Automated Grade Classification and Route Generation with Affordances on Climbing Training Boards

FRANK STAPEL, University of Twente, The Netherlands

Fig. 1. MoonBoard with the 2017 hold setup.

With the rise of training boards for climbing, research exploring the possibilities of classifying the grade of a route and generating new routes is becoming more popular. This thesis describes a framework using machine learning algorithms to generate training board routes tailored to individual climbers' action capabilities and training needs. An attempt is made to improve current grade classification methods by looking at routes as a series of moves. Moves are extracted from a route by a beta finder that compares potential moves in a new route to previously seen routes in a training set. The principles of ordinal regression are implemented in the grade prediction process to use the order in climbing grades. The grade classifier performs similarly to previous related work and human benchmarks with accuracies of respectively 46.5%, 46.7%, and 45%. However, the move-based structure of the grade classifier allows it to be used in the route generator. The route generator uses the reach, finger strength, power, and core strength of a climber to fit a generated route. The generated routes are tested with user studies. Even though some generated routes are of bad quality, other generated routes perform well and show potential for use in training. Unfortunately, the user studies are unreliable due to their small sample size and lack of variety in the demographic of participants. To use this method in practice, it is essential to improve the performance of grade classification, the process of gathering training board route data, and receiving detailed feedback on existing routes.

CCS Concepts: • **Human-centered computing** → *Interaction design*; • **Computing methodologies** → **Artificial intelligence**.

Additional Key Words and Phrases: climbing, climbing training board, MoonBoard, affordances, motion graphs, machine learning, domain-specific language, route generation, difficulty assessment

Author's address: Frank Stapel, University of Twente, Enschede, The Netherlands, research@frankstapel.nl.

1 INTRODUCTION

On the 3rd of August in 2016, the International Olympic Committee announced that climbing would appear as an Olympic sport for the first time at the Olympic games of 2020 in Tokyo [1, 2]. Additionally, the next two Olympic Games have been confirmed to add climbing to their program as well [3, 4]. Climbing being admitted to the Olympics is coherent with the sport's rapid growth and confirms its popularity with younger audiences.

As the sport grows and the best climbers get more competitive, there is an increased interest in effectively training for climbing. Because regular climbing walls are expensive, take up a lot of space, and offer relatively few climbing routes, most climbing and bouldering gyms offer a form of *training boards* to climb on. A training board is a small board, usually a few meters high and wide, filled with various climbing holds and is described as one of the best devices for strength training [23]. Unlike regular climbing walls where a climber follows a climbing route marked with a certain color, routes on a training board are created by taking a specific subset of the climbing holds to reach the top. This structure leads to the possibility of a training board containing thousands of routes, making it a cost-effective way to train for climbing.

This work looks into the classification and generation of climbing routes on training boards and their link with the action capabilities of a climber. This link could be used to provide training regimens on the spot tailored to climbers' unique qualities. The goal is to answer the following research question: *How can machine learning algorithms automatically generate training board routes tailored to individual climbers' action capabilities and training needs?*

1.1 Background

A few phenomena in the climbing world are explained and defined to clarify the rest of this thesis.

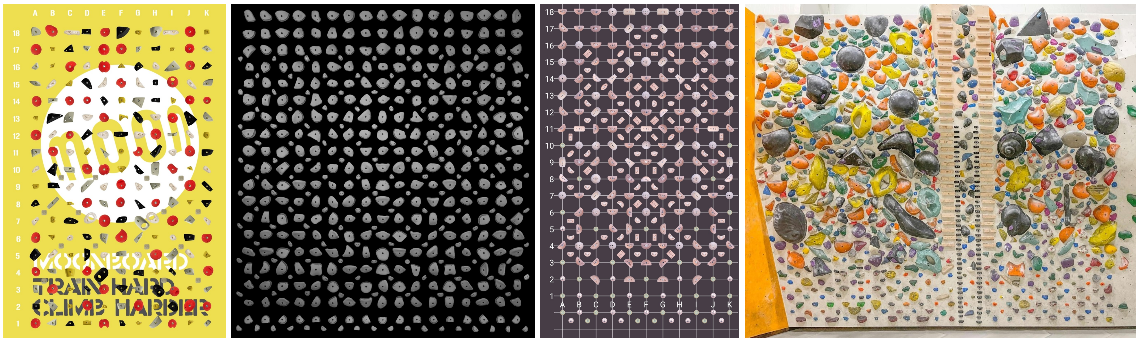
1.1.1 Training Boards. The first training boards originated in 1980 to add to the limited existing climbing training options of going rock climbing and doing pull-ups on a door frame [23]. In the following years, their popularity increased, causing more climbers and climbing gyms to build their own training boards. This has led to some training boards being innovated into the modern human-computer interaction devices they are now. Here routes are stored in mobile applications connected via Bluetooth to LEDs in the training board, marking which climbing holds can be used in a route. Climbers can add their own routes to these applications to contribute to the global route database of that specific board layout. This functions well because these layouts are standardized and sold by certain companies, allowing people all around the world to climb on the same training board with the same database of climbing routes.

Figure 2.a shows the oldest and best-known LED board, the MoonBoard [40]. Spanning the three hold layouts and both options for the angle of the wall (25°and 40°), the MoonBoard contains over 80,000 routes. Some of these routes are called *Benchmarks*, routes with curated quality and a grade confirmed by the MoonBoard team.

Figure 2.b depicts the Kilter Board, which innovates on the MoonBoard by placing the LEDs inside of partially transparent holds, making the usable holds visible during climbing instead of only from the ground [31]. The Kilter Board is roughly twice as large as the MoonBoard and contains ergonomically shaped holds.

Figure 2.c shows the Tension Board, an LED board with wooden holds [12]. The Tension Board is symmetric, meaning a route can be climbed on both sides for symmetrical training. Beastmaker offers wooden holds as well, making training easier on the skin of a climber their fingers [5].

Figure 2.d shows a custom training board in Enschede. To allow the climbing community to save and share routes easily, the application from Stōkt uses computer vision to outline the different climbing holds in the application [52].



(a) MoonBoard (b) Kilter Board (c) Tension Board (d) Regular spray wall

Fig. 2. Multiple training boards.

1.1.2 Difficulty Grades. The difficulty of a training board route is indicated by its grade. There are multiple grade scales, which are used for bouldering and lead climbing [29]. Previous works use the V bouldering scale. The easiest boulders are rated VB, after which difficulty increases with grades ranging from V0 to V17. This research uses the Fontainebleau bouldering scale. Each grade consists of a number, followed by the letter A, B, or C. A "+" denotes a grade between letter grades, so 6C+ is followed in difficulty by 7A. Font grades currently range from 3A to 9A. This means there are 37 grades in the Font scale and 19 in the V scale, making the Font scale more specific.

Draper et al. propose the IRCRA scale to bridge the gap between research related to climbing grades. Conversion from the Fontainebleau scale can be done using their provided table [19].

The grade of a training board route is determined by the consensus of climbers who completed the route. This means that a route with more ascents has a more reliable grade. Research shows that climbers can generally assess the difficulty of a route, where assessment becomes more accurate for difficult routes [15, 18]. Deviation in these difficulty assessments can be attributed to climbers giving routes that fit their strengths an easier evaluation.

Scarff proposes the Whole-History Rating (WHR) system as an objective approach to determining the difficulty of a route [13, 47]. Drummond and Popinga implement this Bradley-Terry model described using statistical analysis to determine the logarithmic grade increase with respect to difficulty [8, 20].

All current grading systems do not take *affordances* (see section 2.1) into account, which place the action capabilities of a climber in the context of a climbing route [19, 25, 56]. For example, the grade of a route with distant moves is the same for short and tall climbers, regardless of the difference in perceived difficulty.

Even though there is a general consensus on how difficult a climb of a certain grade should be, there are notable differences in the grading of routes on different training boards. For example, a 7B on a MoonBoard is considered to be more difficult than a 7B on a Kilter Board. The cause of this phenomenon is not known but could be attributed to a bias in the grade of the first routes on these boards, such as the Benchmarks on the MoonBoard. As the grade of a new route is often based on similar routes on the same board, this bias is probably copied as well.

1.1.3 Training for Climbing. Training boards mostly focus on a dynamic style of movement and improving upper-body, and finger strength, which are found to be the most important aspects of training for climbing [23, 32, 45, 60]. Additionally, training boards allow for the practice context to be similar to the performance context of a climber. This in-situ context is described to promote learning and transfer by Brunswik [10]. Other options with a high skill transfer, such as climbing indoors on regular climbing walls and outdoors on rocks, are less efficient in their cost and use of space.

Training for climbing differentiates strength, power, short endurance, and long endurance training to classify different training exercises [6, 23]. Most training board routes train either power or short endurance, but by climbing multiple routes in quick succession, long endurance can also be trained. Despite the availability of extensive periodization programs for training, finding the right route for certain training remains difficult.

2 RELATED WORK

Climbing is a relatively new field of research where most studies explore the physiological aspect of climbing [49, 60]. Nevertheless, the last decade shows the field widening, covering topics such as the psychology of climbers and the kinematics of climbing, but also grade classification and route generation [39, 61]. This section explores affordances and action capabilities for climbing on training boards and evaluates existing works related to route generation and grade classification.

2.1 Affordances

The theory of *affordances* was developed by Gibson to describe the relation between animals and their environment. Gibson describes affordances himself as follows: "The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill" [25]. Warren interpreted this with relation to action capabilities: "An affordance is the functional utility of an object for an animal with certain action capabilities" [59]. Here the difference between action capabilities and affordances is key. An action capability only tells something about the animal, whereas an affordance gives information about the animal with certain action capabilities in relation to its environment.

In the context of climbing, the animal can be interpreted as the climber and the environment as the climbing wall. An example of an action capability is how far the climber can reach with their arm. When placing the climber in the context of the climbing wall, the affordance of the climber can be described as whether or not they can reach the next hold. Here, the action capability of reaching is placed in the context of a climbing wall by reaching for the next hold, creating an affordance.

This affordance of how far someone can reach on a wall gets noticed most often by beginning climbers. The reach is determined by multiple action capabilities, foremost by a climber's height and arm span. Measuring the relative arm span with respect to height is done using the *ape index* [26, 43]. Being able to reach further on a wall from a certain position can make it easier to grab the next hold. However, narrow and bunched-up climbs tend to be more difficult for long climbers [23].

The size of a climber's hands can play a big role in the perceived difficulty of grabbing small holds. For example, a hold that a climber with large fingers can only fit 3 fingers onto, a climber with small fingers can grab with their full hand, enabling them to put more pressure on the hold. The depth of a hold determines how many phalanges a climber can use to grab it, where the ability to use more phalanges leads to a better grip [43].

The strengths of a climber translate directly to their affordances on a climbing wall. Finger strength is an important factor in controlling holds, allowing climbers to hold onto smaller and more sloping holds [14, 24, 45]. Finger strength can be measured in multiple ways. The most common way is to test the maximal total loading for hanging off a fixed-size edge for a fixed amount of time. Lattice Training proposes to test the one-repetition maximum load by hanging for 7 seconds of a 20mm edge [55]. By calculating the load with respect to body weight as shown in eq. (1) a relative measure

Table 1. Scoring of core strength test

Score	Exercise	Duration (seconds)
1	L-sit (bend knees)	10
2	L-sit (bend knees)	20
3	L-sit (bend knees)	30
4	L-sit	10
5	L-sit	20
6	L-sit	30
7	Front lever	5
8	Front lever	10
9	Front lever	20
10	Front lever	30

is created called the maximal body weight % (MBW%). This can be used for a good indication of the action capabilities of a climber regarding finger strength.

$$\text{maximal body weight } \% = \frac{\text{total load}}{\text{body weight}} * 100\% \tag{1}$$

Mobråten and Christophersen use the same measure to classify what grade a climber can climb based on their strengths [38]. This test considers the maximum load of one repetition of a pull-up. Measuring upper body strength with pull-ups can be done for multiple repetitions as well, focusing more on power endurance instead of maximal strength. Additionally, hang time from a bar and core strength are tested as well. As hang time usually goes into the minutes, it is less relevant for climbing on training boards. Core strength is necessary for climbers to keep their feet on the board. This is measured by how long a static position of an L-sit with bent legs, an L-sit with straight legs, or a front lever can be held. The scoring of the core strength test is shown in table 1.

