



Research papers

Using machine learning to identify karst sinkholes from LiDAR-derived topographic depressions in the Bluegrass Region of Kentucky

Junfeng Zhu^{a,*}, Adam M. Nolte^{a,d}, Nathan Jacobs^b, Ming Ye^c^a Kentucky Geological Survey, University of Kentucky, Lexington, KY 40506, USA^b Department of Computer Science, University of Kentucky, Lexington, KY 40506, USA^c Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, FL 32306, USA^d Division of Water, Kentucky Energy and Environment Cabinet, Frankfort, KY 40601, USA

ARTICLE INFO

This manuscript was handled by A. Bardossy,
Editor-in-Chief

Keywords:

Sinkhole

LiDAR

Machine learning

Topographic depression

Morphometric characteristic

ABSTRACT

Information about the distribution and characteristics of existing sinkholes is critical for understanding karst aquifer systems and evaluating sinkhole hazards. LiDAR provides accurate and high-resolution topographic information and has been used to improve delineation of sinkholes in many karst regions. LiDAR data also reveal many topographic depressions, however, and identifying sinkholes from these depressions through manual visual inspection can be slow and laborious. To improve the efficiency of the identification process, we applied six machine learning methods (logistic regression, naive Bayes, neural network, random forests, RUSBoost, and support vector machine) to a dataset of morphometric characteristics of LiDAR-derived topographic depressions. Sinkhole data from Bourbon, Woodford, and Jessamine Counties in the Bluegrass Region of Kentucky were used to derive the dataset for training and testing the machine learning methods. The dataset consisted of 22,884 records with 10 variables for each record. For each method, a random subset of 80% of the records was used for training and the remaining 20% was used for testing. The test receiver operating characteristic curves showed that all six methods were applicable to the dataset, as demonstrated by all area under the curves (AUCs) being greater than 0.87. Neural network emerged as the method that performed best, with an AUC of 0.95 and a testing average accuracy of 0.85. To further improve the sinkhole mapping process, we subsequently developed a two-step process that combined the trained neural network classifier and manual visual inspection and applied the process to Scott County, also in the Bluegrass region. We were able to locate 97% of the sinkholes in the county by manually inspecting only 27% of the topographic depressions the neural network classified as having relatively high probabilities of being sinkholes. This study showed that machine learning is a promising method for improving sinkhole identification efficiency in karst areas in which high-resolution topographic information is available.

1. Introduction

Approximately 15% of the ice-free land surface globally is underlain by karst terrain and approximately 20–25% of the global populations depends largely or entirely on the aquifers associated with this terrain (Ford and Williams, 2007); many large population centers rely on these aquifers as their primary water supplies (Chen et al., 2017). Karst aquifers typically have distinct physiographic features that result from the dissolution of bedrock (typically carbonate rocks), and as a result, these aquifers are usually characterized by a network of fractures and conduits that connect to the surface water through sinkholes, sinking streams, and springs. Sinkholes are the most well-known features associated with karst (Schwartz and Zhang, 2003). Sinkholes serve as a

major connection between surface water and groundwater by collecting rainfall and funneling it internally to the subsurface, often via fast flow. Sinkholes also cause damage to property and infrastructure in karst regions throughout the world. In the United States alone, the costs of sinkhole-related damage is estimated to range from \$125 million to \$300 million dollars annually (Kuniansky et al., 2016). Dinger et al. (2007) estimated that in Kentucky the damage associated with sinkholes costs approximately \$20 million to \$24 million dollars per year. The U.S. Disaster Mitigation Act of 2000 requires states to have a state mitigation plan approved by the Federal Emergency Management Agency in order to be eligible for federal hazard mitigation funding. The mitigation plan must include the identification and assessment of natural hazards, such as sinkholes. Detailed information on sinkholes is,

* Corresponding author at: 504 Rose Street, 228 MMRB, Kentucky Geological Survey, University of Kentucky, Lexington, KY 40506, USA.

E-mail address: Junfeng.zhu@uky.edu (J. Zhu).

<https://doi.org/10.1016/j.jhydrol.2020.125049>

Received 14 February 2020; Accepted 4 May 2020

Available online 11 May 2020

0022-1694/ © 2020 Elsevier B.V. All rights reserved.

therefore, essential for understanding karst aquifer systems and mitigating sinkhole hazard.

In many karst regions, sinkholes appear as depressions on the Earth's surface and can be identified with topographic information. In Kentucky, a sinkhole database (Paylor et al., 2003) derived from 1:24,000 scale topographic maps was used to assess the statewide sinkhole hazard (Kentucky Emergency Management, 2013). The topographic maps used to compile the database were low-resolution and were mostly created prior to the 1970s. As a result, many small or newly formed sinkholes were not included in the database (Zhu et al., 2014). LiDAR is a remote-sensing technique that provides high-density and high-accuracy data for depicting small topographic features, therefore providing a great opportunity for mapping karst sinkholes in high resolution and with great detail. Some researchers (Seale et al., 2008; Alexander et al., 2013) have manually located sinkholes based on LiDAR-derived digital elevation models (DEMs) or their derivative maps directly. Others have used a sink-fill method (Jenson and Domingue, 1988; Planchon and Darboux, 2001; Wang and Liu, 2006) to automatically extract surface depressions from LiDAR (Zhu et al., 2014; Wu et al., 2016; Wall et al., 2017). Although the sink-fill method can locate sinkholes, it also extracts other surface features, such as stream channels, meander cutoffs, and more commonly man-made structures (e.g., farm ponds, culverts, swimming pools); therefore, an additional step is needed to separate sinkholes from these other features. Zhu et al. (2014) used a manual visual inspection method with auxiliary data (e.g., satellite images), but this can be slow and laborious for large areas in which tens of thousands of surface depressions can be revealed from LiDAR. We decided to apply machine learning methods to tackle this problem.

