Separating Signals in Elevation Data Improves Supervised Machine Learning Predictions for Hydrothermal Favorability

Pascal D. Caraccioli[1,2,5], Stanley P. Mordensky[3], Jacob DeAngelo[4], Erick R. Burns[4], John J. Lipor[1]

[1]Contractor to the U.S. Geological Survey, Portland, OR, 97201, USA

[2]Portland State University, Portland OR 97201, USA

[3]U.S. Geological Survey, Spokane, WA 99201, USA

[4]U.S. Geological Survey, Moffet Field CA 94035, USA

[5]Cornell University, Ithaca NY 14853, USA

Keywords

Machine learning, Geothermal, Nevada, XGBoost, Play fairway analysis, Feature engineering, Geothermal favorability

ABSTRACT

A recent study identified topography (land surface elevation above sea level) as an important input dataset (feature) for predicting the location of hydrothermal systems in the Great Basin in Nevada. Yet, topography is generally a result of more than one geological process and may consequently contain multiple distinct signals. For example, the geologic evolution of the Great Basin has produced both crustal thickening (i.e., regional-scale trends in elevation) and thinning via Basin and Range extensional faulting (i.e., valley-scale topographic relief). We postulate that these geologic processes may affect the occurrence of hydrothermal systems differently. Therefore, we separate the regional trend from the valley-scale signal in the Great Basin, and then use them separately to evaluate the importance of each as predictors for hydrothermal favorability.

Our prior work applying supervised machine learning (ML) using the data from the Nevada Machine Learning Project demonstrated that employing a training strategy that randomly selects negative training sites produces better performing models for predicting hydrothermal favorability than a training strategy that used expert-selected negatives. The models created using both training strategies exhibited a west-east geographic trend in the predictions for the favorability of hydrothermal resources. These models generally predicted higher favorability in western Nevada

and lower favorability in eastern Nevada. This west-east trend in predicted favorability correlates with elevation across the Great Basin, which trends higher from west to east.

By separating the original elevation feature into distinct features for *elevation trend* (i.e., regional-scale topography) and *detrended elevation* (i.e., valley-scale or local relative topography), we find that models using the separated topographic signals consistently outperform competing models that use the original elevation feature. Although western Nevada still exhibits higher favorability than eastern Nevada, using separated signals for regional elevation and local structure reduces the west-east prediction trend in the region and emphasizes structures associated with hydrothermal upflow. This work emphasizes how carefully engineering features to represent geological conditions relevant to hydrothermal systems allows ML algorithms to detect important patterns for predicting hydrothermal resource favorability and leads to better model performance.

1. Introduction

Hydrothermal systems in the Great Basin, a prominent extensional province in the western United States, are generally controlled by basin-bounding faults located in actively subsiding basins (Faulds et al., 2011; Faulds and Hinz, 2015) with many systems having no easily identifiable surface manifestations (i.e., many systems are *blind*; Coolbaugh et al., 2007). Yet, changes in valley-scale topography (i.e., relative positions of basins and adjacent mountain ranges) serve as an indicator of geologic structures that may be associated with hydrothermal upflow.

The Nevada Machine Learning Project (NVML; Brown et al., 2020; Faulds et al., 2021a; Smith, 2021; Faulds et al., 2024), aimed to identify hidden geothermal resources in the Great Basin. Using datasets compiled under the Nevada Play Fairway Analysis (Faulds et al., 2017; Faulds et al., 2021b) as well as data collected specifically for NVML, NVML fit an artificial neural network (ANN) to predict the presence or absence of hydrothermal resources. The NVML research team labeled 83 of the 1,728,000 250-m square cells as positive (i.e., as having a known hydrothermal system) and 62 cells as negative (i.e., as not having a hydrothermal system). The remaining cells were unlabeled (i.e., the presence or absence of a hydrothermal system was and remains unknown). Eleven 250-m resolution evidence layer grids were used as input features (i.e., predictors or unassociated datasets) to the NVML ANN. Ten of these 11 input features consisted of quantitative data (e.g., elevation, strain rate, heat flow, and gravity). One of the 11 input features was engineered by experts as ellipses that were drawn around favorable structural settings. By using expert-knowledge to address fundamental challenges (e.g., selecting known sites without hydrothermal systems), the NVML project demonstrated that machine learning methods can be successfully used in geothermal resource evaluation.

Mordensky et al. (2023) conducted a similar study to predict hydrothermal favorability in the western United States, using data-driven ML approaches and data from the most recent U.S. Geological Survey (USGS) Geothermal Resource Assessment (Williams et al., 2008). Of the approximately 700,000 2-km square cells, 278 were known positives, but there were no known negatives provided in the Williams et al. (2008) data. Because hydrothermal systems are inherently sparse (i.e., < 99.9% of the study area), Mordensky et al. (2023) recognized that a random sample of the unlabeled cells could be labeled as negative with high confidence; therefore, they implemented a training strategy that treated a random sample of the unlabeled cells as negatives during the training process. The random selection of negative cells from unlabeled cells also addressed the mathematically problematic condition of severe class imbalance (i.e., having many

more sites without hydrothermal systems than with hydrothermal systems) that mired the ML algorithms. Mordensky et al. (2023) estimated the natural positive-negative ratio as approximately 1:700, thereby identifying the number of randomly selected negatives needed during fitting. Mordensky et al. (2023) demonstrated that data-driven ML algorithms could be used successfully to remove or minimize the need for expert feature-weighting decisions used previously in Williams et al. (2008).

Recent work by Caraccioli et al. (2023) compared the results of the NVML ANN to other less-complicated ML models (logistic regression and eXtreme Gradient Boosting [XGBoost; Chen and Guestrin, 2016]) using different strategies for selecting negatives (i.e., NVML versus Mordensky et al., 2023) and nearly the same input features. The single difference between the input features used in NVML compared to those in Caraccioli et al. (2023) is that Caraccioli et al. (2023) did not use the ellipses that were drawn around favorable structural settings. Caraccioli et al. (2023) elected not to use these expert-crafted ellipses which tended to dominate ML models because the NVML team had drawn the ellipses around the majority of the positives, arguably engineering each ellipse as a fuzzy (i.e., mathematically vague or imprecise) positive. The influence of the ellipses on the NVML ANN can be seen in Fig. 1.

By comparing the NVML strategy of expert-selected negatives to the strategy of randomly selecting negatives from Mordensky et al. (2023), Caraccioli et al. (2023) demonstrated that employing a training strategy that randomly selects negative training sites consistently produces better-predicting models when using the XGBoost algorithm, which is mathematically simpler than an ANN. Caraccioli et al. (2023) hypothesized that using only expert-selected negative sites may impart bias towards one or two types of negatives while there are really many conditions that can cause hydrothermal systems to be absent. Central to the work described herein, Caraccioli et al. (2023) identified elevation as the most important feature in the best-performing model; however, a low-high, west-east trend in elevation dominated the signal of that feature and contributed to similar favorability predictions at the tops of mountain ranges in the west as in the valley floors in the east. This finding seemed contrary to the geologic interpretation that proximity to relative valley-scale structure controls deep circulation pathways. As a result, Caraccioli et al. (2023) postulated that the contribution of the valley-scale topographic signal in the elevation feature was negligibly small compared with the regional signal. To allow for the possibility that valley-scale topography contains an important signal, we herein use the separate regional trend and valley-scale topographic data from DeAngelo et al. (2023) as separate features to test the importance of local topographic relief when predicting hydrothermal systems.

