Boulder Grade Classification using Machine Learning

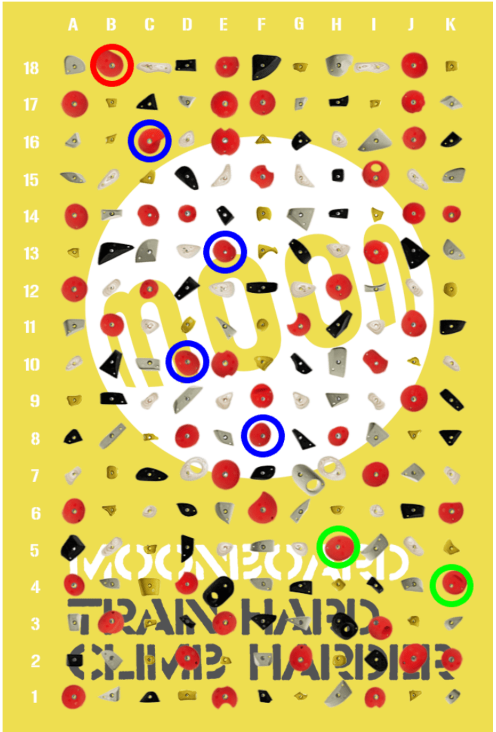
Dylan Wells

INTRODUCTION / AIM

Assigning a grade for the difficulty of a climbing route is a somewhat infamously subjective task in the climbing community. Many climbs are made more harder or easier depending on personal factors. So, coming to a consensus on the grade of a climb can be inconsistent.

This leads to the question; can we construct a model to impartially and accurately decide? This project seeks to explore the applications of machine learning to classify MoonBoard boulder problems in the Hueco Grading System.

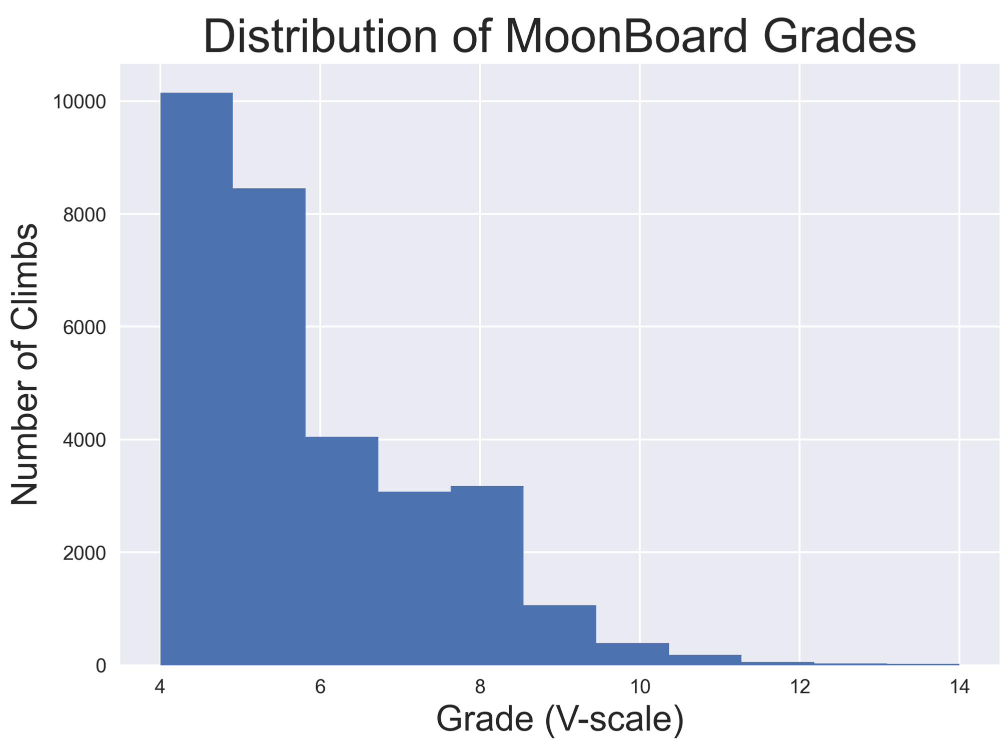
Figure 1.



