



OPEN

Early identification of dropouts during the special forces selection program

Ruud J. R. den Hartigh¹✉, Rik Huijzer¹, Frank J. Blaauw², Age de Wit³ & Peter de Jonge¹

Recruits are exposed to high levels of psychological and physical stress during the special forces selection period, resulting in dropout rates of up to 80%. To identify who likely drops out, we assessed a group of 249 recruits, every week of the selection program, on their self-efficacy, motivation, experienced psychological and physical stress, and recovery. Using linear regression as well as state-of-the-art machine learning techniques, we aimed to build a model that could meaningfully predict dropout while remaining interpretable. Furthermore, we inspected the best-performing model to identify the most important predictors of dropout. Via cross-validation, we found that linear regression had a relatively good predictive performance with an Area Under the Curve of 0.69, and provided interpretable insights. Low levels of self-efficacy and motivation were the significant predictors of dropout. Additionally, we found that dropout could often be predicted multiple weeks in advance. These findings offer novel insights in the use of prediction models on psychological and physical processes, specifically in the context of special forces selection. This offers opportunities for early intervention and support, which may ultimately improve success rates of selection programs.

Keywords Gradient boosting, Stable and interpretable RUle sets, Military, Personnel selection, Recovery, Stress

Special forces are often considered the most elite military units, with the potential to significantly impact strategic military outcomes. They are typically composed of highly trained and motivated individuals who are able to operate in high-stakes environments that are both psychologically and physically demanding. However, dropout rates during the selection process are close to 80%¹. This is a concern for both the recruits and the military as it incurs a personal toll on the recruits and is costly for the organization. Scientifically, a major challenge is identifying potential dropouts early in the selection period via accurate predictive models. Such models could allow for early intervention on the relevant psychological and physical processes related to dropout. The scarce previous research investigated dropout by comparing test scores from before the selection period with the final dropout or graduation decision. Typical psychological tests in the military include personality inventories^{2–7}, and recent research shows, for instance, that graduates tend to score higher on emotional stability and conscientiousness than dropouts³. In other research, psychological hardiness was associated with graduation among 1138 special forces recruits⁸ and 178 Norwegian border patrol soldiers⁹. On the other hand, in a study including 73 South African special forces, hardiness and self-efficacy were not associated with graduation¹⁰. In another study, higher self-efficacy was significantly associated with graduation among 380 special forces recruits¹¹.

Physical scores, typically based on fitness, strength, and endurance tests, have also been related with graduation. For example, in a study among 69 Finnish soldiers, baseline information of aerobic fitness significantly predicted graduation¹². In a study on 160 Swedish police counterterrorism intervention units including various psychological and physical tests, the authors found running capacity was a significant predictor of graduation¹³. A study on 800 special forces recruits showed that both psychological and physical test scores were significantly associated with graduation¹⁴. A follow-up study on 117 special forces soldiers specifically found that physical characteristics of the body, such as a lower percentage body fat and fat mass were predictors of physical performance and graduation¹⁵.

Despite some evidence for the role of psychological and physical factors in predicting dropout, a main issue of previous studies is that they showed limited effects and different predictor combinations. For instance, when comparing agreeableness between military recruits and a civilian control group, one study showed

¹Department of Psychology, Faculty of Behavioural and Social Sciences, University of Groningen, Grote Kruisstraat 2/1, 9712TS Groningen, The Netherlands. ²Research and Innovation, Researchable BV, Groningen, The Netherlands.

³Human Performance Team Commando Corps, Ministry of Defence, The Hague, The Netherlands. ✉email: j.r.den.hartigh@rug.nl

that agreeableness was lower for the military recruits⁴, whereas this pattern was not found in more recent studies^{3,13}. Such contradicting results could be due to theoretical and methodological factors. Theoretically, a commando profile could be composed of different combinations of characteristics that could allow an individual to perform in highly psychologically and physically demanding situations¹⁶. Methodologically, an important factor contributing to dropout is how recruits respond to the stress *during* the heavy selection program. This cannot be derived from psychological and physical measures taken at one point before the selection program. Thus, an important question is: how do recruits actually respond to, and recover from, the stress to which they are exposed? Such a question can be answered by measuring recruits during the selection period on relevant psychological and physical processes of stress and recovery.