Machine learning is a branch of artificial intelligence in which the goal is to construct computer-based systems that improve automatically through training experience (Mitchell, 1997). Over the last two decades, machine learning has advanced dramatically with broad applications across many scientific fields, including biology, cosmology, medical science, and social science (Jordan and Mitchell, 2015). Machine learning has also been applied to sinkhole-related problems. Miao et al. (2013) trained a random forests model with a high training accuracy of 87.9%, but their model was built on a small dataset of 66 records and was not tested in a different area. Zhu and Pierskalla (2016) trained a weighted random forests model from 8427 records, but they found the accuracy decreased considerably when the trained model was tested in a nearby area. Taheri et al. (2019) applied four machine learning methods to evaluate sinkhole susceptibility factors using a small dataset of less than 50 records of existing sinkholes. Identifying sinkholes from a large number of surface depressions quickly and efficiently, however, remains a challenge.

In this study, we aimed to identify sinkholes by exploring multiple machine learning methods using a large dataset and developing an efficient way to apply trained machine learning models.

2. Study area

The study area consists of Bourbon, Woodford, Jessamine, and Scott Counties in Central Kentucky (Fig. 1). Most of the study area is in the Inner Bluegrass Region, the second largest karst region in Kentucky (Currens, 2002). The Inner Bluegrass is mostly underlain by the Middle Ordovician Lexington Limestone (Cressman and Peterson, 1986), which is predominantly a limestone unit with interbedded shales that has developed numerous karst features, such as sinkholes and sinking streams. The Kentucky River flows along the western boundary of Woodford County and the southwestern boundary of Jessamine County. The Kentucky River Palisades and the lower parts of the river's tributaries inside the two counties are underlain by the Middle Ordovician High Bridge Group, which is mainly composed of limestone and dolomite and is the oldest stratigraphic unit exposed in Kentucky (Cressman and Peterson, 1986). Northern Scott County, southeastern Bourbon

County, and southern Jessamine County are in the Eden Shale belt, a band of round hills and ridges around the Inner Bluegrass (Newell, 1986). These parts of the study area are underlain by the Upper Ordovician Garrard Siltstone and Clays Ferry Formations, which are mostly shale with interbedded siltstone and minor limestone and have very few karst features developed on them (Cressman and Peterson, 1986).

3. Methods

3.1. Create a sinkhole dataset

Karst sinkholes in Bourbon, Jessamine, and Woodford Counties were first mapped using the method of Zhu et al. (2014) and the karst sinkhole data were then used to create a dataset of sinkhole morphometric characteristics for training and testing machine learning methods. Sinkhole morphometric characteristics have been used to remove depressions that are not sinkholes (Filin and Baruch, 2010; Rahimi and Alexander, 2013; Doctor and Young, 2013; Wu et al., 2016; Zhu and Pierskalla, 2016). We selected 10 morphometric variables to describe the three-dimensional characteristics of the topographic depressions. The plan-view of a depression (i.e., a polygon representing the closed contour of the depression on the surface) was characterized by three variables: perimeter, area, and compactness. The compactness (also called circularity), C , measures how closely a shape resembles a circle. One way to measure the compactness is by Cole (1964)

$$C = \frac{A}{A_c} \quad (1)$$

where A is the area of the polygon and A_c is the area of the smallest circle circumscribing the polygon. Values of compactness range from 0 to 1 and equal 1 when the polygon is a perfect circle. Natural sinkholes tend to have a circular shape with a high compactness value although large sinkholes are more likely to have complex shapes.

To capture characteristics of the depressions in the vertical direction, we calculated statistics of depths in each depression and statistics of slopes of a 9-m ring (slope ring) surrounding each depression (Fig. 2). Variables selected for characterizing depths were maximum depth, mean depth, depth standard deviation, and depth sum. Depth sum variable is the sum of the depth of every cell within a depression; when multiplied by the cell size of the elevation raster (1.524 m in this study), the depth sum variable becomes the depression volume. Variables selected to describe the slope ring were mean slope and slope standard deviation. To see if there is a relation between depth and the surface area of a sinkhole, we selected a depth index D_i , defined as:

$$D_i = D_{max} / \sqrt{A/\pi} \quad (2)$$

where D_{max} is the maximum depth and A is the surface area (i.e., the area of the polygon on the surface). The depth index reflects the slope by assuming that a sinkhole is an inverted cone (Miao et al., 2013).

