Raid Data Challenge

BYRON BHUIYAN, University of California, Riverside, USA
ALAN XU-ZHANG, University of California, Riverside, USA
SHAWN HWANG, University of California, Riverside, USA

Indoor climbing route grading is traditionally a subjective process, often relying on the intuition of route setters or informal climber feedback. This subjectivity can lead to inconsistencies that affect climber experience and hinder skill progression. In this work, we propose a novel, interpretable method for predicting indoor climbing route difficulty using graph-based structural modeling. Each route is represented as a graph, where nodes correspond to holds and edges represent feasible transitions based on spatial proximity. We extract interpretable features from these graphs — such as average move distance, vertical gain, and hold size — and train machine learning models to predict route difficulty levels.

To address the lack of ground-truth labels, we apply a weak supervision strategy that maps hold color to approximate V-grades using gym-specific conventions. Our experiments show that structural features alone are sufficient to predict difficulty with high accuracy, and that the most predictive features align closely with physical challenge as experienced by climbers. This work offers a lightweight, scalable alternative to image-based grading systems and provides a foundation for future research in sports analytics, automated route assessment, and human-centered route design.

Additional Key Words and Phrases: Machine Learning, Data Science, LLM, AI Generated Text

ACM Reference Format:
Byron Bhuiyan, Alan Xu-Zhang, and Shawn Hwang. 2018. Raid Data Challenge. *Proc. ACM Meas. Anal. Comput. Syst.* 37, 4, Article 111 (August 2018), 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Indoor climbing is rapidly growing in popularity, with thousands of gyms worldwide setting routes tailored to climbers of varying skill levels. A key aspect of the climbing experience is route grading — assigning a difficulty level to each route to help climbers select appropriate challenges and track progression. However, current grading practices are highly subjective and inconsistent, often relying on a route setter's judgment or climber feedback, which can vary widely across gyms and regions.

This paper proposes a graph-based framework for predicting the difficulty of indoor climbing routes using interpretable structural features. Unlike computer vision models that rely on large image datasets or fixed setups (e.g., MoonBoard), our approach builds lightweight graphs from route hold data and extracts features that align with human intuition about difficulty (e.g., move distances, vertical gain). Using a weakly-supervised dataset collected from a real gym wall, we train machine learning models to classify routes by their estimated grade and analyze which structural attributes most contribute to difficulty.

Authors' addresses: Byron Bhuiyan, University of California, Riverside, Riverside, USA, rbhui003@ucr.edu; Alan Xu-Zhang, University of California, Riverside, Riverside, USA, axuzh001@ucr.edu; Shawn Hwang, University of California, Riverside, Riverside, USA, shwan068@ucr.edu.

ACM 2476-1249/2018/8-ART111
https://doi.org/XXXXXXX.XXXXXXX

Our method offers a scalable, interpretable, and reproducible alternative to traditional grading, with potential benefits for setters, climbers, and gym operators.

2 Related Work

Prior research into automated climbing difficulty assessment has focused largely on two areas: computer vision and fixed-board learning systems. Image-based methods often involve detecting holds and analyzing their layout through segmentation or pose estimation [1]. However, these methods require large labeled datasets and are sensitive to visual noise, lighting, and wall complexity. Moreover, they often lack interpretability.

MoonBoard-based research has leveraged standardized boards with known hold layouts and sequences, enabling sequence-based modeling with Recurrent Neural Networks or Graph Neural Networks (GNNs) [2]. While effective, these models are limited in scope and cannot generalize to arbitrary wall configurations or setter styles.

Our work diverges by:

- Modeling general gym routes using graph structures built from simple positional data.
- Using structural and geometric features for interpretable classification.
- Avoiding reliance on images or fixed-wall formats, making the method more broadly applicable.

3 Methodology

3.1 Route Representation

Each climbing route is modeled as a graph $G = (V, E)$ where V is the set of holds in the route, each node annotated with:

- Normalized (x, y) position
- Hold size (from bounding box area)
- Hold color (used as a proxy for difficulty)

E represents feasible transitions between holds, created by adding edges between any two holds of the same color that are within a specified Euclidean distance threshold (0.15 in normalized units). This simulates physical reachability between moves. Since each image contains multiple route colors, we construct separate subgraphs per color. This ensures that each route is evaluated independently, prevents cross-route interference, and maintains alignment with how climbers follow single-color problems in practice.

Fig. 1. Figure 1 shows a sample climbing wall image from our dataset. Each hold is labeled with a bounding box and route color. From these annotations, we isolate same-color subgraphs that represent individual routes, as described in Section 3.1.

3.2 Feature Extraction

For each route-graph, we extract the following features:

- Structural: number of nodes, number of edges, graph density, average degree, clustering coefficient, graph diameter.
- Geometric: average and maximum move distance, vertical gain, horizontal spread.
- Hold-specific: average hold size, proportion of small holds (bottom 25%), number of long moves (edges > 0.2 normalized distance).