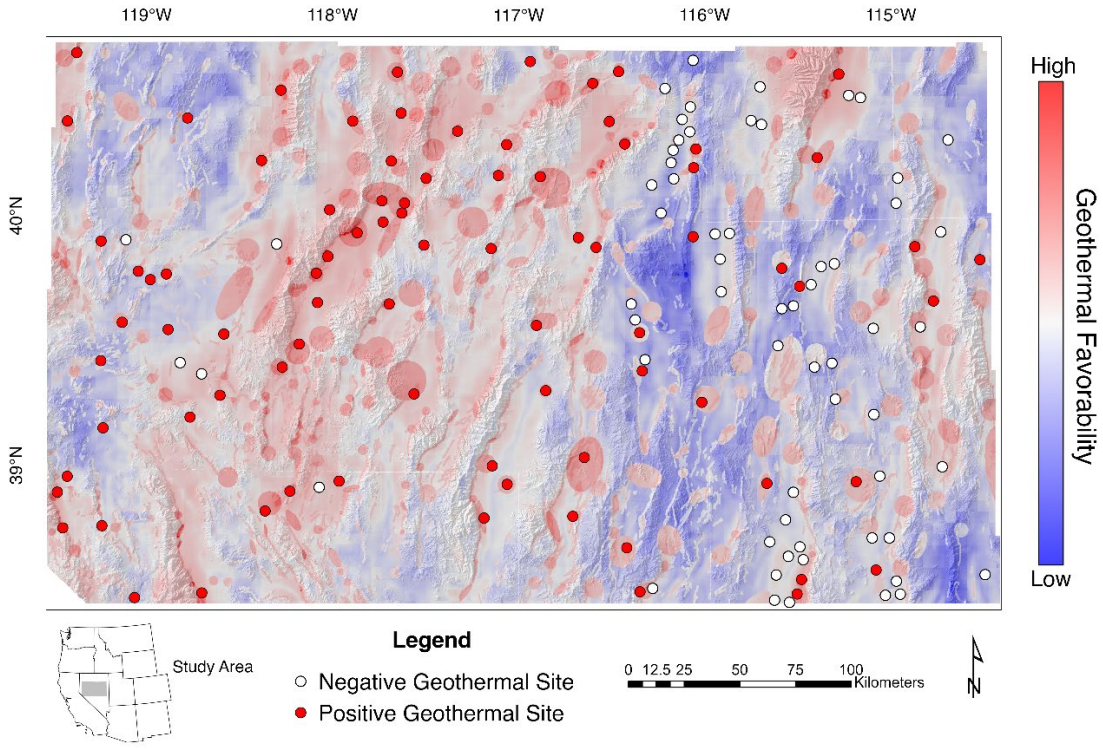

Figure 1: Nevada Machine Learning project (NVML) favorability map from Faulds et al. (2021a). White points are negative training/testing sites defined through expert selection. Red points are the positive sites (i.e., known hydrothermal systems). Favorability scores from the artificial neural network (ANN) have been normal score transformed for easy comparison with the results below from our study. Hillshade from USGS 3D Elevation Program (U.S. Geological Survey, 2019).

2. Methods

We compare XGBoost ML models using the original NVML features with models that replace the original elevation feature with two separate features (i.e., *elevation trend* and *detrended elevation* from DeAngelo et al., 2023). That is, although *elevation trend* and *detrended elevation* sum to the original elevation feature, we consider these two topographic features as two unique signals that we separate to inspect their value as predictors.

XGBoost is a boosted decision-tree ML algorithm in which a series of decision trees (i.e., estimators) predict probabilities for a cell being a positive or negative site (Chen and Guestrin, 2016). The cells are then formally classified as positive or negative by specifying a probability decision threshold. We use the commonly chosen decision threshold of 0.5 (see generally Fernández et al., 2018). ML algorithms (e.g., XGBoost, ANNs) calculate probabilities differently. Therefore, to allow the comparison of the new models with the NVML ANN, all favorability maps presented herein are the normal score transform (see generally Pyrcz and Deutsch, 2018) of the model predictions. Because the normal score transform preserves quantiles, the transformed prediction values are plotted as hydrothermal favorability, allowing for an easy comparison of regions of highest and lowest favorability.

In the remainder of this section, we briefly describe the data processing, exploratory data analysis, training strategies, hyperparameter optimization, and our method for comparing model performance.

2.1 Feature Selection

We use the features from the original NVML project and the separated regional and valley-scale elevation features engineered in DeAngelo et al. (2023). The original NVML ANN used 11 features. Ten of the 11 features were derived from interpolated maps of the following geologic properties: elevation, strain rate, slip rate, fault recency, gravity, slip and dilation tendency, seismic density, heat flow, magnetics, and distance to the nearest fault (Fig. 2). The eleventh feature was created by defining ellipses that contain geologic structures deemed by experts to be favorable for the occurrence of hydrothermal systems (i.e., known favorable structural setting; Fig. 3). Because the experts drew ellipses around known positives, the ML algorithms associate all areas within the ellipses with the occurrence of a hydrothermal system; therefore, the structural ellipses impart an implicit bias. The goal of this study is to predict favorable structural settings where a geologist has not had occasion to draw an ellipse, thus we remove the dataset with the structural ellipses.

Using the features from NVML and then by replacing the original elevation feature with the two separated elevation features from DeAngelo et al. (2023), we create two feature sets that can be compared. The first feature set (the Original Elevation Signal feature set) contains the same input features used in Caraccioli et al. (2023), corresponding to the 10 interpolated maps from NVML (i.e., original elevation, strain rate, slip rate, fault recency, gravity, slip and dilation tendency, seismic density, heat flow, magnetics, and fault). The second dataset (the Separated Elevation Signal feature set) replaces the original elevation feature with *elevation trend* and *detrended elevation*.

2.2 Exploratory Data Analysis

As part of exploratory data analysis, we inspect the linear correlation of the features using the Pearson correlation coefficient (see generally Lee Rodgers and Nicewander, 1988). Comparison of correlation allows us to see the pairwise linear relationships of the features (i.e., if the feature is unique or strongly correlated with one or more other features) and establish if two features represent similar signals. The Pearson correlation is a statistical measure that quantifies the linear relationship between two variables by providing values between -1 and 1, where -1 indicates a strong negative correlation, 1 indicates a strong positive correlation, and 0 indicates no correlation.

For each input feature, we compare the distributions of feature values for the NVML-labeled (i.e., positive and negative) sites to the full range of input-feature values to see if labeled sites mostly have low, intermediate, or high values relative to the unlabeled sites. To allow for an evaluation of whether labeled data are in discrete intervals within the larger range of the input feature values, we plot a cumulative distribution function (CDF) for every combination of features and labels (Fig. 4). When randomly sampling negatives and handling class imbalance, the CDF of the randomly sampled negatives should look similar to the CDF of all values of the input feature, because we are sampling most of the map area except the sparse positives.