3.2. Train and test machine learning models

Training an algorithm to identify sinkholes using a dataset with known sinkhole/non-sinkhole classification is considered as supervised learning in machine learning. Supervised learning systems rely on known responses to learn how to map inputs to predictions. There are many mapping techniques for supervised learning, including Bayesian classifiers, decision trees, decision forests, logistic regression, kernel machines, neural networks, and support vector machines (Hastie et al., 2011). We selected six methods: logistic regression, naive Bayes, support vector machine, neural network, RUSBoost, and random forests. Logistic regression applies the logit function (the logarithm of the odds) with a linear combination of inputs to estimate the probability of response variables (usually binary). Naive Bayes uses the Bayes theorem to predict the posterior probability of each class while assuming that

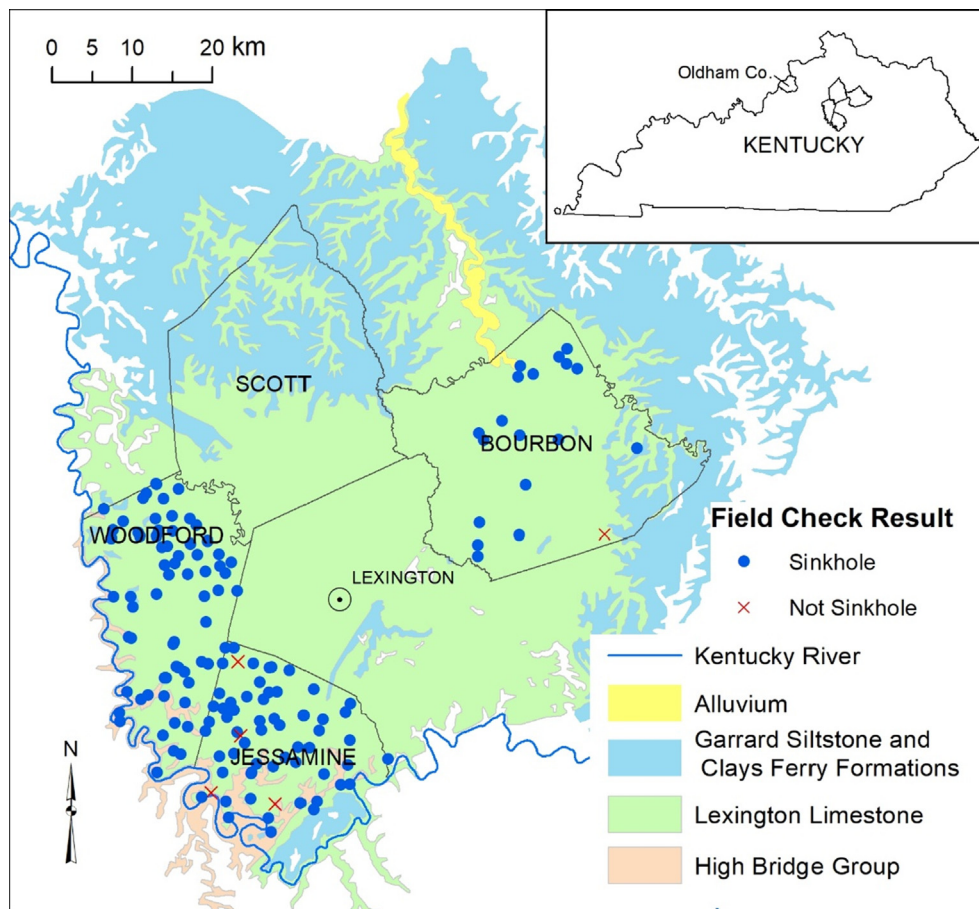


Fig. 1. Study area.

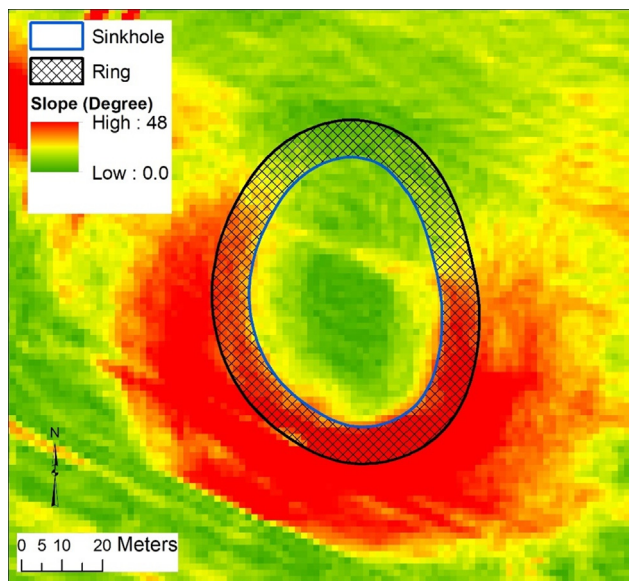


Fig. 2. Example of a slope ring surrounding a sinkhole.

inputs are conditionally independent. A support vector machine (SVM) makes predictions by seeking a hyperplane that provides the maximum separating margin for inputs. SVMs rely on different kernel functions to transform inputs into a separable form. A neural network is a two-stage model; the key idea is to derive features through a linear combination of inputs at the first stage and then model the responses using a non-linear function of these derived features at the second stage (Hastie

et al., 2011). RUSBoost is a decision forest method designed to handle imbalanced data; it uses RUS (random under-sampling) to achieve a balance between classes in the sampled data. A common sampling strategy in RUSBoost is to have the same number of observations from each class. The random forests method was applied to compare this study with our previous study (Zhu and Pierskalla, 2016), which also used the random forests method. Both the random forests and RUSBoost grow multiple decision trees, but in the random forests the trees are independent of each other (i.e., bagging) whereas in RUSBoost the later trees are adaptive to the earlier trees (i.e., boosting). A more detailed description of these methods can be found in Hastie et al. (2011), Seiffert et al. (2010), and Breiman (2001).