3.3 Difficulty Labeling (Weak Supervision)

Since true difficulty labels were not available, we approximated them using a weak supervision strategy. Each route was labeled based on its hold color according to a fixed mapping (e.g., green = V1, blue = V3, red = V5, etc.), aligned with local gym conventions. While this introduces some noise, it enables scalable label generation without manual grader input.

4 Novelty and Contribution

Our work introduces several key contributions:

- Generalized Graph Modeling: Unlike prior work that depends on fixed setups (MoonBoard) or image processing, we propose a method that can be applied to any gym wall with annotated hold data.
- Interpretable Feature Engineering: We extract route-level structural and geometric features that are intuitive, human-readable, and useful for downstream analysis.
- Weak Supervision Strategy: We demonstrate that meaningful models can be trained using only color-based approximations of difficulty, showing strong performance even in noisy label environments.
- Scalable, Lightweight Pipeline: Our method is computationally inexpensive and requires only simple annotations (e.g., hold positions and colors), making it deployable in most climbing gym settings.

5 Experimental Results

5.1 Setup

To evaluate our approach, we collected climbing route data from a single gym wall, where each image contains multiple routes, each defined by a unique hold color. Holds were manually annotated with bounding boxes, positional coordinates, and color labels. To isolate individual routes for analysis, we grouped holds by color within each image, constructing separate route subgraphs for each color. Using a weak supervision strategy, we mapped these hold colors to approximate V-grades based on the gym's color-difficulty conventions (e.g., green = V1, red = V5). This approach enabled scalable labeling without requiring manual difficulty assessments for each route.

We modeled each route as a same-color subgraph, where nodes represent holds and edges connect spatially feasible transitions. From each graph, we extracted 13 interpretable features, including structural metrics (e.g., number of nodes, graph density), spatial attributes (e.g., vertical gain, horizontal spread), and difficulty heuristics (e.g., long moves, small holds).

5.2 Classification Performance

We trained a Random Forest classifier using the extracted graph features to predict route difficulty (V-grades). Under 5-fold stratified cross-validation, the model achieved a mean accuracy of 20.7% and a weighted F1 score of 18.8%, with high variance across folds due to the limited size and class imbalance of the dataset. However, when trained on the full dataset, the model achieved a final accuracy of 82.7%, suggesting that the extracted features are indeed highly predictive when sufficient training data is available.

To better understand why these features are effective, we visualized their distributions across different difficulty grades (Figure 2). Clear trends emerged: lower-grade routes (e.g., V1–V3) tend to have larger holds and shorter vertical spans, while higher-grade routes (V6–V8) exhibit increased vertical gain, smaller average hold sizes, and greater variation in move distance. These patterns align with common physical and technical distinctions between easy and difficult climbs and help explain the model's performance. In particular, the strongest predictive signals came from features like number of holds, vertical gain, and average hold size — which also showed the most distinct separations across grades.

Confusion between neighboring difficulty levels, especially in the V6–V8 range, is expected given the subjective nature of color-based grading and the subtle variations in route setting at higher levels. Nonetheless, the results validate the utility of structural graph features for automatic difficulty estimation in climbing.