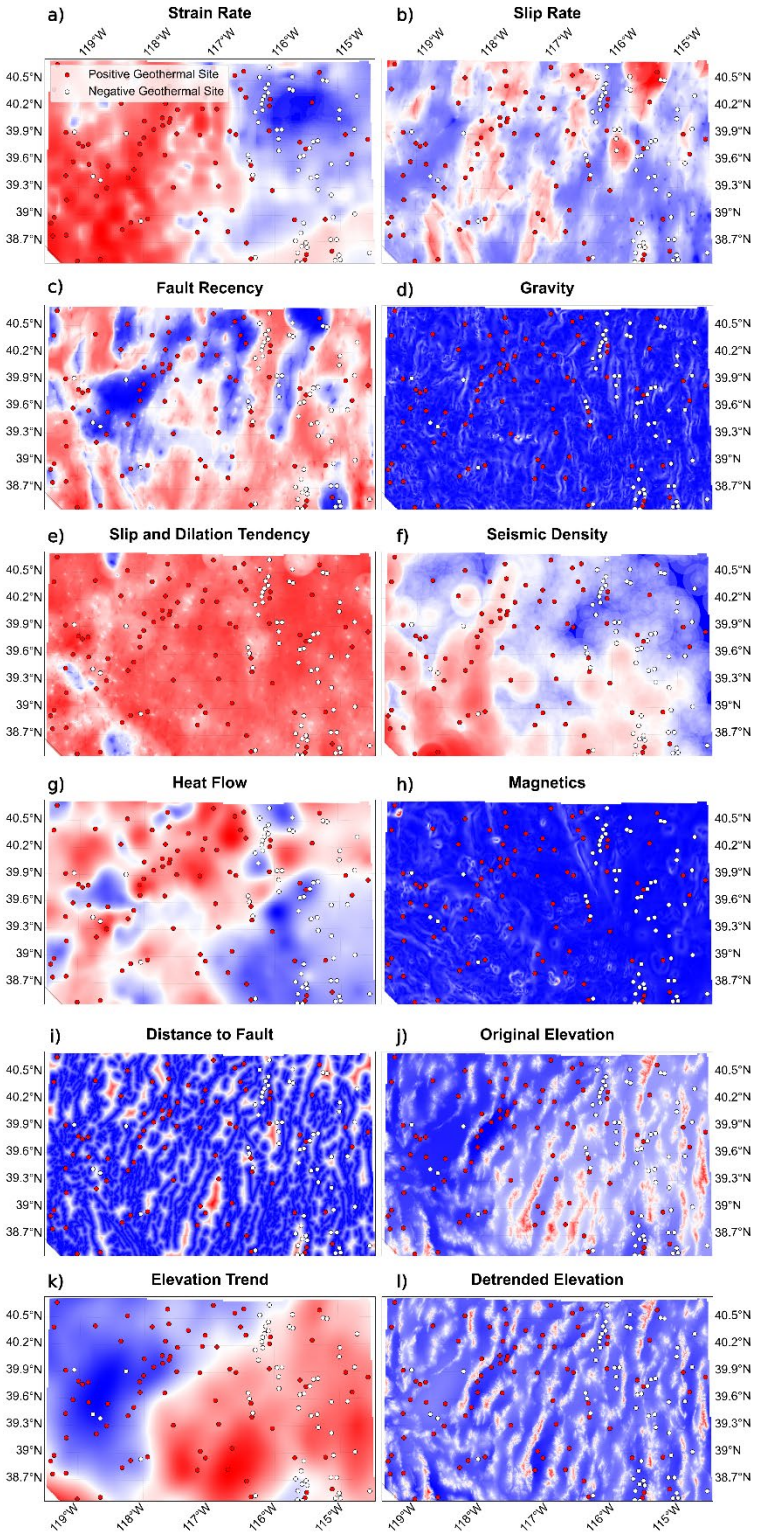

Figure 2: Maps of the ten standardized (i.e., unitless) Nevada Machine Learning project (NVML; Faulds et al., 2021a) features (a-j) and the DeAngelo et al. (2023) separated elevation input features (k, l) used for analysis herein. Blue depicts low values in the unitless scale. Red depicts high values in the unitless scale. White points are negative training/testing sites defined through expert selection. Red points are the positive sites (i.e., known hydrothermal systems).

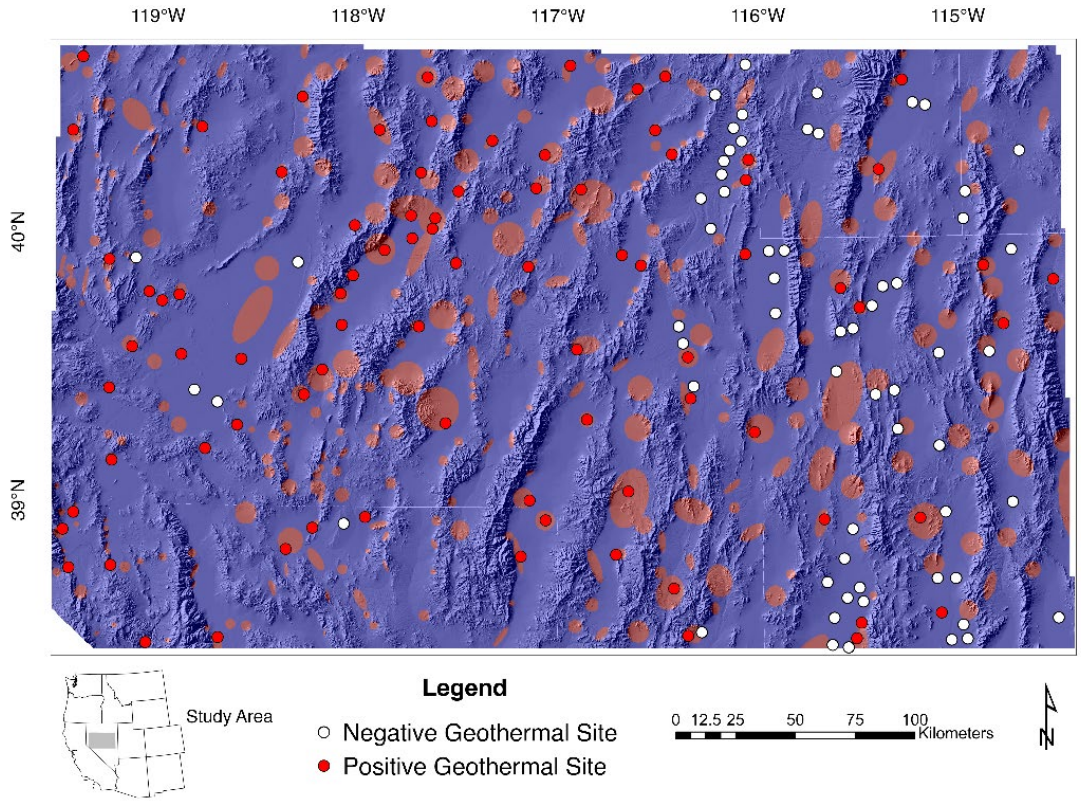

Figure 3: Map of the eleventh Nevada Machine Learning project (NVML) feature (favorable structural setting ellipses [red areas] from Faulds et al., 2021a) used for the artificial neural network (ANN) model (Fig. 1), but not used as an input feature for analyses herein. For strategy subsampling negative sites from unlabeled sites, negative sites were selected from areas outside of ellipses (i.e., blue area). White points are negative training/testing sites defined through expert selection. Red points are the positive sites (i.e., known hydrothermal systems). Hillshade from USGS 3D Elevation Program (U.S. Geological Survey, 2019).

2.3 Training Strategies

We implement two strategies for each training dataset: 1) the NVML training strategy; and 2) the training strategy from Mordensky et al. (2023) that implements random subsampling of negatives from unlabeled sites and accounts for class imbalance (the Natural Class Imbalance training strategy). The NVML strategy uses the same negative and positive sites from NVML (i.e., 62 negative sites and 83 positive sites). The Natural Class Imbalance strategy uses the same known positives as the NVML strategy, but randomly selects negative sites. Because we would like to minimize the possibility of randomly sampling and labeling an unlabeled positive site as negative, we choose to only select negatives from outside the favorable structural ellipses delineated by NVML (blue area in Fig. 3). We do this under the assumption that the highly sparse systems are less likely to occur without a favorable structural setting. Following Mordensky et al. (2023), we estimate that one in four hydrothermal systems have already been discovered in the study area, thereby approximating a roughly 1:5,100 positive-negative ratio. Although this estimate is approximate, Mordensky et al. (2023) demonstrated that the corresponding model predictions are insensitive within the expected range of uncertainty (see Mordensky et al. [2023] for complete details).

2.4 Hyperparameter Optimization

We optimize prediction performance by tuning four hyperparameters: class weight, number of estimators, maximum depth of estimators, and learning rate. Class weight is a way to correct for class imbalance. The greater the class weight, the greater the emphasis the model imparts on correctly identifying positive labels (i.e., the minority class) as positives at the expense of predicting negative labels as negatives (i.e., the majority class). The number of estimators specifies the number of decision trees (i.e., estimators). The maximum depth of estimators determines the number of levels in the estimators. The learning rate controls the amount of information communicated from a previous estimator to a new estimator. We leave the other parameters on XGBoost at the default settings found in the Python XGBoost, version 1.7.3 module as they have only a modest impact on performance (Chen and Guestrin, 2016).