We used MATLAB (2017) to build machine learning classifiers from the sinkhole dataset. MATLAB provides built-in functions for many machine learning methods. For the support vector machine, we used a quadratic polynomial kernel. For RUSBoost, we grew 1000 trees. For random forests, we grew 300 trees – the same number of trees used in Zhu and Pierskalla (2016). For neural network, we used one hidden layer with 10 nodes and scaled conjugate gradient backpropagation training function. For logistic regression and naive Bayes, we used the option requiring no additional parameters besides the dataset. Each method was trained by using 80% of a randomly selected subset of the full dataset. The trained machine learning models were then tested using the remaining 20% of the dataset.

For two-class classification problems, a widely used tool to evaluate the performance of a machine learning method is a receiver operating characteristic (ROC) curve (Fawcett, 2006). A ROC curve plots false positive rate vs. true positive rate under different classification thresholds. The true positive rate is the proportion of positive cases that are correctly classified; the false positive rate is the proportion of

Table 1
Summary of sinkhole mapping results in Bourbon, Jessamine, and Woodford Counties.

County	Number of depressions processed	Number of sinkholes identified	Number of field-confirmed sinkholes	Number of field-confirmed non-sinkholes
Bourbon	7775	1294	18	1
Jessamine	8134	2257	57	4
Woodford	6975	2075	68	0

negative cases that are incorrectly classified as positive. As a result, in a ROC curve, increases in true positive rate are often accompanied by increases in false positive rate. The performance can be evaluated through how well a method separates the true positive rate from the false positive rate; the area under the ROC curve, or AUC, provides a straightforward measure of this. An AUC of 1 represents a perfect test and an AUC of 0.5 represents a worthless test. The closer the AUC is to 1, the better the test. A ROC curve with an AUC above 0.9 generally indicates an excellent classifier.

Although a ROC curve is widely used to measure the performance of binary classifiers (Fawcett, 2006), it is not sensitive to data imbalance. To better evaluate the performance of the classifiers to an imbalanced dataset, such as the sinkhole dataset in this study, we applied precision-recall (PR) curves. Precision is defined as the number of correct positive predictions divided by the sum of true positive cases and false positive cases. The number of false positive predictions depends on the number of actual positive cases. Precision measures the portion of predicted positives that are actually correct. Recall, also known as true positive rate, measures the portion of actual positives that are identified correctly. Consequently, a PR curve is considered a more appropriate evaluating measure than a ROC curve for imbalanced datasets (Saito and Rehmsmeier, 2015). In a PR curve, improving precision typically reduces recall and vice versa. The area under the PR curve (PR-AUC) provides a measure of performance. A high PR-AUC value represents both high recall and high precision, showing that the classifier not only returns high accuracy in positive predictions but also identifies a high portion of actual positives.

The ROC and PR curves evaluate the overall performance of the classifiers. Prediction accuracy of the classifiers was further assessed using confusion matrices from the testing data. A confusion matrix lists actual and predicted classifications in a matrix form. From the confusion matrices of the six classifiers, we calculated overall accuracy, average accuracy, precision, recall, and F-measure. Overall accuracy is the ratio of the total number of correct predictions to the total number of records. The overall accuracy is a widely used evaluating metric, but it can be misleading for very imbalanced data, in which a classification missing all minority classes can still achieve high overall accuracy. The average accuracy is the average between true positive rate and true negative rate, which takes into account class imbalance. The true negative rate is the ratio of correct negative predictions to total actual negatives and measures the portion of actual negatives that are identified correctly. F-measure is the harmonic mean of precision and recall.

3.3. Model application

To investigate how to effectively apply a trained classifier to a new location, we developed a two-step procedure that combined the best-performing classifier and visual inspection. We anticipated that a trained classifier would make some wrong predictions, so the two-step procedure was aimed to combine human intervention and the machine learning classifier to improve sinkhole mapping accuracy.

The accuracy assessment metrics described in Section 3.2 used a confusion matrix calculated from a commonly used classification threshold of 0.5, meaning that a depression is classified as a sinkhole when its predicted probability of being a sinkhole above 0.5. A different threshold can be applied based on the nature of the application. In sinkhole mapping, the mistake of mapping a non-sinkhole as a sinkhole

is a bigger error than the mistake of missing a sinkhole. So the goal is to reduce false positives while maintaining a high true positive rate. Reducing the false positive rate often decreases the true positive rate, however. Thus, the first step of the two-step procedure was to apply a machine learning classifier to predict probability of depressions being sinkholes. Then, a probability threshold was selected so that most true sinkholes were classified into the “sinkhole” category. By doing that, the sinkhole category inevitably contained some non-sinkholes. We then used visual inspection to remove non-sinkholes from the sinkhole category in the second step.

4. Results and discussion

4.1. Sinkhole dataset

Karst sinkholes in Bourbon, Jessamine, and Woodford Counties were used to create the dataset for training and testing machine learning methods. Using the method in Zhu et al. (2014), a total of 22,884 depressions were extracted from LiDAR-derived 1.524-m DEMs, of which 5631 (24.6%) were identified as probable sinkholes. We field-checked 148 randomly selected probable sinkholes and confirmed that 144 (97.3%) of them were actual sinkholes. A summary of the results for each county is listed in Table 1 and the locations of field-checked sinkholes are shown in Fig. 1. For the purposes of training and testing machine learning methods, the sinkhole classification result was considered to be a “true” classification. The “true” classification is not entirely accurate, as demonstrated by the 2.7% of error rate from the field-checked results.

