Obesity Medicine 59 (2026) 100679

Contents lists available at ScienceDirect

Obesity Medicine

journal homepage: www.elsevier.com/locate/obmed

Machine learning prediction of adolescent obesity using physical fitness data

Tatiana Sampaio[a,b], Samuel Encarnação[a,c,d], Bruna Amaro[e], Joana Ribeiro[b,e], Luís Branquinho[b,e], António M. Monteiro[a,d], José E. Teixeira[a,f,g], Soukaina Hattabi[e,h], Andrew Sortwell[b,i], Luciano Bernardes Leite[j,*], Alexandra Malheiro[e], Pedro Rodrigues[a,d], Beat Knechtle[k,l], Pedro Flores[e], Pedro Forte[a,d,e]

[a] Department of Sports Sciences, Instituto Politécnico de Bragança, 5300-253, Bragança, Portugal
[b] Research Center in Sports, Health, and Human Development (CIDESD), Covilhã, Portugal
[c] Department of Sports Sciences, Universidad Autónoma de Madrid, 28049, Madrid, Spain
[d] Research Centre for Active Living and Wellbeing (LiveWell), Instituto Politécnico de Bragança, Bragança, Portugal
[e] CI-ISCE, Higher Institute of Educational Sciences of the Douro (ISCE Douro), Penafiel, Portugal
[f] Department of Sport Sciences, Polytechnic Institute of Guarda, Guarda, Portugal
[g] SPRINT—Sport Physical activity and health Research & Inovation Center, Rio Maior, Portugal
[h] High Institute of Sports and Physical Education of Elkef, University of Jendouba, Kef, 7100, Tunisia
[i] School of Health Sciences and Physiotherapy, University of Notre Dame Australia, Sydney, 2007, Australia
[j] Department of Physical Education, Federal University of Viçosa, Brazil
[k] Institute of Primary Care, University of Zurich, Zurich, Switzerland
[l] Medbase St. GallenAm Vadianplatz, St. Gallen, Switzerland

ARTICLE INFO

Keywords:
Obesity prediction
Adolescents
Machine learning
Physical fitness
Random forest regression
BMI classification

ABSTRACT

The escalating prevalence of obesity among adolescents has emerged as a critical global public health challenge. Machine learning techniques have been used to predict obesity in adolescents. This study aimed to develop and validate a robust obesity prediction model for adolescents using this hybrid approach, leveraging data from a diverse cross-sectional population-based study. The hybrid method combines statistical inference with non-linear machine learning to enhance prediction accuracy. Physical fitness data were collected from the FITescola® tests. Multiple tests were employed to evaluate physical fitness. Multiple Poisson's multiple regression method was applied to identify the most predictive variables set of the adolescent's body mass index (BMI) classification. The model's goodness-of-fit statistics indicate a strong fit, with a log-likelihood of −8068.6 and a Pseudo R-squared value of 0.8853, where the aerobic fitness (AF), upper limb strength (ULS) and lower limb flexibility (LLF) presented an inverse association with the adolescent's BMI. In contrast the adolescent's core strength presented a positive association with their body mass. The random forest regression showed that an average of 35 repetition on the yo-yo test predicted a healthy BMI percentile [predBMIperc = 0.31]. In addition, the model presented good validity [MAE = 0.36, MSE = 0.20, RMSE = 0.45, R2 = 0.54]. The model's strong fit and accurate random forest regression's predictions suggest that physical fitness components,

* Corresponding author.
E-mail addresses: luciano.leite@ufv.br (L.B. Leite), pedromiguelforte@gmail.com (P. Forte).

https://doi.org/10.1016/j.obmed.2025.100679
Received 19 July 2025; Received in revised form 10 November 2025; Accepted 22 December 2025
Available online 26 December 2025

such as aerobic fitness, upper limb strength, lower limb power, and core strength, play a significant role in obesity risk among adolescents.

1. Introduction

The escalating prevalence of obesity among adolescents has emerged as a critical global public health challenge, warranting urgent attention and comprehensive preventive strategies (Kansra et al., 2021). Obesity in adolescence not only poses a serious health concern in its own right but also sets the stage for adverse health outcomes in adulthood, including an increased risk of developing chronic diseases such as cardiovascular disorders, type 2 diabetes, musculoskeletal problems, and mental health issues (Raj and Kumar, 2010). Moreover, the burden of obesity extends beyond individual health, straining healthcare resources and imposing substantial economic costs on society, making it imperative to address this multifaceted issue (Tremmel et al., 2017).

During adolescence, the rapid physiological, behavioural, and environmental changes significantly impact obesity development (Kansra et al., 2021). Hormonal fluctuations, increased autonomy in food choices, and sedentary lifestyle patterns contribute to the vulnerability of this age group to obesity (Cardel et al., 2011). Consequently, identifying effective and efficient methods for predicting adolescent obesity is crucial to implementing timely interventions that can help prevent the adverse health consequences and long-term complications associated with obesity.

Body mass index (BMI) has been widely used as a standard metric for assessing weight status in adolescents due to its simplicity and accessibility (Khanna et al., 2022). BMI is calculated by dividing an individual's weight (in kilograms) by the square of their height (in meters) and is commonly expressed in units of kg/m^2 (Purnell, 2000). However, BMI has limitations, such as its inability to distinguish between fat mass and lean mass, leading to the misclassification of individuals with varying body compositions (Lin et al., 2018). Solely relying on BMI to predict obesity may overlook important contributors to the condition, hampering preventive efforts (Smith et al., 2020).

The physical fitness plays a pivotal role in an individual's overall health and well-being (Dewi et al., 2021). Physical fitness is a multifaceted concept that encompasses various components, including cardiovascular endurance, muscular strength, flexibility, and body composition (Ortega et al., 2013). Evaluating physical fitness allows for a more comprehensive assessment of an adolescent's health beyond weight-based metrics. The integration of physical fitness assessments with BMI in obesity prediction models has shown promise in enhancing the accuracy of predictions (Webb and Chen, 2021). These models provide a more comprehensive evaluation of an individual's risk, offering valuable insights for targeted interventions and preventive measures.