We optimize the hyperparameters using the F1 score (Equation 1), a recommended metric for binary positive-unlabeled classifications (Bekker and Davis, 2020). We minimize bias resulting from any singular train-test split by conducting 60 80:20 train-test splits with five-fold cross-validation. For the Natural Class Imbalance strategy, we select new random negatives with each train-test split. We fit a final model using all the labeled data, the median optimal hyperparameters from the 60 train-test splits, and one last random sample of negative sites for the Natural Class Imbalance strategy.

$$F1\ Score = \frac{True\ Positives}{True\ Positives + \frac{1}{2}(False\ Positives + False\ Negatives)} \quad (1)$$

To test and prevent overfitting, we impose generalization loss early stopping (see generally Prechelt, 2002) in the validation subset of the training data during the five-fold cross-validation. With generalization loss early stopping, the fitting of new estimators stops immediately after the loss (a function that is penalized by decreased model performance) increases. We train a final model for each approach using the median estimator at which early stopping is engaged from the 60 train-test splits.

2.5 Measures of Feature Importance

For every modeling approach, we measure the relative importance of each input feature in making predictions using three measurements of feature importance: 1) sensitivity analysis using an F1 score; 2) sensitivity analysis using the area under the receiver operating characteristic curve (i.e., ROCAUC), and 3) Shapely Additive exPlanation (i.e., SHAP) values (see Mordensky et al. [2023] for a more detailed summary). To allow comparison between the different measures, each measure is min-max normalized to a zero-to-one scale. Using three different measures allows us to explore the variability between the measures.

2.6 Comparing Model Performance

To compare model performance, we perform a normal score transform on the predictions for each approach and compare the transformed distribution of predictions for the known positives with the transformed distribution of predictions for the unlabeled sites (see Mordensky et al. [2023] for additional details). Approaches with greater overlap in the distributions of predictions for known

positives and unlabeled sites indicate lower predictive skill than approaches with distinct distributions of predictions for the known positive and unlabeled sites.

3. Results

In this section, we briefly describe the input feature data, present favorability maps, and plot feature importance. We provide the median optimal hyperparameter values in Appendix A.

3.1 Exploratory Data Analysis

CDFs of feature values for the positive, NVML negative, and remaining unlabeled sites (essentially the distribution of a random sample of negatives) are shown in Fig. 4. The separation of cumulative distribution functions show that distributions are markedly different. The CDFs for the positive and negative sites generally bound that for the unlabeled sites with strain rate, seismic density, heat flow, fault distance, and slip rate, indicating that the corresponding features may have more value for separating positives from NVML negatives when used as predictors. Conversely, the CDFs for positive and unlabeled data show a greater difference in distribution for fault recency, fault distance, original elevation, *elevation trend*, and *detrended elevation*.

The Pearson correlation coefficient (Fig. 5) shows various degrees of correlation among the features with values from 0.78 to -0.59. Overall, strain rate is the feature most correlated with the greatest number of other features (e.g., *elevation trend* and seismic density; -0.59 and 0.64, respectively). *Elevation trend* and *detrended elevation* are minimally correlated. *Elevation trend* has similar correlation to the original elevation. *Detrended elevation* has poor correlation with all other datasets other than the original elevation.

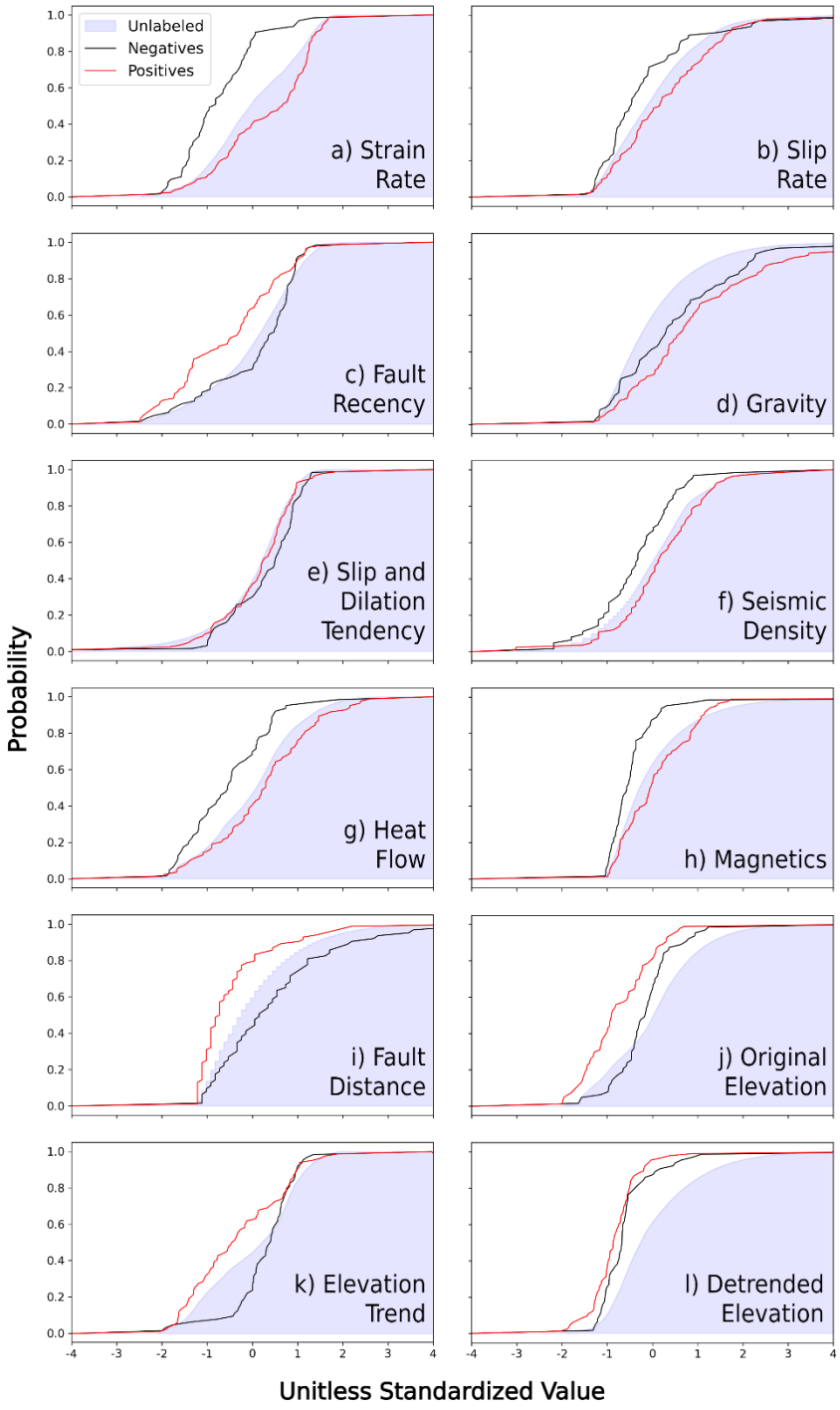

Figure 4: Cumulative distributions of the standardized features for the ten Nevada Machine Learning project (NVML) features (a-j) and the DeAngelo et al. (2023) separated elevation features (k, l). The shaded blue area corresponds to the unlabeled sites in NVML. The red line corresponds to positive-labeled sites in NVML. The black line corresponds to the negative-labeled sites in NVML. Separation of cumulative distribution functions show that distributions are markedly different, so a difference between the positive and negative lines implies the feature may be useful for discriminating between the positive and negative training sites.