The 10 morphometric variables were then calculated for all the 22,884 depressions. A response variable with a binary classification (sinkhole or non-sinkhole) was created using sinkhole mapping result. The resulting dataset was imbalanced as only 24.6% of the records were classified as sinkholes. The dataset was skewed toward the non-sinkhole class.

4.2. Model testing results

The ROC curves (Fig. 3) show that all the methods worked well for the sinkhole dataset with AUCs ranging from 0.874–0.950. The best performing methods were neural network (AUC = 0.950), random forests (AUC = 0.947), and RUSBoost (AUC = 0.942). The PR curves (Fig. 4) show that neural network (PR-AUC = 0.860) performed the best among the six methods, followed by random forests (PR-AUC = 0.851), and support vector machine (PR-AUC = 0.828). Overall, the ROC curves and PR curves gave similar assessment results for these methods. Table 2 shows that neural network had the highest overall accuracy, average accuracy, and F-measure. Random forests and RUSBoost achieved very close performance to neural network.

4.3. Model application results

The performance evaluation of the six classifiers showed that the dataset derived using the morphometry of the depressions from the three counties provides adequate information for machine learning to separate sinkholes from other depressions. The evaluation also showed that neural network worked best among the six methods. However, although the neural network correctly identified 830 sinkholes, it also

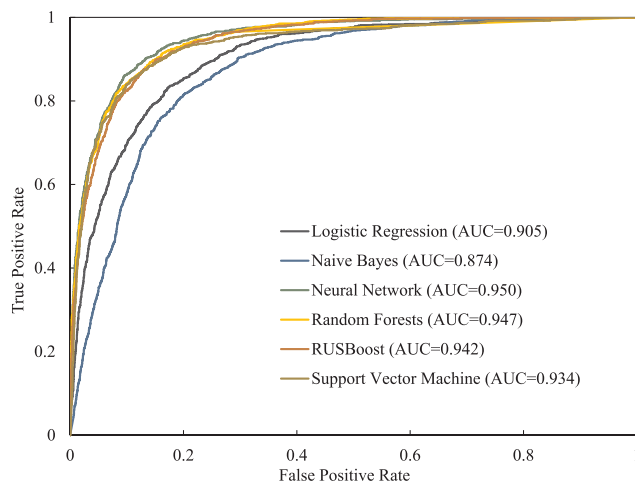


Fig. 3. Receiver operating characteristic curves for the six classifiers.

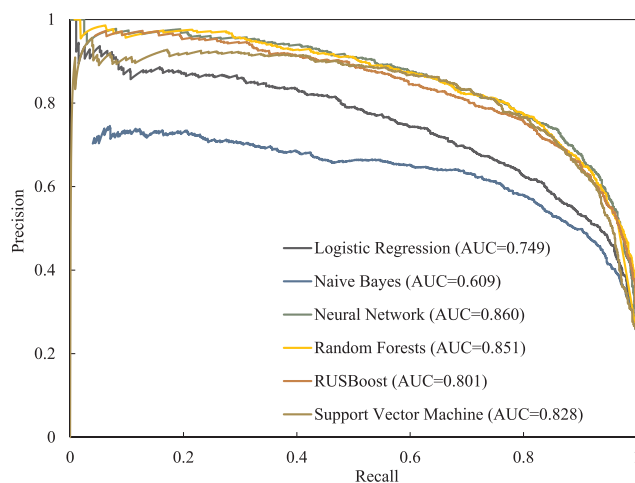


Fig. 4. Precision recall curves for the six classifiers.

Table 2

Performance metrics of the classifiers for the test data.

Method	Recall	Precision	Overall accuracy	Average accuracy	F-measure
Naïve Bayes	0.365	0.690	0.804	0.656	0.478
Logistic regression	0.602	0.743	0.851	0.767	0.665
Support Vector Machine	0.692	0.834	0.890	0.823	0.756
Random forests	0.756	0.804	0.895	0.848	0.779
RUSBoost	0.780	0.769	0.888	0.852	0.774
Neural network	0.765	0.801	0.899	0.853	0.783

Table 3

Confusion matrix of neural network classifier for testing data.

	Predicted non-sinkhole	Predicted sinkhole
True non-sinkhole	3286	206
True sinkhole	255	830

missed 255 actual sinkholes and misclassified 206 non-sinkholes as sinkholes (Table 3). We needed to find a way to take advantage of the classifier while minimizing loss in accuracy.

We applied the two-step procedure to Scott County, another county in the Bluegrass Region (Fig. 1). We retrained a neural network classifier using all 22,884 records collected from Bourbon, Woodford, and

Jessamine Counties. Using the depression extraction method by Zhu et al. (2014), we extracted 10,626 topographic depressions from a LiDAR-derived DEM of Scott County. The same 10 variables describing morphometric characteristics of these depressions were calculated. The trained neural network classifier was then applied to predict the probability of these depressions being sinkholes. Because most of Scott County has similar geology as the other three counties, we expected that the ratio of sinkholes to non-sinkholes in the Scott County dataset would be similar to the ratio in the training data (0.33). Using this ratio with some added safety margins, we used a probability threshold of 0.1 to separate the depressions into a “sinkhole” group and a “non-sinkhole” group, so that all records in the sinkhole group had a predicted sinkhole probability of 0.1 or higher and all records in the non-sinkhole group had a predicted sinkhole probability below 0.1. The resulting sinkhole group had 2889 (27%) records and the non-sinkhole group had 7737 (73%) records, a ratio of 0.37.