The tests above measure the performance of basic muscle groups. Nick et al. created the *powerslap* test, which measures how high a climber can reach in a dynamic footless movement from two designated starting holds [41]. This test is relevant to training board climbing because the dynamic character of moves on a training board is measured well. Unlike the tests above, the powerslap test results do not take the height or reach of a climber into account, meaning the result of the test alone can be hard to interpret. Combining the result with these other physical attributes of a climber can be useful as a measurement. Unfortunately, the test can not be done with regular climbing gym equipment, limiting its use as a simple strength test.

Training board routes generally focus on finger, upper body, and core strength as the moves are primarily dynamic and explosive [23, 38, 51]. A climbing route's difficulty is determined by many factors, most of which relate to these action capabilities. Stapel describes that a climbing route will become more difficult if holds are included that are harder to hold onto, causing more finger strength to be necessary to stay in the route. Increasing the distance between holds makes the route more difficult because more upper body and core strength is required to complete the move. Another factor is the availability of footholds. Worse options for foot placement lead to worse positions on the board which to more pressure on the arms of a climber and require more strength to compensate [50].

ClimbAX uses wearables to analyze climbing performance, measuring power, control, stability, and speed. A test on 53 climbers in a competition setting reported a strong correlation between these parameters and climbing performance [34].

These links between these measurable strengths and the perceived difficulty of climbing routes show the importance of taking a climber's action capabilities and affordances into account within route generation. This allows for routes that are tailored to the training requirements of climbers.

2.2 Grade Classification

The following two sections describe current solutions in grade classification and route generation, show current challenges, and establish general problems in the field.

A decade ago, Phillips et al. created a variation generator designed to support human setters in building interesting routes [44]. By developing a formalized system used to transcribe routes, a domain-specific language (DSL) was created that attempts to express the abstraction of the climbing domain [57]. This generator gave route setters a list of moves described in the DSL, to aid them in the building process. The exact placement of holds was still left to the route setters themselves. This became the first step toward fully automatizing climbing route generation.

Kempen used the DSL described by Phillips et al. in classifying the difficulty of climbing routes by training a variable-order Markov model (VOMM). This model differentiated climbs into the two categories "easy" and "hard", which achieved an accuracy not much higher than a trivial classifier. This showed the issue of missing relevant information in the DSL such as distance between holds and a more accurate difficulty assessment of individual holds.

Dobles et al. were the first to use machine learning in the classification of MoonBoard climbing routes [17]. Based on the 13,871 available routes at that time Naive Bayes, softmax regression, and a convolutional neural network (CNN) were applied and compared. All models performed comparably at approximately 35% top-1 accuracy. Dobles et al. also describes the frequency at which holds are used for a certain difficulty, showing a clear difference between holds used in easy and difficult routes. Training performed a similar analysis, creating a map with for each hold the average difficulty of climbing routes that include that hold. This map confirmed the spread of difficulties and also shows that certain places on the board are used more often than others, for example, the middle columns and the top row see an increased frequency in usage.

Ebert et al. used a sensor-based approach in an attempt to classify the difficulty of a route by measuring asynchronous non-recurrent motions of a climber [22]. Unfortunately, the tests were performed on mainly inexperienced climbers. By limiting the results to the easiest grades 1a-5c, there is no information on intermediate and difficult climbs. These limited tests make it hard to compare the achieved results to other works.

Tai et al. apply a Graph Convolutional Network (GCN) to routes on the MoonBoard, resulting in an AUC of 0.73 [53]. The authors found that the GCN is less susceptible to data imbalance, resulting in better results on the skewed dataset than other models had achieved in the past.

Duh and Chang developed a pipeline that can assess and generate routes called *BetaMove* [21]. The grade classifier, called GradeNet, reaches an accuracy of 46.7% and an accuracy of 84.7% when counting classifications that are only one grade off as well. For example, this would also count a grade of 6C+ to be correct when the grade is in fact 6C. These scores reach roughly the same performance as humans (45.0% and 87.5%) in classifying grades. The main reason this model outperforms other models is the pre-processing of the routes that is done before applying the model. A move sequence is created as input for the machine learning models by figuring out in what order the holds are used. This move sequence is also called the *beta* of a route. The beta is found by performing a beam search over scores based on the sum of two two-dimensional Gaussian functions representing the position of a hold with respect to the last hold [46]. As this move sequence representation approaches a realistic, abstract form, patterns in the data that correlate with the difficulty of a route become clearer.

Breunig et al. created the outlier detection method Local Outlier Factors LOF that gives a score based on how isolated a data point is compared to its neighbors [9]. Kriegel et al. build on LOF by outputting the probability a data point is an outlier instead of a difficult-to-interpret number [33]. These probabilities are called Local Outlier Probabilities (LoOP). This method could improve the currently used sum of two two-dimensional Gaussian functions to score the location of the next hold in BetaMove, because it removes the authors' bias of creating two two-dimensional Gaussian functions by taking this data-driven approach of comparing a new move to previously seen moves.

Difficulty grades are ordered classes. Using this aspect remains unexplored but could prove beneficial. Cheng et al. show how training a model that recognizes individual dividing lines in the ordered data instead of predicting a probability for each class leads to a reduced number of decisions and an increase in prediction accuracy [11, 36].

The main problem with most current solutions is approaching the problem space as an image. These representations do not come close to the abstraction of climbing moves and therefore struggle to accurately predict the grade of a route. Duh and Chang has successfully taken the first steps in approaching the problem space as a Natural Language Processing (NLP) problem, closing in on a better representation of climbing moves. This trend is promising and should be continued by adding LoOP to the pre-processing of routes. Using the order in difficulty grades by implementing ordinal regression into the decision-making process looks promising as well.

2.3 Route Generation

After Phillips et al. used chaotic variations to generate route transcriptions, Stapel created the first algorithm to generate climbing routes for the 2017 MoonBoard setup [44, 50]. A route was generated by building a tree of possible moves matching the requested difficulty and move length. Move length could be adjusted with a parameter directly used when calculating the optimal distance between two holds, missing a defined or clear translation to the climber's length or wing span. Additionally, the quality of generated routes was found to be worse than existing MoonBoard Benchmarks by judging the flow and enjoyability of each route. The simple heuristics used in the generation were not advanced enough to encapsulate the important details in a climbing route.

Lo generated new routes for the 2017 MoonBoard setup by applying a variational autoencoder to the MoonBoard routes using a one-hot encoding in a matrix form [35]. Unfortunately, the quality assessment of generated routes in this work is limited to the author's opinion, who claimed the algorithm could generate high-quality routes. An attempt at user testing was done by selecting the 22 out of a total of 50 generated routes that passed the requirements set by the author. These problems were uploaded to the MoonBoard platform, awaiting user reviews. The result of the user study remains unknown.

BoulderBot uses machine learning to generate routes on custom training boards. The user first enters the holds on their available training board, indicating their location, orientation, and difficulty, as well as their own length, wing span, and the desired difficulty of the generated route [7]. The generated routes are presented in a user-friendly way, making the application well received by training board owners [1]. Sadly, details of the generative model are not public and therefore can not be used to build upon. Comparing the quality of generated routes to those of other route generators could be done by adding the MoonBoard to the application as a custom training board.

Seal and Seal generates boulders based on random points generated in a plane [48]. Generated routes depend on the *stride length* of a climber, which is a factor of 0.42 of the body height. The generated routes are logical but do not consider footholds, hand order, information about a hold, or climbing-related technique. Nevertheless, the model shows

[1]A Reddit post by the creator of BoulderBot showing positive reactions from the community: https://www.reddit.com/r/climbharder/comments/miitt8/i_trained_a_procedural_generation_model_to_create/ (accessed April 2022)

a good combination of human action capabilities and computer interaction with the direct link between stride length and moves on a climbing wall.

Duh and Chang produced the best-performing route generator thus far with BetaMove. By training a long short-term memory (LSTM) recurrent neural network (RNN) on pre-processed move orders instead of a raw one-hot encoded matrix [21]. The 2016 MoonBoard setup was used, which uses 29% fewer holds than the 2017 setup. The quality of generated routes was tested by subjecting 50 generated routes, 50 routes generated by the LSTM developed by Houghton et al. and the latest 50 routes submitted to the MoonBoard platform to the opinion of experienced climbers [28]. BetaMove generated more high-quality problems (80%) than both the latest MoonBoard problems (60%) and the routes generated by Houghton et al. (20%), showing that generated routes can outperform randomly chosen human-generated routes. Unfortunately, the routes were not tested against MoonBoard Benchmarks.

Other promising techniques for generating routes include transformers and Generative Adversarial Networks (GAN) [27, 58]. The latter has been expanded multiple times to take conditions into account in later works, showing potential for combining the generative model with the limitations posed by the affordances of a climber [16, 37, 42].

Over the years, the models for generating climbing routes became more advanced but follow the same trends as the grade classifiers. Initial solutions used computer vision techniques to generate new problems, looking at the MoonBoard problem space as an image. Later solutions become more advanced by emphasizing pre-processing and bringing the data closer to human movements by viewing the problem as an NLP problem. This trend is beneficial because it allows for a better representation of climbing movements.

The other large problem with most route generators is that they do not consider a climber's affordances, making it hard for the climber to find a route that fits their action capabilities and climbing style. If affordances were taken into account, climbers would be able to select routes that would fit their training program accurately. This allows for more effective training closer to the climber's limit, reckoning with the unique qualities of the climber.

3 AIMS AND OBSTACLES

Multiple obstacles become clear from related work. Current grade classification solutions view the problem space as an image. By using NLP techniques the models come closer to the abstraction of climbing moves. Improving pre-processing with the data-driven LoOP and implementing ordinal regression into the decision-making process are logical steps to improve the performance of grade classification. With a structured framework to compare the performance of different models next to each other, developing new models becomes more accessible as well.

Current route generation models struggle with the same problems as grade classification and would benefit from the improvements mentioned above. Additionally, it is clear that taking the action capabilities of a climber into account has the potential to improve the personalization of generated routes by adjusting moves to fit the climber's training needs.

This means that to answer the research question *"How can machine learning algorithms automatically generate training board routes tailored to individual climbers' action capabilities and training needs?"*, the following issues have to be addressed:

(1) What features can be extracted from climbing moves to be effectively used in machine learning models for grade classification and route generation?
(2) How can data-driven pre-processing and ordinal regression improve the performance of grade classification?
(3) Which specific action capabilities can be identified and utilized to generate personalized training board routes for individual climbers using machine learning algorithms?

To address this, a new framework for grade classification and route generation is described, looking into what features can be used and how LoOP and ordinal regression can be implemented. The accuracy of the predictions of the new grade classifier is compared to that of previous work and to human benchmarks. User studies test how action capabilities can be used to generate routes and how these generated routes are experienced by climbers. Finally, the results and feasible future improvements are discussed, allowing this work to be used as a stepping stone in future work as well.

4 CONTRIBUTION

Finding answers to the research question contributes to both science and society. A scientific contribution is made by improving machine learning solutions and forming a better understanding of the connection between the biomechanical foundation of climbing and abstract route representations. This understanding can be the foundation of future machine learning solutions that bridge the gap to 3D climbing routes with complex constraints, for example, generating a new route on a regular climbing wall with multiple planes.

