

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Joensuu, Laura; Rautiainen, Ilkka; Äyrämö, Sami; Syväoja, Heidi J; Kauppi, Jukka-Pekka; Kujala, Urho M; Tammelin, Tuija H

Title: Precision exercise medicine : predicting unfavourable status and development in the 20-m shuttle run test performance in adolescence with machine learning

Year: 2021

Version: Published version

Copyright: © Author(s) (or their employer(s)) 2021. Published by BMJ.


Rights: CC BY-NC 4.0

Rights url: <https://creativecommons.org/licenses/by-nc/4.0/>

Please cite the original version:

Joensuu, L., Rautiainen, I., Äyrämö, S., Syväoja, H. J., Kauppi, J.-P., Kujala, U. M., & Tammelin, T. H. (2021). Precision exercise medicine : predicting unfavourable status and development in the 20-m shuttle run test performance in adolescence with machine learning. *BMJ Open Sport & Exercise Medicine*, 7(2), Article e001053. <https://doi.org/10.1136/bmjsem-2021-001053>

Precision exercise medicine: predicting unfavourable status and development in the 20-m shuttle run test performance in adolescence with machine learning

Laura Joensuu ^{1,2}, Ilkka Rautiainen,³ Sami Äyrämö,³ Heidi J Syväoja,² Jukka-Pekka Kauppi,³ Urho M Kujala,¹ Tuija H Tammelin²

To cite: Joensuu L, Rautiainen I, Äyrämö S, *et al.* Precision exercise medicine: predicting unfavourable status and development in the 20-m shuttle run test performance in adolescence with machine learning. *BMJ Open Sport & Exercise Medicine* 2021;**7**:e001053. doi:10.1136/bmjsem-2021-001053

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjsem-2021-001053>).

LJ and IR contributed equally.

Accepted 7 May 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland
²LIKES Research Centre for Physical Activity and Health, Jyväskylä, Finland
³Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland

Correspondence to
Mrs Laura Joensuu;
laura.p.joensuu@jyu.fi

ABSTRACT

Objectives To assess the ability to predict individual unfavourable future status and development in the 20m shuttle run test (20MSRT) during adolescence with machine learning (random forest (RF) classifier).

Methods Data from a 2-year observational study (2013–2015, 12.4±1.3 years, n=633, 50% girls), with 48 baseline characteristics (questionnaires (demographics, physical, psychological, social and lifestyle factors), objective measurements (anthropometrics, fitness characteristics, physical activity, body composition and academic scores)) were used to predict: (Task 1) unfavourable future 20MSRT status (identification of individuals in the lowest 20MSRT tertile after 2 years), and (Task 2) unfavourable 20MSRT development (identification of individuals with 20MSRT development in the lowest tertile among adolescents with baseline 20MSRT below median level).

Results Prediction performance for future 20MSRT status (Task 1) was (area under the receiver operating characteristic curve, AUC) 83% and 76%, sensitivity 80% and 60%, and specificity 78% and 79% in girls and boys, respectively. Twenty variables showed predictive power in boys, 14 in girls, including fitness characteristics, physical activity, academic scores, adiposity, life enjoyment, parental support, social status in school and perceived fitness.

Prediction performance for future development (Task 2) was lower and differed statistically from random level only in girls (AUC 68% and 40% in girls and boys).

Conclusion RF classifier predicted future unfavourable status in 20MSRT and identified potential individuals for interventions based on a holistic profile (14–20 baseline characteristics). The MATLAB script and functions employing the RF classifier of this study are available for future precision exercise medicine research.

INTRODUCTION

Precision medicine is prevention and treatment strategies of diseases taking the individual variability into account.¹ Recently, a similar concept called precision exercise medicine was brought forward where the role of physical activity (PA) and cardiorespiratory

Key messages

What is already known

- The 20-m shuttle run test is commonly used in adolescents to estimate unfavourable cardiorespiratory fitness
- Currently used methods for assigning interventions based on the 20-m shuttle run test have limitations in individual level accuracy

What are the new findings

- Machine learning algorithm was able to identify adolescents with unfavourable future 20 m shuttle run test (20MSRT) status based on 14 baseline characteristics in girls, and 20 in boys.
- This study provides an example with attached MATLAB script and functions how to use machine learning in precision exercise medicine.
- Adolescents' overall physical, psychological and social status are recommended to be assessed before deciding on interventions based on the 20MSRT score.

fitness (CRF) in health enhancement was acknowledged.² However, currently, the focus in precision exercise medicine is mainly on exploring treatment procedures and exercise response variability in adults.^{2–3} Nevertheless, many chronic diseases have origins already in early childhood.⁴ Prevention strategies warrant more focus on children and adolescents, especially as health risks have associations with CRF⁵ and reversibility with exercise interventions in this age group.⁶

The 20-m shuttle run test (20MSRT) is the most commonly used field test to estimate CRF.⁷ Low 20MSRT score has adverse associations with many aspects of children's and adolescents' daily lives. Previous studies have reported 20MSRT associated with lower overall physical performance,⁸ poorer tissue health (including adiposity,⁸ brain⁹ and bone tissue¹⁰), lower cardiometabolic and psychosocial health, and cognitive performance.⁸

However, currently used methods to assign interventions based on the 20MSRT have limitations by their individual level accuracy.^{7,11} The ability to predict 20MSRT prospects during adolescence would enhance the identification of potential individuals for lifestyle interventions.