Recent research provided initial evidence that measurements taken during a training or selection period can be used to predict dropout. For instance, one longitudinal study in the context of elite military training found that recruits who voluntarily dropped out exhibited an increase in emotional or physical pain, and a decrease in confidence in course completion up to three days before dropping out¹⁷. Similarly, in a study on 46 recruits in the Australian Army basic military training course, higher stress and recovery, as measured via the Short Recovery and Stress Scale¹⁸, were associated with a higher risk of delayed completion¹⁹. Comparable results have been found in sports: In a study on 135 adolescent elite athletes, lower recovery and higher stress states as measured by the Acute Recovery and Stress Scale (ARSS) were followed by depressive, burnout, and insomnia symptoms²⁰. In a study on 74 middle and long-distance runners, recovery and exertion were considered some of the most important variables for predicting injuries²¹. These findings are promising as they suggest that dropout, either voluntary or involuntary (e.g., due to injury), can be predicted in advance based on psychological and physical assessments during selection or training periods.

Building upon first efforts of predicting dropout from military programs and the increasing interest in the psychological and physical stress monitoring during army training, important statistical strides can be made. Most notably, while previous studies often applied traditional statistical methods, that is, how variables were associated with dropout or graduation, they often did not report the predictive performance. This means that associations between variables could be too small to be useful in practice or they could be wrong due to overfitting²². Ideally, a study would report predictive performance for multiple models to avoid overfitting and dependence on one model, and use repeated measures to allow for prediction of dropout in advance (for a recent example in the context of the marine corps, see²³).

The current study therefore aimed to assess the experienced psychological and physical states of recruits during the selection weeks, while improving upon the statistical methods used in previous research. In line with recommendations from previous literature, we specifically focused on the experiences of self-efficacy, motivation, and psychological and physical stress and recovery²⁴. We compared various classical and state-of-the-art machine learning models via cross-validation to determine which statistical model can predict dropout best. In addition, we explored the moment at which valid predictions of dropout could be made (e.g., one week or more weeks in advance). Such knowledge could lead to a better understanding of the dropout process, and to targeted interventions in practice.

Method

Participants

The sample for this study consisted of 249 male special forces recruits, ranging in age from 18 to 35 years. Prior to their involvement in the study, active informed consent was obtained from each recruit. The information letter informed participants about the study's purpose, procedures, and potential risks, as well as their right to withdraw from the study at any time. The participants were diverse in terms of their military experience, with some being new recruits while others had prior experience in different branches of the armed forces. Due to the sensitive nature of the data, more detailed information about the participants could not be made available.

Measures and procedure

During the selection period that lasted up to 16 weeks, we asked the following self efficacy and motivation questions, both in Dutch: "How confident are you in your capabilities to pass the training program?" (0 = not at all confident, 100 = very confident) and "How motivated are you to pass the training program?" (0 = not at all motivated, 100 = very motivated). Furthermore, we used a Dutch version of the Short Recovery and Stress Scale (SRSS), a self-report questionnaire assessing perceived stress and recovery levels¹⁸. The SRSS was validated in a group of 385 Dutch and Flemish athletes²⁵, and consists of 8 items divided into two subscales: Recovery and Stress. Items were rated on a seven-point Likert scale, with higher scores indicating greater levels of recovery or stress. The Recovery subscale evaluates an individual's current state in comparison to their best recovery state ever, with items such as "Physical performance capacity" and "Mental performance capacity". The Stress subscale assesses an individual's current state relative to their highest stress state ever, including items like "Muscle stress" and "Lack of inspiration" (see¹⁸ for the manual). Over the course of the study, the recruits completed the questionnaire weekly, resulting in a total of 1652 responses.

On average, we received about 6 responses per person. The number of responses per participant varied due to individuals dropping out of the selection process before the end of the study. The data was collected using an electronic questionnaire, which was administered via a web-based platform that we developed for this project. The collection occurred at the start of the training week, which was typically on Monday morning at 8AM.

Our study was conducted according to the requirements of the Declaration of Helsinki. It was approved by the ethics committee of the Faculty of Behavioral and Social Sciences of the University of Groningen, The Netherlands (research code: PSY-1920-S-0512).

Analysis

We processed the data to include the following 13 variables columns: participant id, week, self-efficacy, motivation, 8 SRSS items, and whether the individual drops out in the week after the response. Here, we truncated the data at 13 weeks, considering the data was collected for 14 out of 16 weeks. Next, we analyzed the data in three steps to evaluate the model^{26,27}. First, we applied principles and techniques from machine learning to estimate the model's ability to predict future outcomes. We used 12-fold cross-validation and the area under the receiver operating characteristic curve (AUC) as a performance metric, for which we used the MLJ.jl software package²⁸. The AUC is a measure of the performance of a binary classifier, where a value of 0.5 indicates random guessing and a value of 1.0 indicates perfect predictions. We used the AUC because it is a robust metric that is not sensitive to class imbalance and is a common metric in the literature.

