Transforming MoonBoard Climbing Route Classification

Joshua Grant, Michael Kirkton, Aiden Miller, Aydin Ruppe, Benjamin Weber, and Ryan Kruk

Department of Electrical Engineering and Computer Science

Milwaukee School of Engineering

{grantj, kirktonm, milleraa, ruppea, weberbw, krukr}@msoe.edu

March 18, 2023

Abstract

This study endeavors to showcase the viability of leveraging machine learning to automate the procedure of rating rock climbing routes. The focus of this investigation lies in the MoonBoard, a globally standardized rock climbing wall, which simplifies the challenge of locating relevant data and reduces the inherent complexity involved in climbing. Our proposed methodology involves utilizing attention-based models to classify routes. In our experiments, we employed an encoder predicated on the transformer architecture, and have thus far attained an accuracy of 48.8%, alongside a ± 1 accuracy of 85.3%. Our empirical findings propose that deep learning models exhibit the potential to predict difficulty ratings more effectively than humans, thereby opening avenues for future work in route generation and bouldering training via data-driven approaches.

1 Introduction

The emergence of machine learning in sports has witnessed a tremendous surge in its application, ranging from forecasting game outcomes to analyzing player performance. However, the utilization of machine learning in climbing remains largely unexplored. In the realm of rock climbing, route setting represents a daunting and highly subjective task that necessitates expertise and proficiency. Currently, experienced climbers manually set the routes, while the difficulty rating of the routes remains highly subjective, as it is contingent upon individual factors such as strength, height, and experience. This subjectivity poses a significant challenge in establishing an objective rating standard for climbing problems and accurately predicting ratings based on data. Furthermore, the complexity of climbing problems and the limited availability of data compounds the difficulty of automating route setting and difficulty rating. By leveraging the MoonBoard, a standardized rock-climbing wall utilized globally in gyms, we can make several of the issues above obsolete.

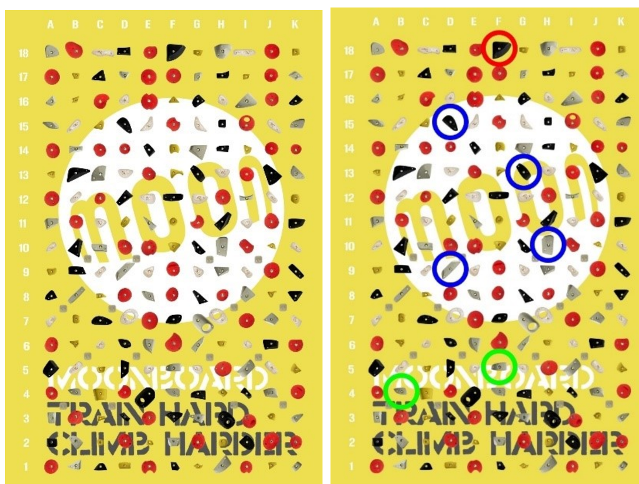


Figure 1: The 2017 MoonBoard's base configuration is on the left, with a route displayed on the right. The circles in green are the start holds, blue are the intermediate holds, and red is the end hold

The MoonBoard, a standardized training board with a fixed set of holds, has gained widespread popularity among climbers for enhancing their skills and strength. It offers three distinct variations, namely the 2016 MoonBoard, the 2017 MoonBoard, and the 2019 MoonBoard. The MoonBoard app has garnered a substantial community of climbers who have developed routes and labeled their difficulty levels, thus providing us with an abundant dataset to train a machine learning model. We utilized data from the 2017 MoonBoard as it presented the largest collection of routes at our disposal.

Several past models have represented the MoonBoard as an 11x18 matrix, with each cell indicating a hold's position. Initially, we adopted this approach; however, we observed that the information available to the learning model was insufficient, and models encountered challenges in establishing connections between distinct holds. The sparse nature of features in the matrix meant that CNN models struggled to extract features, while classical models found it challenging to establish associations between the features. Drawing upon previous model types, data representation methods, and their success (detailed further in section II), we propose leveraging a transformer-based model, entirely composed of attention mechanisms, to classify routes represented as a sequence, rather than a matrix, on the MoonBoard. The ability to classify routes presents practical applications for route setters and climbers, assisting with route setting and enabling climbers to choose routes that align with their skill levels.