By automating the process of flexibly creating training routes on the spot that are tailored to a climber's strengths and weaknesses, the climber's limits can be accurately pushed. By providing training regimens that consider an athlete's unique qualities, a societal contribution is made. A scientific contribution is made by showcasing how these training regimes can be personalized using machine learning algorithms.

Additionally, new boards can be filled rapidly with only a limited set of routes on the new board to indicate the difficulty of individual climbing holds. This contributes to the climbing community as a hurdle in developing new training boards is removed.

5 A NEW FRAMEWORK FOR GRADE CLASSIFICATION AND ROUTE GENERATION

This section describes the design of the data structure and the models used for grade classification and route generation.

5.1 Data Set

The routes from the 2016 MoonBoard setup at 40° are used for the grade classification model, which is trained on 75% of the dataset and its performance is tested on the remaining 25%. This setup is chosen such that the performance of the grade classifier can be compared to previous works. The route generation model is based on the 2017 setup, chosen due to its local availability for testing generated routes. As the grade classifier is used by the route generator, it is trained again on the 2017 setup. Routes from the Kilter and Tension Board are not used because retrieving that data is more difficult. Only a single set of routes is necessary to show the functionality of the models described in this thesis.

5.2 Grade Classification

As seen in section 2.2, pre-processing the routes is crucial to achieving an accurate grade classification. The process of predicting the grade of a route is shown in fig. 3. First, the beta of a route is determined. This beta is used to extract moves from a route that are scored individually. The individual move scores are combined into a grade prediction for the whole route.

Multiple models use the hold score of a route, which is calculated by taking the average grade of all routes containing that hold. Normalization is applied to move all scores to [0, 1], where 0 indicates the easiest hold and 1 the most difficult.

5.2.1 Beta Finder. The beta finder uses the position and scores of the holds in a route to find the easiest route. Additionally, moves are scored on their flow to encourage finding a logical beta. This is done by comparing a potential

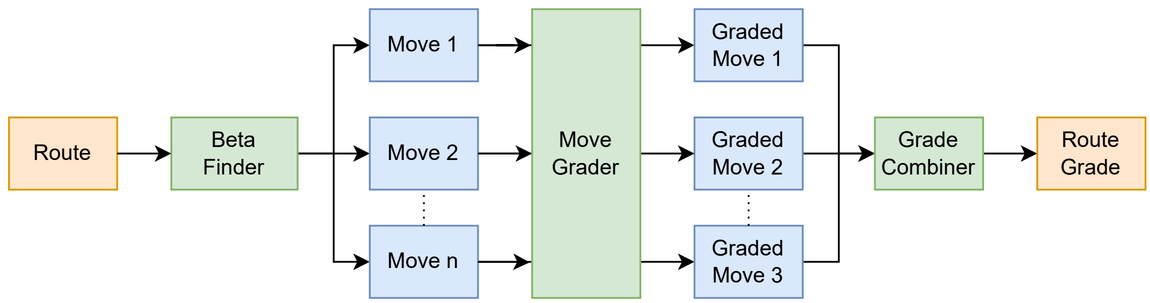


Fig. 3. The grade classification process of a route.

move to previously seen moves. The available dataset contains whole routes and no individual moves, so a dataset is created with manually labeled moves. The author labeled the routes by noting the beta of 250 routes, resulting in 1524 individual moves to which new moves are compared. Each move consists of three holds: the hold of the moving hand, the hold of the stationary hand, and the next hold as shown in fig. 4. All moves made with the left hand are mirrored to double the number of useful data points. This means that, for the model, the moving hand to the next hold is always the right hand.

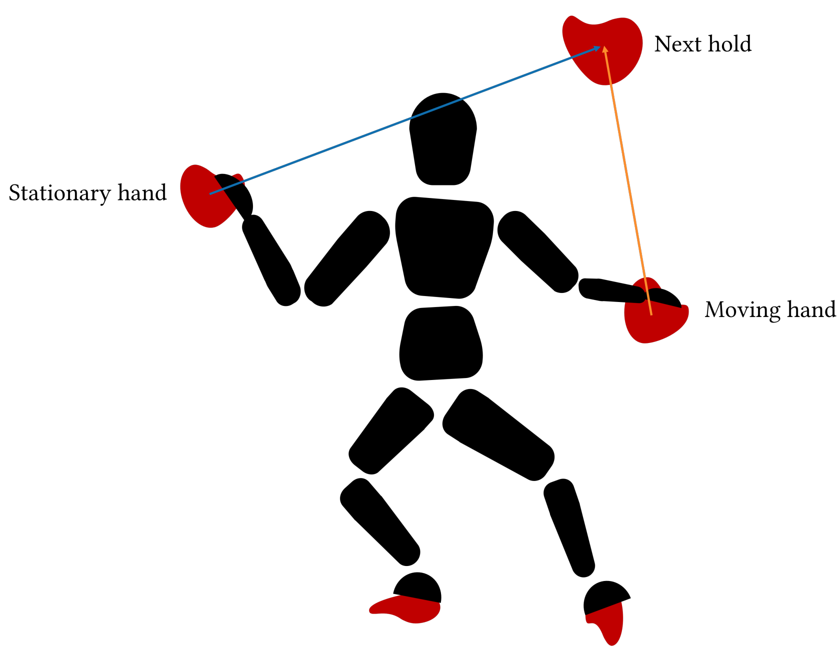


Fig. 4. The moving hand, stationary hand, and next hold. The orange arrow shows the moving hand's trajectory to the next hold while moving. The blue arrow denotes the distance between the hold held by the stationary hand and the next hold.

The beta of a route is found by applying beam search to the set of available holds with a branching factor of 3 and a beam width of 5. The beam search is initiated with all pairs of possible starting holds, all of which are positioned on or below the sixth row of the MoonBoard. A search is finished once a hold on the top row is reached. The total cost of a path is based on four costs. Each cost is in the range [0, 1]:

- $distance$ is calculated by determining how much the distance between the center point between the current two holds and the next hold deviates to an optimal distance. If the distance is greater than humanly possible, the cost becomes 1.0. To encourage the beta-finding model to converge, downward moves are penalized by doubling the cost if the next hold is placed lower on the board than the center point.
- $finger$ depends on the hold scores and is calculated by taking the average hold score of three holds in the move.
- $flow$ describes how well the move matches the 1524 previously seen moves. LoOP is used to determine the level of similarity of the vectors noted with a blue and orange arrow in fig. 4.
- $footholds$ is based on the two best available footholds. Available footholds include all holds in the route and on the kickboard that are within reach and underneath the currently held holds. For these footholds, an individual cost is calculated based on how close the foothold is to the optimal foothold position and the hold score of the foothold. Because any foothold is assumed to be better than no footholds, this score is halved to give the model incentive to choose a bad foothold over no foothold. The two lowest costs are averaged for the total $foothold$.

These costs are combined into the total move as shown in eq. (2).

$$movecost = \begin{cases} 1.0, & \text{if } distance = 1.0, \\ 0.3 * distance + 0.3 * finger + 0.2 * flow + 0.2 * foothold, & \text{otherwise.} \end{cases} \tag{2}$$

5.2.2 Move Scores. Individual moves are extracted from the found beta and are used to train a model that predicts the difficulty of a single move. Following the ideas of ordinal regression, a separate model is trained to decide between each of the ten grade-splitting points. For example, the model $P(> V7)$ predicts the probability that the move is more difficult than V7. The following features are used:

- The individual and average hold scores of the holds in the move.
- The vector noted with the blue and orange arrows in fig. 4 and their distances.
- The distance between the next hold and the center point of the holds of the moving and stationary hand.
- The coordinates of the rotations of the holds in the move on the unit circle.
- The score of the available footholds is calculated as the inverse of the $footholds$ calculated at the beta finding.

Different models are tested in the framework to see which yields the best performance. All models fit the data structure of a binary classification with the addition of calculating the probability of the classification being correct. Sixteen models are tested:

- Dummy Classification is used to set a baseline that can be used to compare the other models. This model will always predict the most frequent label in the training set.
- AdaBoost Classification.
- Bagging Classification.
- Decision Tree Classification.
- Extra-trees Classification.
- Gradient Boosting Classification Tree (GBC).
- Gaussian Naive Bayes Classification (Gaussian NB).
- Histogram-based Gradient Boosting Classification Tree (HGBC).
- K-nearest Neighbors Classification (KNN).
- Linear Discriminant Analysis (Linear DA).
- Logistic Regression Classification.
- Multi-layer Perceptron Classification (MLP).

- Multivariate Bernoulli Naive Bayes Classification (Bernoulli NB).
- Quadratic Discriminant Analysis (Quadratic DA).
- Random Forest Classification.
- Stochastic Gradient Descent (SGD).

5.2.3 Grade Prediction. The grade of a single move is predicted by taking the largest difference between the predicted probabilities of two adjacent decision models. Combining the probabilities of all moves in a route can be done with two methods:

- The **horizontal** combiner first predicts the difficulty grade for each move in the route and then averages the predicted grades.
- The **vertical** combiner first averages the probabilities for a single decision model over all the moves and then predicts the grade of the whole route by taking the largest difference between these average probabilities.

Variations of both methods are tested by only using the top 5 most difficult moves and skipping the most difficult move in the combination process. A grid search over all variants is carried out to find the best-performing model combiner pair.

5.3 Action Capabilities

Even though action capabilities related to lower body strength, balance, stability, mobility, and flexibility are important to climbing in general, they are less relevant for climbing on a training board. Finger, upper body, and core strength are the primary focus of training board routes [23, 38], leaving the following action capabilities as the indications for affordances that are easiest to measure:

- **Grade** is self-indicated by the climber based on their most difficult previously ascended training board routes.
- **Reach** is measured by taking two holds on the board and trying to touch a hold on a row that is as low as possible. This way of measuring combines the climber's height, arm length, and shoulder flexibility. Also, this measurement does not require measuring tape or any additional equipment. As rows on boards are usually placed with 15 to 20 centimeters in between, measurements achieved this way are less accurate than creating a measurement in centimeters. This is not expected to be a problem for generating suitable routes, as differences in centimeters are not expected to make noticeable changes.
- **Finger strength** is measured by the MBW% for one-repetition hanging for 7 seconds of a 20mm edge.
- **Power** is measured by the MBW% for one pull-up.
- **Core strength** is measured by how long a static position of an L-sit with bent legs, an L-sit with straight legs, or a front lever can be held, using the scoring shown in table 1.

To summarize, measuring all action capabilities needed in the route generation process requires a training board, a pull-up bar, an edge roughly 20mm deep, and a weight belt with weights. All these items are standard in most climbing gyms.

5.4 Route Generation

The route generator functions similarly to the beta finder as both use beam search to find an optimal way to the top. However, for the route generator, the most optimal path is not the easiest path but the path that best fits the action capabilities of the climber. To accommodate this, the costs are calculated differently as well:

- *reach_cost* is calculated by determining how much the distance the moving hand travels deviates from the optimal reach distance. The optimal reach distance is based on the climber's reach, difficulty grade, and upper body strength action capabilities. To prevent moves that are out of reach, moves are penalized with the maximum cost of 1, where the stationary or moving distance is too large. To improve the flow of the generated routes, moves that are downwards or only cover a small distance are also penalized with the maximum cost. Routes are encouraged to converge by slightly reducing the cost of moves to the eighteenth row.
- *finger_cost* depends on the distance between the hold score of the next hold and the optimal hold score. The optimal hold score is equally based on the finger strength and difficulty grade action capabilities.
- *footholds_cost* is based on the difference between the cost of the two best available footholds and the optimal footholds cost. The foothold cost of the best available footholds is calculated as in the beta finder. The optimal footholds cost is based on the core strength, difficulty grade, and reach action capabilities. The core strength and difficulty grade are used to determine how good the footholds should be. The footholds become worse if a climber has a stronger core or can climb a more difficult grade. The reach of the climber determines how far away the footholds should be placed.

