GEOCLIMB: ADVENTURE IS OUT THERE!

Ian Christie, Ankur Varma
{ian.christie,ankur.varma}@colorado.edu

ABSTRACT

GeoClimb leverages geospatial data and machine learning to predict rock climbing locations, significantly reducing the search space for new areas of exploration. We integrate Sentinel-2 imagery, Digital Elevation Models (DEMs), and lithologic data, utilizing unsupervised learning techniques to create embeddings. These embeddings are evaluated using a Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), and Random Forest (RF), with ensemble techniques applied to reduce noise. Our model achieves a strong predictive performance, with a precision of 0.92 and a false negative rate of 0.09. Cross-validation using current Mountain Project data reveals that 30% of the top positive predictions have had climbing locations developed since 2019, underscoring the model's effectiveness.

1 INTRODUCTION

The climbing industry relies on exploration, but manually searching for high-quality climbing areas is both costly and time-consuming. By developing a model that predicts rock climbing locations and generalizing it on a global scale, we aim to significantly reduce the search space for exploration. We hypothesize that a combination of satellite imagery [1], elevation data [2], and lithologic information [3] provides sufficient signals to predict rock climbing potential in a given area. To train this model, we will use Mountain Project data [4] from the United States as labels.

All code can be found at our github and data can be downloaded from our google drive.

2 METHODOLOGY

For each location datum, we employed a data type-specific encoder to extract meaningful features. To achieve optimal results, we experimented with multiple encoders for each data type. After obtaining individual embeddings, we concatenated them to construct combined embeddings. These combined embeddings were then divided into training, validation, and test splits. We trained several machine learning models on the training set and evaluated their performance using the test split. The top 10 predictions from the test split, generated using an ensemble model, were manually reviewed to cross-validate the model. The overall process is illustrated in Fig 1.

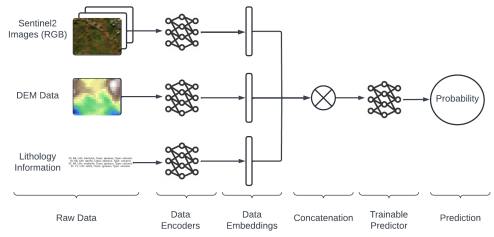

Figure 1: The Geoclimb architecture encodes data into embeddings, which are then processed by a machine learning model to generate a probability prediction for climbing at a given location.

2.1 Dataset and Preprocessing

We created a novel GeoClimb Dataset, containing Sentinel-2 RGB imagery, Digital Elevation Model (DEM) data, and lithologic information for 12,742 climbing locations across the continental United States. The dataset also includes the same information for 12,692 unlabeled locations randomly distributed across the continental US. For more information on the creation and composition of the dataset please refer to our midpoint report paper.

2.2 Data Encoding

For each location datum, we used a data type-specific encoder to extract meaningful features, experimenting with multiple encoders per data type to achieve the best results.

For Sentinel-2 RGB imagery, we evaluated several approaches, including MOSAIKS[5], flattened imagery, Random Convolution Features (RCFs) using Gaussian and empirical modes, and a pretrained model developed with Momentum Contrast (MoCo)[6]. For Digital Elevation Models (DEMs), we tested flattened embeddings and RCFs in Gaussian and empirical modes[5]. For lithology data, we encoded the text using SciBERT[7], a BERT language model trained on scientific text, with and without the inclusion of the description field from the JSON data, see Appendix for more details. Mean pooling was applied to the processed input tokens to generate the final lithology embedding.

We trained the Multi-Layer Perceptron (MLP) model, detailed in the following section, on each singular embedding type to evaluate which approach extracted the most useful features for the machine learning model. Using the best-performing embeddings from each dataset, we concatenated them to form the final embeddings for each location. Additionally, we performed an ablation study by combinatorially concatenating embeddings to assess the individual contributions of each embedding type to the model's performance.

2.3 Model Training

We split the locations into 70:15:15 training, validation, and test sets, ensuring an even distribution between labeled and unlabeled locations to avoid class imbalance.

To predict climbing locations, we employed a simple Multi-Layer Perceptron (MLP) designed for binary classification tasks. The model processes an input embedding vector, passing it through two hidden layers with sizes 512 and 256, each using ReLU activation for non-linearity. The final layer outputs a single value, which is passed through a sigmoid activation to interpret it as a probability. Binary Cross-Entropy Loss was used to measure how well the predicted probabilities aligned with the target binary labels.