Second, we used multiple different models to determine which one performed best in terms of predictive performance. We fitted a binary logistic model with no intercept as our baseline model. Next, we fitted two Stable and Interpretable RULE Sets (SIRUS) models to the training data as the SIRUS model has shown to perform well in similar situations with relatively few samples and binary outcomes^{29–31}. SIRUS is based on random forests and is non-parametric, meaning that it does not make assumptions about the distributions of the data. Random forest-based models are robust to outliers, do not require scaling of the data, and perform very well in general³². Furthermore, we fitted a modern gradient boosting model called EvoTrees.jl³³. Gradient boosting models are not fully interpretable due to the large amounts of trees³⁰, but they are known to have high predictive performance in many situations^{34,35}. In the context of military selection, we prefer models with an optimal trade-off between predictive accuracy and interpretability. Therefore, to combine predictive performance and interpretation²⁶, we inspected the model that scored best on this trade-off. Specifically, we trained the model on the full dataset and inspected the resulting model.

Third, we evaluated the predictive performance in practice. To achieve this, we converted the model's predictions, initially in the range of 0 to 1, back to binary outcomes. We did this by selecting a threshold and using it to classify the outcomes into dropout and graduate groups. Finally, we visualized the predictions of the model for different thresholds. This visualization assists researchers and practitioners in selecting the right balance between the number of false positives and false negatives, and thereby provides an indication of the predictive performance in practice.

Results

Figure 1 displays the results for the evaluation runs on the cross-validation data.

In Fig. 1, the bottom two graphs both have a max tree depth of 2. This higher depth allows these models to capture more complex interactions between variables. However, as can be seen in the figure, these models do not perform markedly better than the simpler models. This is likely caused by more complex models overfitting the data and could potentially be solved by using more data.

Overall, the logistic regression model performs best since it reveals the best trade-off between predictive performance and interpretability. The interpretability of this model is very high because the algorithm is much simpler compared to the thousands of trees in gradient boosting models, yet its performance is similar to that of the gradient boosting model. Therefore, we inspect the logistic regression model in more detail below. The coefficients of the logistic model, when fitted on the full dataset, are shown in Table 1. When interpreting this model, note that there is variation in performance for the different cross-validation folds (see Fig. 1). This is why we decided post hoc to set our alpha level conservatively to 0.001 instead of the commonly used 0.05. This lower alpha level means that we are less likely to find significant results. Setting this level post hoc seemed reasonable as we use the *p*-value as just one of the many tools to interpret the model²⁷. From Table 1, we can see that the variables “Self-Efficacy” and “Motivation” were significant. The positive coefficients indicate that recruits who score higher of self-efficacy and higher on motivation are less likely to drop out.

We visualized the predictions of the logistic regression model for different thresholds (see Fig. 2). The figure shows that many of the dropouts were predicted correctly (events marked in blue in subfigures **b**, **c**, **d**), which is in line with the AUC score as reported in Fig. 1. The higher the threshold, the more likely it is for the model to predict dropout. As a consequence, the number of false positive predictions increased for higher thresholds. An interesting observation is that, across all three thresholds, several dropouts were predicted weeks before the actual dropout event. These early predictions are marked in purple (Note here that purple events in the last week indicate a prediction that recruits dropout in the very last phase of the selection period. This could, however, not be verified, as we only studied the data of the first 14 selection weeks).

Discussion

The current study aimed to predict dropout during the special forces selection period. To that end, we assessed recruits' psychological and physical states during this period. We applied simple logistic regression models as well as more complex machine learning models on the recruits' data. Next, we evaluated how well each model performed, we interpreted the best model, and explored the predictive performance in practice. We found that a logistic regression model scored best on the trade-off between predictive performance and interpretability, because it performed relatively well with an area under the curve (AUC) of 0.69, and the contributions of the features can be easily extracted through the regression coefficients. The most complex models scored only slightly better on the AUC, which suggests we had insufficient data for more complex models.