The total cost is a weighted average of these three costs where the weights are decided by the preferred climbing style of the climber. The beam search starts with pairs of starting holds that fit the finger strength action capability of the climber and finishes once a hold on the top row has been reached. When the generation is finished, the grade of each generated route is validated by predicting the grade with the grade classifier. The route with the grade closest to the target grade and the best score is presented to the climber.

6 GRADE CLASSIFICATION PERFORMANCE

This section shows the performance of the grade classification model. 30,641 routes remained after removing duplicates and routes with illegal moves. Appendix A.1 shows how the routes are distributed over the difficulty grades. The hold scores can be found in appendix A.2. The beta finder uses LoOP to compare new moves to manually labeled moves. The features of the manually labeled data are shown in appendix A.3. The plot shows how the moves of the moving hand and distances to the next hold of the stationary hand are positioned on the wall, as depicted in fig. 4. The performance of the move score models that give a probability of whether the route is harder or easier than a certain grade are shown in table 2. HGBC has the highest accuracy in all decisions. The accuracy of the final grade prediction for each pair of models and combiners can be seen in table 3. GBC achieved the highest accuracy of **46.5%** using the vertical combination method with all available probabilities. The one-off accuracy for this model and combination method was **80.6%**.

7 VALIDATION WITH USER STUDIES

Qualitative user studies testing the generated routes have been conducted to answer the following questions:

(1) How does the quality of the generated routes compare to that of MoonBoard benchmarks?
(2) Are the generated routes useful for training purposes?
(3) Does the elicited motor behavior emerge when climbing generated routes?
(4) To what extent can humans predict the grade of a MoonBoard benchmark route?

Table 2. Accuracy of each model's ordinal regression decision.

Model	P(>V4)	P(>V5)	P(>V6)	P(>V7)	P(>V8)	P(>V9)	P(>V10)	P(>V11)	P(>V12)	P(>V13)
Dummy	63.9%	63.4%	76.3%	85.9%	95.3%	**98.1%**	**99.3%**	**99.7%**	**99.8%**	**99.9%**
AdaBoost	76.0%	75.7%	80.2%	86.8%	95.3%	98.0%	**99.3%**	**99.7%**	**99.8%**	**99.9%**
Bagging	75.7%	75.6%	80.1%	86.7%	95.3%	**98.1%**	**99.3%**	**99.7%**	**99.8%**	**99.9%**
Bernoulli NB	66.9%	62.6%	73.5%	85.9%	95.3%	**98.1%**	**99.3%**	**99.7%**	**99.8%**	**99.9%**
Decision Tree	71.5%	72.2%	77.3%	84.2%	93.8%	97.3%	98.9%	99.6%	99.7%	99.8%
Extra Trees	71.6%	72.1%	77.4%	84.3%	93.9%	97.4%	98.9%	99.6%	**99.8%**	**99.9%**
Gaussian NB	71.6%	73.7%	77.4%	82.8%	90.1%	93.3%	95.3%	97.0%	97.5%	97.9%
GBC	74.4%	74.6%	79.3%	86.0%	94.9%	97.8%	99.2%	**99.7%**	**99.8%**	**99.9%**
HGBC	**76.1%**	**75.9%**	**80.5%**	**86.9%**	**95.4%**	**98.1%**	**99.3%**	**99.7%**	**99.8%**	**99.9%**
KNN	71.8%	71.4%	77.3%	84.8%	95.1%	98.0%	**99.3%**	**99.7%**	**99.8%**	**99.9%**
Linear DA	73.0%	73.2%	79.0%	86.1%	95.3%	98.0%	**99.3%**	**99.7%**	**99.8%**	**99.9%**
Logistic Regression	73.0%	73.1%	79.1%	86.4%	95.3%	**98.1%**	**99.3%**	**99.7%**	**99.8%**	**99.9%**
MLP	74.4%	73.8%	79.0%	85.7%	94.8%	97.9%	99.2%	**99.7%**	**99.8%**	**99.9%**
Quadratic DA	70.5%	71.9%	77.7%	83.7%	92.2%	95.4%	96.9%	97.9%	98.5%	**99.3%**
Random Forest	74.7%	74.9%	79.6%	86.0%	95.0%	97.9%	99.2%	**99.7%**	**99.8%**	**99.9%**
SGD	69.8%	72.0%	76.6%	86.3%	95.3%	**98.1%**	**99.3%**	**99.7%**	**99.8%**	**99.9%**

Table 3. Accuracy of final grade prediction for model and combiner pairs.

Combiner	Horizontal				Vertical			
Top	All		5		All		5	
Skip	0	1	0	1	0	1	0	1
Dummy	28.3%	28.3%	28.3%	28.3%	28.3%	28.3%	28.3%	28.3%
AdaBoost	32.6%	32.6%	32.6%	32.6%	32.6%	32.6%	32.6%	32.6%
Bagging	44.5%	44.5%	44.1%	44.6%	45.2%	44.2%	44.6%	44.0%
Bernoulli NB	34.4%	32.7%	34.2%	32.7%	33.0%	32.8%	33.0%	32.7%
Decision Tree	43.2%	42.9%	41.4%	42.7%	41.8%	41.9%	40.0%	41.4%
Extra Trees	42.6%	43.2%	40.9%	43.1%	42.4%	42.5%	40.1%	41.9%
GBC	43.6%	44.8%	42.5%	44.5%	44.0%	44.2%	42.6%	43.6%
Gaussian NB	38.8%	40.7%	37.4%	40.2%	39.6%	41.3%	39.1%	41.1%
HGBC	46.2%	45.0%	45.5%	44.8%	**46.5%**	45.5%	46.2%	45.3%
KNN	41.8%	42.2%	40.9%	41.8%	43.7%	43.2%	42.7%	43.0%
Linear DA	42.0%	41.6%	41.1%	41.4%	42.2%	41.6%	41.8%	41.4%
Logistic Regression	41.9%	41.0%	41.0%	40.8%	42.1%	41.3%	41.7%	41.1%
MLP	43.6%	44.4%	41.8%	44.1%	45.0%	44.2%	43.5%	43.7%
Quadratic DA	40.6%	40.6%	39.9%	40.7%	41.3%	40.5%	41.2%	40.4%
Random Forest	44.2%	45.0%	42.9%	44.8%	44.7%	44.3%	43.6%	44.1%
SGD	35.6%	38.1%	32.9%	37.4%	34.7%	35.7%	33.2%	35.2%

7.1 Participants

Eight people participated in the user studies. The participants' demographic can be seen in table 4. To not waste the time and ensure the safety of the participants, a minimum self-estimated climbing grade on MoonBoard benchmarks has been set at 6C. Participants with this climbing level are expected to be able to complete half of the routes in the user studies. Being able to complete less reduces the amount of useful data gathered from their responses. To maximize

Table 4. Demographic of participants in the user studies.

Participant	Age	Gender	Focus	Grade	Reach	Pull-up	Finger Strength	Core
1	25	Male	Finger Strength	7A	11	127.4%	141.1%	3
2	23	Male	Finger Strength	7A+	10	181.1%	170.9%	8
3	20	Male	Finger Strength	7A+	11	145.5%	151.5%	6
4	22	Male	Finger Strength	7B	11	166.7%	189.4%	6
5	26	Male	Finger Strength	7B+	11	182.6%	172.5%	9
6	27	Male	Finger Strength	7C+	10	179.4%	176.5%	6
7	17	Male	Finger Strength	8A	9	194.4%	226.4%	6
8	31	Male	Power	7A	10	161.9%	131.0%	6
9	25	Male	Power	7C	11	177.5%	163.4%	7
10	20	Male	Power	7C+	9	216.7%	200.0%	10
Average	23.6	Male	Finger Strength	7B	10.3	173.3%	172.3%	6.7

the number of respondents, it is possible to fill in the questionnaire without the author's presence. The questionnaire is shared via social media platforms such as Instagram and Reddit, by hanging posters in climbing gyms with the required equipment, and through the social contacts of the author. The user studies are approved by the ethical committee of Electrical Engineering, Mathematics, and Computer Science (EEMCS) of the University of Twente (RP 2022-176).

7.2 Setup

The user studies are filled in via a Google Form. A pull-up bar, an edge between 18 and 20mm deep, and a weight belt with weights are needed for the strength measurements. For climbing and judging the routes, participants need access to a MoonBoard with the 2017 hold setup at 40°.

7.3 Procedure

The user studies are filled in individually and consist of two sequential parts. The two parts don't have to be filled in back to back to accommodate participants who do not have the tools for the strength measurement available close to a MoonBoard.

7.3.1 Strength Measurements. After filling in a consent form, the participants are asked about their sex and age to help understand the demographic. Afterward, the participants are asked to measure the action capabilities described in section 5.3. For the tests, participants have three attempts to set a maximum. Additionally, participants are asked about their preferred boulder style, choosing between boulders focusing on finger strength, dynamic moves, and complicated (foot) movements. The participants' demographics are shown in table 4.

7.3.2 Route Testing. In the second part, the participants attempt to climb ten routes on the MoonBoard, of which six are generated, and four are MoonBoard benchmarks. The action capabilities used to generate routes are shown in table 5. They are chosen to have an easy and difficult route for each focus while covering various grades. No route is generated for grades 7B and up as these routes are not expected to be climbed by enough participants to get valuable insights. Besides, the generated routes are not targeted at climbers who perform at an extremely high level, so the chosen grades should represent the target audience. The benchmarks are randomly selected from all benchmarks with the corresponding grade to reduce the bias introduced by the author.

Table 5. The used action capabilities and parameters to generate routes and randomly pick benchmarks for user studies.

Type	Grade	Focus	Power	Finger Strength	Core Strength
Generated	6B+	Power	150%	110%	2
Generated	7A+	Power	190%	110%	2
Generated	6A+	Fingers	110%	150%	2
Generated	6C+	Fingers	110%	190%	2
Generated	6B+	Core	110%	110%	5
Generated	6C+	Core	110%	110%	9
Benchmark	6A+	-	-	-	-
Benchmark	6B+	-	-	-	-
Benchmark	6C+	-	-	-	-
Benchmark	7A+	-	-	-	-

The ten routes are presented randomly, so the participants do not know whether the route in front of them is generated by a computer or made by a human climber. Participants get three attempts to complete a route, after which they are asked to answer a set of questions. The participants are asked to grade each route to see how accurately human climbers can predict the difficulty grade of benchmarks. The participants are also presented with numerous statements, for which they need to state whether they strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree:

- Climbing the route is enjoyable.
- The route is of high quality.
- The route has a good flow.
- The route is unsafe or could be harmful.
- The route fits my climbing style.
- The design of the route promotes the training of finger strength.
- The design of the route promotes the training of dynamic movements (power).
- The design of the route promotes the training of core strength.

When the author is present, each attempt to climb a route will be filmed. This is done to see whether behavior elicited by the action capabilities in the generated routes emerges in the climbing style of the participant.

7.4 Results

The results of the user studies are shown respectively to their topic, matching the questions posed at the start of this section. As multiple participants indicated they found it difficult to give their opinion on a route they did not complete, only the results given by participants who completed the climb are used.

7.4.1 Quality of Generated Routes. Table 6 shows how the participants judged the routes in the user studies. A score of five shows the participant strongly agrees and a score of one shows the participant strongly disagrees. NC stands for "not climbed", meaning none of the participants completed the route. On average, the completed generated routes received similar or slightly lower scores for all questions.

7.4.2 Generated Routes for Training Purposes. Table 7 shows how well routes fit certain training purposes. Each training purpose has two results per route: finger strength, dynamic movements (power), and core strength. The first, "Trains ...", shows how well the design of the route promotes that training purpose. The second, "Fits ...", shows how the route

Table 6. Route quality scores per route.