Machine learning (ML)-based pattern recognition approaches have emerged as promising alternatives to traditional statistical methods in precision exercise medicine.³ Random forest (RF) is a commonly used ML algorithm. Contrary to other high learning capacity methods, such as neural networks and support vector machines, major advantages of RF include that the extensive tuning of hyperparameters is not required and overfitting the model is usually of lesser concern. An additional benefit especially suited for our research goals is extracting the estimates of importance for each variable in the data.^{12,13} The main aim of this study was to evaluate the performance of RF on predicting future individual unfavourable 20MSRT status and development during adolescence based on 48 baseline variables, including physical, psychological and social indicators. Two prediction tasks were implemented: (Task 1) prediction of unfavourable future 20MSRT status (identification of individuals in the lowest 20MSRT tertile after 2 years), and (Task 2) prediction of unfavourable 20MSRT development in adolescents with limitations in their 20MSRT performance (identification of individuals with 20MSRT development in the lowest tertile among adolescents with baseline 20MSRT below median level). Task 1 focuses on the normal population, while Task 2 focuses specifically on children and adolescents who are more likely to experience the adverse outcomes related to lower 20MSRT performance.

We hypothesised that the baseline data contain variables that can predict future 20MSRT status and development. A secondary aim was to evaluate with a data-driven approach the best predictors of unfavourable 20MSRT prospects out of a wide range of baseline characteristics. We furthermore provide the predictive modelling algorithms used in this study for future research.

METHODS

Study design and participants

Secondary data analyses were performed for data collected in a 2-year longitudinal observational study (2013–2015) related to the Finnish Schools on the Move programme.¹⁴ Data contained information from 971 students (mean 12.5±1.3 years, min 9.2 years, max 15.3 years, 52% girls). The sample of this study was further reduced to 633 (50% girls) (Task 1) and 300 subjects (50% girls) (Task 2), described in more detail in the Predictive modelling section. The data were collected at baseline during Spring and Fall semesters (1 May 2013 and 8 November 2013) and at follow-up during the Spring semester (1 May 2015) in nine Finnish public schools. The baseline and follow-up measurements during the Spring semester were performed within the same calendar week in each school.