The logistic regression model revealed that self-efficacy was the strongest predictor of dropout. This provides support for earlier research showing that decreases in confidence about graduating relate to dropout in a military context¹⁷. The prominent role of self-efficacy in the special forces selection context may not be surprising, given the saliency of the four typical sources of self-efficacy in this environment. Indeed, this context provides a novel

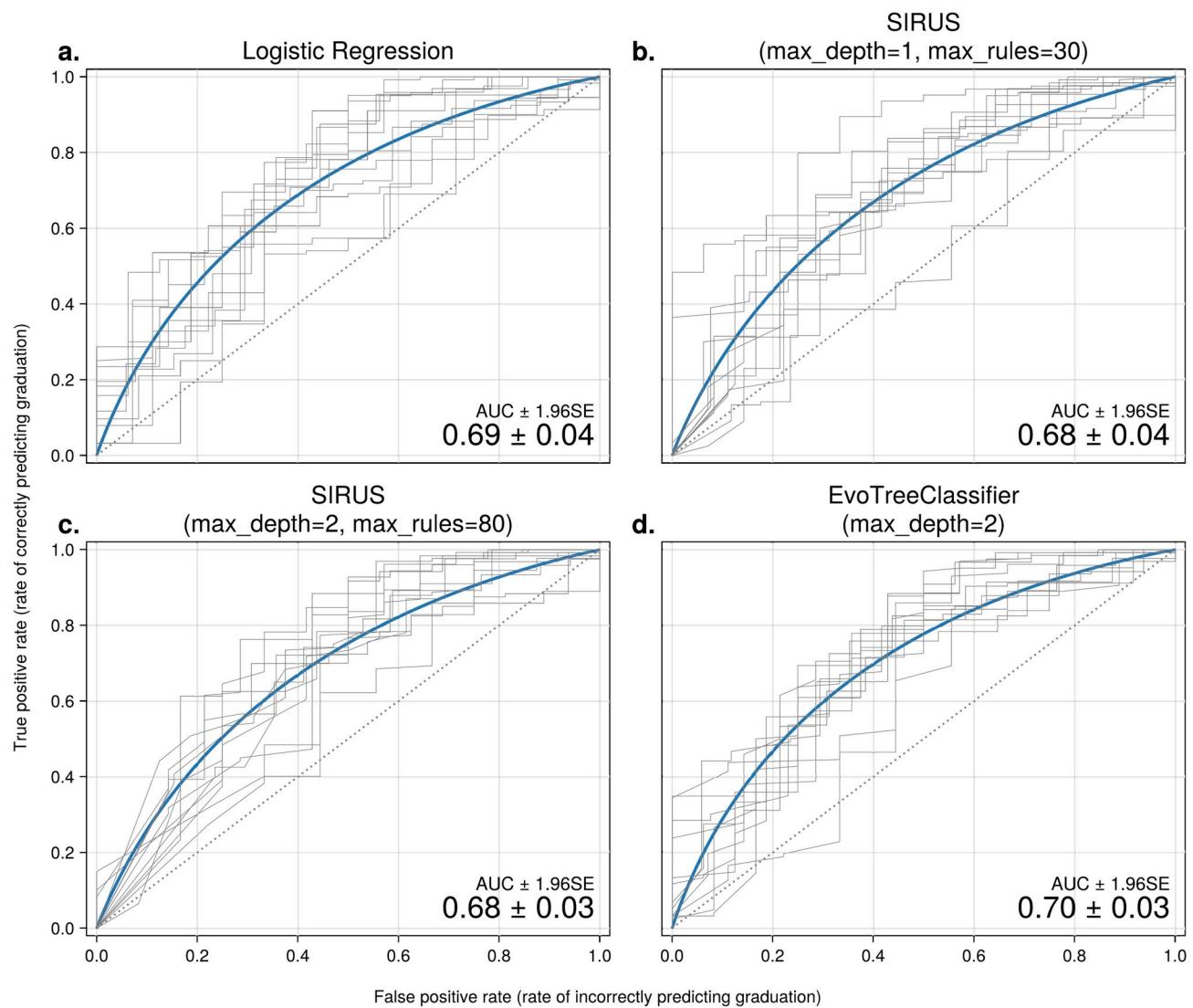


Fig. 1. Receiver Operating Characteristic (ROC) Curves. The thick blue line represents the estimate of the average ROC curve over all folds. The thinner lines in gray display all individual folds of the 12-fold cross-validation. The average Area Under the Curve (AUC) and $1.96 \times$ standard error scores are shown in the bottom right of each graph.