Example picture of a MoonBoard boulder problem, circling the holds used.

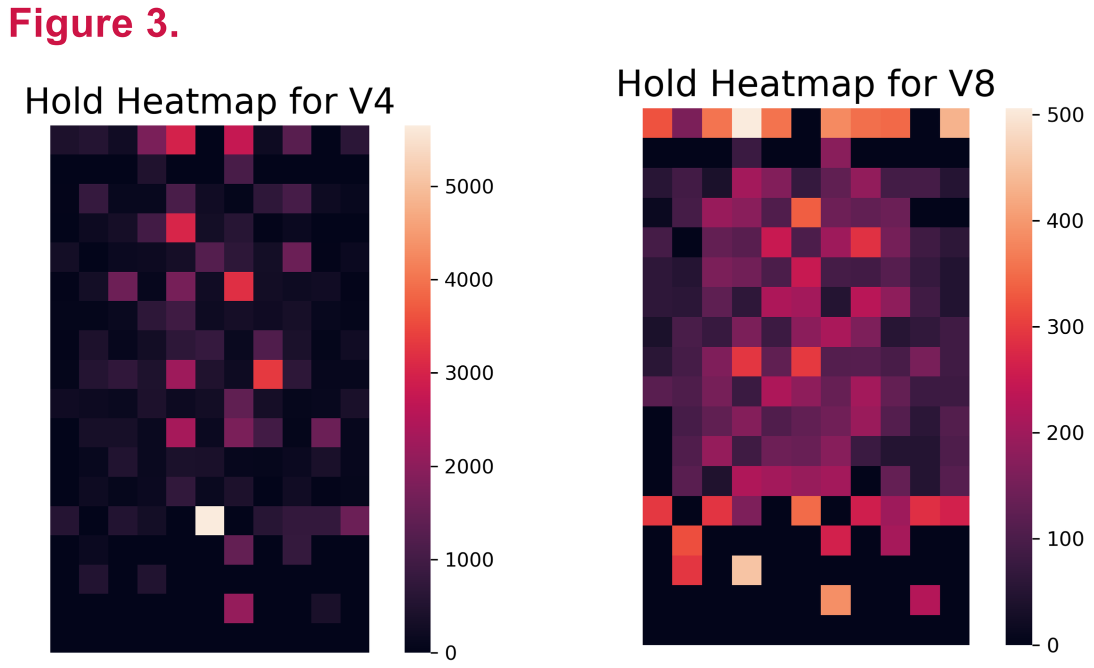
VISUALIZATION

Figure 2.



This graph shows the distribution of all MoonBoard boulders by V-Grade. There are less climbs as the difficulty increases. V8 is the exception as it combines 2 Font Grades in the conversion instead of just 1 like V7 and V9 do.

Figure 3.



Heatmaps showing the most used holds on all V4 and V8 MoonBoard climbs respectively.