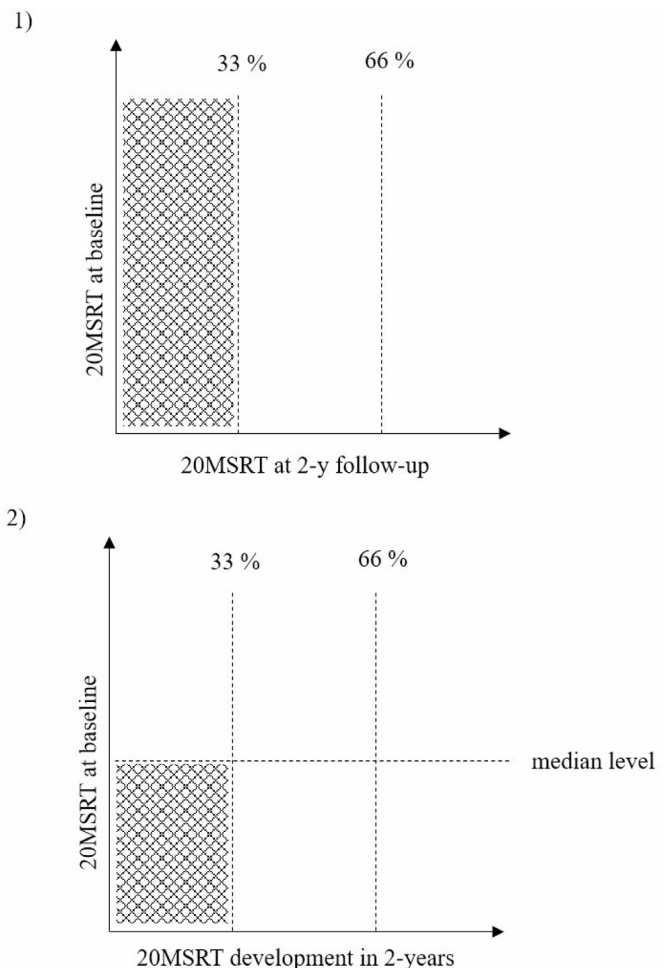


Figure 1 Prediction tasks were (A) unfavourable future 20MSRT status (identification of individuals in the lowest 20MSRT tertile after 2 years), and (B) unfavourable 20MSRT development in adolescents with limitations in their 20MSRT performance (identification of individuals with 20MSRT development in the lowest tertile among adolescents with baseline 20MSRT below median level). Both of these target tertile groups are highlighted in grey. The exact outcome variables to be predicted were (A) status of 20MSRT at follow-up (laps) and (B) absolute change between baseline and follow-up (in laps). The median level refers to the 50% performance level that was determined for each age cohort and both sexes separately to select the study sample in Task 2. The 33%, 66% cut-offs represent the tertiles used in Tasks 1 and 2. In both tasks, the outcome tertiles were determined for each age cohort and both sexes separately. 20MSRT, 20-m shuttle run test.

Forty-eight baseline variables (see the full list in online supplemental information document 1) were used in the prediction tasks (figure 1). Information regarding participants' demographics, physical, psychological and social factors was obtained from self-assessment questionnaires and non-invasive objective measurements.

Self-assessment questionnaires

Participants completed two web-based questionnaires at baseline. Due to the extensiveness of the questionnaires, the data were collected in two parts: a first round

during the Spring 2013 and a second during the Fall 2013 semester (see division in online supplemental information document 1). In addition to basic demographic information (age and sex), the questionnaires assessed student's perceptions of their physical, psychological, and social status and health-related behaviour, for example, subjective evaluation of PA,¹⁵ pubertal status on Tanner scale,¹⁶ societal status of the family,¹⁷ perceived health,¹⁸ and cigarette, alcohol, and unhealthy food consumption.

Objective measurements

All objective measurements were performed during the Spring semester of 2013. Body height was measured with an accuracy of 0.1 cm (Charder HM 200P scale). Body composition and mass were measured in light clothing using a bioelectrical impedance analysis device (InBody 720, Biospace Co.). Waist circumference was measured according to WHO guidelines.¹⁹

Physical fitness measurements were conducted in schools during the school day, with measurements included in the Finnish national Move!—monitoring system for physical functional capacity²⁰: 20MSRT, push-up, curl-up, 5-leaps test, throwing–catching combination test and flexibility. Procedures for fitness measurements are described in detail in our previous baseline article.²¹ The 20MSRT followed the Eurofit protocol and was recorded as laps run until voluntary exhaustion.

Device-based PA was evaluated using a hip-worn accelerometer (ActiGraph GT3X+, wGT3X+, Pensacola, Florida, USA) during a 7-day measurement period with raw 30 Hz acceleration, standard filtering and 15 s epoch conversion. Evenson criteria were used to define sedentary (<100 counts/min (cpm)), light (101–2295 cpm), moderate-to-vigorous (2296–20 000 cpm) physical activity (MVPA).²² The valid amount of data was set for at least 500 min/day (between 07:00 and 23:00),²³ including at least 2 weekdays and 1 weekend day. Activity intensities were converted into weighted mean values per day (eg, $MVPA = ((\text{average MVPA min/day of weekdays} \times 5 + \text{average MVPA min/day of weekend days} \times 2) / 7)$).

Academic scores (teacher-rated grade points) included grade point average (GPA) and grade point in physical education. Regional education services provided the data.

Predictive modelling

The predictive modelling algorithms are provided in a data file (online supplemental information document 2) and available for future studies. All analyses were performed using MATLAB R2018a with the Statistics and Machine Learning Toolbox and conducted separately for both sexes.

The flow chart of predictive modelling is presented in figure 2. Please see the full details of the analyses in the online supplemental information document 3.