Route type	Generated						Benchmark					
Focus	Power		Fingers		Core							
Grade	6B+	7A+	6A+	6C+	6B+	6C+	6A+	6B+	6C+	7A+	Average	
Times Completed	8	0	9	0	7	4	8	8	6	6	Generated	Benchmarks
Enjoyable	4.3	NC	4.6	NC	3.3	3.0	3.9	3.9	4.2	4.5	3.9	**4.1**
Quality	4.1	NC	4.2	NC	3.3	3.3	3.9	3.0	3.8	4.5	**3.8**	**3.8**
Flow	3.9	NC	4.1	NC	2.9	3.0	3.9	3.5	3.5	3.7	**3.6**	**3.6**
Unsafe	1.5	NC	2.1	NC	1.9	1.0	1.0	1.3	1.0	1.3	1.7	**1.1**

fits the style of those climbers that indicated at the start of the user studies that the training purpose describes their bouldering style. MoonBoard benchmarks show higher scores for training finger strength, whereas the generated routes receive higher scores for training dynamic movements and power. The scores for core strength are similar. No participants described their bouldering style to focus on core strength, so "Fits Core" remains empty. The participants' responses show no clear link between the elicited and experienced training focus.

Table 7. Fit for Training Purposes.

Route type	Generated						Benchmark					
Focus	Power		Fingers		Core							
Grade	6B+	7A+	6A+	6C+	6B+	6C+	6A+	6B+	6C+	7A+	Average	
Times Completed	8	0	9	0	7	4	8	8	6	6	Generated	Benchmarks
Trains Fingers	2.3	NC	2.7	NC	2.3	2.8	2.5	3.4	5.0	4.8	2.5	**3.8**
Fits Fingers	4.0	NC	3.7	NC	2.8	3.3	2.8	3.6	4.3	4.5	3.5	**3.7**
Trains Power	4.5	NC	4.6	NC	3.4	4.3	2.4	3.0	3.3	3.2	**4.2**	2.9
Fits Power	4.0	NC	4.3	NC	3.5	4.0	3.0	3.3	4.0	4.5	**4.0**	3.6
Trains Core	3.0	NC	3.8	NC	3.4	3.0	3.3	3.6	3.7	3.3	3.4	**3.5**
Fits Core	-	-	-	-	-	-	-	-	-	-	-	-

7.4.3 Emergent Elicited Motor Behavior. The video analysis showed that the generated 7A+ focusing on power was not climbed because the moves were too far apart. The idea of the moves was clear to the climbers. This was not the case for generated 6C+ focusing on finger strength. The video showed the participants did not know how to approach climbing the route. The 6B+ focusing on power and the 6A+ focusing on finger strength showed flowing dynamic movements. The remaining generated routes proved difficult at the top of the route. The participants indicated this move was significantly harder than the rest of the route.

7.4.4 Human Prediction Performance. Participants were asked to grade each climbed route. As shown in table 8, humans correctly predict the grade of MoonBoard benchmarks 32.1% of the time and have a one-off accuracy of 60.7%. The results also show that the participants found predicting the grade of generated routes harder.

8 DISCUSSION

This section discusses the model design, the model performance, and the user studies.

Table 8. Performance of grade prediction by human climbers.

Route type	Generated						Benchmark					
Focus	Power		Fingers		Core							
Grade	6B+	7A+	6A+	6C+	6B+	6C+	6A+	6B+	6C+	7A+	Average	
Times Completed	8	0	9	0	7	4	8	8	6	6	Generated	Benchmarks
Predicted Correctly	1	0	0	0	1	1	1	1	3	4	10.7%	32.1%
Predicted One-off	5	0	1	0	6	2	3	5	4	5	50.0%	60.7%

8.1 Grade Classification

The grade classifier achieved an accuracy of 46.5%, similar to the highest-scoring grade classifier in previous works and the best-performing human classifier, with accuracies of respectively 46.7% and 45%. Even though the performance is similar, the grade classification model is easier to use by the route generator due to its move-based abstraction. The beta finder uses LoOP to compare new moves to previously seen moves. These manually labeled moves in appendix A.3 show two two-dimensional Gaussian distributions. This is a good step toward making the grade classification fully data-driven. Certain parameters in the method remain hard-coded. To improve the performance, make the employability of the model more flexible, and reduce the bias induced by the author, more (hyper)parameters must follow the data-driven way of being based on previously seen climbing moves.

Another source of errors is the inconsistency of grade labels in the training set. Even though the grade for some routes is validated hundreds of times by climbers repeating it and giving their opinion, the majority of the training data consists of routes that have only been climbed and checked a few times, resulting in inaccurate grades. Climbers not agreeing with each other because they experience the difficulty of a route differently due to having different action capabilities also causes inaccurate grading.

The grade classifier is trained on moves that share the grade of a parent route. However, the grade of the parent route is not always accurate for all the moves in it. For example, a route that consists of four easy moves at the start that finishes with a difficult move will be graded by the difficulty of the last move, the crux of the route. This means that the four easy moves will receive the label of the difficult move, making the labels used in training less accurate.

Drummond and Popinga show a logistic pattern in climbing grades [20]. The grade classifier described in this thesis assumes linearity in the data points' features. Answering how this non-linearity affects performance needs more investigation and could lead to better-performing grade classification.

8.2 Route Generation

The quality of the majority of the routes is good and comparable to that of MoonBoard benchmarks. One route took its action capabilities to extreme levels resulting in long and unclimbable moves. Another route did not find a good flow to the top, which also made the route unclimbable. Based on the user studies, the power action capability should have a smaller influence on the distance to the next hold, whereas the finger strength action capability should have a bigger effect on the difficulty of the holds in the generated routes. Even though this can be adjusted manually in the models, a more resistant solution is desirable. The action capability of core strength was not noticeable in generated routes or benchmarks. This action capability can be excluded in future models, as the action capabilities of difficulty grade, reach, power, and finger strength seem to have a more significant effect on training board routes.

8.3 User Studies

Even though the user studies were extensively advertised under climbers, the number of participants remained low. This is expected to be caused by the number of requirements for filling in the user studies. The most given reasons for not participating were the entry level of participating and not having access to the right MoonBoard setup. The minimum climbing level also reduced the variety in the demographic of participants. As there are few female board climbers with a MoonBoard benchmark grade of 6C or higher, it is logical women are missing in the results. To improve the size and the variety of participants, the minimum level of generated routes needs to become easier.

The predictions of MoonBoard benchmark grades were lower than the human-level performance described by Duh and Chang [21]. It is possible that the climbers tested by Duh and Chang were more experienced in finding the grade. The difference can also be attributed to the environment in which the climbers had to grade the boulders.

9 FUTURE WORK

This section investigates the steps that can be taken to improve the presented work, but also how to build upon the insights gained in this thesis to create new usages for the climbing community.

Both the difficulty classifier and the route generator need better data quality. The grade classifier can benefit from more training data with difficulty annotations on a move level and the route generator could improve its hyperparameters with a more structured feedback system. To solve this problem, a data framework can be set up that supports multiple climbing boards. Ideally, such a framework would be supported by the mobile applications of popular training board companies. This allows for a bigger training set for the models, but also for crowd-sourced data gathering and labeling, using the action capabilities of the users to fine-tune the hyperparameters of the route generator. For the best performance, this framework should have a standardized format for submitting routes and labeling betas, and support for different hold types and wall angles. This could also incentivize progress in automatic route and beta recognition using computer vision, simplifying the process of adding new routes or betas to the framework.

Grading systems remain a hurdle for machine learning solutions in climbing. Climbers having a different experience on a routebased on their action capabilities, and therefore giving it a different grade stands in the way of accurate grades. A grading system that takes the action capabilities of the climber into account could solve this problem. The relation between different action capabilities, for example, how not being able to statically reach a hold can be compensated by explosive power, needs to be further investigated for such a grading system to function properly.

Grade classification could be used to suggest grades for routes created by novice climbers. This makes climbing on training boards more attractive and can lead to more beginners using training boards. Grade suggestions could also prevent climbers from giving climbers a new route a grade that is easier than it really is.

Once the hyperparameters of the route generator are tuned it could be applied in training generation. An example is generating four different routes based on the action capabilities of a climber that can be used in 4x4 endurance training. The route generator could also be applied partially, giving a few options for the next move to a climber that is stuck creating a new route. Finally, route generation could find an application in filling the set of available routes for a new hold setup, only needing a limited number of routes to calibrate its hold scores.

Moving the problems of difficulty estimation and route generation from the 2D problem space to 3D climbing walls remains a major challenge. The main reason the models become complicated is the difficulty of detecting and preventing shortcuts in a route. To solve this problem, many of the current models and cost estimators need to be revised to be able to handle the problems of 3D climbing. Additionally, 3D climbing brings a whole new set of action capabilities into play. Balance, flexibility, technique, and endurance are all ignored in the models for training boards but become major influences on routes that are longer or more vertical.

10 CONCLUSION

Multiple issues were addressed to answer the research question of this thesis: *How can machine learning algorithms be used to automatically generate training board routes tailored to individual climbers' action capabilities and training needs?* Firstly, a set of features was established that closely resembles the abstract form of climbing movement. These features can be used to describe a single climbing move on a training board, allowing models to make decisions on a local move level instead of a global level. This data structure suits both grade classification and route generation.

The established features are used to create a beta-finding algorithm, which extracts single moves out of routes. These moves train a model that predicts a difficulty score for a single move. Combining the difficulty scores of all moves in a route results in the predicted grade. The grade classification model has an accuracy of 46.5%, performing similarly to the best previous work with an accuracy of 46.7% and the best human benchmark with an accuracy of 45% [21]. Because of the similar performance, it can be used to suggest a grade to climbers creating a new route. The data-driven route sequencing and ordinal regression represent the climbing moves closer to reality, making it easier to use the structure in the route generation model.

Action capabilities measuring the difficulty grade, finger strength, power, and core strength are used to tailor generated routes to the training needs of an individual climber. The distance between the holds in generated routes is adjusted based on the power and reach of the climber. The difficulty of chosen holds in generated routes is based on the finger strength of the climber. These properties also change with the desired difficulty grade of the generated route.

Unfortunately, the results of the user studies are unreliable because of the small sample size and the lack of diversity in the demographic of participants. Nevertheless, the video analysis shows elicited behavior on generated powerful moves emerge from the climbers. The link between finger strength and the generated routes must be fine-tuned with more data. Some generated routes proved unclimbable, either through the route generator making moves too large or generating an illogical sequence.

Despite the route generation not performing optimally, this route generation method takes a step forward by weaving the grade classification model into the route generation process and bridging the gap to the action capabilities of a climber. To use this method in practice, we must look at improving the performance of grade classification, the process of gathering training board route data, and receiving detailed feedback on existing routes.

11 ACKNOWLEDGEMENTS

I thank everyone who has supported me in the completion of this thesis. I'm incredibly grateful to Dees Postma for his enthusiasm and for bridging the gap between climbing and technology, to Nicola Strisciuglio for helping me structure and improve the framework, and to Dennis Reidsma for the invaluable discussions and feedback on this thesis. I appreciate all participants who made the effort to test the generated routes in the user studies and want to give a special thanks to Kin Ho (Lucien) Lo for providing the data of the 2017 MoonBoard hold set. Finally, I thank my friends and family for their support and encouragement, which helped me to stay motivated and focused.