Variable	Coefficient	Z-score	p-value	Lower 95%	Upper 95%
Self-efficacy*	1.734	4.78	<0.001	1.022	2.446
Motivation*	1.205	3.6	<0.001	0.549	1.86
Physical performance capability	-0.343	-0.83	0.407	-1.157	0.47
Mental performance capability	-0.245	-0.6	0.549	-1.049	0.56
Emotional balance	0.406	1.24	0.215	-0.233	1.044
Overall recovery	0.341	0.89	0.373	0.413	1.096
Muscular stress	0.561	2.18	0.029	0.056	1.066
Lack of activation	0.513	1.38	0.168	-0.221	1.282
Negative emotional state	-0.493	-1.38	0.168	-1.194	0.209
Overall stress	-0.701	-1.95	0.051	-1.405	0.003

Table 1. Fitted binary logistic regression statistics.

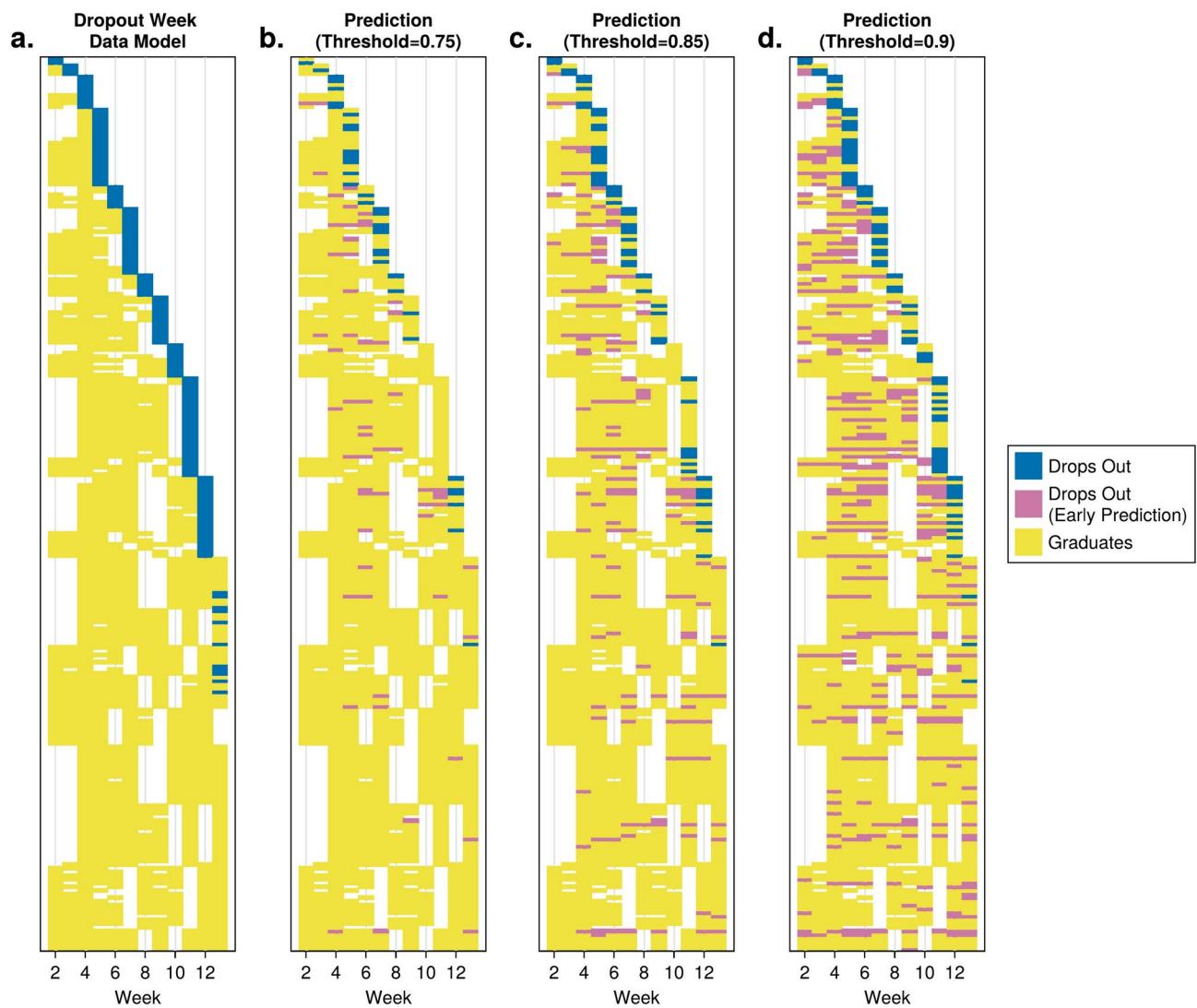


Fig. 2. Dropout Data and Predictions of the Model. This figure shows how the data was modeled, thereby marking the true points of drop out for each participant in blue (Graph **a**). The other three graphs (**b–d**) show the predictions according to the logistic regression model for different thresholds. Exploring different thresholds allows practitioners to select the right balance between the number of false positives and false negatives. This, together with the AUC, provides an indication of the predictive performance in practice.

situation to recruits in which they are exposed to 1) (un)successful mastery experiences, 2) vicarious experiences in the form of (un)successful attainments of other recruits, 3) verbal persuasion by the instructors who lead the selection process, and 4) variations in physiological and affective states throughout the exercises. Across achievement contexts, robust evidence suggests a positive relationship between the self-efficacy levels resulting from such sources and performance outcomes³⁶, especially when performance feedback is salient³⁷. Moreover, individuals' sense of self-efficacy determines their strength of motivation in an achievement situation³⁶, which was the second-strongest predictor in the regression model.

Our findings are also in accordance with the perspective that temporal measures of self-efficacy and motivation can provide important information on an individual's resilience. That is, motivation and self-efficacy are important psychological performance factors that ideally return to normal levels following psychological and physical stress. When individuals loose resilience, as reflected in their self-efficacy and motivation levels, this could be a warning signal for negative outcomes such as psychological problems or dropout (cf.,²⁴). Interesting in this regard is that more direct measures of stress and recovery experiences, as assessed using the SRSS, were less predictive of dropout. One reason for this could be that the SRSS has, so far, only been validated in the sports context²⁵. Despite the parallels between the sport and military context, individuals are typically exposed to more extreme psychological and physical stress during a special forces selection program. It could be that the experience of stress and recovery are so high for the recruits, that it cannot account for much variance in the outcome anymore.