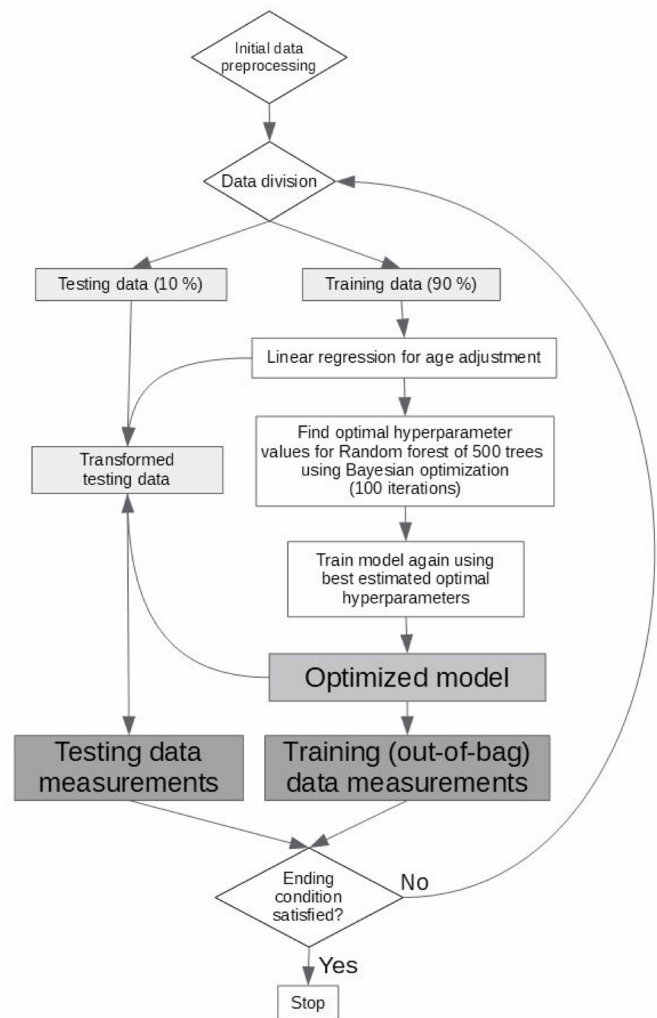


Figure 2 The flow chart of predictive modelling.

Initial data preprocessing

Target variable formatting

The target variables to be predicted were (1) status of 20MSRT at follow-up and (2) absolute change in 20MSRT test result (laps) between the baseline and the follow-up (figure 1). The tertile groups were determined for both sexes and each age cohort separately. From a total of 971 observations, the 20MSRT baseline level could be determined for 871 students. A total of 633 participants were included in the Task 1 analysis. Exclusion criteria included participants with no result from the 20MSRT follow-up test. Here the missing mechanism was assumed to be missing completely at random. Altogether 300 adolescents were included in the Task 2 analyses. These participants had a recorded result for both 20MSRT tests, and their baseline 20MSRT result was below the age-specific and sex-specific median level. Here participants with no results from either of the two 20MSRT tests were excluded from the analysis.

Variables heavily dependent on age (see online supplemental information document 3 for a list) were age-adjusted using linear regression. The age-adjustment was first performed for the training data, and the residual

information was thereafter used to age adjust the corresponding variables in the testing data.

Data division

The 10-fold cross-validation (CV) was used for model assessment where the data set (eg, in Task 2: $n=150$ boys, $n=150$ girls) was divided into 10 subsamples ($n=15$ participants per subsample) called folds. Nine folds were then used as the training data (90% of the whole data set, to fit the tree model and estimate the variable importance values) and one fold as the testing data (10% of the whole data set, to evaluate the prediction accuracy on an independent sample). The procedures of training and prediction were then performed for these folds in a rotating manner, where eventually, all the folds had been used for training and testing. These procedures provided in total a set of 10 data-driven prediction models. The average performance of these 10 prediction models is shown in the Results section.

Training and prediction

RF is an ML method that grows a forest of multiple de-correlated decision trees.¹³ This forest of trees is thereafter employed as a voting ensemble, where each tree votes for the group of a single student (ie, does the individual belong to the lowest, middle or highest tertile group). The final predicted group for the student has the most votes in the whole forest.^{12 13} For each of the 10 folds, the trained model was employed to predict the testing portion of data. The area under the receiver operating characteristic curve (AUC), sensitivity and specificity metrics were recorded. A t-test in MATLAB was performed for AUC results to determine if the mean was significantly ($p<0.05$) above the random level of 0.5.

The prediction strength of each feature is estimated using the out-of-bag (OOB) samples of each tree, that is, training data samples that have not been used when forming the tree. The OOB samples are shown to the tree, and the F1-score measure (online supplemental information document 3) of the predictions are recorded. Then the values of each feature are permuted one-by-one randomly, and after each permutation, the classification error is calculated again. This procedure is applied to all the trees in the forest. The final estimate of individual feature importance is the difference between the original classification error and the randomly permuted feature classification error, averaged for all the trees.¹² The final list of statistically significant ($p<0.05$) predictors (online supplemental information document 5) was then formed, using MATLAB's t-test function. T-test was again performed for each predictor to determine which feature importance estimates were significantly above the mean of zero, indicating that they had predictive power.