The FITescola® tests, a comprehensive battery of fitness assessments, have emerged as a valuable tool for evaluating physical fitness in adolescents (Henriques-Neto et al., 2020). These tests provide objective and standardized measurements of an individual's fitness levels, offering a more nuanced evaluation of their overall health (Ortega et al., 2023). Moreover, physical fitness, as measured by the FITescola® tests, has been associated with a reduced risk of obesity and related chronic diseases (Ortega et al., 2013). Finally, it is important to highlight that adolescence is a critical period for the development of obesity due to numerous physiological, behavioural, and environmental factors that interact in complex ways (Narciso et al., 2019).

Despite these advancements, accurately predicting adolescent obesity remains challenging due to the complex interplay of various factors, including genetics, environment, lifestyle, and socioeconomic status (Lee et al., 2000). However, machine learning algorithms have demonstrated promising potential in tackling complex prediction tasks, including obesity prediction (Thamrin et al., 2021). Machine learning algorithms, such as support vector machines (SVM), random forests, and neural networks, have shown potential in predicting obesity in different age groups and populations (Ferdowsy et al., 2021). These algorithms can handle complex and high-dimensional data, providing the ability to consider a wide array of predictor variables, which is particularly valuable in obesity prediction due to the multifactorial nature of the condition (Colmenarejo, 2020). Furthermore, the availability of large-scale data and advancements in computing power have facilitated the development of machine learning models that can accurately predict obesity in different populations (Colmenarejo, 2020). These models have shown particular promise in improving predictive accuracy by incorporating multiple factors, such as genetics, dietary habits, physical activity levels, and socioeconomic factors (An et al., 2022).

Recent works employing advanced machine learning architectures have further improved predictive accuracy for adolescent obesity (Li et al., 2023; Park and Lee, 2024). The pursuit for a comprehensive predictive model integrating diverse variables to anticipate adolescent obesity persists within the scientific community. A pressing need exists for the development of precise and reliable predictive models capable of discerning the propensity for obesity among adolescents. While Colmenarejo (2020) and Ferdowsy et al. (2021) demonstrated early ML models for childhood obesity, these lacked standardized physical fitness input, which this study introduces. Such models hold the potential to inform targeted interventions and personalized strategies for individuals at elevated risk, thereby mitigating the adverse health consequences associated with adolescent obesity (Sanyaolu et al., 2019). This research endeavours to significantly contribute to the field of adolescent obesity prediction and advance preventive efforts in public health. It introduces a novel methodology characterized by the synergistic fusion of statistical methodologies and artificial intelligence techniques to forecast the incidence of obesity in adolescents (Colmenarejo, 2020).

The primary aim of this study is to develop and validate a robust obesity prediction model for adolescents using this hybrid approach, leveraging data from a diverse cross-sectional population-based study (Assis et al., 2022). Unlike previous studies, this research integrates standardized physical fitness testing with machine learning for obesity prediction in adolescents, a relatively unexplored intersection of public health and AI. The model will incorporate multiple regression analysis alongside machine learning techniques, optimizing the prediction accuracy of adolescent obesity based on identified risk factors (Colmenarejo, 2020). The hypothesis of this study is that integrating physical fitness assessments, measured by Fitescola tests, with traditional body mass index percentiles will result in a more accurate and comprehensive obesity prediction model for adolescents using multiple regression analysis and machine learning techniques.

2. Methods

2.1. Study subjects

This is an observational, prospective, and cross-sectional study which aims to evaluate the prevalence of obesity by body mass index percentiles in adolescents from both sexes, and the impact of physical fitness in the adolescents' obesity status. Data from FITescola®, a Portuguese project which aims to promote healthy behaviour in children and adolescents, was used. The data collection included the BMI percentiles, abdominal adiposity, and physical fitness tests from the FITescola® that took place between September 14, 2021 (start date) and September 30, 2021 (end date). Thus, in total, 654 adolescents aged between 10 and 19 years old male 334 (51 %) and female: n = 320 (49 %) were selected to participate in this study. Furthermore, the eligibility criteria considered adolescents of both sexes, without any incapacitating conditions, and aged between 10 and 19 years, using the WHO cohort cut-offs for adolescents. The World Health Organization (WHO) considers these cut-offs to categorize adolescents into the following three categories: 1) pre-adolescence (10–14 years old) and 2) adolescence (15–19 years old); thus, this age range was selected respecting the WHO cut-offs (World Health Organization, 2024). A cross-sectional design was chosen to provide a snapshot of obesity-related fitness associations in adolescents, aligning with the FITescola® annual data structure.

2.2. Ethical Aspects

The participants signed an informed consent form, and all procedures were in accordance with the Declaration of Helsinki respecting research involving humans. The Higher Institute of Educational Sciences of the Douro's Scientific Board gave its approval to this study (PF:10.2021).

2.3. Data collection

Physical fitness data were collected from the FITescola® tests, a Portuguese project developed to assess and promote healthy behaviours in children and adolescents. Multiple tests were employed to evaluate physical fitness, including the Yo-Yo test to measure aerobic fitness (AF); abdominal curl to evaluate abdominal strength (AS); push-up test to assess upper limb strength (ULS); horizontal jump to evaluate lower limb power (LLP); the 40 m sprint time test to determine maximal running speed (RS); and the seat and reach to measure the lower limb flexibility (LLF). The FITescola® battery tests have been previously validated in the context of physical education and sport The FITescola® protocols have demonstrated strong inter-rater reliability (ICC >0.85) and standardized administration across Portuguese schools (Henriques-Neto et al., 2020). The employed protocol is described as follows.