We then visually inspected the sinkhole group and identified 888 of them as sinkholes. To see how many sinkholes could be missed by this two-step procedure, we also visually inspected the non-sinkhole group and found that 31 of them were sinkholes. This suggested that, with the assistance of a machine learning classifier, we would be able to identify 97% of sinkholes by inspecting only 27% of the total depressions for Scott County. The procedure saved more than 70% of the manual labor in visual inspection with a cost of missing only a small number of sinkholes.

When inspecting the 31 sinkholes the neural network classifier predicted with low probability of being sinkholes (< 0.1), we found their morphometric characteristics were not typical as sinkholes. For example, some of them had a naturally irregular surficial shape that was very different from the circular shape of typical sinkholes (Fig. 5a). Some had been partially modified by human activities, such as road constructions (Fig. 5b). Some were groups of small sinkholes that were too small to be classified as sinkholes individually (Fig. 5c). Fig. 5c shows the limits of the dataset, in which each depression was considered independently spatially. We have observed that sinkholes tend to form in clusters along major rivers or joints and lineaments (Zhu et al., 2014). Visual inspection can naturally take this spatial pattern into consideration, but extracting variables to reflect that pattern can be difficult. These cases showed a small number of sinkholes that cannot be characterized by their morphometric characteristics alone.

4.4. Discussion

In this study, the trained neural network model was applied to an area that has similar geologic and geographic conditions as the area for which the training dataset was developed. Development of sinkholes is influenced by many geologic, topographic, and climatologic factors and morphometry of sinkholes may not be equivalent from one region to another (Taylor and Doctor, 2016). To see if the trained neural network model is applicable to another karst region, we tested the model in Oldham County, Kentucky where sinkholes have been mapped previously using LiDAR data (Zhu and Pierskalla, 2016). Oldham County is approximately 100 km northwest of the study area and is underlain by limestone and dolomite of Late Ordovician and Silurian age (Newell, 1986). The application of the two-step procedure to Oldham County showed that a sinkhole probability threshold of 0.1 classified 60% of depressions as “non-sinkhole”, but the “sinkhole” group missed 14% of actual sinkholes. When a threshold of 0.025 was used, 43% of the polygons were classified as “non-sinkhole” whereas the “sinkhole” group missed only 4% of actual sinkholes. This suggests that the trained classification model may yield less accurate results when applied to other karst regions, which is not surprising. To apply the method described in this paper to other karst regions, a better way would be to re-train the machine learning models using data from the same region. On the other hand, the model trained in this study can still be potentially useful for another karst region if used as a screening tool to remove

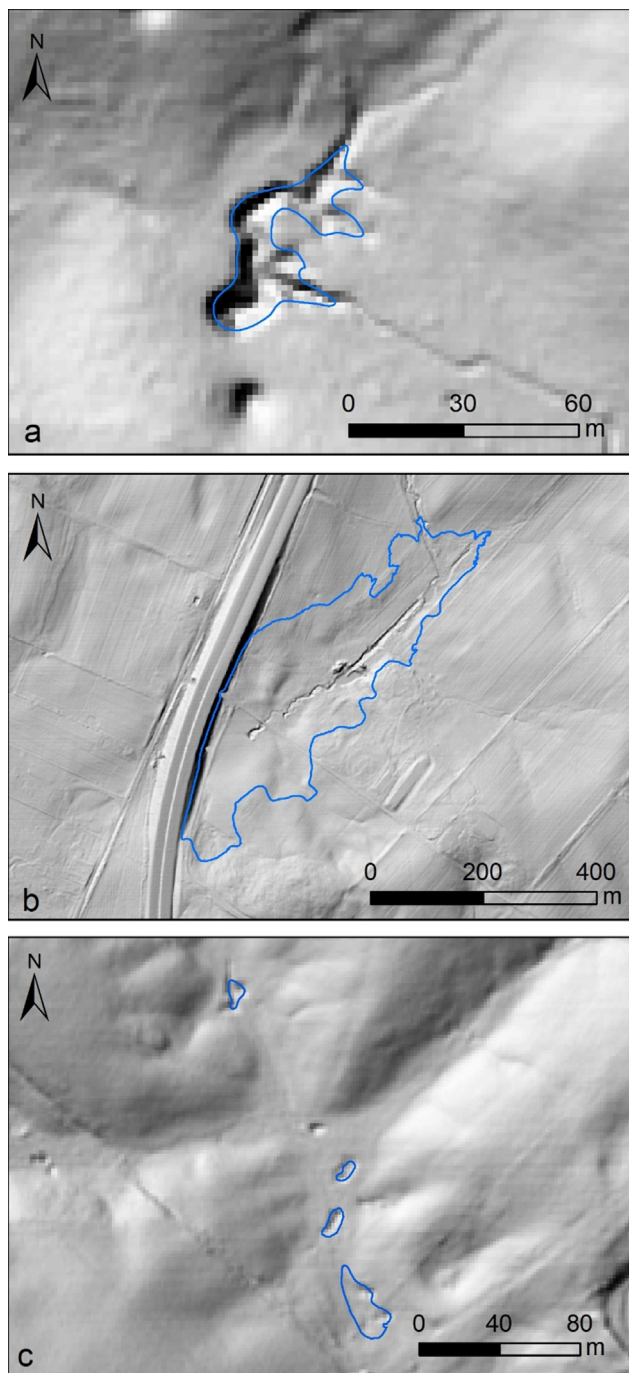


Fig. 5. Examples of true sinkholes with low predicted probabilities of sinkholes by the neural network classifier. Blue lines depict surficial polygons of these sinkholes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

obvious non-sinkholes. In this case, the threshold used for screening depressions should be tested and adjusted for the region.