The direction of the associations

The directions for the significant variables (significance set at $p<0.05$, presented in figures 3 and 4) were estimated using a separate receiver operating characteristic

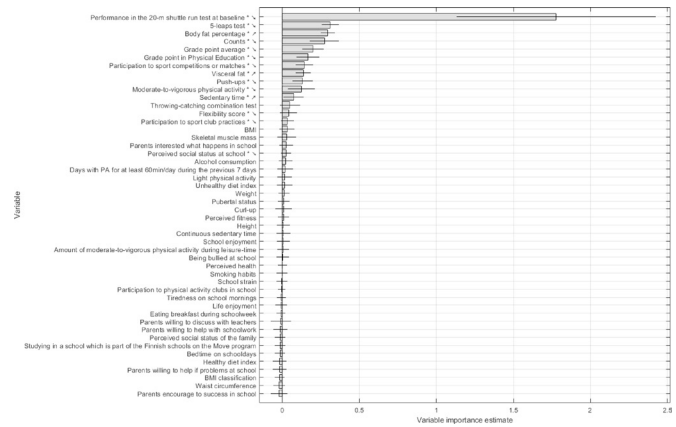


Figure 3 Best predictors for Task 1 in girls (20MSRT performance in the lowest tertile at 2-year follow-up). Statistically significant predictors are marked with * ($p<0.05$). Descending arrow (\searrow): low values are associated with 20MSRT in the lowest tertile. Ascending arrow (\nearrow): high values are associated with 20MSRT in the lowest tertile. The solid line represents the 95% CI. Variable importance estimate indicates the significance of the predictor. 20MSRT, 20-m shuttle run test.

(ROC) analysis.²⁴ The analysis was performed for the two prediction tasks, separately for girls and boys. Here, the whole data were employed without separation to training and testing data sets. Each variable in the data was then used one by one. The idea was to see how well a single variable can separate the data into two groups: the first group contained the lowest tertile and the second group contained the two upper tertiles. The separation threshold in the analysis is then changed step-by-step. At each step, two metrics need for the ROC curve, sensitivity and specificity, are recorded. For each variable, we recorded the AUC value. The AUC value was then compared with

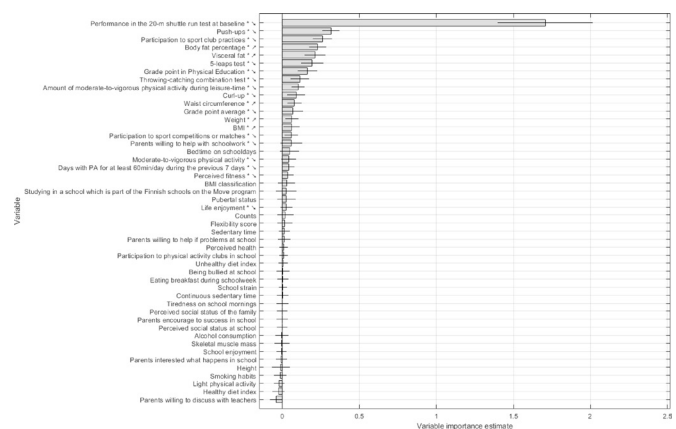


Figure 4 Best predictors for Task 1 in boys (20MSRT performance in the lowest tertile at 2-year follow-up). Statistically significant predictors are marked with * ($p<0.05$). Descending arrow (\searrow): low values are associated with 20MSRT in the lowest tertile. Ascending arrow (\nearrow): high values are associated with 20MSRT in the lowest tertile. The solid line represents the 95% CI. Variable importance estimate indicates the significance of the predictor. 20MSRT, 20-m shuttle run test.

Table 1 Descriptives of the study sample at baseline

	Boys (n=319)	Girls (n=314)
Age (years)	12.5±1.3	12.3±1.3
Height (cm)	156.1±11.7	154.1±9.6
Weight (kg)	46.1±12.9	44.8±10.5
BMI (kg/m ²)	18.6±3.3	18.7±3.1
20MSRT (laps)	45.3±19.0	36.4±15.2
20MSRT centile*	60th	70th
MVPA (min/day)	58.0±22.4	48.3±17.9
Pubertal status [†]	2.6±1.0	2.5±0.9

Units are means and SD unless other mentioned.

*International normative values by Tomkinson *et al*, 2016.

[†]Classification is based on self-assessment questionnaire and Tanner's scale.

BMI, body mass index; 20MSRT, 20-m shuttle run test; MVPA, accelerometry-based moderate-to-vigorous physical activity.

the random level (0.5). If the value was higher than the random level, we assumed that the variable information is applied correctly. The associated direction was that the higher the variable value, the higher the probability of the student belonging to the lowest tertile. Additionally, if the AUC value was lower than 0.5, a simple transformation of multiplying all the variable values with the number -1 was made, and the AUC was then calculated again. In this case, the associated direction was inverted: the lower the variable value, the higher the probability of belonging to the lowest tertile. The results of the ROC analysis are presented in online supplemental information document 4.

Patient and public involvement

Patients or the public were not involved in designing, analysing or interpreting this study.

RESULTS

The characteristics of the study sample are described in [table 1](#). Participants' average performance in the 20MSRT was 45.3 and 36.4 laps at baseline in boys and girls, representing the 60th and 70th centile in the international normative values for 20MSRT.