Additionally, we experimented with other machine learning models in order to compare results. For the Support Vector Machine (SVM), we optimized the model using linear, polynomial, radial basis function (RBF), and sigmoid kernels. For the Random Forest (RF), we tested the following configurations: (10 estimators, no max depth), (50 estimators, max depth = 10), and (100 estimators, max depth = 20).

2.4 Evaluation Principles

We evaluate the model using two metrics: precision and false negative rate (FNR). A high performing model will have a precision near 1 and a FNR near 0.0175, which is explained below.

Because the labeled points are known positives, we aim for a precision of 1. However, since the unlabeled points may or may not contain climbing areas, the FNR should not be 0. Instead, it should reflect prior knowledge of the dataset. To estimate this, we consider the probability that a random location in the U.S. has climbing within 1 km.

We calculate this probability by drawing circles with a radius of 1 km around each known climbing location in our dataset. After merging overlapping circles, we calculate the total explored landmass, which is approximately 14,057.06 km^2. Assuming, unquantitatively, that only 10% of all potential climbing areas in the U.S. have been discovered, the total climbing landmass is estimated to be

140,576 km^2. Given the total land area of the continental U.S. (8,080,464.3 km^2), we calculate the probability of climbing within 1 km of a random location to be approximately 1.75%.

3 Experiments

3.1 Individual Dataset Experiments

Figure 2 shows the accuracy results when individual encodings were used in the MLP model. The encodings derived from Sentinel-2 images in our dataset did not achieve an accuracy higher than 0.5. However, using MOSAIKS[5] embeddings for these locations improved accuracy to 0.84. We hypothesize that this task is too challenging for images with a 10 m resolution, and MOSAIKS[5] performs better because its embeddings are derived from images with a 3 m resolution.

Flattened DEM, DEM RCF Gaussian[6], and DEM RCF Empirical[6] encodings all performed similarly. We selected DEM RCF Empirical due to its balanced performance across accuracy, precision, and recall.

For lithology embeddings derived from SciBERT[7], both the version with and without the description field performed equally well, with the embedding containing the description showing a slight advantage. However, we chose to use the embedding without the description, as we felt the description was too survey-specific and might not generalize well to unseen geographical regions. We explore this in the Appendix in Fig 8.

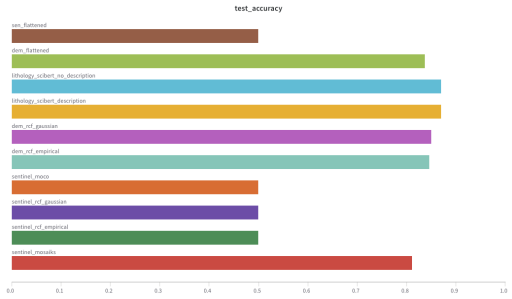

Figure 2: displays the test accuracy for all of the single embedding experiments.

3.2 Ablation Study

To assess the contribution of individual features, we conducted an ablation study by excluding Sentinel-2, DEM and Lithology embeddings in different combinations.

Results from test precision in the Appendix in Fig 7 show that integrating all three features consistently outperformed any subset, validating the importance of combining multi-modal data for accurate probability predictions. Lithology contributed the most to model performance, followed by DEM and then Sentinel Imagery.

3.3 Model Type Experiments

We leverage deep learning during the encoding phase but rely on simpler machine learning models during the prediction phase. Using the combined embeddings, we reduced the task to cluster classification in a multidimensional space. We hypothesized that a simpler ML model could perform comparably to the MLP model. To test this, we optimized Support Vector Machine (SVM) and Random Forest (RF) models with varying hyperparameters.

Figure 3 validates this hypothesis: all RF models achieved precision comparable to the MLP, with the linear kernel SVM demonstrating the best precision. Additionally, the SVM with the linear kernel had a false negative rate (FNR) closest to 1.75%. This indicates that the linear kernel SVM was our best performing model.