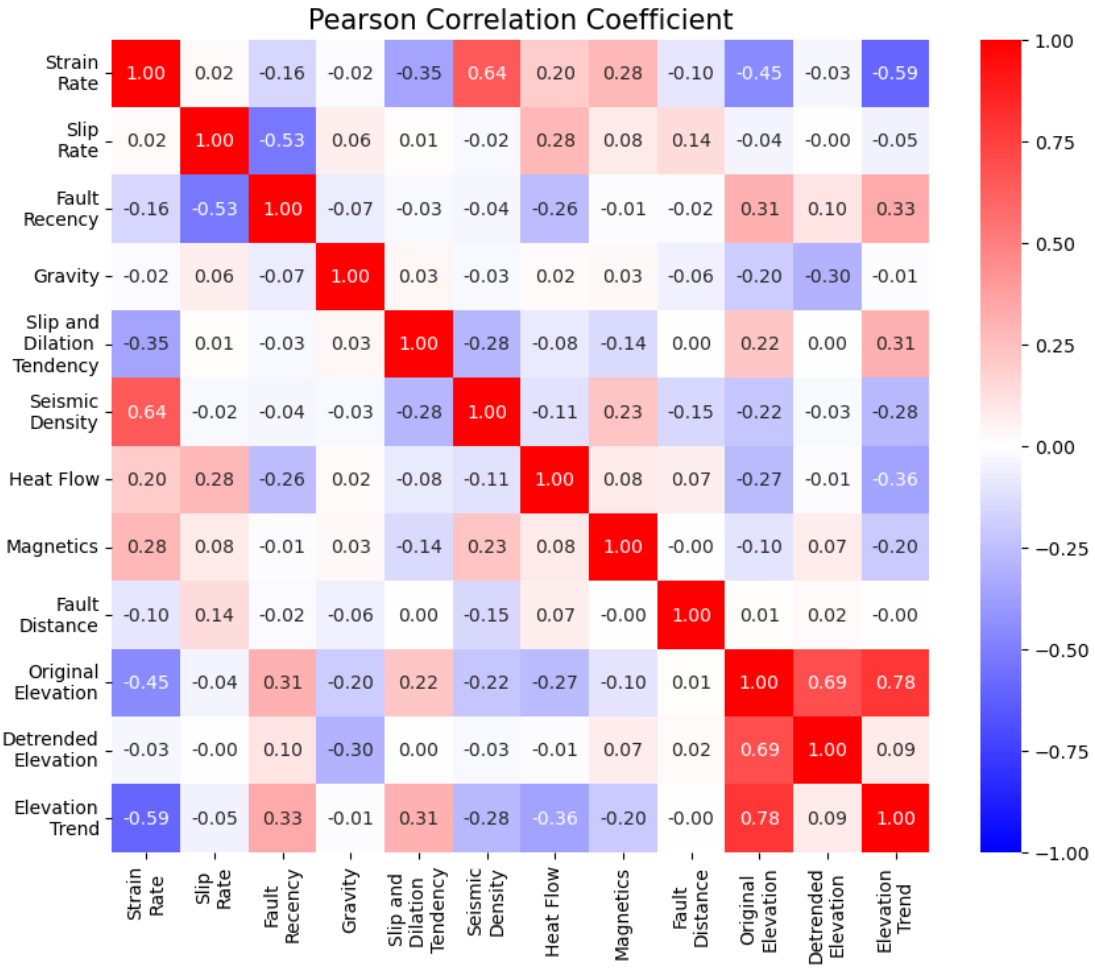

Figure 5: Pearson correlation matrix of the different features for all examples (i.e., sites or cells). Brighter colors indicate a higher absolute correlation between feature pairs. Blue indicates a negative correlation, and red indicates a positive correlation.

3.2 Favorability Maps

When using the Original Elevation Signal feature set there is a greater west-east trend between high and low favorability (compare Fig 6a to 6b, and 7a to 7b) in the models produced by both the strategy that used expert-selected negatives (i.e., NVML; Fig. 6) and the strategy that used randomly selected negatives (i.e., Natural Class Imbalance; Fig. 7). The model using the Separated Elevation Signal feature set and the randomly selected negatives (Fig. 7b) places greater emphasis on low relative topography (i.e., basins) when predicting high favorability.

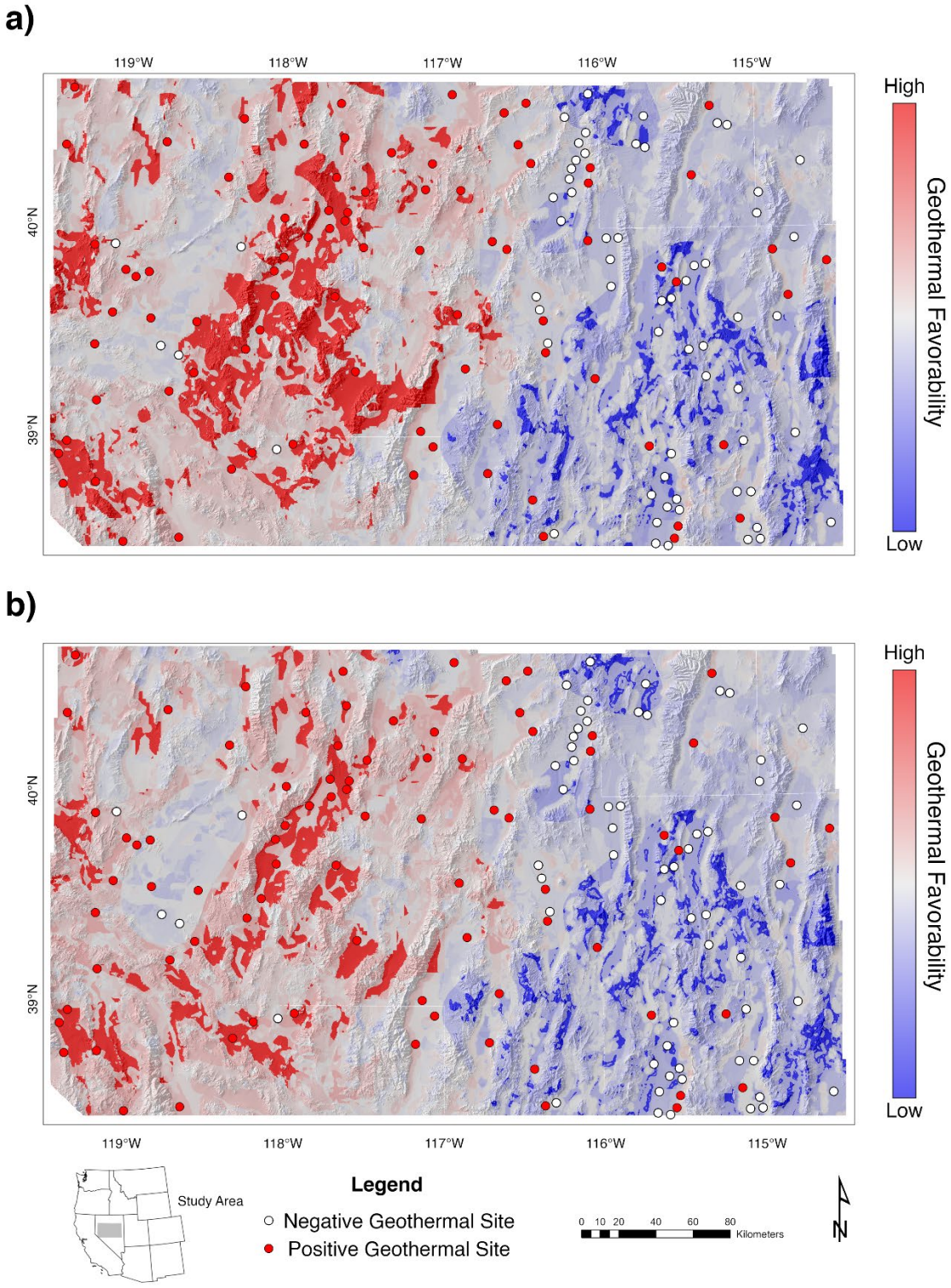

Figure 6: Geothermal favorability maps using the NVML negatives with a) the Original Elevation Signal feature set; and b) the Separated Elevation Signal (separated regional-scale and valley-scale elevation) feature set. Geothermal favorability is the normal score transform of XGBoost-computed probability. Hillshade from USGS 3D Elevation Program (U.S. Geological Survey, 2019).