Prediction performance

The ability of the RF method to predict unfavourable future 20MSRT status (Task 1) is presented in [table 2](#). The AUC values were higher in girls (0.83) than in boys (0.76), both statistically higher than the random level of 0.5 ($p<0.001$). Sensitivity (individuals correctly predicted to belong to the lowest performance tertile) was higher in girls (0.80) than in boys (0.60). Specificity (individuals correctly predicted not to belong to the lowest performance tertile) was 0.78 in girls and 0.79 in boys.

The ability of the RF method to predict unfavourable 20MSRT development in a group of adolescents with baseline 20MSRT below the median level (Task 2) is presented in [table 2](#). The prediction performance of ML was lower in these analyses. The AUC values were higher in girls (0.68) than boys (0.40), but only girls' predictions statistically differed from the random level ($p=0.001$). Sensitivity (individuals correctly predicted to belong to the lowest development group) was higher in girls (0.59) than in boys (0.13). Specificity (individuals correctly predicted not to belong to the lowest development group) was 0.70 in girls and 0.79 in boys.

Best predictors of 20MSRT prospects

The statistically significant predictors for Tasks 1 and 2 are represented in [figures 3 and 4](#). The x-axis in the figures gives the estimate for variable importance, calculated using the increase or decrease in classification error when the predictor values are randomly permuted separately for each predictor. The higher the estimate, the higher is the significance of the predictor. Please see detailed information related to the direction and statistical significance of the variables in online supplemental information document 4. The top predictor for Task 1 was 20MSRT performance at baseline, both in boys and girls ($p<0.001$, [figures 3 and 4](#)), indicating that low initial 20MSRT performance predicts low performance also after 2 years.

Girls had 13 additional predictors ([figure 3](#)): low performance in other physical fitness tests (5-leaps test ($p<0.001$), push-ups ($p<0.001$) and flexibility score ($p=0.049$)), high markers of adiposity (body fat percentage ($p<0.001$) and visceral fat ($p<0.001$)), low

Table 2 The overall prediction performance of the unfavourable future 20MSRT status and development

	AUC	95 % CI	P value	Sensitivity	95 % CI	Specificity	95 % CI
Task 1	Unfavourable future 20MSRT status (identification of individuals in the lowest 20MSRT tertile after 2 years)						
Girls	0.83	0.76 to 0.90	<0.001	0.80	0.69 to 0.91	0.78	0.74 to 0.82
Boys	0.76	0.71 to 0.81	<0.001	0.60	0.52 to 0.68	0.79	0.74 to 0.84
Task 2	Unfavourable 20MSRT development (identification of individuals with 20MSRT development in the lowest tertile among adolescents with baseline 20MSRT below median level)						
Girls	0.68	0.60 to 0.76	0.001	0.59	0.50 to 0.68	0.70	0.59 to 0.81
Boys	0.40	0.29 to 0.51	0.108	0.13	0.04 to 0.22	0.79	0.70 to 0.88

P value: statistical difference of the AUC value from the random level of 0.5; Sensitivity: individuals correctly predicted to belong to the explored group; Specificity: individuals correctly predicted not to belong to the explored group.

AUC, area under the receiver operating characteristic curve; ;20MSRT, 20-m shuttle run test.

markers of PA (accelerometry-based counts ($p<0.001$), MVPA ($p=0.003$), participation to sport club practices ($p=0.025$) or competitions ($p<0.001$) and high percentage of accelerometry-based sedentary time ($p=0.009$)), low academic scores (GPA and grade point in physical education (both $p<0.001$)) and low perceived social status in school ($p=0.015$), all predicting placement in the lowest 20MSRT tertile after 2 years.

In addition to the baseline 20MSRT performance, boys had 19 additional predictors (figure 4): low performance in other physical fitness tests (push-ups ($p<0.001$), 5-leaps test ($p<0.001$), throwing–catching combination test ($p<0.001$) and curl-up ($p=0.001$)), high markers of adiposity (body fat percentage ($p<0.001$), visceral fat ($p<0.001$), waist circumference ($p<0.001$), weight ($p<0.001$) and BMI ($p=0.005$)), low academic scores (grade point in physical education ($p<0.001$), and GPA ($p=0.015$)), low markers of PA (participation to sport club practices ($p<0.001$) or competitions ($p=0.001$), self-reported PA status (two questions: $p<0.001$ and $p=0.006$) and accelerometry-based MVPA ($p=0.020$)), low parents' willingness to help with schoolwork ($p=0.045$), low perceived fitness ($p=0.007$) and low life enjoyment ($p=0.042$), all predicting future placement in the lowest 20MSRT performance tertile after 2 years.

As prediction performance for 20MSRT development was below 0.7 for both sexes, the best predictors are recommended to be interpreted with caution. These results are described in online supplemental information document 5.