2 Related Works

In recent years, there has been a growing interest in using machine learning techniques for climbing route classification on the MoonBoard. Dobles et al. [2] evaluated three different types of models, including Naive Bayes, Softmax Regression Classifiers, and a CNN to classify MoonBoard routes. They represented each route as an 11x18 matrix, and as previously mentioned, they struggled to produce satisfactory results. The use of stratified data and weighted training to combat class imbalance failed to improve their results.

In contrast, Duh and Chang [3] proposed that a sequence model is a more natural representation of a MoonBoard problem than the 11x18 matrix. They preprocessed their data into a list of vectors that represented a sequence of left and right-hand moves, combined with various other details about each hold. Their main model, GradeNet, is an RNN, specifically an LSTM, with an accuracy of 47.5% and a ± 1 accuracy of 84.8%. In a similar vein, Tai et al. [6] used a GCN with attention mechanisms to classify rock climbing difficulties. They used oversampling and undersampling to limit class imbalance and achieved an accuracy of 21.9%, with a ± 1 accuracy of 56.3%, which is on par with other models that used the matrix as a form of data representation.

3 Methodology

Our study extends prior research by utilizing a transformer-based encoder model, in contrast to LSTM models. The emergence of transformers in 2017 through the publication of "Attention is All You Need" [7] has led to their widespread use in deep learning. This model type has proven to be highly effective in addressing sequence-based problems, due to its self and global attention mechanisms, efficient parallelization, and rapid training. Given its success in natural language processing tasks, we opted for this model type to handle the processing of our route sequence data.

num_layers	3
d_model	128
dff	736
num_heads	8
dropout_rate	0.2
warmup_steps	3250
beta_1	0.79
beta_2	0.98
epsilon	6.3581e-08