Finally, we estimated the predictive performance in practice. We visualized the predictions of the model for different thresholds, and found that the model could often predict dropout multiple weeks in advance. In practice, this means that the calculated AUC scores may underestimate the predictive performance due to the way the data was modeled. Note that choosing the right threshold is important as it determines the balance between the number of false positives and false negatives. We showed multiple thresholds which could be used by practitioners to select this balance, where higher thresholds led to more false positives. Yet, since the cost of missing a dropout is high, it could be recommendable to pick a higher threshold that results in more early warnings of dropout.

The current study thus introduced a new methodological approach that can be further exploited in the field of performance prediction in high-stakes contexts, such as the military or sports. Future work could account for some limitations in the design of our study. First, the sample size was relatively small for machine learning models. With a higher sample size, the variation in the cross-validation folds would most likely decrease. Second, the frequency of measurements could be increased. More frequent measurements allows for more fine grained analytic opportunities, and for earlier intervention and support in practice. Third, future research could be complemented with qualitative measures to gain deeper insights into the personal experiences, coping strategies, and psychological states of recruits. To address these limitations, practical feasibility should be considered, as the duration and frequency with which individuals can be measured are often restricted in high-stakes training or selection programs. As a final suggestion for future research, intervention studies could be conducted that may have an (positive) influence on dropout rates. Examples of such interventions are targeted physical conditioning programs, as well as evidence-based resilience or mental toughness training^{38,39}.

Taken together, our study builds on previous research that has highlighted the importance of psychological and physical factors in predicting dropout from special forces programs. The longitudinal design of our study allowed to identify future dropouts based on psychological and physical processes during the stressful selection period. Besides, by picking the right threshold, individuals at risk of dropout could sometimes be identified weeks in advance. This allows for targeted interventions and support, which could subsequently improve success rates and reduce the personal and human resource costs associated with high dropout rates.

Data availability

The analytic code can be accessed at the Open Science Framework repository: <https://osf.io/7xp9v/>. The data contains sensitive information about the subjects and is therefore restricted from openly sharing it. Requests can be made by researchers affiliated with universities or independent, non-commercial research institutes via the same repository.