DISCUSSION

Main findings

ML approach was able to predict, based on baseline characteristics, unfavourable future 20MSRT status with 0.76–0.83 (AUC) accuracy. Prediction performance was better in girls than in boys (eg, sensitivity values 0.80 in girls and 0.60 in boys). The prediction performance declined when predicting unfavourable 20MSRT development in a group of adolescents with an initial 20MSRT below the median level. These findings indicate that ML was able to identify potential individuals for interventions. Additionally, future fitness status might be easier to predict than development, at least in a group of adolescents with more homogeneous 20MSRT performance capacity.

Best predictors of individual fitness development

Our findings showed that baseline 20MSRT performance was the best predictor of future performance in a large group of adolescents. However, this study highlighted 13–19 variables (out of 48 variables) with predictive power. These variables included a low performance in other field-based physical fitness tests, low perceived fitness, high markers of adiposity, low markers of PA, low academic achievement in school, low grade in physical education, low life enjoyment, low parental support and low perceived social status at school. These findings

indicate that multiple factors, that is, adolescents' overall physical, psychological and social well-being, contribute to the trajectory of the 20MSRT during adolescence. This information adds to the previous body of research where performance development is typically examined through growth and maturation ignited morphological changes.²⁵

Precision exercise medicine prospects

These promising findings also provide new prospects for precision exercise medicine in adolescents. Findings suggest that preventive measures linked to the 20MSRT score benefit from the ML-enabled holistic approach. In ML, patterns are explored from the data. This has benefits as data-driven characteristic profiles can be recognised if such exist in the data. Furthermore, the CV technique helps overcome a phenomenon where models or thresholds created with traditional statistics tend to fit poorly with other data sets or future individual observations.²⁶ An ML approach is recommended to be considered in future precision exercise medicine studies aiming to identify potential individuals for interventions.

Our findings indicated that information from adolescents' overall physical, psychological and social status provides additional value over evaluating only an individual's 20MSRT score. Potential use-cases are, for example, the national or regional fitness monitoring systems where a large number of children and adolescents are tested (up to >90% of age-cohort). Resources for interventions are typically limited and necessary to be directed for correct individuals. The next steps to use this method in practice would be to train the final model with selected feasible variables and to collect independent test data that the model could be evaluated against. To reduce the number of variables, for example, to indicate PA, it is possible to employ a stepwise variable elimination method to RF to select only the best variable.²⁷

It is, however, important to use ML methods and computational power robustly. The availability of ML libraries and computational power lead easily to data fishing. This means that a fair application of CV techniques must assess the generalisation ability of the models, and the risk of chance findings should be eliminated using permutation testing or other relevant techniques. In the present framework, these aspects of ML application have been considered carefully.

Strengths and limitations

The strengths of this study were the novel application for RF and the approach to predict individual fitness development in apparently healthy adolescents, the extensiveness of the variables in the data sample, robust analyses and measurements performed by educated professionals. Limitations include the 2-year duration of the study—more prominent changes could have potentially emerged with a longer follow-up period. The data sample was limited by its size (eg, $n=50$ in the lowest tertile in Task 2), possibly influencing prediction performance. There is also room for improvement in handling

the importance of variables. For example, it is possible to employ a stepwise variable elimination method to RF to reduce the effect of multicollinearity in data. The study used a sample from an observational study. Despite the efforts, sampling bias might exist and affect the generalisability of the findings to the adolescent population.

Conclusion

With the ML approach, we could predict unfavourable future 20MSRT status based on 14–20 baseline characteristics and identify potential individuals for interventions. These promising findings support adopting a more holistic approach, taking physical and psychological and social factors into account in large-scale fitness monitoring systems. The ML algorithms used in this study are provided for future research.

Acknowledgements The authors would like to thank the schools, children and their guardians who helped us to facilitate this research. We would also like to thank Dr Alan Barker from Children's Health and Exercise Research Centre, University of Exeter, UK, for valuable comments during the early manuscript preparation phase.

Contributors LJ, IR, SÅ, HJS, UMK and THT contributed to planning this work. IR, J-PK and SÅ contributed to analyses. All authors contributed to the interpretation of the data, drafting and reporting the work, and revising critically the intellectual content. All authors have given final approval for this version, agreed to be accountable and are committed to resolving possible questions related to its content.