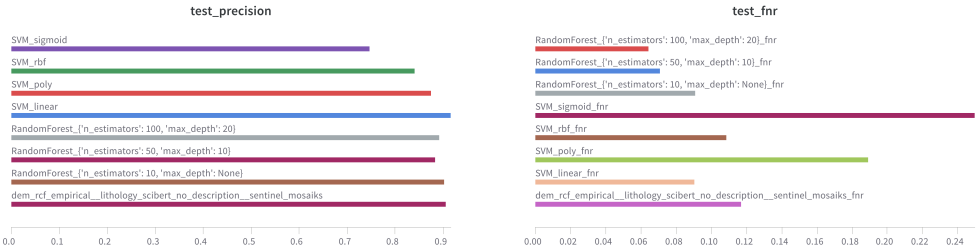

Figure 3: (left) shows the test precision for the models trained on the finalized combined embeddings. (right) shows the test FNR for the models trained on the finalized combined embeddings.

4 PREDICTIONS

To predict climbing locations, we used the unlabeled locations from the test split. We calculated Cohen's Kappa (K) to assess the agreement between the best-performing model types: $K(\text{SVM, RF}) = 0.4814$, $K(\text{SVM, MLP}) = 0.5773$, and $K(\text{RF, MLP}) = 0.5094$. These values indicate a substantial amount of disagreement between the models regarding positive predictions.

To mitigate this noise, we ensembled the models by averaging their probability predictions for each location. Figure 4 displays the locations with the highest probability predictions from the ensemble model.

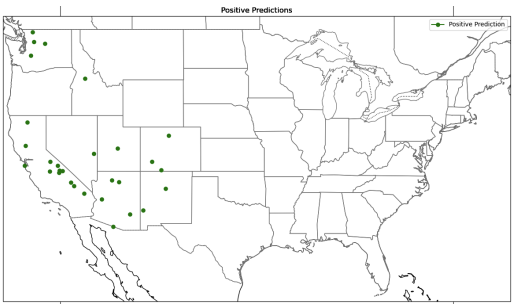

Figure 4: The locations of the predictions with a probability over 0.9 given by the ensemble model.

The top 10 locations with the highest climbing probabilities were cross-verified using Mountain Project and Google Earth. It is possible that the initial Mountain Project scrape missed some locations or that climbing areas in the predicted regions have been developed since the 2019 scrape. Our analysis found that **30**% of the locations had climbing, supporting the validity of our model. Snapshots of some of these sites are shown in Figure 5.

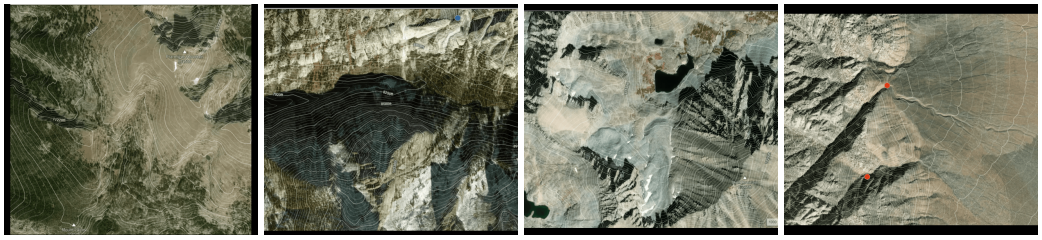

Figure 5: A sample of snapshots from the Mountain Project of our top 10 predicted climbing areas. Pictures 2 &4 indicate that there is a climbing area has developed since 2019 in the area.

REFERENCES

[1] https://developers.google.com/earth-engine/datasets/catalog/sentinel-2

[2] https://developers.google.com/earth-engine/datasets/tags/dem

[3] https://macrostrat.org/

[4] https://www.kaggle.com/datasets/pdegner/mountain-project-rotues-and-forums/discussion?sort=hotness

[5] Rolf, E., Proctor, J., Carleton, T. et al. A generalizable and accessible approach to machine learning with global satellite imagery. Nat Commun 12, 4392 (2021). https://doi.org/10.1038/s41467-021-24638-z

[6] Stewart, A. J., Robinson, C., Corley, I. A., Ortiz, A., Ferres, J. M. L., & Banerjee, A. (2022, November). Torchgeo: deep learning with geospatial data. In Proceedings of the 30th international conference on advances in geographic information systems (pp. 1-12).

[7] Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. arXiv preprint arXiv:1903.10676.