Received: 5 September 2024; Accepted: 20 January 2025

Published online: 25 January 2025

References

1. Gayton, S. D. & Kehoe, E. J. A prospective study of character strengths as predictors of selection into the australian army special force. *Mil. Med.* **180**, 151–157 (2015).
2. Beattie, S., Du Preez, T., Hardy, L. & Arthur, C. What do you bring to the table? Exploring psychological attributes that predict successful military training. *Mil. Psychol.* **37**, 50–61 (2023).
3. Huijzer, R. et al. Personality traits of special forces operators: Comparing commandos, candidates, and controls. *Sport Exerc. Perform. Psychol.* **11**, 369–381 (2022).
4. Jackson, J. J., Thoemmes, F., Jonckmann, K., Lüdtke, O. & Trautwein, U. Military training and personality trait development: Does the military make the man, or does the man make the military?. *Psychol. Sci.* **23**, 270–277 (2012).
5. Rolland, J. P., Parker, W. D. & Stumpf, H. A psychometric examination of the french translations of neo-pi-r and neo-ffi. *J. Pers. Assess.* **71**, 269–291 (1998).
6. Sørlie, H. O., Hetland, J., Dysvik, A., Fosse, T. H. & Martinsen, Ø. L. Person organization fit in a military selection context. *Mil. Psychol.* **32**, 237–246 (2020).
7. Tedefholm, P. G., Sjöberg, A. & Larsson, A. C. Personality traits among swedish counterterrorism intervention unit police officers: A comparison with the general population. *Pers. Indiv. Dif.* **168**, 110411. <https://doi.org/10.1016/j.paid.2020.110411> (2021).
8. Bartone, P. T., Roland, R. R., Picano, J. J. & Williams, T. J. Psychological hardness predicts success in us army special forces candidates. *Int. J. Select. Assess.* **16**, 78–81 (2008).
9. Johnsen, B. H. et al. Psychological hardness predicts success in a Norwegian armed forces border patrol selection course. *Int. J. Select. Assess.* **21**, 368–375 (2013).
10. De Beer, M. & van Heerden, A. Exploring the role of motivational and coping resources in a special forces selection process. *J. Ind. Psychol.* **40**, 1–13 (2014).
11. Gruber, K. A., Kilcullen, R. & Iso-ahola, S. Effects of psychosocial resources on elite soldiers' completion of a demanding military selection program. *Mil. Psychol.* **21**, 427–444 (2009).
12. Vaara, J. P. et al. Can physiological and psychological factors predict dropout from intense 10-day winter military survival training?. *Int. J. Environ. Res. Public Health* **17**, 9064. <https://doi.org/10.3390/ijerph17239064> (2020).
13. Tedefholm, P. G., Larsson, A. C. & Sjöberg, A. Predictors in the Swedish counterterrorism intervention unit selection process. *Scand. J. Work Organ. Psychol.* **8**, 1–13. <https://doi.org/10.16993/sjwop.194> (2023).
14. Farina, E. K. et al. Physical performance, demographic, psychological, and physiological predictors of success in the us army special forces assessment and selection course. *Physiol. Behav.* **210**, 112647. <https://doi.org/10.1016/j.physbeh.2019.112647> (2019).
15. Farina, E. K. et al. Anthropometrics and body composition predict physical performance and selection to attend special forces training in united states army soldiers. *Mil. Med.* **187**, 1381–1388 (2022).
16. Den Hartigh, R. J. R., Van Dijk, M. W., Steenbeek, H. W. & Van Geert, P. L. C. A dynamic network model to explain the development of excellent human performance. *Front. Psychol.* **7**, 532. <https://doi.org/10.3389/fpsyg.2016.00532> (2016).
17. Saxon, L. et al. Continuous measurement of reconnaissance marines in training with custom smartphone app and watch: Observational cohort study. *JMIR mHealth and uHealth* **8**, e14116. <https://doi.org/10.2196/14116> (2020).
18. Kellmann, M. & Kölling, S. *Recovery and Stress in Sport: A Manual for Testing and Assessment* (Routledge, 2019).
19. Tait, J. L., Drain, J. R., Bulmer, S., Gastin, P. B. & Main, L. C. Factors predicting training delays and attrition of recruits during basic military training. *Int. J. Environ. Res. Public Health* **19**, 7271. <https://doi.org/10.3390/ijerph19127271> (2022).