REFERENCES

[1] 2016. Climbing Officially Approved for 2020 Olympics. https://www.climbing.com/news/climbing-officially-approved-for-2020-olympics/

[2] 2016. IOC approves five new sports for Olympic Games Tokyo 2020. https://olympics.com/ioc/news/ioc-approves-five-new-sports-for-olympic-games-tokyo-2020

[3] 2020. Breaking, sport climbing, surfing, skateboarding confirmed as additional sports for Paris 2024. https://olympics.com/en/news/breaking-sport-climbing-surfing-skateboarding-paris-2024

[4] 2022. IOC approves surfing, skateboarding, and sport climbing for Los Angeles 2028 Games. https://olympics.com/en/news/surfing-skateboarding-sport-climbing-approved-los-angeles-la2028-games

[5] Beastmaker. 2022. Whole board sets. https://www.beastmaker.co.uk/products/whole-board-sets

[6] Tudor Bompa and Buzzichelli Carlo. 2019. *Periodization-6th Edition: Theory and Methodology of Training*. Human Kinetics. Google-Books-ID: 2f9QDwAAQBAJ.

[7] BoulderBot. 2021. BoulderBot. https://www.boulderbot.io/

[8] Ralph Allan Bradley and Milton E. Terry. 1952. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika* 39, 3/4 (1952), 324–345. https://doi.org/10.2307/2334029 Publisher: [Oxford University Press, Biometrika Trust].

[9] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. 2000. LOF: Identifying Density-Based Local Outliers. (2000).

[10] Egon Brunswik. 1956. *Perception and the Representative Design of Psychological Experiments*. University of California Press. Google-Books-ID: xTwwQvk6XCUC.

[11] Jianlin Cheng, Zheng Wang, and Gianluca Pollastri. 2008. A neural network approach to ordinal regression. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. 1279–1284. https://doi.org/10.1109/IJCNN.2008.4633963 ISSN: 2161-4407.

[12] Tension Climbing. 2022. Tension Board Sets – Tension Climbing. https://www.tensionclimbing.com/product/tension-board-sets/

[13] Rémi Coulom. 2008. Whole-History Rating: A Bayesian Rating System for Players of Time-Varying Strength. In *Computers and Games*, H. Jaap van den Herik, Xinhe Xu, Zongmin Ma, and Mark H. M. Winands (Eds.). Springer, Berlin, Heidelberg, 113–124. https://doi.org/10.1007/978-3-540-87608-3_11

[14] Colin Matthew Davis. 2004. *A comparison of training methods for enhancing climbing performance*. Thesis. Montana State University - Bozeman, College of Education, Health & Human Development. https://scholarworks.montana.edu/xmlui/handle/1/1144 Accepted: 2013-06-25T18:37:11Z Pages: 1-154.

[15] Didier Delignières, Jean-Pierre Famose, Chantalle Thépaut-Mathieu, Philippe Fleurance, and Université Paris-Sud. 1993. A Psychophysical Study of Difficulty Rating in Rock Climbing. (1993), 11.

[16] Xin Ding, Yongwei Wang, Zuheng Xu, William J. Welch, and Z. Jane Wang. 2020. CcGAN: Continuous Conditional Generative Adversarial Networks for Image Generation. https://openreview.net/forum?id=PrzjugOsDeE

[17] Alejandro Dobles, Juan Carlos Sarmiento, and Peter Satterthwaite. 2017. Machine Learning Methods for Climbing Route Classification. (2017), 6.

[18] Nick Draper, Tabitha Dickson, Gavin Blackwell, Simon Fryer, Sefton Priestley, David Winter, and Greg Ellis. 2011. Self-reported ability assessment in rock climbing. *Journal of Sports Sciences* 29, 8 (May 2011), 851–858. https://doi.org/10.1080/02640414.2011.565362 Publisher: Routledge _eprint: https://doi.org/10.1080/02640414.2011.565362.

[19] Nick Draper, David Giles, Volker Schöffl, Franz Konstantin Fuss, Phillip Watts, Peter Wolf, Jiří Baláš, Vanesa Espana-Romero, Gina Blunt Gonzalez, Simon Fryer, Maurizio Fanchini, Laurent Vigouroux, Ludovic Seifert, Lars Donath, Manuel Spoerri, Kelios Bonetti, Kevin Phillips, Urs Stöcker, Felix Bourassa-Moreau, Inmaculada Garrido, Scott Drum, Stuart Beekmeyer, Jean-Luc Ziltener, Nicola Taylor, Ina Beeretz, Franziska Mally, Arif Mithat Amca, Caroline Linhart, and Edgardo Abreu. 2015. Comparative grading scales, statistical analyses, climber descriptors and ability grouping: International Rock Climbing Research Association position statement. *Sports Technology* 8, 3-4 (Oct. 2015), 88–94. https://doi.org/10.1080/19346182.2015.1107081 Publisher: Routledge _eprint: https://doi.org/10.1080/19346182.2015.1107081.

[20] Alexei Drummond and Alex Popinga. 2021. Bayesian inference of the climbing grade scale. *arXiv:2111.08140 [cs, stat]* (Nov. 2021). http://arxiv.org/abs/2111.08140 arXiv: 2111.08140.

[21] Yi-Shiou Duh and Ray Chang. 2021. Recurrent Neural Network for MoonBoard Climbing Route Classification and Generation. *arXiv:2102.01788 [cs]* (Feb. 2021). http://arxiv.org/abs/2102.01788 arXiv: 2102.01788.

[22] André Ebert, Kyrill Schmid, Chadly Marouane, and Claudia Linnhoff-Popien. 2018. Automated Recognition and Difficulty Assessment of Boulder Routes. In *Internet of Things (IoT) Technologies for HealthCare (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering)*, Mobyen Uddin Ahmed, Shahina Begum, and Jean-Baptiste Fasquel (Eds.). Springer International Publishing, Cham, 62–68. https://doi.org/10.1007/978-3-319-76213-5_9

[23] Ned Feehally and Shauna Coxsey. 2021. *Beastmaking: a fingers-first approach to becoming a better climber*. OCLC: 1285703978.

[24] Franz Konstantin Fuss, Yehuda Weizman, Lisa Burr, and Günther Niegl. 2013. Assessment of grip difficulty of a smart climbing hold with increasing slope and decreasing depth. *Sports Technology* 6, 3 (Aug. 2013), 122–129. https://doi.org/10.1080/19346182.2013.854800 Publisher: Routledge _eprint: https://doi.org/10.1080/19346182.2013.854800.

[25] James J. Gibson. 1977. The theory of affordances. *Hilldale, USA* 1, 2 (1977), 67–82.

[26] Luisa V. Giles, Edward C. Rhodes, and Jack E. Taunton. 2006. The Physiology of Rock Climbing. *Sports Medicine* 36, 6 (June 2006), 529–545. https://doi.org/10.2165/00007256-200636060-00006

[27] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, Vol. 27. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html

[28] Andrew Houghton, Ryan Mann, and Jemma Herbert. 2020. Moon Board Climbing. https://github.com/andrew-houghton/moon-board-climbing original-date: 2017-11-03T20:46:03Z.

[29] Jakub Kasse, Mick Ryan, Thomas Mager, Tom Briggs, Charlie Boscoe, and Neil Margetts. 2020. Grade Conversions. https://rockfax.com/climbing-guides/grades/

[30] Lindsay Kempen. 2018. A fair grade: assessing difficulty of climbing routes through machine learning. *Formal methods and tools, University of Twente* (2018).

[31] Kilter. 2022. The Kilter Board – Setter Closet. https://settercloset.com/pages/the-kilter-board

[32] T. D. Koukoubis, L. W. Cooper, R. R. Glisson, A. V. Seaber, and J. A. Feagin. 1995. An electromyographic study of arm muscles during climbing. *Knee Surgery, Sports Traumatology, Arthroscopy* 3, 2 (July 1995), 121–124. https://doi.org/10.1007/BF01552388

[33] Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek. 2009. LoOP: local outlier probabilities. In *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM, Hong Kong China, 1649–1652. https://doi.org/10.1145/1645953.1646195

[34] Cassim Ladha, Nils Y. Hammerla, Patrick Olivier, and Thomas Plötz. 2013. ClimbAX: skill assessment for climbing enthusiasts. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing (UbiComp '13)*. Association for Computing Machinery, New York, NY, USA, 235–244. https://doi.org/10.1145/2493432.2493492

[35] K. H. Lo. 2020. Embedding and generation of indoor climbing routes with variational autoencoder. *arXiv:2009.13271 [cs, eess]* (Sept. 2020). http://arxiv.org/abs/2009.13271 arXiv: 2009.13271.

[36] Peter McCullagh. 1980. Regression Models for Ordinal Data. *Journal of the Royal Statistical Society: Series B (Methodological)* 42, 2 (1980), 109–127. https://doi.org/10.1111/j.2517-6161.1980.tb01109.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1980.tb01109.x.

[37] Mehdi Mirza and Simon Osindero. 2014. Conditional Generative Adversarial Nets. *arXiv:1411.1784 [cs, stat]* (Nov. 2014). http://arxiv.org/abs/1411.1784 arXiv: 1411.1784.

[38] Martin Mobråten and Stian Christophersen. 2020. *The Climbing bible: technical, physical and mental training for rock climbing*. Vertebrate Publishing, Sheffield, United Kingdom.

[39] JERRY MOFFATT. 2019. *MASTERMIND: mental training for climbers*. VERTEBRATE PUBLISHING, S.l. OCLC: 1127750870.

[40] MoonBoard. 2012. Welcome to training on the MoonBoard, climb on the same problems as other climbers from around the world. https://www.moonboard.com

[41] Draper Nick, Tabitha Dickson, Gavin Blackwell, S Priestley, Simon Fryer, Helen Marshall, J Shearman, Michael Hamlin, D Winter, and Greg Ellis. 2011. Sport-specific power assessment for rock climbing. *The Journal of sports medicine and physical fitness* 51 (Sept. 2011), 417–25.

[42] Amin Heyrani Nobari, Wei Chen, and Faez Ahmed. 2021. PcDGAN: A Continuous Conditional Diverse Generative Adversarial Network For Inverse Design. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (Aug. 2021), 606–616. https://doi.org/10.1145/3447548.3467414 arXiv: 2106.03620.

[43] Adam Ondra. 2019. What is the best body type for climbing? https://www.adamondra.com/tokyo/what-best-body-type-climbing

[44] C. Phillips, L. Becker, and E. Bradley. 2012. strange beta: An assistance system for indoor rock climbing route setting. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 22, 1 (March 2012), 013130. https://doi.org/10.1063/1.3693047 Publisher: American Institute of Physics.

[45] Kevin C. Phillips, Joseph M. Sassaman, and James M. Smoliga. 2012. Optimizing Rock Climbing Performance Through Sport-Specific Strength and Conditioning. *Strength & Conditioning Journal* 34, 3 (June 2012), 1–18. https://doi.org/10.1519/SSC.0b013e318255f012

[46] D. R. Reddy, M. F. Medress, F. S. Cooper, J. W. Forgie, C. C. Green, D. H. Klatt, M. H. O'Malley, E. P. Neuburg, A. Newell, B. Ritea, J. E. Shoup-Hummel, D. E. Walker, and W. A. Woods. 1977. Speech understanding systems: Report of a steering committee. *Artificial Intelligence* 9, 3 (Dec. 1977), 307–316. https://doi.org/10.1016/0004-3702(77)90026-1

[47] Dean Scarff. 2020. Estimation of Climbing Route Difficulty using Whole-History Rating. *arXiv:2001.05388 [stat]* (July 2020). http://arxiv.org/abs/2001.05388 arXiv: 2001.05388.

[48] Dibyendu Seal and Reshmita Seal. 2022. Optimum Route Computation in a Chaotic Artificial Climbing Wall. In *ICT Systems and Sustainability*. Springer, 671–680.

[49] A. W. Sheel. 2004. Physiology of sport rock climbing. *British Journal of Sports Medicine* 38, 3 (June 2004), 355–359. https://doi.org/10.1136/bjsm.2003.008169 Publisher: British Association of Sport and Excercise Medicine Section: Review.

[50] F. T. A. Stapel. 2020. A Heuristic Approach to Indoor Rock Climbing Route Generation. http://essay.utwente.nl/80579/ Publisher: University of Twente.