2.4. Body mass index percentiles

The adolescents BMI was determined by following certain measurement protocols. Initially, the patients were weighed while wearing light clothing and without shoes. They straightened up and waited their time till the reading on the brand scale stabilized. We calibrated a scale to within of 100 g precision. Next, we measured the height of the adolescents. Their backs were against the stadiometer scale as they stood there barefoot and with their feet together. We positioned the stadiometer's headpiece at the top of the subject's head, compressing the higher part of the head (vertex). We set a stadiometer with a precision of millimetres. We utilized the body weight divided by the height squared (in kg/m^2) to calculate the BMI. We considered the BMI percentile cutoffs according to the Centers for Disease Control and Prevention, 2023 2023 growth reference charts: underweight: <5th percentile; healthy weight: 5th to <85th percentile; overweight: 85th to <95th percentile; and obesity: ≥95th percentile. Then, we classified the adolescents with a BMI percentile: ≥85th as under obesity risk (Centers for Disease Control and Prevention, 2023).

2.5. Horizontal jump

To evaluate the horizontal jump, a horizontal line was drawn at the starting point and reference lines were drawn every 10 cm (1 m after the starting line). A measuring tape with an accuracy of millimetres was placed perpendicular to the horizontal lines to facilitate the measurement of the distance reached. The subject was positioned standing behind the line that marked the starting point, with feet shoulder-width apart. The participant was required to jump as far as they could while bending their knees and pulling their arms behind them during a continuous motion that began with standing. Distances were measured from the starting point to the heel. Two attempts were performed to take the best result of the two evaluations in cm FITescola.

2.5.1. 40 m maximal velocity

For the evaluation of the running speed of 40 m at maximal velocity, the test of maximum speed in 40 m was applied. A 3 min warm-up for general muscle activation was performed to avoid injuries during the test. Two signalling cones were used to identify the initial and final points of the test. The subject was positioned standing behind the line that marked the starting point, with the lower limbs in the anteroposterior distance and the trunk slightly inclined forward. At the sign of "prepare, now!" from the evaluator, the stopwatch was started with an accuracy of 0.1 s, and the evaluated subject had to start a run at the highest possible speed. When the student crossed the finish line, the timer was stopped. Two trials were given, and the value recorded was the best result of the two trials (FITescola).

2.6. Aerobic fitness

For the assessment of aerobic fitness, the Yo-Yo test was applied. Two cones were used to delimit the space of 20 m between the beginning and the end of the test. The subject was positioned behind the starting line, and at the signal of "prepare, now!", the subject began the race and was required to touch the 20-m finish line upon hearing the sound signal produced by the audio. At the sound signal, the subject had to then reverse the running direction and run to the other end. If the subject reached the line before the beep, he had to wait for the new beep to run in the opposite direction. In the ideal situation, the subject would regulate their running speed to finish the 20 m right before the beep. During the course, the audio signal assisted the student in keeping track of their speed. Initially, the speed was lower (8.5 km/h) and progressively increased (0.5 km/h every minute; 1 min is equal to one stage) up to a maximum of 120 routes. The sound signal indicated the end of a 20 m course, and a triple sound signal indicated the end of each stage. The subject was required to promptly change his running direction even if he had not reached the finish line when he was unable to cross it. The subject had to remain in the test as long as possible and stop when he could no longer reach the line before the audio signal on two, not necessarily consecutive, occasions. The first foul counted towards the final score. The result was the highest number of laps performed (FITescola).

2.7. Statistical procedures

Data were stored in a Microsoft Excel® 365 spreadsheet. Continuous data were reported in average and standard deviation, and categoric data were reported in absolute and percentual values. Outputs were reported in tables and figures. Following the multiple poisson regression's statistical assumptions, we tested the predictive variable's multicollinearity. Then, we found that the variables age, body weight, abdominal perimeter, lower limb power, and 40m-sprint time scored more than 10 points for VIF of the adolescent's BMI classification. In this way, we removed the multicollinear variables from the dataset, maintaining only the variables with VIF values bellow 10 points (aerobic fitness, core strength, upper limb strength, and lower limb flexibility). In addition, we did not verify strong dependence between the final predictive variables set. After that, we applied the multiple Poisson's multiple regression method to identify the most predictive variables set of the adolescent's BMI classification (Muñoz-Pichardo et al., 2021). Poisson regression was selected due to the count-like nature of BMI percentile scores (bounded between 0 and 100) and the presence of non-normal residuals that violated linear regression assumptions. As the primary regression's outputs, we reported the coefficient of estimation (β), standard error (SE), Z scores (Z), p-value for Z statistics, and confidence intervals (CI) for the overall model and individual predictors. Exponentiated regression coefficients (IRR) allow to interpret how each fitness variable increases or decreases the likelihood of higher BMI percentiles and should be interpreted as: IRR <1 means the variable reduces obesity risk; IRR >1 means the variable increases obesity risk. The coefficient of determination (R^2) as the Poison multiple regression's effect size (ES), and the significance level of $p < 0.05$ was standardized for all analyses. Cohen's cutoffs of ≤ 0.10 = small ES, ≥ 0.30 = moderate ES, ≥ 0.50 = large ES (Cohen, 1988) were used. A post-hoc power analysis ($\alpha = 0.05$, power $= 0.80$) indicated that $n = 654$ was sufficient to detect small-to-moderate effect sizes (Cohen's $f^2 = 0.02$).