20. Gerber, M. et al. Perceived recovery and stress states as predictors of depressive, burnout, and insomnia symptoms among adolescent elite athletes. *Sports Psychiatry* **2**, 13–22 (2022).
21. Lövdal, S. S., Den Hartigh, R. J. R. & Azzopardi, G. Injury prediction in competitive runners with machine learning. *Int. J. Sports Physiol. Perform.* **16**, 1522–1531 (2021).
22. Yarkoni, T. & Westfall, J. Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspect. Psychol. Sci.* **12**, 1100–1122 (2017).
23. Dijksma, I., Hof, M. H., Lucas, C. & Stuiver, M. M. Development and validation of a dynamically updated prediction model for attrition from marine recruit training. *J. Strength Cond. Res.* **36**, 2523. <https://doi.org/10.1519/JSC.00000000000003910> (2022).
24. Den Hartigh, R. J. R. et al. Resilience in sports: A multidisciplinary, dynamic, and personalized perspective. *Int. Rev. Sport Exerc. Psychol.* **17**, 564–586 (2024).
25. Brauers, J. J. et al. Monitoring the recovery-stress states of athletes: Psychometric properties of the acute recovery and stress scale and short recovery and stress scale among dutch and flemish athletes. *J. Sports Sci.* **42**, 189–199 (2024).
26. Hofman, J. M. et al. Integrating explanation and prediction in computational social science. *Nature* **595**, 181–188 (2021).
27. McShane, B. B., Gal, D., Gelman, A., Robert, C. & Tackett, J. L. Abandon statistical significance. *Am. Stat.* **73**, 235–245. <https://doi.org/10.1080/00031305.2018.1527253> (2019).
28. Blaom, A. D. et al. MLj: A julia package for composable machine learning. *J. Open Source Softw.* **5**, 2704. <https://doi.org/10.21105/joss.02704> (2020).
29. Bénard, C., Bliau, G., Da Veiga, S. & Scornet, E. SIRUS: Stable and interpretable rule set for classification. *Electron. J. Stat.* **15**, 427–505 (2021).
30. Huijzer, R., Blaauw, F. & den Hartigh, R. J. R. SIRUS.jl: Interpretable machine learning via rule extraction. *J. Open Source Softw.* **8**, 5786. <https://doi.org/10.21105/joss.05786> (2023).
31. Huijzer, R., De Jonge, P., Blaauw, F. J., Baatenburg de Jong, M., De Wit, A. & Den Hartigh, R. J. R. Predicting special forces dropout via explainable machine learning. *Eur. J. Sport Sci.* <https://doi.org/10.1002/ejsc.12162> (2024).
32. Bliau, G. & Scornet, E. A random forest guided tour. *Test* **25**, 197–227 (2016).
33. Desgagne-Bouchard, J. et al. *EvoTrees.jl*; <https://doi.org/10.5281/zenodo.10569605> (2024).
34. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* **22**, 785–794. <https://doi.org/10.1145/2939672.2939785> (2016).
35. Ke, G. et al. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **30**, 3146–3154 (2017).
36. Bandura, A. *Self-efficacy: The Exercise of Control* (Freeman, 1997).
37. Beattie, S., Woodman, T., Fakhey, M. & Dempsey, C. The role of performance feedback on the self-efficacy–performance relationship. *Sport Exerc. Perform. Psychol.* **5**, 1–13 (2016).
38. Fletcher, D. & Sarkar, M. Mental fortitude training: An evidence-based approach to developing psychological resilience for sustained success. *J. Sport Psychol. Acti.* **7**, 135–157 (2016).
39. Fitzwater, J. P. J., Arthur, C. A. & Hardy, L. “The tough get tougher”: Mental skills training with elite military recruits. *Sport Exerc. Perform. Psychol.* **7**, 93–107 (2018).

Acknowledgements

The authors would like to thank Maurits Baatenburg de Jong for his contribution in the development of the research project.

Author contributions

RdH, RH, FJB, AdW, and PdJ contributed to the conceptualization and implementation of the study. RH, RdH, PdJ, and FJB contributed to the data analysis. RdH, RH, FJB, AdW, and PdJ contributed to the writing of the manuscript.

Declarations

Competing interests

Our research received funding from the Ministry of Defence, the Netherlands. This organization had no role in the conceptualization, design, analysis, or preparation of the manuscript. AdW is researcher of the Human Performance Team at the Commando Corps and receives salary from the Ministry of Defence. RdH, RH, FJB, and PdJ declare no potential conflict of interest.

Additional information

Correspondence and requests for materials should be addressed to R.J.R.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025