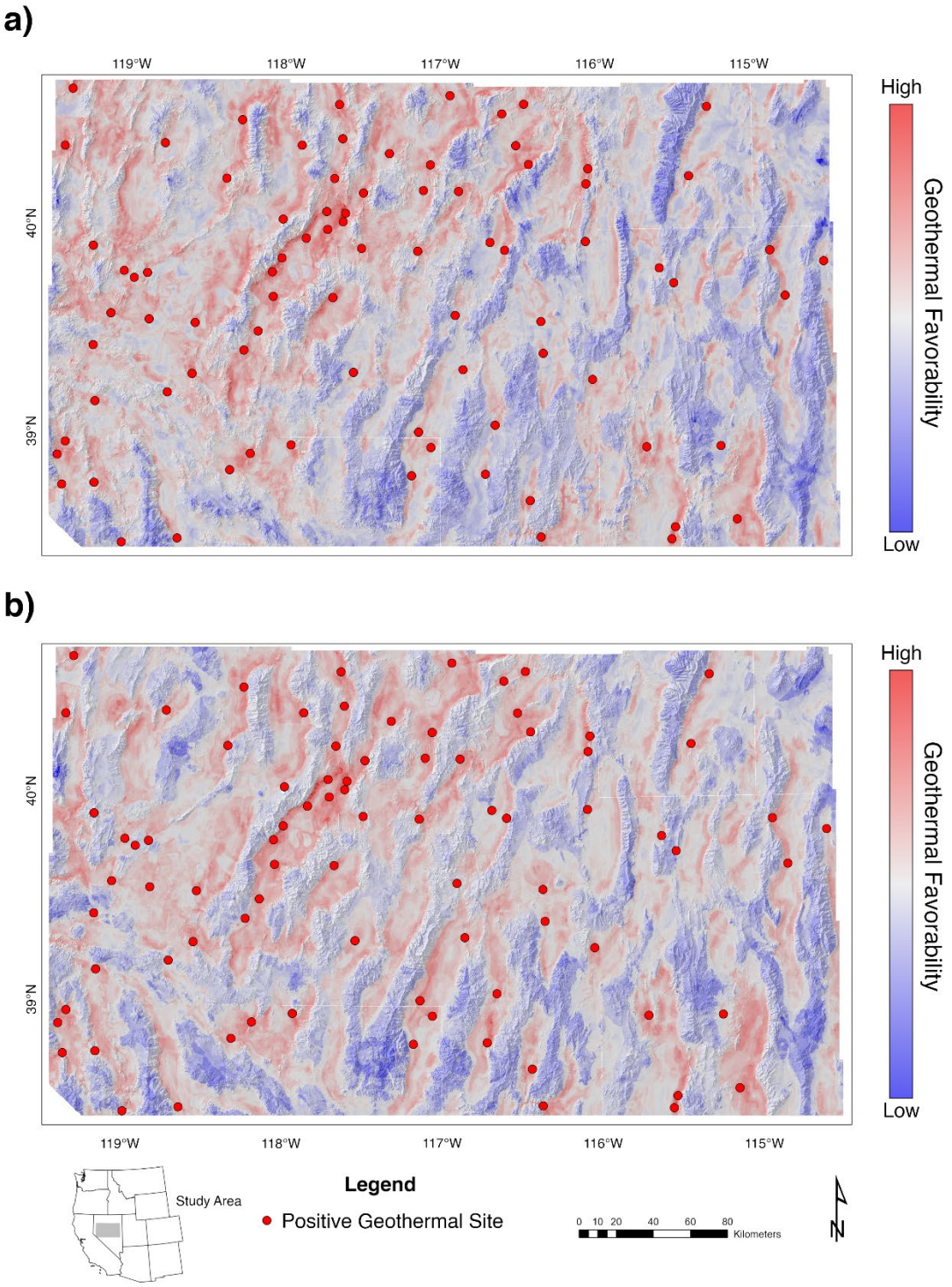

Figure 7: Geothermal favorability maps using random negatives and accounting for Natural Class Imbalance with a) the Original Elevation Signal feature set; and b) the Separated Elevation Signal (separated regional and valley-scale elevation) feature set. Because each random sampling produces different negatives, negatives are not shown. Geothermal favorability is the normal score transform of XGBoost-computed probability. Hillshade from USGS 3D Elevation Program (U.S. Geological Survey, 2019).

3.4 Feature Importance

The relative ranking of feature importance by different measures (Fig. 8) shows that models are primarily dominated by five features (i.e., strain rate, original elevation or *detrended elevation,* fault distance, and heat flow). For the NVML strategy (Figs. 8a, 8c), strain rate is the dominant feature for both models, regardless of the topographic dataset used. For the Natural Class Imbalance strategy (Figs. 8b, 8d), the most important feature varies depending on the input features used. If the model is trained using the feature set containing the original elevation, then the original elevation is the most important feature. Similarly, if the model is trained using the dataset containing *detrended elevation*, then *detrended elevation* ranks as the most important feature.

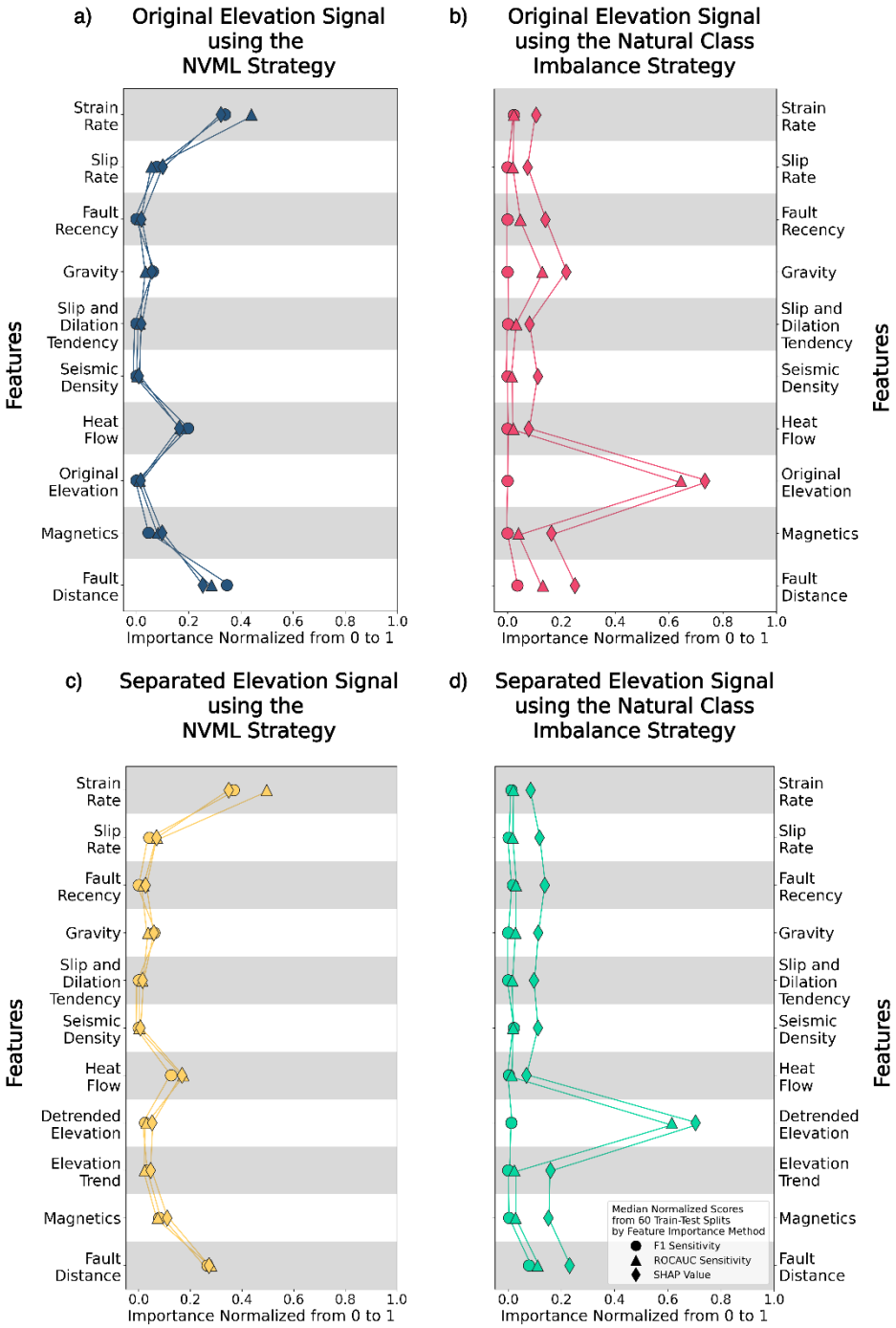

Figure 8: Median normalized feature importance values from the 60 train-test splits using a) the Original Elevation Signal feature set using the Nevada Machine Learning (NVML) strategy (blue); b) the Original Elevation Signal feature set using the Natural Class Imbalance strategy (red); c) the Separated Elevation Signal feature set using the NVML strategy (yellow); d) the Separated Elevation Signal feature set using the Natural Class Imbalance strategy (green). Abbreviation: ROCAUC (triangle) – Area Under the Receiver Operating Characteristic Curve; SHAP (diamond) – Shapely Additive explanation. The legend shown in 8d also applies to 8a, 8b, and 8c.