2.8. Machine learning support

The Machine Learning pipeline steps included: (1) data normalization (min–max scaling), (2) feature selection using regression-derived predictors, (3) model training with grid search, and (4) performance evaluation using R^2, MAE, MSE, and RMSE metrics. Additionally, we applied a machine learning algorithm called Random Forest regression to predict the adolescent's obesity based on the most explanatory variable/set variables provided by Poison's multiple regression model. To measure the percentual influence of the independent variable in the dependent variable, we calculated the R^2 (Hamilton et al., 2015). To measure the precision between real and predicted data, we calculated the mean absolute error (MAE), the mean square error (MSE), and the root mean square error (RMSE) (Chicco et al., 2021). Hyperparameter tuning was conducted using a grid search (n_estimators = [100–500], max_depth =

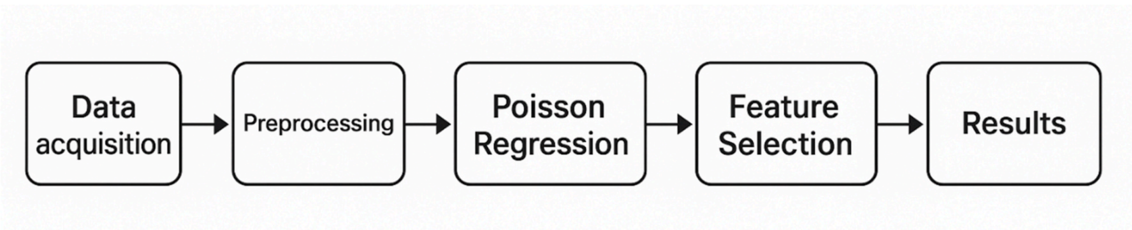

Fig. 1. Flowchart of study design and ML pipeline.

[3–10], min_samples_split = [2–5]) with 5-fold cross-validation to optimize model performance. The model validation followed a 70/30 train-test split with stratified sampling to preserve BMI category balance. We performed all analysis in PythonTM, computational programming language (Python software foundation. Python, 2023). The procedures flowchart is described in Fig. 1.

3. Results

The model's goodness-of-fit statistics indicate a strong fit, with a log-likelihood of −8068.6 and a Pseudo R-squared value of 0.8853, [Intercept = 4.3370, (SE = 0.019, $p < 0.001$), where the AF, ULS and LLF presented an inverse association with the adolescent's BMI [AF: $\beta = -0.0110$, SE = 0.000, $p < 0.0001$, Z = −26.590, CI = 0.012 - 0.010; ULS: $\beta = -0.0085$, SE = 0.001, $p < 0.0001$, Z = −10.046, CI = −0.010 -0.007; LLF: $\beta = -0.0015$, SE = 0.001, Z = −2.096, $p = 0.036$, CI = −0.003-9.83e-05]. In contrast the adolescent's core strength presented a positive association with their body mass [core strength: $\beta = 0.0012$, SE = 0.000, $p < 0.0001$, Z = 3.736, CI = 0.001–0.002]. The results are presented in Table 1 below. In addition, the bivariate coefficient of correlations is described in Fig. 2.

3.1. Machine learning algorithm results

The random forest regression showed that an average of 35 repetition on the yo-yo test predicted a healthy BMI percentile [predBMIperc = 0.31]. Feature importance (Gini index) ranked predictors as follows: AF (0.46), CS (0.31), ULS (0.15), LLF (0.08). The value predBMIperc = 0.31 represents the predicted normalized BMI percentile ($R^2 = 0.54$). In addition, the model presented good validity [MAE = 0.36, MSE = 0.20, RMSE = 0.45, R2 = 0.54]. These results strengthened the preliminary study's results, showing that the AF average values protected the adolescents from obesity in 54 %, as shown in Fig. 3.

Comparatively, the Random Forest model achieved higher predictive accuracy ($R^2 = 0.54$) relative to the regression model (pseudo-$R^2 = 0.48$). This reinforces the ML model's capacity to capture non-linear patterns.

4. Discussion

The present study aimed to develop an obesity prediction model for adolescents by integrating physical fitness assessments, measured by FITescola® tests, with traditional body mass index percentiles using multiple regression analysis and machine learning techniques. Our main findings indicated that physical fitness components, such as AF, ULS, LLF, and core strength, played a significant role in obesity risk among adolescents. Our main hypothesis was confirmed when the multiple regression model indicated a strong fit, as evidenced by a log-likelihood of −8068.6 and a Pseudo R-squared value of 0.8853. In addition, the algorithm's outputs were well-validated, predicting the adolescent's BMI in a healthy cut-off of BMI (predBMIperc = 0.31) basing on their AF average values in the Yo-yo test (35 repetitions), proving good replicability of our findings in real life. These findings suggest that the combined approach of incorporating physical fitness assessments with BMI percentiles holds promise in accurately predicting adolescent obesity.

The observed inverse associations between adolescent BMI and aerobic fitness, upper limb strength, and lower limb flexibility underscore the importance of these physical fitness components in predicting obesity risk. Adolescents with higher levels of AF, ULS, and LLF demonstrated lower BMIs, indicating that greater physical fitness may serve as a protective factor against obesity. These results align with previous research highlighting the role of physical fitness in overall health and its association with a reduced risk of obesity (Ortega et al., 2023; Ruiz et al., 2019).

Clinically, the Yo-Yo test threshold of 35 repetitions could serve as a screening cut-off for identifying adolescents at lower obesity risk. The physical fitness, particularly aerobic fitness, is considered a powerful marker of overall health and well-being in children and adolescents (Ortega et al., 2013). Higher levels of aerobic fitness have been associated with improved cardiovascular health, better metabolic profiles, and reduced adiposity, including lower BMI and body fat percentages (Marandi et al., 2013). The findings of our study are consistent with these previous reports, suggesting that aerobic fitness plays a crucial role in mitigating obesity risk among adolescents (Forte et al., 2023). Moreover, upper limb strength and lower limb flexibility have also been linked to various health outcomes in adolescents. Upper limb strength is an essential component of overall physical fitness, contributing to functional capacity and muscular health (Miguel-Etayo et al., 2014). Our study's observation of an inverse association between ULS and BMI supports the notion that adequate muscular strength may be protective against obesity. Similarly, lower limb flexibility is associated with improved joint health and mobility, potentially influencing physical activity levels and weight management (Stathokostas et al., 2012). The inverse relationship between LLF and BMI in our study further highlights the importance of flexibility as a potential protective factor

Table 1
Poison's multiple regression results.