Table 1: Hyperparamters

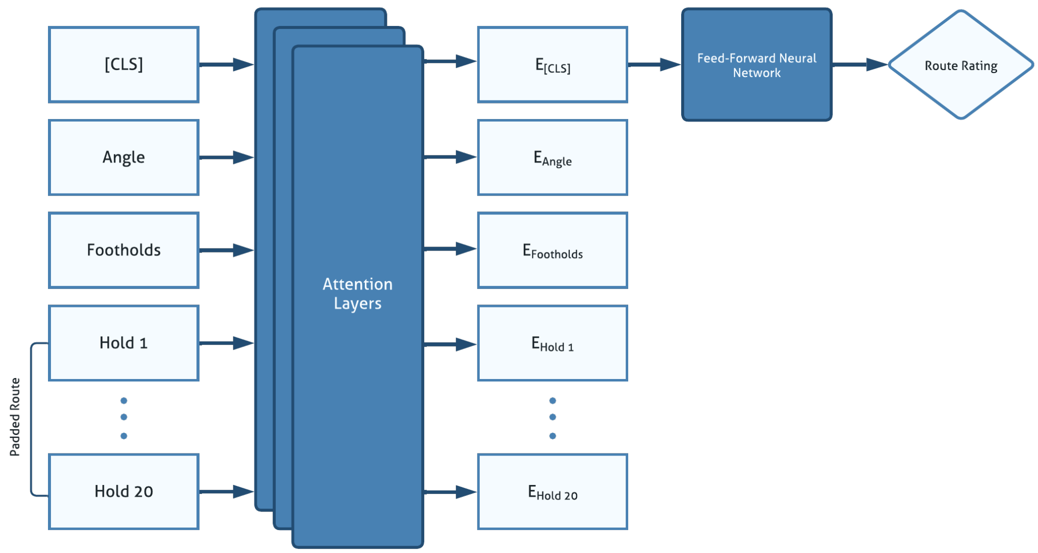


Figure 2: Model Architecture

3.1 Sequence

The sequence used in our model consisted of three distinct components. The first component was a singular token that denoted the angle of the MoonBoard. The second component was a token that captured the footholds that were available for use during the climbing route. Finally, the sequence was composed of a series of tokens that captured the specific holds present along the route. To ensure that the sequence was presented in a logical and comprehensible order, we applied a heuristic approach that approximated the order in which a climber would tackle the route. This approach was implemented with the aim of enhancing the overall effectiveness of the sequence by incorporating a degree of human insight into its design.

3.2 Class Token

As part of our model design, we implemented a class token inspired by BERT [1]. The class token is a special token that is prepended to the input sequence and represents the classification label of the input. The class token allows the model to incorporate the classification label into its training and prediction processes. This allows our model to use an attention-based architecture for the use of classifying the routes rating by running inference on the output of the attention encoding on the class token only.

4 Dataset

The class imbalance problem inherent in the dataset was addressed through a combination of weighted training as well as oversampling and undersampling techniques during the training process. Specifically, the benchmark routes, which were considered to be high-quality problems defined by expert climbers at MoonBoard, were oversampled at a higher rate than the other samples, and the lower-rated routes were undersampled to balance the distribution of ratings in the dataset. In addition, routes that had no repeated ascents were removed, as they were considered to have low quality, and routes with ratings ranging from V11 to V14 were also excluded due to the lack of accurately rated data points in those classes. The final grade distribution is shown in Appendix A2. The remaining 21695 problems were divided into training, validation, and test splits, with 17356, 2169, and 2170 problems, respectively before sampling.

4.1 Default Rating

A significant finding was that a subset of the MoonBoard database contained a high degree of error. Notably, the default rating used when a user creates a route is 6B+ on the font scale [5]. However, any user who is just playing around with the route creator or does not end up changing the rating creates a route with an improper classification [4]. Since both 6B and 6B+ from the font scale correspond to V4 on the V-scale, all 6B+ routes were removed from our dataset due to the inability to verify their rating. We believe that this issue has not been addressed in any previous studies.

5 Results

We evaluated our model using Accuracy (the percentage of true predictions out of the dataset), ±1 accuracy (the percentage of predictions within one grade of the true rating), and ±2 Accuracy (predictions within two grades). Other models have used the F1 score (a measure of accuracy and precision from 0-1.0, higher is better), and the AUC (Area under the curve from 0-1.0, with values over 0.8 considered good), which were not used in training, but have been calculated (found in Table 1), both exceeding any previous results. We found that our encoder has so far achieved an accuracy of 48.8%, and a ±1 accuracy of 85.3%, which is a little better than GradeNet, the most accurate classifier among previous studies for the MoonBoard.

	HLP	Attention (Ours)	GradeNet (LSTM)	GCN	Naive RNN	MLP	CNN
Accuracy	45%	48.8%	47.5%	21.9%	34.7%	35.6%	34%
±1 Accuracy	87.5%	85.3%	84.8%	56.3%	-	74.5%	-
±2 Accuracy	-	97.2%	-	75.6%	-	88%	-
F1 Score	-	0.362	0.242	0.310	0.165	-	-
AUC	-	0.875	0.764	0.73	-	-	-

Table 2: Our models vs. previous models

Our model's performance is on par with the best-performing models created thus far. Our model clearly outperforms the Naive Bayes, Softmax Regression, and CNN models, as well the GCN model. While our model's performance only slightly exceeds the LSTM GradeNet model, it is important to note that these models were evaluated using the 2016 MoonBoard, not the 2017 MoonBoard that we used. To compare our models completely accurately with the others, we would need to train a copy of our model on the 2016 dataset as well as tune all the hyperparameters. Despite this, we believe the accuracy demonstrated by our model is enough to prove that transformers are competitive with the best-performing models published thus far and are a viable option for MoonBoard route classification.