THE DATASET

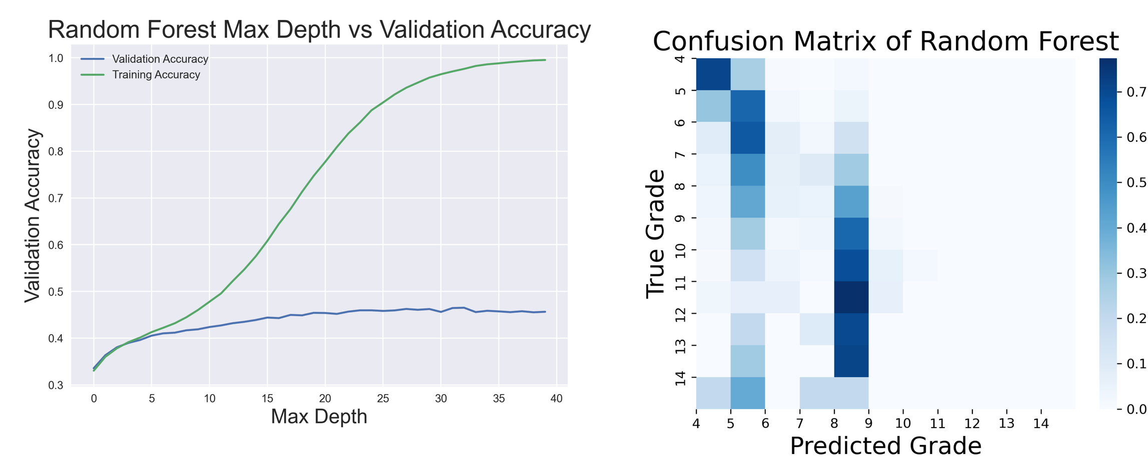
The MoonBoard dataset, sourced from [2], contains over 30,000 user-made and graded climbing routes. Each route is formatted as a 141-dimensional array of binary values, determining which of the 141 unique holds are used for the climb. Figure 2 shows the distribution of grades for all MoonBoard routes

RANDOM FOREST

Figure 3 shows how many times each hold is used in all V4 climbs and all V8 climbs. The stark difference in most common holds indicated that simple cuts in the data by holds used with a Random Forest might be a strong approach to the problem

The initial test of the RF showed, however, the downside of these cuts in the data. According to Figure 4, the classifier found that relegating most climbs into V4, V5, or V8 as most profitable – probably due to the prevalence of V4s and V5s as shown in Figure 2.

Figure 4.



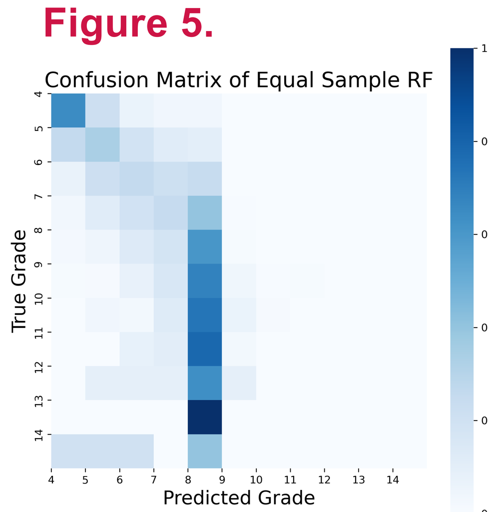
First Random Forest Model showing the accuracies against max depth and the Confusion Matrix of the model trained to max depth 30, when the validation accuracy starts to plateau.

Random Forest With Even Sampling

While initial model's confusion matrix showed poor performance, there are still ways to solve the issue. To get the RF to predict a more even spread of grades, I limited the number of V4 and V5 climbs in the training set.

This forced the model to attempt to learn how to predict V6 and V7 climbs as well, leading to the more diagonal confusion matrix in Figure 5.

Figure 5.



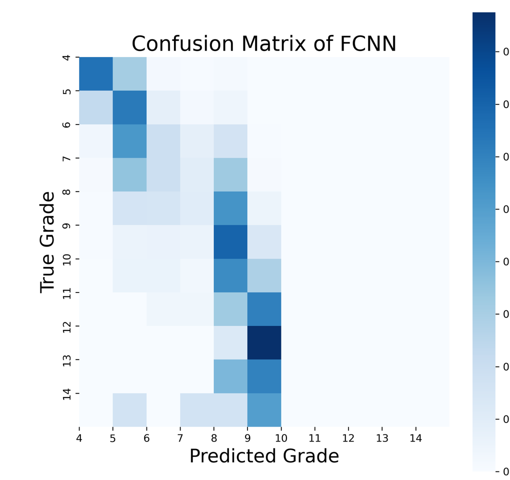
This confusion matrix shows the results of the RF model trained on equal number of V4-V8 climbs, instead of the skewed distribution.