[51] Nicolay Stien, Atle Hole Saeterbakken, Espen Hermans, Vegard Albert Vereide, Elias Olsen, and Vidar Andersen. 2019. Comparison of climbing-specific strength and endurance between lead and boulder climbers. *PLOS ONE* 14, 9 (Sept. 2019), e0222529. https://doi.org/10.1371/journal.pone.0222529 Publisher: Public Library of Science.

[52] Get Stōkt. 2021. Stōkt. https://www.getstokt.com

[53] Cheng-Hao Tai, Aaron Wu, and Rafael Hinojosa. 2020. Graph Neural Networks in Classifying Rock Climbing Difficulties. (2020), 6.

[54] Lattice Training. 2018. Every Moonboard Problem Ever, Analysed. https://latticetraining.com/2018/01/26/every-moonboard-problem-ever-analysed/

[55] Lattice Training. 2021. 2 Arm Finger Strength Test. https://latticetraining.com/2-arm-fs-test/

[56] Colwyn Trevarthen. 1986. Development of Intersubjective Motor Control in Infants. 209–261. https://doi.org/10.1007/978-94-009-4460-2_14

[57] Arie van Deursen, Paul Klint, and Joost Visser. 2000. Domain-specific languages: an annotated bibliography. *ACM SIGPLAN Notices* 35, 6 (June 2000), 26–36. https://doi.org/10.1145/352029.352035

[58] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[59] William H. Warren. 1984. Perceiving affordances: Visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance* 10, 5 (Oct. 1984), 683–703. https://doi.org/10.1037/0096-1523.10.5.683 Publisher: American Psychological Association.

[60] Phillip B. Watts. 2004. Physiology of difficult rock climbing. *European Journal of Applied Physiology* 91, 4 (April 2004), 361–372. https://doi.org/10.1007/s00421-003-1036-7

[61] Iupe Werner, Walter Gebert, and Birpit Kauer. 1999. Three Dimensional Analysis of Rock Climbing Techniques. *Proceedings of the 1st International Conference on Science and Technology in Climbing and Mountaineering* (1999), 7–9.

A APPENDIX

A.1 Grade Distribution

Figure 5 shows the grade distribution of the routes in the MoonBoard 2016 hold setup.

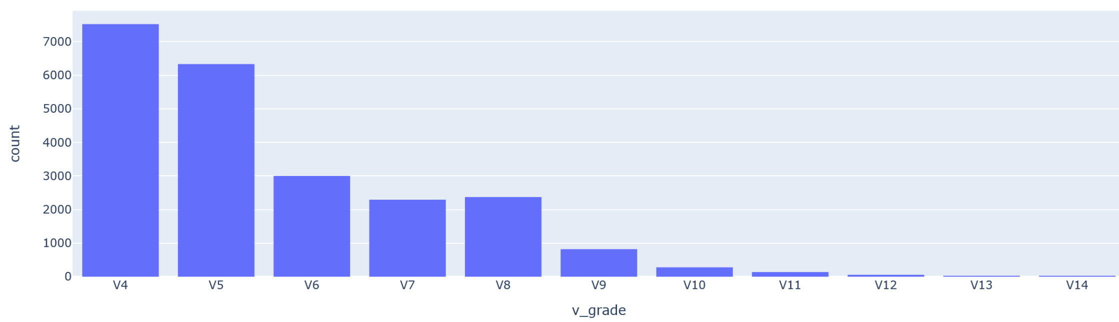


Fig. 5. MoonBoard 2016 routes grade distribution after pre-processing.

A.2 Hold Scores

The hold scores of the 2016 hold set are shown in fig. 6.

A.3 Flow of Manually Labeled Moves

Figure 7 displays the movement in the manually labeled routes.

A.4 Confusion Matrix

Figure 8 depicts the confusion matrix of the grade classifier's predictions.

A.5 Routes User Studies

The routes in fig. 9 are used in the user studies.

Received 15 January 2023

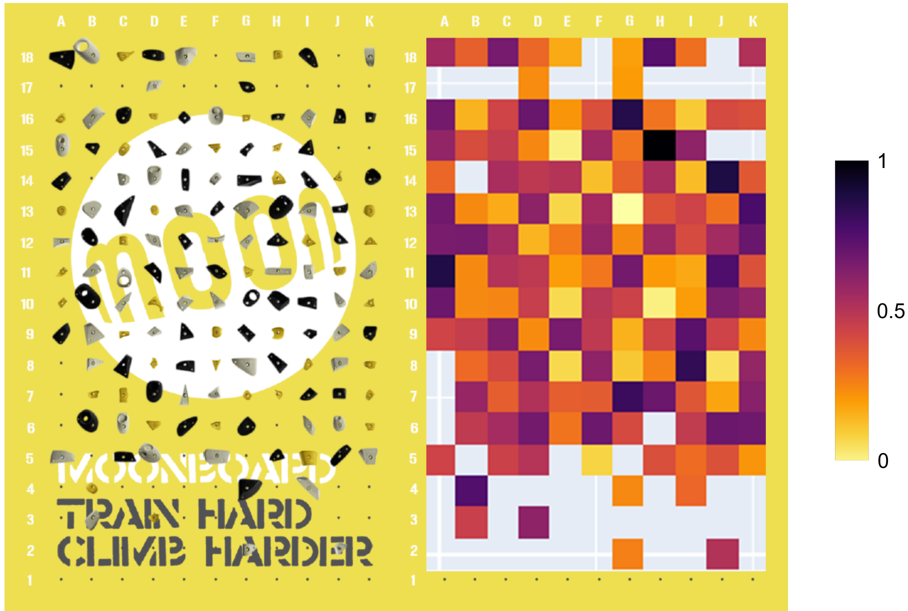


Fig. 6. MoonBoard with its hold scores.

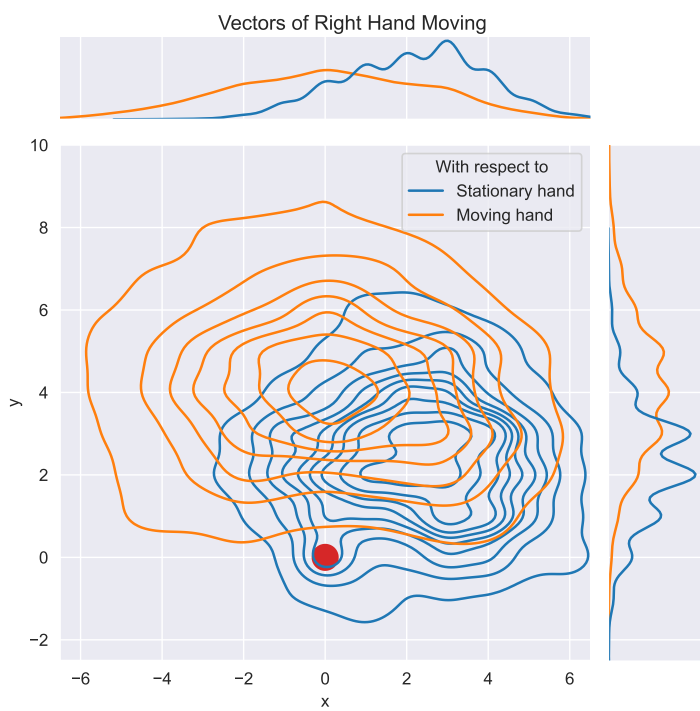


Fig. 7. Distribution of all moves in the manually labeled routes.

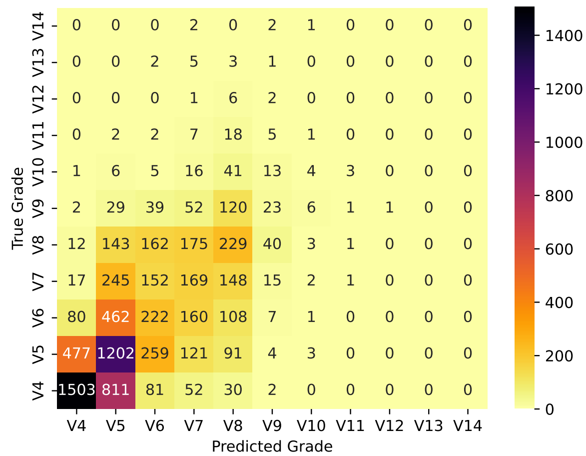
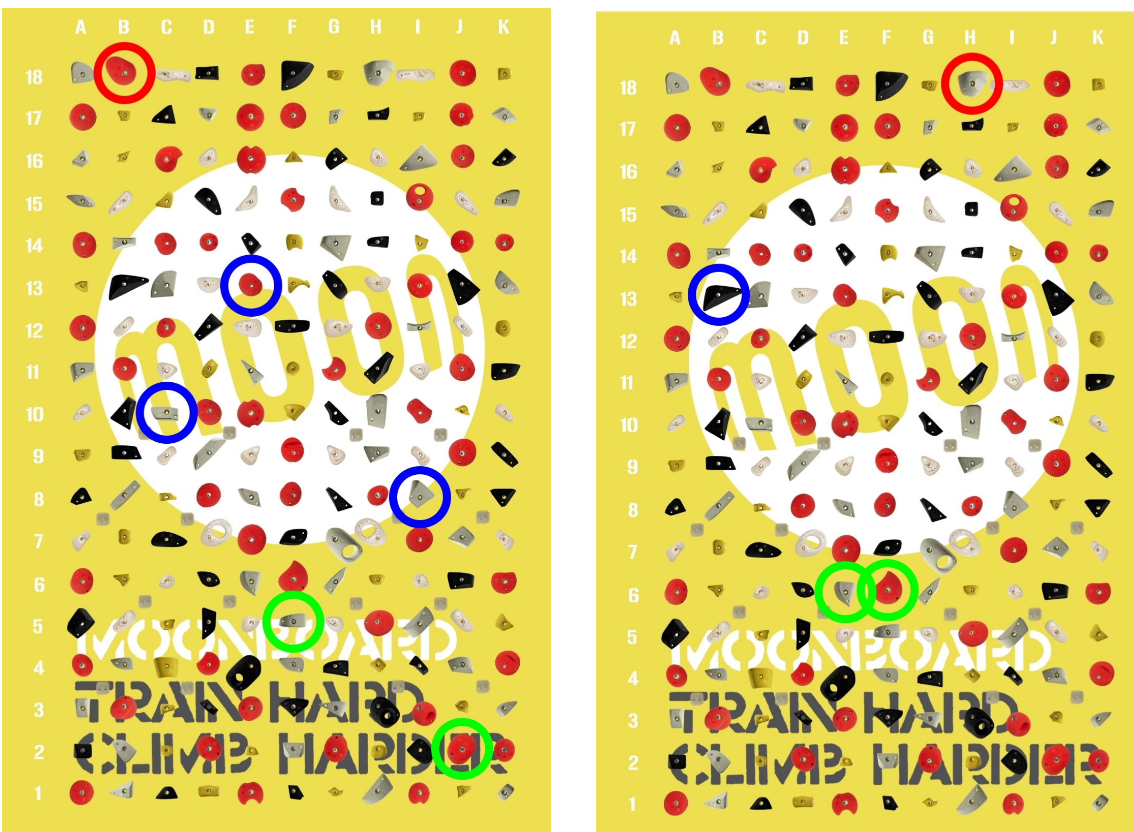


