine learning methods for digital soil mapping. *Applied Mathematical Modelling*, 81, 401–418. Available from: https://doi.org/10.1016/j.apm.2019.12.016

Kirkwood, C. (2022) Geological mapping in the age of artificial intelligence. *Geoscientist*, 32(3), 16–23. Available from: https://doi.org/10.1144/geosci2022-023

Kirkwood, C., Cave, M., Beamish, D., Grebby, S. & Ferreira, A. (2016) A machine learning approach to geochemical mapping. *Journal of Geochemical Exploration*, 167, 49–61. Available from: https://doi.org/10.1016/j.gexplo.2016.05.003

Kirkwood, C., Economou, T., Pugeault, N. & Odbert, H. (2022) Bayesian deep learning for spatial interpolation in the presence of auxiliary information. *Mathematical Geosciences*, 54(3), 507–531. Available from: https://doi.org/10.1007/s11004-021-09988-0

Kovačević, M., Bajat, B. & Gajić, B. (2010) Soil type classification and estimation of soil properties using support vector machines. *Geoderma*, 154(3–4), 340–347. Available from: https://doi.org/10.1016/j.geoderma.2009.11.005

Kuhn, S., Cracknell, M.J. & Reading, A.M. (2018) Lithologic mapping using random forests applied to geophysical and remote-sensing data: a demonstration study from the eastern goldfields of Australia. *Geophysics*, 83(4), B183–B193. Available from: https://doi.org/10.1190/geo2017-0590.1

Lark, R.M., Lawley, R.S., Barron, A.J.M., Aldiss, D.T., Ambrose, K., Cooper, A.H., et al. (2015) Uncertainty in mapped geological boundaries held by a National Geological Survey:eliciting the Geologists' tacit error model. *Solid Earth*, 6(2), 727–745. Available from: https://doi.org/10.5194/se-6-727-2015

Lee, S. & Park, I. (2013) Application of decision tree model for the ground subsidence Hazard mapping near abandoned underground coal mines. *Journal of Environmental Management*, 127, 166–176. Available from: https://doi.org/10.1016/j.jenvman.2013.04.010

Liu, D.C. & Nocedal, J. (1989) On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1–3), 503–528. Available from: https://doi.org/10.1007/BF01589116

Mansuy, Nicolas, Evelyne Thiffault, David Paré, Pierre Bernier, Luc Guindon, Philippe Villemaire, Vincent Poirier, and André Beaudoin. 2014. "Digital mapping of soil properties in Canadian managed forests at 250m of resolution using the K-nearest neighbor method." *Geoderma* 235–236 (December):59–73. https://doi.org/10.1016/j.geoderma.2014.06.032

Massey, M.A. (2017) Surficial geologic map of the De Mossville 7.5-minute quadrangle, North-Central Kentucky (1:24,000). In: *Kentucky geological survey contract report 8–13, 1 sheet*. Lexington, KY: Kentucky Geological Survey.

Massey, M.A., M. III Hammond, A.E. Bottoms, and S.L. Martin. 2019. "Surficial Geologic Map of the Pitts Point 7.5-Minute Quadrangle, North-Central Kentucky (1:24,000)." Kentucky geological survey contract report 20–13, 1 sheet. Lexington, KY: Kentucky Geological Survey.

"Mathematica." 2022. *Mathematica*. Champaign, Illinois: Wolfram Research, Inc. www.wolfram.com/mathematica

McDowell, R.C. (1986) The Geology of Kentucky - A Text to Accompany the Geologic Map of Kentucky. In: *Professional paper 1151-H*. U.S. Geological Survey Professional Paper. Denver, CO: United States Geological Survey. Available from: https://pubs.usgs.gov/pp/p1151h/about.html

Metelka, Václav, Lenka Baratoux, Mark W. Jessell, Andreas Barth, Josef Ježek, and Séta Naba. 2018. "Automated regolith landform mapping using airborne geophysics and remote sensing data, Burkina Faso, West Africa." *Remote Sensing of Environment* 204 (January):964–978. https://doi.org/10.1016/j.rse.2017.08.004

Muñoz, V.A. & Valeriano, M.D.M. (2014) Mapping of flood-plain by processing of elevation data from remote sensing. In: Pardo-Igúzquiza, E., Guardiola-Albert, C., Heredia, J., Moreno-Merino, L., Durán, J.J. & Vargas-Guzmán, J.A. (Eds.) *Mathematics of planet earth*, Lecture Notes in Earth System Sciences. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 481–484 https://doi.org/10.1007/978-3-642-32408-6_106

Ng, A.Y. (2004) Feature Selection, L1 vs. L2 Regularization, and Rotational Invariance. In: *Proceedings of the twenty-first international conference on machine learning*, ICML'04, Vol. 78. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1015330.1015435

Pham, B.T., Bui, D.T., Pourghasemi, H.R., Indra, P. & Dholakia, M.B. (2017) Landslide susceptibility Assesssment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of Naïve Bayes, multilayer perceptron neural networks, and functional trees methods. *Theoretical and Applied Climatology*, 128(1–2), 255–273. Available from: https://doi.org/10.1007/s00704-015-1702-9

Potter, P.E. (2007) *Exploring the geology of the Cincinnati/northern Kentucky region*. Lexington, KY: Kentucky Geological Survey, Special Publication, p. 8.

Reu, D., Jeroen, J.B., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P., et al. (2013) Application of the topographic position index to heterogeneous landscapes. *Geomorphology*, 186 (March, 39–49. Available from: https://doi.org/10.1016/j.geomorph.2012.12.015

Sahin, E.K. (2020) Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random Forest. *SN Applied Sciences*, 2(7), 1308. Available from: https://doi.org/10.1007/s42452-020-3060-1

Shanker, M., Hu, M.Y. & Hung, M.S. (1996) Effect of data standardization on neural network training. *Omega*, 24(4), 385–397. Available from: https://doi.org/10.1016/0305-0483(96)00010-2

Smith, William. 1815. "A delineation of the strata of England and Wales, with part of Scotland; exhibiting the collieries and mines, the marshes and fen lands originally overflowed by the sea, and the varieties of soil according to the variations in the substrata, illustrated by the Most descriptive names."

Sun, D., Wen, H., Wang, D. & Jiahui, X. (2020) A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology*, 362, 107201. Available from: https://doi.org/10.1016/j.geomorph.2020.107201

Tikoff, B., Shipley, T.F., Nelson, E.M., Williams, R.T., Barshi, N. & Wilson, C. (2023) Improving the practice of geology through explicit inclusion of scientific uncertainty for data and models. *GSA Today*, 33(7), 4–9. Available from: https://doi.org/10.1130/GSATG560A.1

Weiss, A.D. (2001) Topographic Position and Landforms Analysis. In: *Presented at the ESRI user conference, San Diego, CA*. Redlands, CA: ESRI.

Winchester, S. (2009) The map that changed the world: William Smith and the birth of modern geology. In: *Repr*. New York, NY: Harper Perennial.

Yeon, Y.-K., Han, J.-G. & Ryu, K.H. (2010) Landslide susceptibility mapping in Injae, Korea, using a decision tree. *Engineering Geology*, 116(3–4), 274–283. Available from: https://doi.org/10.1016/j.enggeo.2010.09.009

Zhu, J. & Pierskalla, W.P. (2016) Applying a weighted random forests method to extract karst sinkholes from LiDAR data. *Journal of Hydrology*, 533 (February, 343–352. Available from: https://doi.org/10.1016/j.jhydrol.2015.12.012

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.