6 Conclusion

In conclusion, we have shown that an attention-based model applied to the MoonBoard 2017 dataset, with a combination of weighted training, oversampling and undersampling techniques and removal of unverified and high difficulty routes, achieved an accuracy of 48.8% and a ± 1 accuracy of 85.3%. Our model outperformed previously explored models such as the Naive Bayes, CNN, and GCN models, and was slightly better than the GradeNet LSTM model.

Based on our findings, we suggest the adoption of attention-based models for MoonBoard climbing route classification, as they have the potential to offer improved accuracy and are more suitable for sequence modeling. Additionally, we recommend the continued implementation of sample weighting techniques, with a particular emphasis on benchmark routes. As a prospect for future research, we propose the investigation of route generation methods. Our model may be utilized to generate new climbing routes for the MoonBoard, an accomplishment that would be a valuable contribution to the climbing community. While we have commenced exploring the usage of attention mechanisms for route generation, the exploration of this calls for further research.

References

[1] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding — arxiv.org*. https://arxiv.org/abs/1810.04805. 2019.

[2] A. Dobles, J. C. Sarmiento, and P. Satterthwaite. *[PDF] Machine Learning Methods for Climbing Route Classification — Semantic Scholar — semanticscholar.org*. https://www.semanticscholar.org/paper/Machine-Learning-Methods-for-Climbing-Route-Dobles/dd282a66afd38b92698dd56ac812122403b1629a.

[3] Y.-S. Duh and J.-R. Chang. *Recurrent Neural Network for MoonBoard Climbing Route Classification and Generation — arxiv.org*. https://arxiv.org/abs/2102.01788.

[4] Remus Knowles. *Every Moonboard Problem Ever, Analysed — latticetraining.com*. https://latticetraining.com/2018/01/26/every-moonboard-problem-ever-analysed/.

[5] M. *Free Climbing Grades Conversion Chart + Complete Rating Guide — cruxrange.com*. https://www.cruxrange.com/blog/climbing-ratings-explained/.

[6] C.-H. Tai, A. Wu, and R. Hinojosa. *[PDF] Graph Neural Networks in Classifying Rock Climbing Difficulties — Semantic Scholar — semanticscholar.org*. https://www.semanticscholar.org/paper/Graph-Neural-Networks-in-Classifying-Rock-Climbing-Tai-Wu/77e2b78ed51b59558f9d48187a36bd56571553bd. 2020.

[7] A. Vaswani et al. *Attention Is All You Need — arxiv.org*. https://arxiv.org/abs/1706.03762. 2017.

A Appendix

A.1 Appendix A1: Human-level Performance

In a previous study conducted by Duh and Chang, the use of user-rated grades was found to be unfair as climbers generally adhere to the original grade, unless there is a significant discrepancy. As an alternative, the study sought to determine the accuracy of grades estimated by climbing experts who had not climbed the problems themselves. Specifically, three climbing experts were asked to estimate the grades of 40 climbing problems without actual climbing the route itself. The results demonstrated that even for experts, it is challenging to determine the grade without firsthand experience, with an accuracy of up to 87.5% within a tolerance of one grade off. The experts cited several reasons for the difficulty in accurate grading, including the challenge of grading without personal climbing experience, the suitability of problems for different body types, and the uncertainty when a problem falls between two grades.

	Accuracy	±1 Accuracy
Climbing Expert 1	47.6%	82.5%
Climbing Expert 1 (second try)	30%	77.5%
Climbing Expert 2	42.5%	87.5%
Climbing Expert 3	45%	87.5%
Estimated HLP	45.0%	87.5%

Table 3: Human level performance of estimating the grade of a MoonBoard problem without actually climbing it.