Variable	*β*	*IRR*	*SE*	*Z*	*p*	95 % CI
Intercept	4.3370	76.510	0.019	232.942	<0.001	[4.301; 4.374]
AF	−0.0110	0.989	0.000	−26.590	<0.001	[-0.012; −0.010]
CS	0.0012	1.001	0.000	3.736	<0.001	[0.001; 0.002]
ULS	−0.0085	0.992	0.001	−10.046	<0.001	[-0.010; −0.007]
LLF	−0.0015	0.999	0.001	−2.096	0.036	[-0.003; −9.83e-05]

Note – AF, aerobic fitness; AS, abdominal strength; ULS, upper limb strength; LLF, lower limb flexibility; *β*, coefficient of variation, *SE*, standard error; *Z*, Z-statistics, *p*, significance level for alternative hypothesis at 95 %; CI, confidence interval.

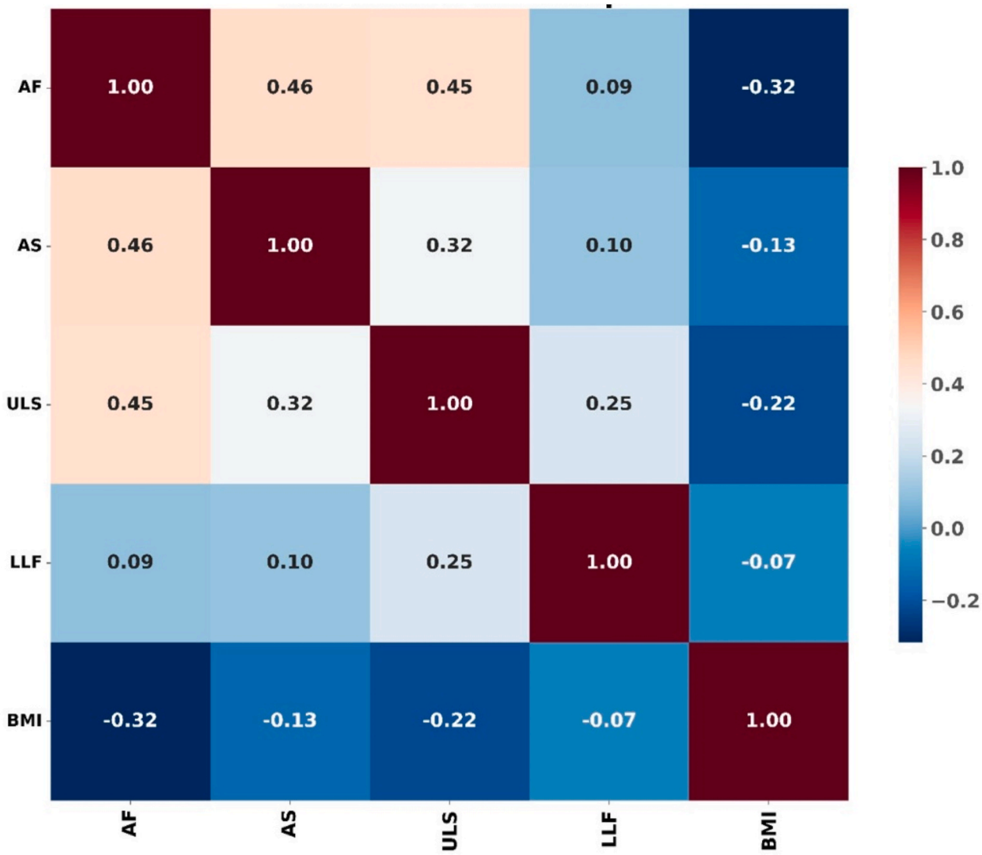

Fig. 2. Coefficient of correlation between the final model's predictors and the target variable (adolescent's BMI percentiles).

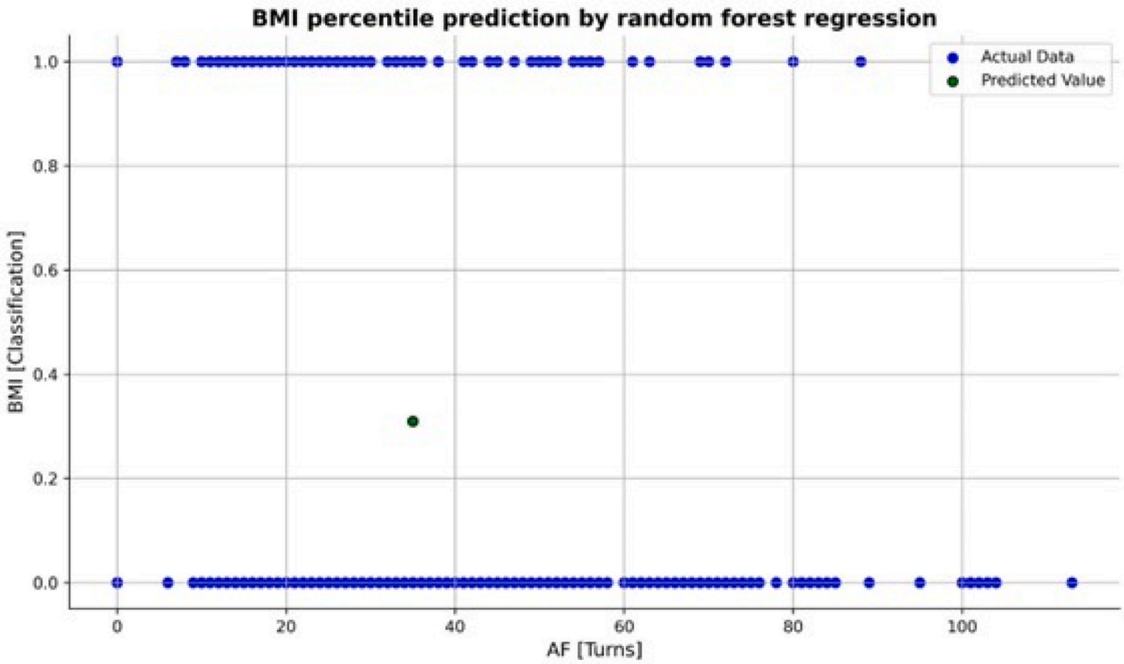

Fig. 3. BMI percentiles prediction based on the adolescent's AF.

against obesity (Shi et al., 2022).