4. Discussion

All the resulting resource favorability maps (Figs. 6, 7) exhibit some degree of a west-east trend in favorability predictions, but the models using the separated signals for regional elevation and valley-scale elevation (Figs. 6b, 7b) fit models with spatially narrower predicted zones of high hydrothermal favorability that correspond to low relative topography and reduce the west-east trend in predicted favorability regardless of the dominant signal in the model (i.e., Fig. 8c, 8d). In the context of resource assessments, these narrower highly favorable areas minimize exploration efforts to smaller regions.

Recognizing that hydrothermal systems are sparse, we postulate that the models with the most known positives in the smallest area with the highest favorability are the best-performing models. A CDF of predictions for known positives serves as a summary of this information, such that CDF curves of predictions for known positives further to the right indicate that the predictions for the positives have higher favorability scores (Fig. 9). The models from each training strategy using the separated topographic features (i.e., *detrended elevation* and *elevation trend*) consistently outperform the models using the same respective training strategy and the original elevation. Despite the increased model performance when using separated topographic signals, utilizing a training strategy that randomly selects negatives (i.e., the Natural Class Imbalance strategy) rather than using expert-selected negatives (i.e., the NVML strategy) has a greater impact on improved model performance than using the Separated Elevation Signal feature set rather than the Original Elevation Signal feature set. The improvement in performance when using the Separated Elevation Signal feature set may be because adding a *detrended elevation* feature (i.e., valley-scale topographic signal) isolates relevant basin and range structural patterns and may contribute to the algorithm capturing patterns indicative of hydrothermal upflow (Faulds et al., 2011; Faulds and Hinz, 2015). We note, however, that in systems where basin and range faulting is not a controlling feature for hydrothermal circulation, detrending elevation may not be necessary or the scale of the trend may need to be tuned.

The models using the Natural Class Imbalance training strategy outperform NVML ANN regardless of which version of the feature is used, even though XGBoost is mathematically simpler than an ANN. The most important feature in the top-performing approach (i.e., *detrended elevation* in the approach using the Separated Elevation Signal feature set and the Natural Class Imbalance training strategy) shows that the most important signal for predicting hydrothermal systems we use is valley-scale topography. That signal is partially obfuscated when in the original elevation feature because of the inclusion of the signal for regional trend. The second-best performing model (i.e., the approach using the Natural Class Imbalance with the Original Elevation Signal feature set), assigns greater importance to the original elevation and gravity features compared to the top-performing model. This difference may arise from the higher pairwise correlation between gravity and *detrended elevation* (-0.3; Fig. 5) used in the top-performing model compared to the correlation between gravity and original elevation (-0.2; Fig. 5) used in the second-best performing model. Because *detrended elevation* better partially captures the signal for gravity than the original elevation feature, the model using the separated topographic signals lessens the overall contribution of gravity to the model. As the importance of gravity decreases when going from the approach using detrended elevation to the approach using original elevation, the importance of fault distance increases in two of the three metrics of feature importance, emphasizing the importance of local structure for predicting hydrothermal systems (Fig. 8b vs 8d). Hence, the mixed signal in the feature for original elevation with higher elevation in the east and lower

elevation in the west combined with the geospatial distribution of known positives (being primarily in the west) biases the model, causing the model to fail at identifying local topographic signals and predict high favorability at high local elevation (on ranges; Fig. 7a).

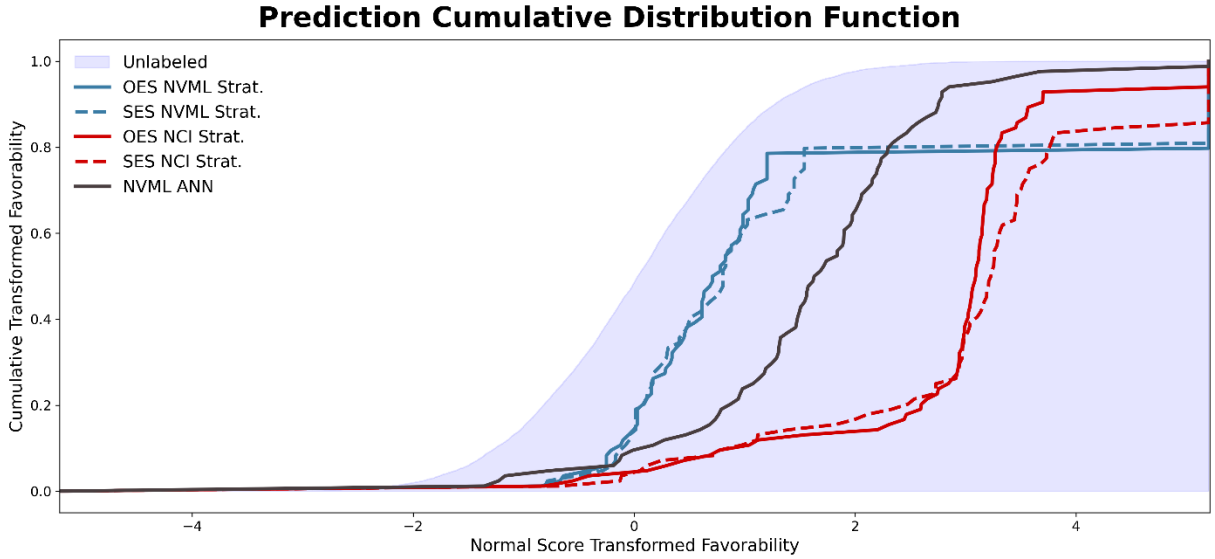

Figure 9: Cumulative distribution function (CDFs) of favorability scores for known geothermal systems relative to other map locations (i.e., unlabeled sites). Shaded blue provides the cumulative distribution of predictions for unlabeled sites. Lines represent the cumulative distributions of predictions for positive-labeled sites from the different approaches. Abbreviations: OES NVML Strat. (solid blue line) – Original Elevation Signal feature set using the Nevada Machine Learning strategy; SES NVML Strat. (dashed blue line) – Separated Elevation Signal feature set using the Nevada Machine Learning Strategy; OES NCI Strat. (solid red line) – Original Elevation Signal feature set using the Natural Class Imbalance; SES NCI Strat. (dashed red line) – Separated Elevation Signal feature set using the Natural Class Imbalance; NVML ANN (solid black line) - Nevada Machine Learning Artificial Neural Network.

The most important feature in the models using the NVML training strategy is strain rate, but adding the *detrended elevation* feature lessens the intensity of the west-east trend in the predicted hydrothermal favorability (Fig. 6). Although both models using the NVML training strategies rank last (Fig. 9), adding the *detrended elevation* feature improves model performance, demonstrating the importance of engineering input features that emphasize relevant geological information.