A APPENDIX

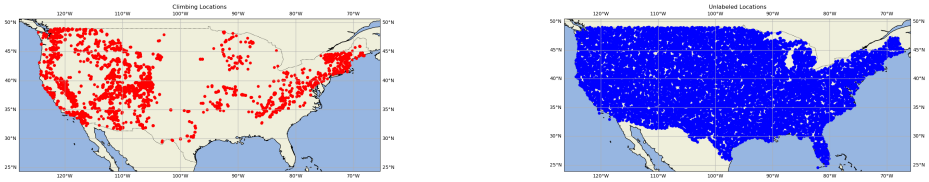

Figure 6: (left): Shows the spatial distribution of scraped climbing locations in the continental US. (right): Shows the spatial distribution of the generated unlabeled locations in the continental US. Note: these have the same number of points demonstrating the tight clustering of climbing locations.

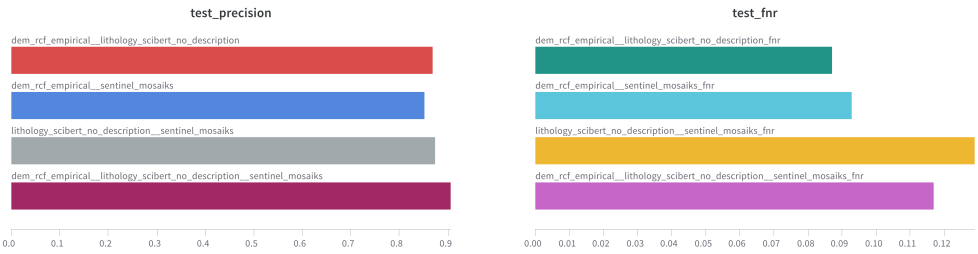

Figure 7: The test precision and test false negative rate for the ablation study. The worse the precision for the experiment, the more important the left out data type was for the overall performance of the model.

```
{
    "name": "Edwards Limestone",
    "age": "Albian",
    "strat_name": "Edwards Limestone",
    "lith": "Major: {limestone}, Minor: {dolostone},
             Incidental: {chert}",
    "descrip": "Bottom to top: Bee Cave Marl Member, Cedar
        Park Limestone Member, Port Terrett Member, Segovia
        Member contains Kirschberg evaporite; In Eastern part
        of Trans-Pecos and High Plains - Limestone, dolostone,
        and chert; limestone aphanitic to fine-grained, massive
        to thin-bedded, hard, brittle, in part rudistid biostromes,
        much miliolid biosparite; dolostone fine to very fine-
        grained, porous, chert nodules common. Important aquifer.
        350-400 ft thick. In North, Central, and South Texas,
        including Quaternary for all of West Texas - Limestone,
        dolostone, and chert; Upper part -- mostly limestone,
        argillaceous, beds of even thickness, some nodular;
        alternating hard and soft units, Exogyra texana and miliolids
        locally abundant; thickness 40 ft. Lower part -- limestone,
        dolostone, and chert; limestone aphanitic to fine-grained,
        argillaceous, locally siliceous, thin beds to nodular and
        massive, hard, gray to light brown, miliolids, gastropods,
        and rudistids, chert nodules and plates common; dolostone
        fine to very fine-grained, porous, chert nodules common.
        Important aquifer. Lower part up to 100 ft thick. Total
        thickness of Edwards in Brownwood Sheet as much as 140 ft.",
    "liths": [
        {
            "lith": "limestone",
            "lith_type": "carbonate",
            "lith_class": "sedimentary",
        },
        {
            "lith": "dolomite",
            "lith_type": "carbonate",
            "lith_class": "sedimentary",
        },
        {
            "lith": "chert",
            "lith_type": "chemical",
            "lith_class": "sedimentary",
        }
    ]
}
```


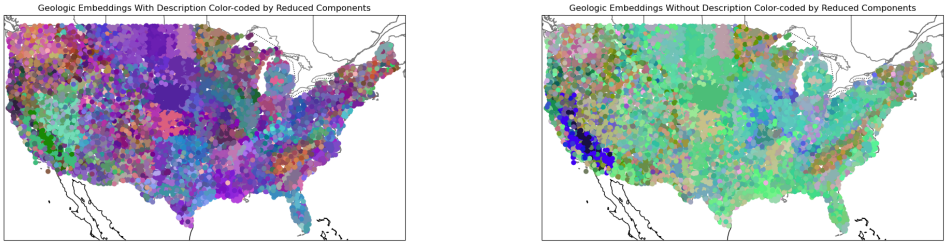

Figure 8: A sample of the JSON data received from the Macrostrat API (top) and the comparison of lithology embeddings with (left) and without (right) descriptions. States lines and survey delineations are more clearly visible in the embeddings with the description whereas geological features are more visible in the embeddings without the description.