A.2 Appendix A2: Grade Distribution

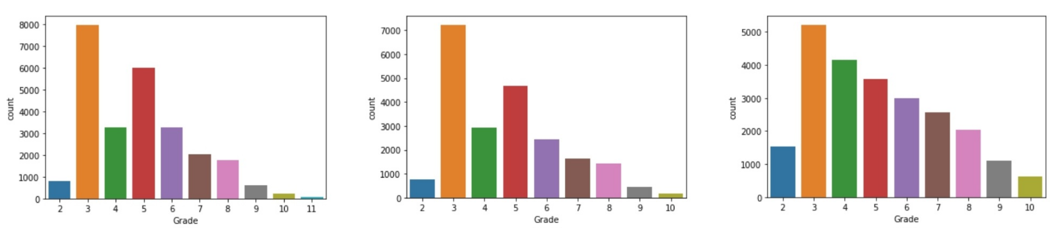


Figure 3: The distribution of our data. On the left is our initial dataset, in the center is after repeats have been removed, and on the right is after sampling has been performed.

A.3 Appendix A3: Training Plots

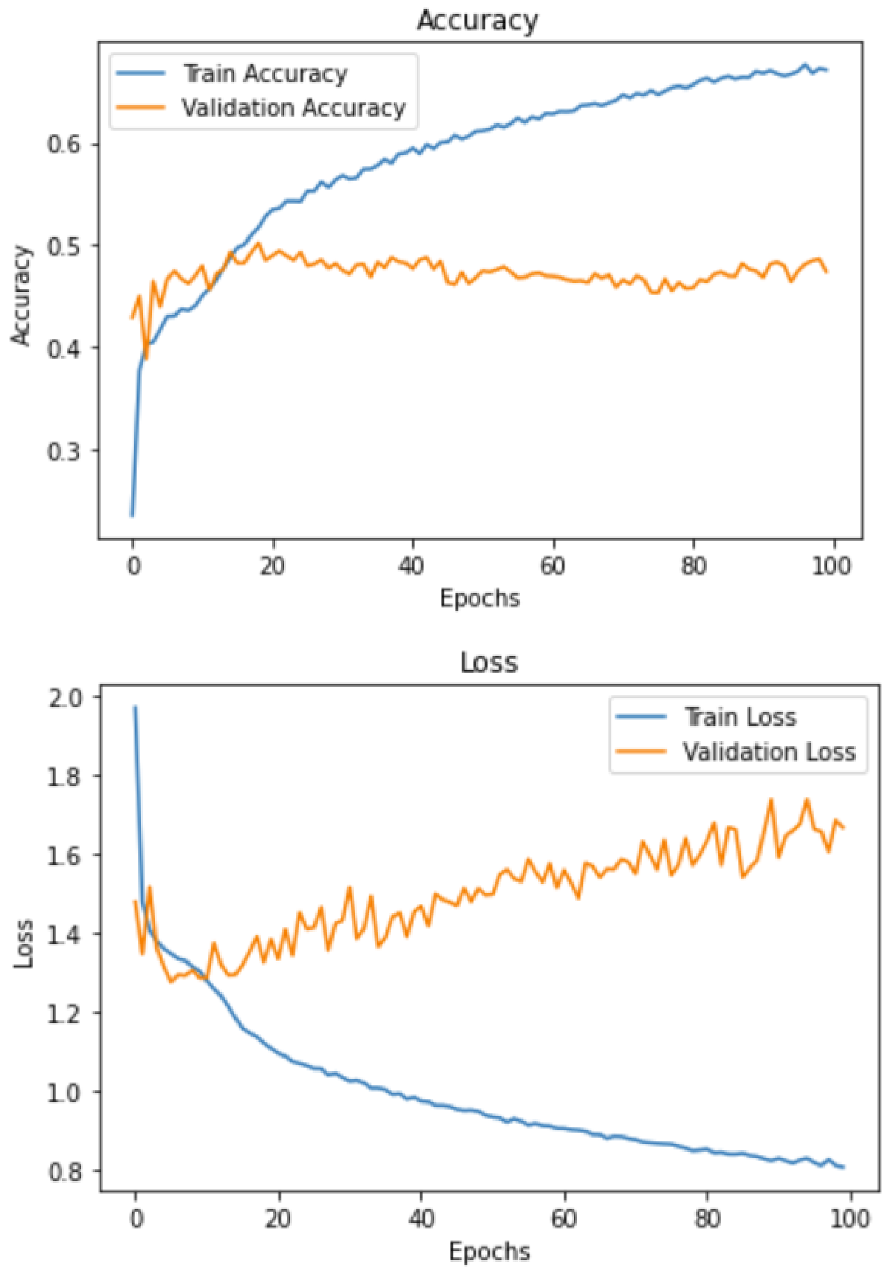


Figure 4: Training curves of the Attention model.