5. Conclusion

In this study, we evaluate the importance of regional-scale and valley-scale topographic relief as separated input features used to fit models that predict hydrothermal favorability. Separating the topographic signals allows the machine learning (ML) algorithm to fit better to the geological structures associated with hydrothermal systems in the Great Basin as evidenced by the models fit using the separated topographic signals consistently outperforming the models fit using the original topographic feature with the combined signals. Separating the topographic signals into two features also lessens the intensity of a west-east trend in the hydrothermal favorability maps. Hence, careful feature engineering through domain expertise with geothermal ML can deconvolute signals in geologic datasets and consequently improve model performance when predicting

favorability for hydrothermal resources. We explore the impact of separated topographic signals using two training strategies: 1) using expert-selected sites with no hydrothermal systems; and 2) using randomly selected sites from unlabeled sites as having no hydrothermal system. As is consistent in our prior works, the approach using randomly selected sites as having no hydrothermal system vastly outperforms the approach using expert-selected sites with no hydrothermal systems.

Acknowledgments

This work was done under contract to the U.S. Geological Survey. This work was supported by the Portland State University Louis Stokes Alliance for Minority Participation Program, Portland State University Institute for Sustainable Solutions Program, and the U.S. Geological Survey-Portland State University Partnership. Additional support for John Lipor was provided by the National Science Foundation award NSF CAREER CIF-2046175. This work was also supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE), Geothermal Technologies Office (GTO) under Contract No. DEAC02-05CH11231 with Lawrence Berkeley National Laboratory, Conformed Federal Order No. 7520443 between Lawrence Berkeley National Laboratory and the U.S. Geological Survey (Award Number DE-EE0008105), and Standard Research Subcontract No. 7572843 between Lawrence Berkeley National Laboratory and Portland State University. This study was funded by the U.S. Geological Survey Energy Resources Program and the U.S. Department of Energy - Geothermal Technologies Office under award DE-EE0009254 to the University of Nevada, Reno for the INnovative Geothermal Exploration through Novel Investigations Of Undiscovered Systems (INGENIOUS). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A: Hyperparameters and Early Stopping

The median optimal hyperparameters from the 60 train-test splits (Table A1) are used to fit the final models using all the data. We use the median estimator for early stopping in the 60 train-test splits in the final models to prevent overfitting.

Table A1: Median optimal hyperparameters (± one standard deviation) from the 60 train-test splits. Abbreviations: OES: Original Elevation Signal feature set model, SES: Separated Elevation Signal feature set model, NVML Strat: Nevada Machine Learning Strategy, NCI: Natural Class Imbalance Strategy

Strategy & Algorithm	Class Weight	Learning Rate	Maximum Depth	Number of Estimators
OES NVML	2 ± 0.8	0.10 ± 0.06	4 ± 1.0	13 ± 15.8
SES NVML	2 ± 0.8	0.11± 0.05	4 ± 1.0	11 ± 13.9
OES NCI	1790 ± 94.0	0.14 ± 0.10	4 ± 0.5	76 ± 30.3
SES NCI	1815 ± 95.0	0.15 ± 0.06	4 ± 0.4	80 ± 26.8

REFERENCES

Bekker, J., and Davis, J. (2020). Learning from positive and unlabeled data: A survey. *Machine Learning*, *109*(4), 719-760.

Brown, S. R., Coolbaugh, M. F., DeAngelo, J., Faulds, J. E., Fehler, M., Gu, C., Queen, J., Treitel, S., Smith, C., and Mlawsky, E. (2020). Machine learning for natural resource assessment: An application to the blind geothermal systems of Nevada. *Geothermal Resources Council Transactions*, *44*, 920-932.

Caraccioli, P. D., Mordensky, S. P., Lindsey, C. R., DeAngelo, J., Burns, E., and Lipor, J. (2023). Don't Let Negatives Hold You Back: Accounting for Underlying Physics and Natural Distributions of Hydrothermal Systems When Selecting Negative Training Sites Leads to Better Machine Learning Predictions. *Geothermal Resources Council Transactions*, *47*, 1672-1693. https://pubs.usgs.gov/publication/70251035

Chen, T., and Guestrin, C. (2016). *XBGoost: A Scalable Tree Boosting System*. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, United States.

Coolbaugh, M. F., Raines, G. L., and Zehner, R. E. (2007). Assessment of exploration bias in data-driven predictive models and the estimation of undiscovered resources. *Natural Resources Research*, *16*, 199-207.

DeAngelo, J., Burns, E., Mordensky, S. P., and Lindsey, C. R. (2023). Detrending Great Basin elevation to identify structural patterns for identifying geothermal favorability. *Geothermal Resources Council Transactions*, *47*, 1694-1702. https://pubs.usgs.gov/publication/70250639

Faulds, J. E., Brown, S. R., Smith, C. M., Queen, J., and Treitel, S. (2021a). *Machine Learning Model Geotiffs-Applications of Machine Learning Techniques to Geothermal Play Fairway Analysis in the Great Basin Region, Nevada*. https://doi.org/https://doi.org/10.15121/1897036

Faulds, J. E., and Hinz, N. H. (2015). *Favorable Tectonic and Structural Settings of Geothermal Systems in the Great Basin Region, Western USA: Proxies for Discovering Blind Geothermal Systems*. Proceedings of the World Geothermal Congress, Melbourne, Australia.

Faulds, J. E., Hinz, N. H., Coolbaugh, M., Ayling, B., Glen, J., Craig, J. W., McConville, E., Siler, D., Queen, J., and Witter, J. (2021b). *Discovering Blind Geothermal Systems in the Great Basin Region: An Integrated Geologic and Geophysical Approach for Establishing Geothermal Play Fairways: All Phases* (DOE-UNR-06731-01).

Faulds, J. E., Hinz, N. H., Coolbaugh, M. F., Cashman, P. H., Kratt, C., Dering, G., Edwards, J., Mayhew, B., and McLachlan, H. (2011). Assessment of Favorable Structural Settings of Geothermal Systems in the Great Basin, Western USA. *Geothermal Resources Council Transactions*, *35*, 777-783.

Faulds, J. E., Hinz, N. H., Coolbaugh, M. F., Sadowski, A., Shevenell, L. A., McConville, E., Craig, J., Sladek, C., and Siler, D. (2017). *Progress Report on the Nevada Play Fairway Project: Integrated Geological, Geochemical, and Geophysical Analyses of Possible New Geothermal Systems in the Great Basin Region*. Proceedings of 42nd Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California, United States.

Faulds, J. E., Smith, C. M., Brown, S., Coolbaugh, M., DeAngelo, J., Glen, J., Ayling, B., Siler, D., Mlawsky, E., and Queen, J. (2024). *Final Technical Report-Applications of Machine Learning Techniques to Geothermal Play Fairway Analysis in the Great Basin Region, Nevada*.

Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., and Herrera, F. (2018). *Learning from imbalanced data sets* (Vol. 10). Springer.

Lee Rodgers, J., and Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, *42*(1), 59-66.

Mordensky, S. P., Lipor, J. J., DeAngelo, J., Burns, E. R., and Lindsey, C. R. (2023). When less is more: How increasing the complexity of machine learning strategies for geothermal energy assessments may not lead toward better estimates. *Geothermics*, *110*, 102662.

Prechelt, L. (2002). Early stopping-but when? In *Neural Networks: Tricks of the trade* (pp. 55-69). Springer.

Pyrcz, M. J., and Deutsch, C. V. (2018). Transforming data to a gaussian distribution. *Geostatistics Lessons*.

Smith, C. M. (2021). *Machine Learning Techniques Applied to the Nevada Geothermal Play Fairway Analysis* University of Nevada, Reno]. Reno, Nevada. https://scholarworks.unr.edu//handle/11714/7940

U.S. Geological Survey. (2019). *3D Elevation Program 1-Meter Resolution Digital Elevation Model*. https://www.usgs.gov/the-nationalmap-data-delivery

Williams, C. F., Reed, M. J., Mariner, R. H., DeAngelo, J., and Galanis, S. P. (2008). Assessment of Moderate-and High-Temperature Geothermal Resources of the United States. *U.S. Geological Survey Fact Sheet 2008-3082*, 1-4.