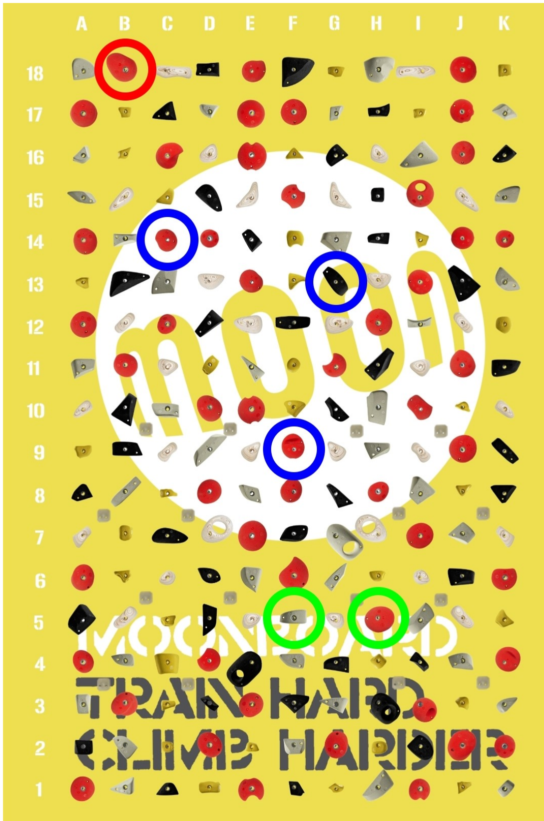
Fig. 8. Predicted grades



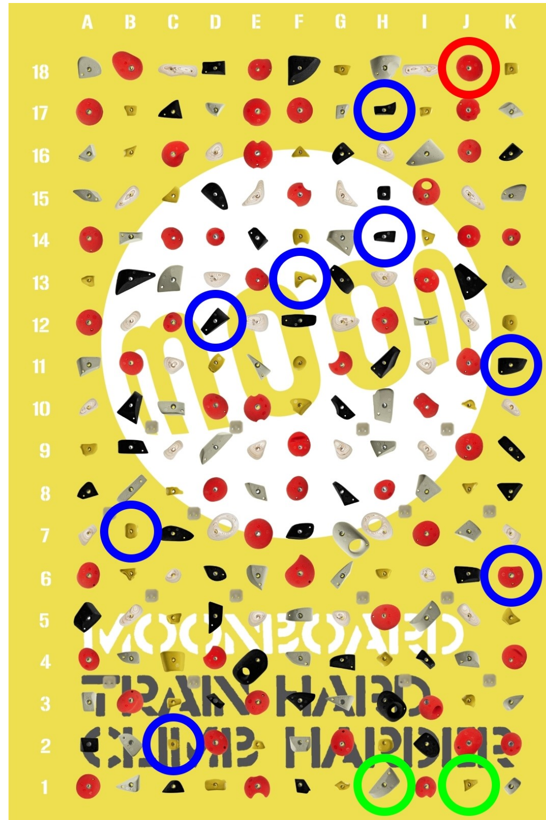
(a) Generated 6B+ focused on power.

(b) Generated 7A+ focused on power.

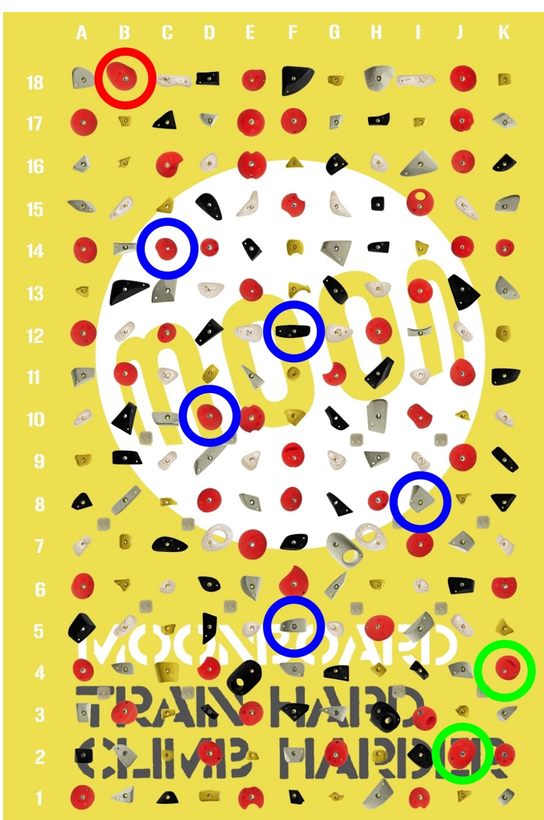
Fig. 9. Routes used in user studies.



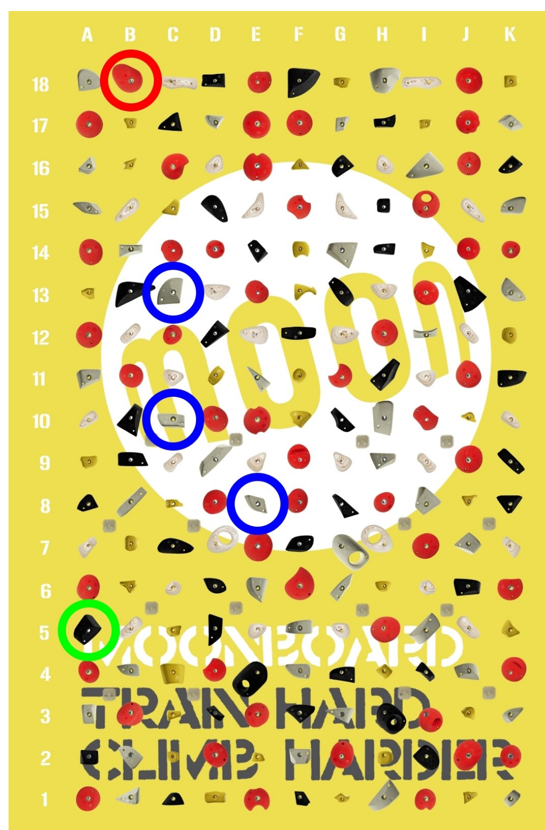
(c) Generated 6A+ focused on finger strength.



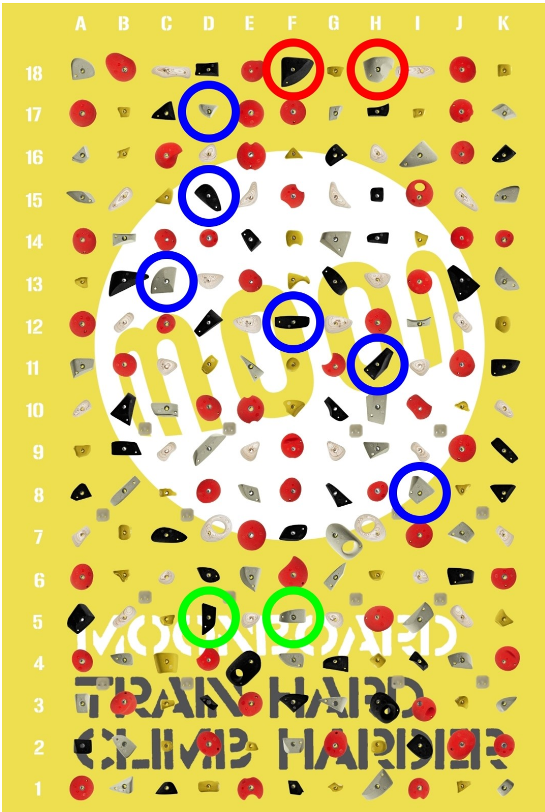
(d) Generated 6C+ focused on finger strength.



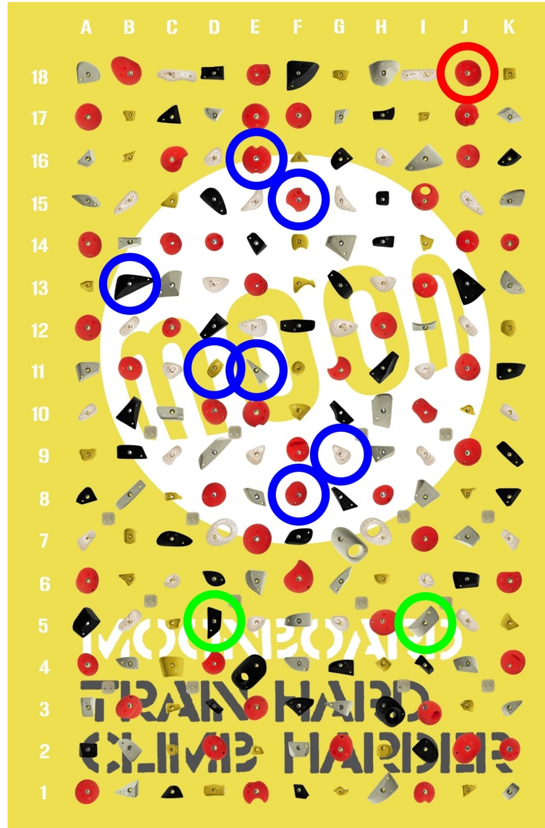
(e) Generated 6B+ focused on core strength.



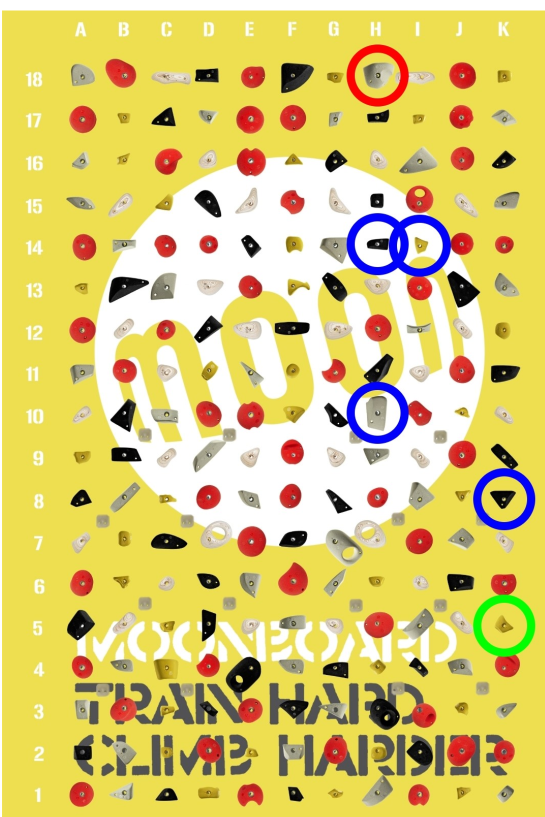
(f) Generated 6C+ focused on core strength.



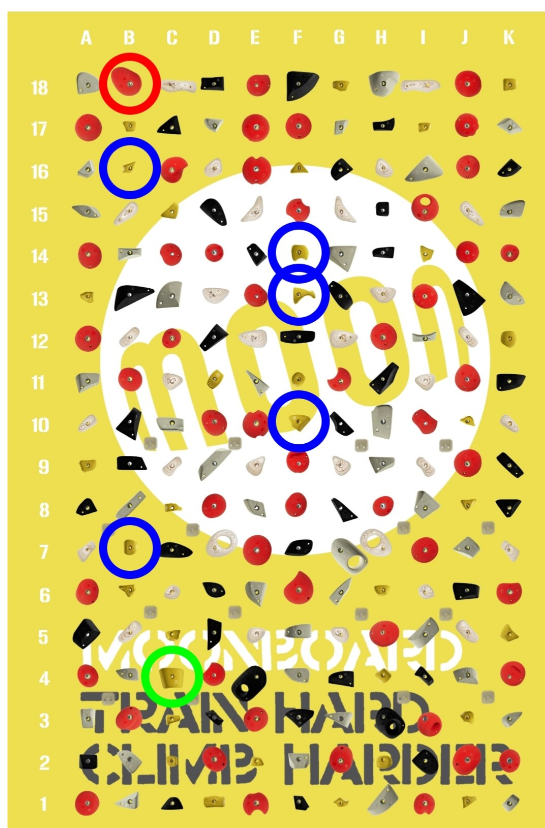
(g) 6A+ MoonBoard Benchmark.



(h) 6B+ MoonBoard Benchmark.



(i) 6C+ MoonBoard Benchmark.



(j) 7A+ MoonBoard Benchmark.