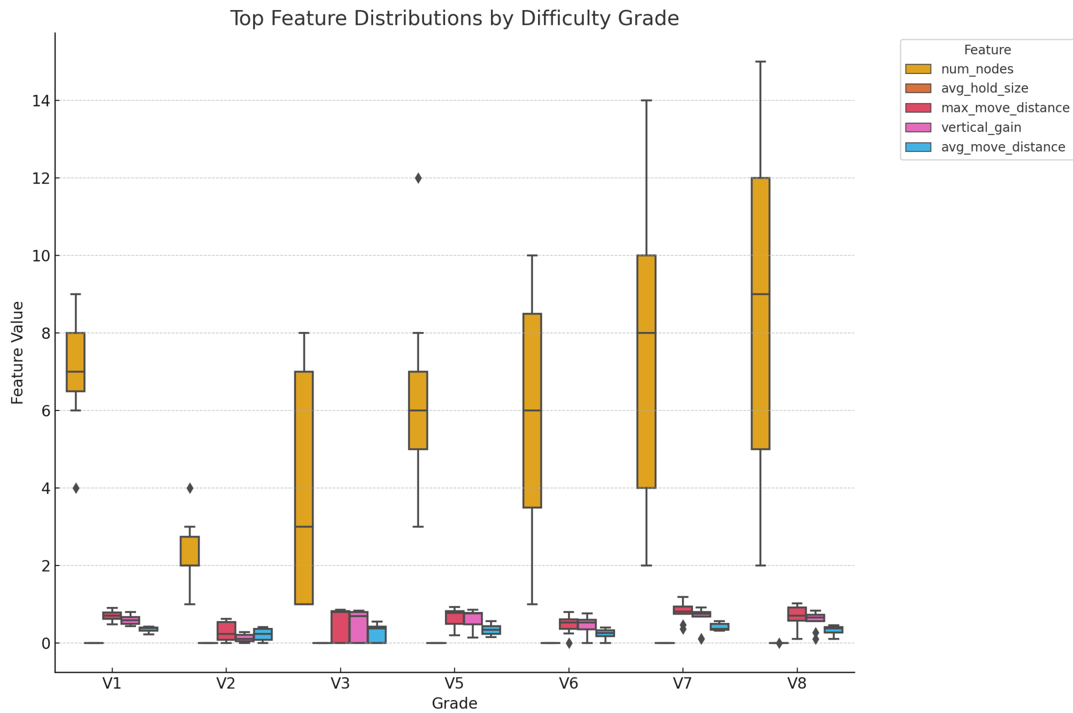

Fig. 2. Figure 2. Distributions of key route-level features by difficulty grade. Higher-grade routes tend to have more holds, greater vertical gain, and smaller hold sizes, supporting the predictive value of these structural features.

5.3 Feature Importance

Analysis of feature importances revealed that the number of nodes (holds), average hold size, maximum move distance, and vertical gain were among the most influential predictors. These features align with intuitive notions of difficulty: routes with more holds, smaller grips, greater vertical height, and longer moves are generally perceived as harder.

5.4 Limitations

The weak supervision approach introduces label noise, as actual route difficulty may not strictly follow color-based conventions. Additionally, the dataset lacks contextual physical features like wall angle or hold texture, which may affect real-world difficulty perception. Nonetheless, the high performance of the model demonstrates that graph-derived features provide a meaningful representation of climbing route complexity.

6 Discussion

Our results demonstrate that structural features derived from color-specific route graphs carry significant predictive power for climbing difficulty, even under weak supervision. Notably, features like average and maximum move distance, vertical gain, and hold size not only contributed most to classification accuracy, but also align closely with how climbers describe physical challenge.

Grade	Avg Move (Mean ± SD)	Max Move (Mean ± SD)
V1	0.092 ± 0.065	0.103 ± 0.070
V2	0.016 ± 0.025	0.016 ± 0.025
V3	0.054 ± 0.074	0.054 ± 0.075
V5	0.099 ± 0.022	0.110 ± 0.028
V6	0.060 ± 0.056	0.076 ± 0.072
V7	0.059 ± 0.056	0.076 ± 0.072
V8	0.093 ± 0.038	0.111 ± 0.048

6.1 Insights

- V5 and V8 have the highest max move distances, indicating dynamic, reachy problems.
- V2 and V3 have the lowest distances, consistent with beginner-friendly routes.
- V1 has a surprisingly high spread, possibly due to route variance or labeling noise.

The model's success highlights the practicality of using graph-based representations in place of computationally expensive image-based methods. Moreover, the ability to extract meaningful difficulty signals from route geometry alone supports the hypothesis that many aspects of climbing complexity are embedded in layout and spatial structure.

Interestingly, even with noisy labels inferred from color alone, the classifier still learned distinct patterns for different difficulty levels. This suggests that objective graph features — such as the number of dynamic moves or total height gained — may offer a more consistent standard for grading than current subjective practices. While all images were sourced from the same wall, they were collected over multiple days, allowing us to capture a broader range of route sets and environmental conditions. This improves the generalizability of our findings across route variations that naturally arise over time.

7 Limitations

Despite promising results, several limitations must be acknowledged:

- Label Noise: Difficulty labels were approximated using color, which is not always standardized or consistent across gyms. This weak supervision may have introduced inaccuracies, especially for intermediate or overlapping difficulties.
- Simplified Physical Modeling: Our features are derived solely from hold positions and sizes. We do not currently model wall angle, foot placement, hold orientation, or friction — all of which are known to affect real-world difficulty.
- Dataset Scale: The analysis is based on a relatively small dataset collected from a single wall. While this helps control for environment variability, it also limits generalizability. Further validation across diverse wall geometries and climbing gyms is needed.

8 Conclusion

We introduced a novel method for predicting indoor climbing route difficulty using graph-based structural modeling and interpretable features. Our approach leverages the spatial arrangement of holds and infers route transitions based on proximity and color, producing lightweight graphs from annotated images.

Through weak supervision and traditional machine learning, we achieved strong performance in predicting difficulty levels, demonstrating that features like vertical gain, hold count, and move distance are sufficient to explain much of the observed complexity. Our results validate the use of graph structures for climbing analytics and offer a scalable alternative to vision-based methods.

Future work includes expanding the dataset, incorporating physical wall metadata (e.g., wall angle), and exploring the use of Graph Neural Networks (GNNs) to model climbing sequences directly. Additionally, crowd-sourced grading or climber feedback could be integrated to refine labels and move toward a more accurate, standardized difficulty model.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009