Funding This work was supported by the Juho Vainio Foundation (201410342) and the Finnish Ministry of Education and Culture (OKM/92/626/2013). IR and SÅ received funding from Business Finland and IR a grant from the Jenny and Antti Wihuri Fund.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval The original study setting was approved by the ethics committee of the University of Jyväskylä. Participants and their guardians delivered a signed informed consent. All measurements were carried out by trained personnel and in accordance with the Declaration of Helsinki.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Raw is agreed not to be shared with third parties. In other cases, data are available upon reasonable request. Please contact THT for data sharing.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Laura Joensuu <http://orcid.org/0000-0002-9544-6552>

REFERENCES

- Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015;372:793–5.
- Ross R, Goodpaster BH, Koch LG, *et al*. Precision exercise medicine: understanding exercise response variability. *Br J Sports Med* 2019;53:1141–53.
- Gevaert AB, Adams V, Bahls M, *et al*. Towards a personalised approach in exercise-based cardiovascular rehabilitation: How can translational research help? A 'call to action' from the Section on Secondary Prevention and Cardiac Rehabilitation of the European Association of Preventive Cardiology. *Eur J Prev Cardiol* 2020;27:1369–85.
- Gluckman PD, Hanson MA. The Developmental Origins of Health and Disease. In: Wintour EM, Owens JA, eds. *Early life origins of health and disease. advances in experimental medicine and biology*. Springer, 2006.
- Veijalainen A, Tompuri T, Haapala EA, *et al*. Associations of cardiorespiratory fitness, physical activity, and adiposity with arterial stiffness in children. *Scand J Med Sci Sports* 2016;26:943–50.
- Faria WF, Mendonça FR, Santos GC, *et al*. Effects of 2 methods of combined training on cardiometabolic risk factors in adolescents: a randomized controlled trial. *Pediatr Exerc Sci* 2020;32:217–26.
- Tomkinson GR, Lang JJ, Tremblay MS, *et al*. International normative 20 m shuttle run values from 1 142 026 children and youth representing 50 countries. *Br J Sports Med* 2017;51:1545–54.
- Lang JJ, Belanger K, Poitras V, *et al*. Systematic review of the relationship between 20m shuttle run performance and health indicators among children and youth. *J Sci Med Sport* 2018;21:383–97.
- Ruotsalainen I, Renvall V, Gorbach T, *et al*. Aerobic fitness, but not physical activity, is associated with grey matter volume in adolescents. *Behav Brain Res* 2019;362:122–30.
- Gracia-Marco L, Vicente-Rodríguez G, Casajús JA, *et al*. Effect of fitness and physical activity on bone mass in adolescents: the Helena study. *Eur J Appl Physiol* 2011;111:2671–80.
- Ruiz JR, Cavaero-Redondo I, Ortega FB, *et al*. Cardiorespiratory fitness cut points to avoid cardiovascular disease risk in children and adolescents; what level of fitness should raise a red flag? A systematic review and meta-analysis. *Br J Sports Med* 2016;50:1451–8.
- Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.
- Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
- Blom A, Tammelin T, Laine K, *et al*. Bright spots, physical activity investments that work: the Finnish schools on the move programme. *Br J Sports Med* 2018;52:820–2.
- Booth ML, Okely AD, Chey T. The reliability and validity of the physical activity questions in the who health behaviour in schoolchildren (HBSC) survey: a population study. *Br J Sports Med* 2001;35:263–7.
- Taylor SJ, Whincup PH, Hindmarsh PC, *et al*. Performance of a new pubertal self-assessment questionnaire: a preliminary study. *Paediatr Perinat Epidemiol* 2001;15:88–94.
- Rajala K, Kankaanpää A, Laine K, *et al*. Associations of subjective social status with accelerometer-based physical activity and sedentary time among adolescents. *J Sports Sci* 2019;37:123–30.
- Piko BF. Self-Perceived health among adolescents: the role of gender and psychosocial factors. *Eur J Pediatr* 2007;166:701–8.
- Waist Circumference and Waist-Hip Ratio: Report of a WHO Expert Consultation 2008.
- Opetushallitus. M toimintakyvyn seurantajärjestelmä. No title.
- Joensuu L, Syväoja H, Kallio J, *et al*. Objectively measured physical activity, body composition and physical fitness: cross-sectional associations in 9- to 15-year-old children. *Eur J Sport Sci* 2018;18:882–92.
- Evenson KR, Catellier DJ, Gill K, *et al*. Calibration of two objective measures of physical activity for children. *J Sports Sci* 2008;26:1557–65.
- Cooper AR, Goodman A, Page AS, *et al*. Objectively measured physical activity and sedentary time in youth: the International children's accelerometry database (ICAD). *Int J Behav Nutr Phys Act* 2015;12:113.
- Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* 2006;27:861–74.
- Armstrong N, Van Mechelen W, Backx FJ. *Oxford Textbook of Children's Sport and Exercise Medicine*. 3rd Ed. Oxford University Press, 2017.
- Shmueli, Koppius. Predictive analytics in information systems research. *MIS Quarterly* 2011;35:553–72.
- Genuer R, Poggi J-M, Tuleau-Malot C. Variable selection using random forests. *Pattern Recognit Lett* 2010;31:2225–36.

© 2021 Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ. This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>. Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.