The positive association observed between adolescent BMI and core strength suggests that a weaker core musculature may be linked to an increased risk of obesity. Additionally, this association may reflect confounding by lean mass, warranting future body composition-adjusted analyses. Core strength is crucial for stability, posture, and movement coordination (Zemková and Zapletalová, 2021). Previous studies have demonstrated the role of core strength in preventing injuries and improving physical performance (Dong et al., 2023). However, limited research exists regarding the specific relationship between core strength and obesity risk in adolescents (Jebeile et al., 2022). Further investigations are warranted to explore the potential mechanisms underlying this association and to understand the implications of core strength on adolescent obesity prevention.

Methodological innovation in this study lies in the hybrid integration of Poisson regression and Random Forest, bridging interpretability and predictive power, an uncommon combination in adolescent health analytics. The model's ability to predict the adolescent's BMI percentile at the 50th percentile, with an average AF being protective against obesity based on BMI percentiles, underscores the practical implications of our findings. This predictive capacity allows for the identification of individuals at a higher risk of obesity, enabling targeted interventions and preventive measures to be implemented proactively. By focusing on improving specific physical fitness components, such as AF, ULS, and LLF, it may be possible to reduce obesity prevalence and associated health risks in the adolescent population.

The integration of machine learning algorithms alongside traditional regression analysis provides a novel and comprehensive approach to obesity prediction (Cheng et al., 2021). Machine learning techniques offer several advantages, including the ability to handle high-dimensional data, model complex relationships, and improve prediction accuracy (Janiesch et al., 2021). In the context of obesity prediction, machine learning models have shown promise in identifying relevant risk factors and improving the accuracy of predictive models (Gupta et al., 2022). The utilization of machine learning algorithms in our study allowed for the consideration of a wide array of predictor variables, contributing to the model's strong fit and accurate predictions (Chang et al., 2022). Machine learning techniques have become increasingly popular in various fields of research, including obesity prediction (Ferreras et al., 2023). These techniques have demonstrated their capacity to handle complex data and identify non-linear relationships between predictors and outcomes, providing valuable insights into obesity risk factors (Chatterjee et al., 2020). The use of machine learning algorithms alongside traditional regression analysis in our study adds to the growing body of literature on advanced predictive modelling in obesity research (Ferreras et al., 2023).

Nonetheless, it is essential to recognize specific constraints within our research. Firstly, the cross-sectional design limits the ability to establish causal relationships between physical fitness components and adolescent obesity. Longitudinal studies are needed to examine the temporal relationship between physical fitness and obesity development. Secondly, the study's sample predominantly consists of adolescents from a specific geographic region, potentially limiting the generalizability of the results to other populations. Third, the Portuguese sample may limit geographic generalizability. Moreover, the absence of dietary and socioeconomic variables restricts interpretation. While Random Forest mitigates overfitting through bootstrapping, small high-dimensional datasets still risk bias. Future studies with more diverse and representative samples are necessary to validate the findings across different demographic groups. Longitudinal designs and integrate biological biomarkers will allow better interpretability. Finally, integrating biochemical biomarkers (e.g., leptin, insulin, adiponectin) will enhance physiological interpretation.

Nonetheless, the comprehensive assessment of physical fitness using Fitescola tests, and the inclusion of multiple regression analysis and machine learning techniques are strengths of this study. The findings contribute valuable evidence to the growing body of literature on obesity prediction and prevention in adolescents, particularly by highlighting the importance of physical fitness as a potential protective factor against obesity. These findings underscore the importance of incorporating physical fitness assessments in obesity prediction models and highlight the potential of targeted interventions to combat the escalating prevalence of adolescent obesity and its associated health consequences.

5. Practical applications

The integration of various modelling techniques has emerged as a crucial strategy in the quest to unravel the intricate dynamics underlying adolescent obesity prediction. This comprehensive approach has not only unveiled key physical fitness components pivotal in forecasting adolescent obesity but has also laid down a robust evidential foundation. This evidence stands as a cornerstone upon which governmental bodies and health institutions can confidently advocate for the adoption of more precise, targeted actions, and screening-intervention solutions in the ongoing battle against obesity among adolescents. By connecting the insights derived from this integrated modelling approach, policymakers and public health practitioners are better equipped to tailor interventions that are finely attuned to the specific needs and vulnerabilities of at risk adolescent populations. Moreover, the identification of key physical fitness components associated with adolescent obesity empowers health institutions to develop more effective screening tools and intervention protocols, thereby optimizing resource allocation and enhancing the efficacy of preventive measures. Furthermore, the successful implementation of this integrated modelling initiative serves as example for other European countries grappling with similar challenges in combatting adolescent obesity.

6. Conclusion

This study presents a robust obesity prediction model for adolescents that incorporates physical fitness assessments measured by Fitescola tests with traditional BMI percentiles and machine learning techniques. The model's strong fit and accurate Randon Forest regression's predictions suggest that physical fitness components, such as AF, ULS, LLF, and core strength, play a significant role in obesity risk among adolescents. These findings emphasize the importance of promoting physical fitness and targeted interventions to combat the escalating prevalence of adolescent obesity and its associated health. By integrating physical fitness assessments with traditional BMI-based metrics, our study provides a valuable framework for developing personalized and effective strategies for obesity prevention in the adolescent population.

CRediT authorship contribution statement

Tatiana Sampaio: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Samuel Encarnação:** Project administration, Methodology, Investigation, Formal analysis, Data curation. **Bruna Amaro:** Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Joana Ribeiro:** Software, Resources, Project administration, Funding acquisition, Formal analysis. **Luís Branquinho:** Resources, Project administration, Methodology, Investigation, Funding acquisition. **António M. Monteiro:** Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **José E. Teixeira:** Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Soukaina Hattabi:** Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Andrew Sortwell:** Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Luciano Bernardes Leite:** Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Alexandra Malheiro:** Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Pedro Rodrigues:** Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Beat Knechtle:** Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Pedro Flores:** Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Pedro Forte:** Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that there are no known financial, personal, or institutional conflicts of interest that could have influenced the work reported in this manuscript. All authors have contributed significantly to the research and preparation of the manuscript and have approved the final version for submission. No funding bodies had any role in study design, data collection, analysis, interpretation, or the decision to submit the article for publication.