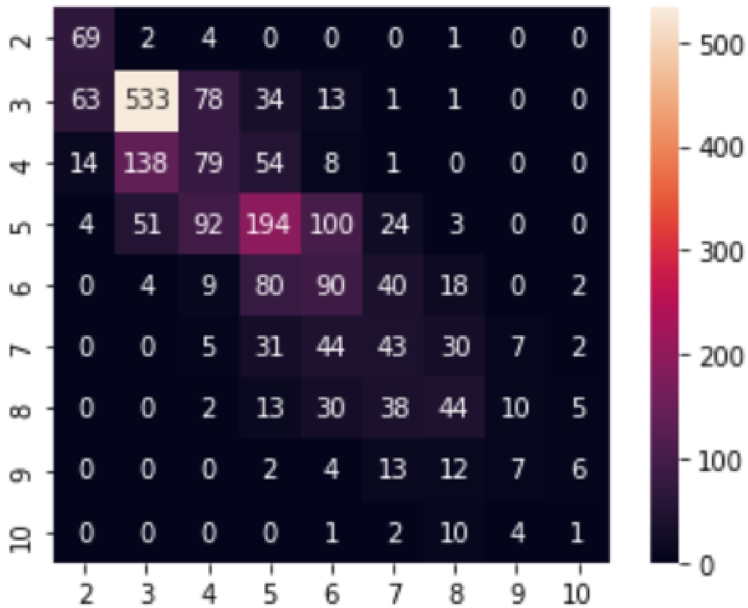
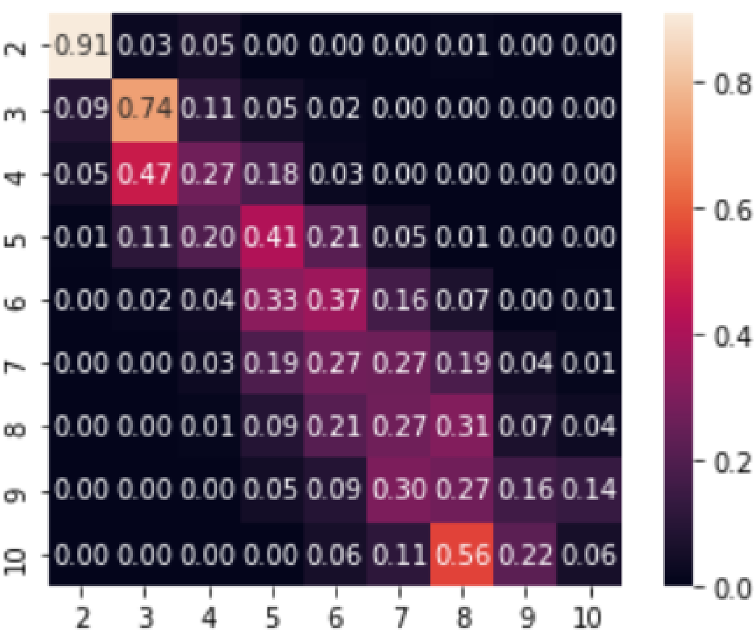


Figure 5: The confusion matrix of the Attention model. The predicted grade (x) and actual grade (y) are distributed along the main diagonal.



w

Figure 6: The normalized confusion matrix of Attention model. The predicted grade (x) and actual grade (y) are distributed along the main diagonal. Shows the percentage of each the predicted routes compared to the total number of routes belonging to that class.