A comparison of performance metrics between this study and our previous study (Zhu and Pierskalla, 2016) showed the random forests model in this study achieved higher accuracy than our previous random forests model. The improvement may be attributed to the larger training dataset used in this study (18,308 records compared to 8427 records used previously). The performance difference may also be attributed to how the two models were tested. In Zhu and Pierskalla (2016), the trained model was tested in an area adjacent to the area where the training data were extracted. In this study, the testing data

and training data were from the same area, and separation of the two subsets was done randomly.

Sinkholes have different sizes and shapes. The 10 morphometric variables certainly did not capture all the details of the three-dimensional nature of sinkholes. The machine learning models may be improved if additional morphometric variables can be developed and added to the training data. On the other hand, deep learning can use images of sinkholes directly as training data and offers an alternate approach that may further improve classification accuracy and expedite the sinkhole mapping process.

5. Conclusions

We tested six machine learning methods to locate karst sinkholes from LiDAR data in the Bluegrass Region of Kentucky. We built a dataset of morphometric characteristics of mapped sinkholes in Bourbon, Jessamine, and Woodford Counties and trained and tested classifiers using logistic regression, naive Bayes, support vector machine, neural network, random forests, and RUSBoost. We then used a two-step procedure that combined the best performing classifier with manual inspection in Scott County to improve sinkhole mapping efficiency. Our study concluded:

1. Morphometric characteristics of sinkholes provided sufficient information for separating most sinkholes from other forms of surface depressions.
2. Neural network performed the best among the six machine learning methods in identifying sinkholes from other depressions. Neural network achieved the highest AUC values from the receiver operating characteristic curves and the precision-recall curves, and best values for overall accuracy, average accuracy, and F-measure.
3. The two-step procedure of combining a machine learning classifier with manual visual inspection improved efficiency while maintaining accuracy. With the assistance of a neural network classifier, we located 97% of sinkholes by inspecting only 27% of the topographic depressions in Scott County.

CRediT authorship contribution statement

Junfeng Zhu: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration. **Adam M. Nolte:** Investigation, Resources, Data curation, Writing - review & editing. **Nathan Jacobs:** Methodology, Validation, Formal analysis, Writing - review & editing. **Ming Ye:** Conceptualization, Writing - review & editing, Funding acquisition.

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.

Acknowledgments

We would like to thank Dr. Anke Meyer-Baese, and Amirhessam Tahmassebi from Florida State University for their collaboration with us on machine learning. We also thank Kalli Beasley, Ethan Davis, Megan Hackett, and Steve Webb for their help in mapping sinkholes for the study area. Our appreciation also goes to the associate editor and the anonymous reviewers for their constructive comments. We thank Meg Smath for her editorial review. The last author was supported by National Science Foundation grant EAR-1828827.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhydrol.2020.125049>.