References

An, R., Shen, J., Xiao, Y., 2022. Applications of artificial intelligence to obesity research: scoping review of methodologies. J. Med. Internet Res. 24 (12), e40589. https://doi.org/10.2196/40589.

Assis, M.M., Gratão, L.H.A., da Silva, T.P.R., Cordeiro, N.G., do Carmo, A.S., de Freitas Cunha, C., et al., 2022. School environment and obesity in adolescents from a Brazilian metropolis: cross-Sectional study. BMC Public Health 22 (1), 1229. https://doi.org/10.1186/s12889-022-13592-0.

Cardel, M., Dulin-Keita, A., Casazza, K., 2011. Contributors to pediatric obesity in adolescence: more than just energy imbalance. Open Obes. J. 3 (2), 17. https://doi.org/10.2174/1876823701103010017.

Centers for Disease Control and Prevention (CDC), 2023. BMI for children and teens. https://www.cdc.gov/obesity/basics/childhood-defining.html.

Chang, V., Ganatra, M.A., Hall, K., Golightly, L., Xu, Q.A., 2022. An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators. Healthcare Analytics 2, 100118. https://doi.org/10.1016/j.health.2022.100118.

Chatterjee, A., Gerdes, M.W., Martinez, S.G., 2020. Identification of risk factors associated with obesity and overweight—A machine learning overview. Sensors 20 (9), 2734. https://doi.org/10.3390/s20092734.

Cheng, X., Lin, S.Y., Liu, J., Liu, S., Zhang, J., Nie, P., et al., 2021. Does physical activity predict obesity—A machine learning and statistical method-based analysis. Int. J. Environ. Res. Publ. Health 18 (8), 3966. https://doi.org/10.3390/ijerph18083966.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. 7, e623. https://doi.org/10.7717/peerj-cs.623.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, second ed. Routledge, New York.

Colmenarejo, G., 2020. Machine learning models to predict childhood and adolescent obesity: a review. Nutrients 12 (8), 2466. https://doi.org/10.3390/nu12082466.

Dewi, R.C., Rimawati, N., Purbodjati, 2021. Body mass index, physical activity, and physical fitness of adolescence. Journal of Public Health Research 10 (2), e2021. https://doi.org/10.4081/jphr.2021.22.

Dong, K., Yu, T., Chun, B., 2023. Effects of core training on sport-specific performance of athletes: a meta-analysis of randomized controlled trials. Behav. Sci. 13 (2), 148. https://doi.org/10.3390/bs13020148.

Ferdowsy, F., Rahi, K.S.A., Jabiullah, M.I., Habib, M.T., 2021. A machine learning approach for obesity risk prediction. Current Research in Behavioral Sciences 2, 100053. https://doi.org/10.1016/j.crbeha.2021.100053.

Ferreras, A., Sumalla-Cano, S., Martínez-Licort, R., Elío, I., Tutusaus, K., Prola, T., et al., 2023. Systematic review of machine learning applied to the prediction of obesity and overweight. J. Med. Syst. 47 (1), 8. https://doi.org/10.1007/s10916-022-01904-1.

FITESCOLA. FITescola: o programa dos alunos ativos. Disponível em: https://fitescola.dge.mec.pt/hometestes.aspx. (Accessed 28 November 2023).

Forte, P., Encarnação, S., Monteiro, A.M., Teixeira, J.E., Hattabi, S., Sortwell, A., et al., 2023. A deep learning neural network to classify obesity risk in Portuguese adolescents based on physical fitness levels and body mass index percentiles: insights for national health policies. Behav. Sci. 13 (7), 522. https://doi.org/10.3390/bs13070522.

Gupta, M., Phan, T.L.T., Bunnell, H.T., Beheshti, R., 2022. Obesity prediction with EHR data: a deep learning approach with interpretable elements. ACM Transactions on Computing for Healthcare (HEALTH) 3 (3), 1–19. https://doi.org/10.1145/3506719.

Hamilton, D.F., Ghert, M., Simpson, A.H.R.W., 2015. Interpreting regression models in clinical outcome studies. Bone & Joint Research 4 (9), 152–153. https://doi.org/10.1302/2046-3758.49.2000571.

Henriques-Neto, D., Minderico, C., Peralta, M., Marques, A., Sardinha, L.B., 2020. Test–retest reliability of physical fitness tests among young athletes: the FITescola® battery. Clin. Physiol. Funct. Imag. 40 (3), 173–182. https://doi.org/10.1111/cpf.12624.

Janiesch, C., Zschech, P., Heinrich, K., 2021. Machine learning and deep learning. Electron. Mark. 31 (3), 685–695. https://doi.org/10.1007/s12525-021-00475-2.

Jebeile, H., Kelly, A.S., O'Malley, G., Baur, L.A., 2022. Obesity in children and adolescents: epidemiology, causes, assessment, and management. Lancet Diabetes Endocrinol. 10 (5), 351–365. https://doi.org/10.1016/S2213-8587(22)00047-X.