MULTI LAYER PERCEPTRON

After the RF classifier, I attempted to create a multi-layer perceptron model to classify the climbs. The advantage of the MLP is the ability to account for how different holds interact with each other rather than simple cuts in the data.

The first model tested with randomly created layers is shown in Figure 6. While less severe, it showed the same issue as the first RF classifier with very few climbs predicted as V6 or V7. So, I employed the same even sampling technique from before combined with using a Genetic Algorithm to tune the hyperparameters of the MLP.

Figure 6.



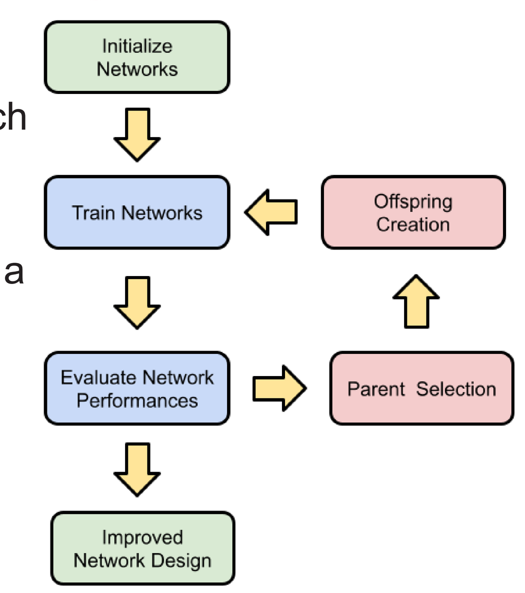
Confusion Matrix of Fully Connected Neural Network with randomly chosen hyperparameters

GENETIC ALGORITHM TUNING

A simple 8-layer Fully Connected Neural Network can have over 12 trillion different combinations of hyperparameters. To efficiently search through this parameter space, I created a Genetic Algorithm to simulate the process of evolution on a population of neural networks. The goal was to arrive at networks more "fit" to the environment of predicting MoonBoard climbs accurately.

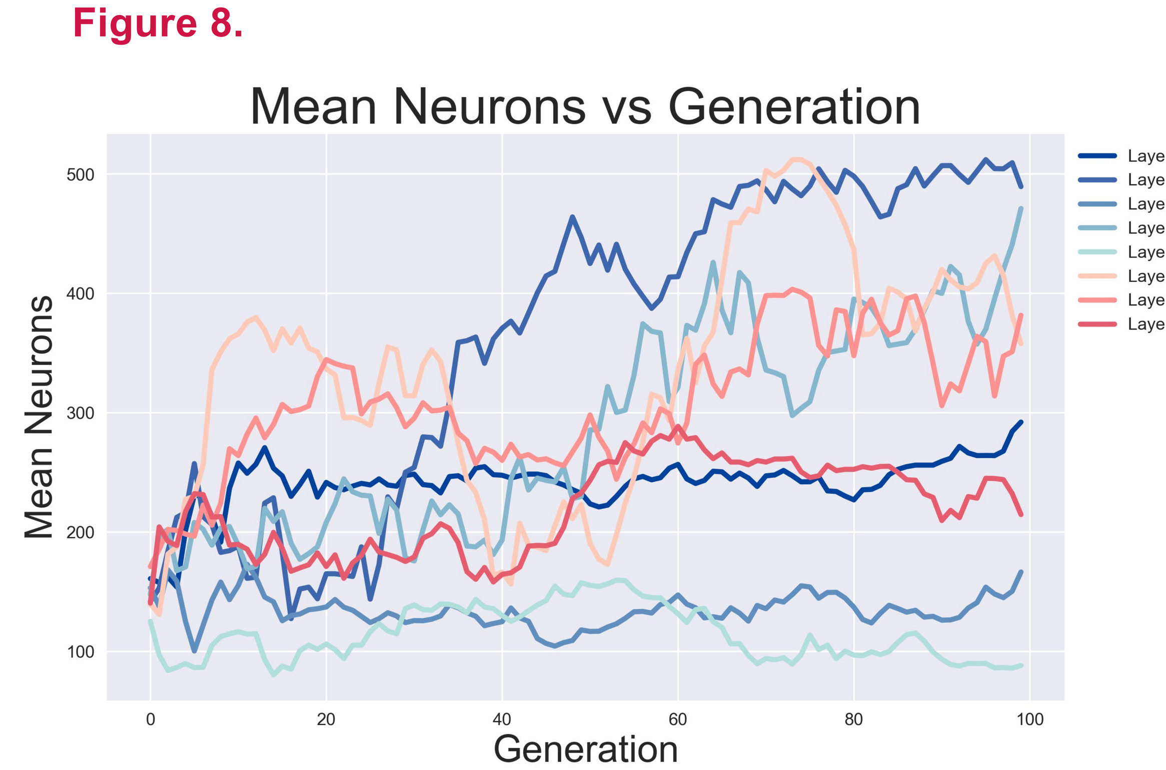
Figure 7 shows the pseudocode for the Genetic Algorithm.