References

- Alexander, S.C., Larson, E., Bomberger, C., Greenwaldt, B., Alexander Jr., E.C., Rahimi, M., 2013. Combining LiDAR, aerial photography, and PictometryH tools for karst features database management. In: Land, L., Doctor, D.H., Stephenson, J.B. (Eds.), *Proceedings of the Thirteenth Multidisciplinary Conference on Sinkholes and the Engineering and Environmental Impacts of Karst: Carlsbad, National Cave and Karst Research Institute, Symposium 2*, pp. 441–448.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Chen, Z., Auler, A.S., Bakalowicz, D., Griger, F., Hartmann, J., Jiang, G., Moosdorf, N., Richts, A., Stevanovic, Z., Veni, G., Goldschieder, N., 2017. The world karst aquifer mapping project: concept, mapping procedure and map of Europe. In: *Hydrogeol. J.* <https://doi.org/10.1007/s10040-016-1519-3>.
- Cole, J.P., 1964. Study of major and minor civil divisions in political geography. In: *The 20th International Geographical Congress*, London.
- Cressman, E.R., Peterson, W.L., 1986. Ordovician system. In: McDowell, R.C. (Ed.), *The Geology of Kentucky – A Text to Accompany the Geologic Map of Kentucky*, U.S. Geological Survey Professional Paper 1151-H.
- Currens, J.C., 2002. Kentucky is karst country! What you should know about sinkholes and springs. Kentucky Geological Survey, information Circular 4, Series XII, 35p.
- Dinger, J.S., Zourarakis, D.P., Currens, J.C., 2007. Spectral enhancement and automated extraction of potential sinkhole features from NAIP imagery—Initial investigations. *Journal of Environmental Informatics* 10 (1), 22–29.
- Doctor, D.H. and Young, J.A., 2013. An evaluation of automated GIS tools for delineating karst sinkholes and closed depressions from 1-meter LiDAR-derived digital elevation data. In: Land, L., Doctor, D.H., Stephenson, J.B. eds., *Sinkholes and the Engineering and Environmental Impacts of Karst: Proceedings of the Thirteenth Multidisciplinary Conference*, May 6–10, Carlsbad, New Mexico: NCKRI Symposium 2. National Cave and Karst Research Institute, Carlsbad (NM), pp. 449–458.
- Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recognit. Lett.* 27 (8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>.
- Filin, S., Baruch, A., 2010. Detection of sinkhole hazards using airborne laser scanning data Photogram. Eng. Remote Sens. 76 (5), 577–587. <https://doi.org/10.14358/PERS.76.5.577>.
- Ford, D.C., Williams, P., 2007. *Karst Hydrogeology and Geomorphology*. John Wiley and Sons Ltd, Chichester, England, pp. 562p.
- Hastie, T., Tibshirani, R., Friedman, J., 2011. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York.
- Jenson, S.K., Domingue, J.O., 1988. Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogram. Eng. Remote Sens.* 54 (11), 1593–1600.
- Jordan, M.I., Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. *Science* 349, 255–260.
- Kentucky Emergency Management, 2013. Commonwealth of Kentucky enhanced hazard mitigation plan: Risk assessment [2013 version]. Kentucky Emergency Management, 173p.
- Kuniansky, E.L., Weary, D.J., Kaufmann, J.E., 2016. The current status of mapping karst areas and availability of public sinkhole-risk resources in karst terrains of the United States. *Hydrogeol. J.* 24, 613–662. <https://doi.org/10.1007/s10040-015-1333-3>.
- MATLAB, 2017. MATLAB and Statistics Toolbox Release 2017a. The MathWorks Inc, Natick, Massachusetts, United States.
- Miao, X., Qiu, X., Wu, S., Luo, J., Gouzie, D.R., Xie, H., 2013. Developing efficient procedures for automated sinkhole extraction from LiDAR DEMs. *Photogram. Eng. Remote Sens.* 79 (6), 545–554. <https://doi.org/10.14358/PERS.79.6.545>.
- Mitchell, T.M., 1997. In: *Machine Learning*. McGraw Hill, pp. 414p.
- Newell, W.L., 1986. Physiography. In: McDowell, R.C. (Ed.), *The Geology of Kentucky – A Text to Accompany the Geologic Map of Kentucky*, U.S. Geological Survey Professional Paper 1151-H.
- Paylor, R.L., Florea, L.J., Caudill, M.J., Currens, J.C., 2003. A GIS coverage of sinkholes in the karst areas of Kentucky: Kentucky Geological Survey, metadata file and shapefiles of highest elevation closed contours, 1 CDROM, <http://kgs.uky.edu/kgsweb/download/karst/ksinks.zip>.
- Planchon, O., Darboux, F., 2001. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. *Catena* 46, 159–176.
- Rahimi, M., Alexander Jr., E.C., 2013. Locating sinkholes in LiDAR coverage of a glacio-fluvial karst, Winona County, MN. In: *Conference Proceedings Thirteenth Multidisciplinary Conference on Sinkholes and the Engineering and Environmental Impacts of Karst*. University of South Florida, pp. 469–480.
- Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 10 (3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>.
- Schwartz, F.W., Zhang, H., 2003. *Fundamentals of Ground Water*. John Wiley & Sons, New York, NY, United States, pp. 583.
- Seale, L.D., Florea, L.J., Brinkmann, R., Vacher, H.L., 2008. Using ALSM to identify closed depressions in the urbanized, covered karst of Pinellas County, Florida—I, Methodological considerations. *Environ. Geol.* 54, 995–1005. <https://doi.org/10.1007/s00254-007-0890-8>.
- Seiffert, C., Khoshgoftaar, T.M., Van Hulse, J., Napolitano, A., 2010. RUSBoost: a hybrid approach to alleviating class imbalance. *IEEE Trans. Syst. Man Cybernet.—Part A: Syst. Humans* 40 (1).
- Taheri, K., Shahabi, H., Chapi, K., Shirzadi, A., Gutiérrez, F., Khosravi, K., 2019. Sinkhole susceptibility mapping: a comparison between Bayes-based machine learning algorithms. *Land Degrad. Dev.* 30, 730–745. <https://doi.org/10.1002/ldr.3255>.
- Taylor, C.J., Doctor, D.H., 2016. Karst: Chapter 89. In: Singh, V.P. (Ed.), *Handbook of Applied Hydrology*, second ed. McGraw-Hill Education, pp. 89-1–89-14.
- Wall, J., Bohnenstiehl, D.R., Wegmann, K.W., Levine, N.S., 2017. Morphometric comparisons between automated and manual karst depression inventories in Apalachicola National Forest, Florida, and Mammoth Cave National Park, Kentucky, USA. *Nat. Hazards* 85, 729–749.
- Wang, L., Liu, H., 2006. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. *Int. J. Geogr. Inf. Sci.* 20 (2), 193–213.
- Wu, Q., Deng, C., Chen, Z., 2016. Automated delineation of karst sinkholes from LiDAR-derived digital elevation models. *Geomorphology* 266, 1–10. <https://doi.org/10.1016/j.geomorph.2016.05.006>.
- Zhu, J., Pierskalla Jr., W.P., 2016. Applying a weighted random forests method to extract karst sinkholes from LiDAR data. *J. Hydrol.* 533, 343–352. <https://doi.org/10.1016/j.jhydrol.2015.12.012>.
- Zhu, J., Taylor, T.P., Currens, J.C., Crawford, M.M., 2014. Improved karst sinkhole mapping in Kentucky using LiDAR techniques: a pilot study in Floyds Fork Watershed. *J. Cave Karst Stud.* 76 (3), 207–216. <https://doi.org/10.4311/2013ES0135>.