Kansra, A.R., Lakkunarajah, S., Jay, M.S., 2021. Childhood and adolescent obesity: a review. Front. Pediatr. 8, 581461. https://doi.org/10.3389/fped.2020.581461.
Khanna, D., Peltzer, C., Kahar, P., Parmar, M.S., 2022. Body mass index (BMI): a screening tool analysis. Cureus 14 (2), e22119. https://doi.org/10.7759/cureus.22119.
Lee, A., Cardel, M., Donahoo, W.T., 2000. Social and environmental factors influencing obesity. In: Feingold, K.R., Anawalt, B., Boyce, A., Chrousos, G., Dungan, K., Grossman, A., Hershman, J.M., Kaltsas, G., Koch, C., Kopp, P., et al. (Eds.), Endotext. MDText.com, Inc.
Li, Y., Zhang, H., Zhao, T., 2023. Machine learning models for obesity prediction using physical and behavioral data. Sci. Rep. 13 (17421). https://doi.org/10.1038/s41598-023-17421-x.
Lin, T.Y., Lim, P.S., Hung, S.C., 2018. Impact of misclassification of obesity by body mass index on mortality in patients with CKD. Kidney Int. Rep. 3 (2), 447–455. https://doi.org/10.1016/j.ekir.2017.12.009.
Marandi, S.M., Abadi, N.G.B., Esfarjani, F., Mojtahedi, H., Ghasemi, G., 2013. Effects of intensity of aerobics on body composition and blood lipid profile in obese/overweight females. Int. J. Prev. Med. 4 (Suppl. 1), S118.
Miguel-Etayo, D., Gracia-Marco, L., Ortega, F.B., Intemann, T., Foraita, R., Lissner, L., et al., 2014. Physical fitness reference standards in European children: the IDEFICS study. Int. J. Obes. 38 (2), S57–S66. https://doi.org/10.1038/ijo.2014.136.
Muñoz-Pichardo, J.M., Pino-Mejías, R., García-Heras, J., Ruiz-Muñoz, F., Luz González-Regalado, M., 2021. A multivariate poisson regression model for count data. J. Appl. Stat. 48 (13–15), 2525–2541. https://doi.org/10.1080/02664763.2021.1877637.
Narciso, J., Silva, A.J., Rodrigues, V., Monteiro, M.J., Almeida, A., Saavedra, R., Costa, A.M., 2019. Behavioral, contextual and biological factors associated with obesity during adolescence: a systematic review. PLoS One 14 (4), e0214941. https://doi.org/10.1371/journal.pone.0214941.
Ortega, F.B., Leskošek, B., Gil-Cosano, J.J., Mäestu, J., Tomkinson, G.R., Ruiz, J.R., et al., 2023. European fitness landscape for children and adolescents: updated reference values, fitness maps and country rankings based on nearly 8 million test results from 34 countries gathered by the FitBack network. Br. J. Sports Med. 57 (5), 299–310. https://doi.org/10.1136/bjsports-2022-106176.
Ortega, F.B., Ruiz, J.R., Castillo, M.J., 2013. Physical activity, physical fitness, and overweight in children and adolescents: evidence from epidemiologic studies. Endocrinol. Nutr. 60 (8), 458–469. https://doi.org/10.1016/j.endoen.2013.10.007.
Park, S., Lee, K., 2024. Predictive modeling of adolescent obesity using deep neural networks and lifestyle data. Front. Public Health 12, 1360047. https://doi.org/10.3389/fpubh.2024.1360047.
Purnell, J.Q., 2000. Definitions, classification, and epidemiology of obesity. In: Feingold, K.R., Anawalt, B., Blackman, M.R., Boyce, A., Chrousos, G., Corpas, E., de Herder, W.W., Dhatariya, K., Dungan, K., Hofland, J., et al. (Eds.), *Endotext* (South Dartmouth. MDText.com, Inc., MA.
Python software foundation. *Python*, 2023. Disponível em. https://www.python.org/. (Accessed 28 November 2023).
Raj, M., Kumar, R.K., 2010. Obesity in children & adolescents. Indian J. Med. Res. 132 (5), 598–607. https://doi.org/10.4103/IJMR.2010_132_05_598.
Ruiz, L.D., Zuelch, M.L., Dimitratos, S.M., Scherr, R.E., 2019. Adolescent obesity: diet quality, psychosocial health, and cardiometabolic risk factors. Nutrients 12 (1), 43. https://doi.org/10.3390/nu12010043.
Sanyaolu, A., Okorie, C., Qi, X., Locke, J., Rehman, S., 2019. Childhood and adolescent obesity in the United States: a public health concern. Global Pediatric Health 6, 2333794X19891305. https://doi.org/10.1177/2333794X19891305.
Shi, Y., Yu, H., Di, S., Ma, C., 2022. Body mass index and academic achievement among Chinese secondary school students: the mediating effect of inhibitory control and the moderating effect of social support. Front. Psychol. 13, 835171. https://doi.org/10.3389/fpsyg.2022.835171.
Smith, J.D., Fu, E., Kobayashi, M.A., 2020. Prevention and management of childhood obesity and its psychological and health comorbidities. Annu. Rev. Clin. Psychol. 16 (1), 351–378. https://doi.org/10.1146/annurev-clinpsy-100219-060201.
Stathokostas, L., Little, R.M., Vandervoort, A.A., Paterson, D.H., 2012. Flexibility training and functional ability in older adults: a systematic review. Journal of Aging Research 2012, 306818. https://doi.org/10.1155/2012/306818.
Thamrin, S.A., Arsyad, D.S., Kuswanto, H., Lawi, A., Nasir, S., 2021. Predicting obesity in adults using machine learning techniques: an analysis of Indonesian basic health research 2018. Front. Nutr. 8, 669155. https://doi.org/10.3389/fnut.2021.669155.
Tremmel, M., Gerdtham, U.G., Nilsson, P.M., Saha, S., 2017. Economic burden of obesity: a systematic literature review. Int. J. Environ. Res. Publ. Health 14 (4), 435. https://doi.org/10.3390/ijerph14040435.
Webb, L.M., Chen, C.Y., 2021. The COVID-19 pandemic's impact on older adults' mental health: contributing factors, coping strategies, and opportunities for improvement. Int. J. Geriatr. Psychiatr. 37 (1), e1002. https://doi.org/10.1002/gps.5647.
World Health Organization (WHO). *World Health Organization*, 2024. Disponível em: https://www.who.int.
Zemková, E., Zapletalová, L., 2021. Back problems: pros and cons of core strengthening exercises as a part of athlete training. Int. J. Environ. Res. Publ. Health 18 (10), 5400. https://doi.org/10.3390/ijerph18105400.