Figure 7.



Overarching design of the Genetic Algorithm architecture.

Figure 8.



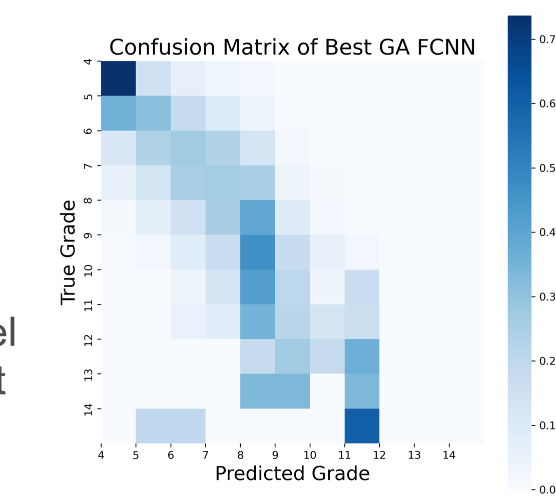
This plot shows the mean number of neurons in the population for each separate layer of the MLP neural network. Possible values were 8, 16, 32, 64, 128, 256, and 512.

Best Performing GA MLP

The Genetic Algorithm converged on the best design by Generation 93 of the evolution, preferring higher number of neurons per layer, relu activations, and rmsprop optimization.

The confusion matrix for the best-performing evolved model appears the most diagonal out of any model and produced the highest validation accuracy.

Figure 9.



Confusion Matrix for best-performing evolved FCNN model.

RESULTS / FUTURE

Figure 10.

	RF	RF Even Sampling	MLP	Evolved MLP
Validation Accuracy	46.8%	47.0%	48.5%	49.8%

After testing these 4 different models for MoonBoard classification, the Genetic Algorithm Tuned MLP network performed with the highest accuracy of 49.8% on the validation data.

Previous work on this subject has yielded validation accuracies of 36.5% using Softmax Regression on the Font Grade System [1] and 46.5% using a Recurrent Neural Network design on the Hueco Grade System [2]. So, the best model my method found has surpassed the accuracy of previously done work!

In the future, I could consider performing a full analysis only considering climbs with multi-repeats as seen in [2]. Early tests showed improved model accuracy, but I decided to include all user-graded climbs for this study.

Additionally, I could add more information to the dataset for the model to use in its predictions. For example; differentiating starting, middle, and end holds, creating features for the physical dimensions of the holds, or designing an algorithm akin to the Beta Move image analysis used in [2] to predict the move sequence before classifying the climb by grade.

BIBLIOGRAPHY

[1] Machine Learning Methods for Climbing Route Classification. Alejandro Dobles, Juan Carlos Sarmiento, and Peter Satterthwaite, 2017.

[2] Recurrent Neural Network for MoonBoard Climbing Route Classification and Generation. Yi-Shiou Duh and Ray Chang, 2021.

The Ohio State University

https://github.com/dylanwells37/HuecoClassification