

Article

Sex-Specific Physical Fitness Profiles in Chilean Adolescents: An Unsupervised Machine Learning and Explainable Artificial Intelligence Approach

Rodrigo Yáñez-Sepúlveda ¹, Juan Pablo Zavala-Crichton ¹, Claudio Hinojosa-Torres ¹, Rodrigo Olivares ², Pablo Olivares ², Exal Garcia-Carrillo ^{3,4}, Guillermo Cortés-Roco ⁵, Jacqueline Páez-Herrera ⁶, Jorge Olivares-Arancibia ⁷, Boryi A. Becerra-Patiño ⁸, Juan David Paucar-Urbe ⁸, Eduardo Guzmán-Muñoz ^{9,10} and José Francisco López-Gil ^{11,12,*}

¹ Faculty of Education and Social Sciences, Universidad Andrés Bello, Viña del Mar 2560377, Chile; rodrigo.yanez.s@unab.cl (R.Y.-S.); jzavala@unab.cl (J.P.Z.-C.); claudio.hinojosa@unab.cl (C.H.-T.)

² Escuela de Ingeniería Informática, Universidad de Valparaíso, Valparaíso 2362905, Chile; rodrigo.olivares@uv.cl (R.O.); pablo.olivareszu@uv.cl (P.O.)

³ Department of Physical Activity Sciences, Faculty of Education Sciences, Universidad Católica del Maule, Talca 3480112, Chile; exal.garcia@gmail.com

⁴ School of Education, Faculty of Human Sciences, Universidad Bernardo O'Higgins, Santiago 8370993, Chile

⁵ Faculty of Life Sciences, Universidad Viña del Mar, Viña del Mar 2580022, Chile; guillermo.cortes@uvm.cl

⁶ eFidac Research Group, Physical Education School, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362807, Chile; jacqueline.paez@pucv.cl

⁷ Grupo AFySE, Investigación en Actividad Física y Salud Escolar, Escuela de Educación Física, Facultad de Educación, Universidad de las Américas, Santiago 8370150, Chile; jolivares@udla.cl

⁸ Faculty of Physical Education, National Pedagogical University, Bogotá 110221, Colombia; babecerrap@pedagogica.edu.co (B.A.B.-P.); jdpaucaru@upn.edu.co (J.D.P.-U.)

⁹ Escuela de Kinesiología, Facultad de Salud, Universidad Santo Tomás, Talca 3460000, Chile; eguzmanm@santotomas.cl

¹⁰ Escuela de Kinesiología, Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Talca 3460000, Chile

¹¹ School of Medicine, Universidad Espíritu Santo, Samborondón 09-01-952, Ecuador

¹² Vicerrectoría de Investigación y Postgrado, Universidad de Los Lagos, Osorno 5290000, Chile

* Correspondence: josefranciscolopezgil@gmail.com

Abstract

Background and objective: During adolescence, physical fitness attributes are a key factor positively influencing not only physical health, but also cognitive, psychological and social development. The objective was to derive data-driven physical fitness profiles and to characterize their sex-specific structure in Chilean adolescents using unsupervised machine learning and an explainable artificial intelligence (AI) framework. **Methods:** This cross-sectional study included a nationally representative sample of 25,912 Chilean schoolchildren (14,184 boys and 11,728 girls) aged 13 to 17 years. Six unsupervised clustering algorithms were applied separately by sex, each configured to partition participants into three clusters representing low, medium, and high physical performance levels. **Results:** K-means achieved the highest composite internal validity score for both sexes (boys: 0.959, Calinski–Harabasz 5757.7, Davies–Bouldin 1.203; girls: 0.976, Calinski–Harabasz 4156.1, Davies–Bouldin 1.349), and the partition was highly stable across 500 age-stratified bootstrap resamples (adjusted Rand index 0.973 in boys and 0.967 in girls). A Random-Forest surrogate reproduced the cluster labels with high fidelity (macro-averaged F1 0.984 in both sexes; Cohen's κ 0.973 in boys and 0.975 in girls). **Conclusions:** Unsupervised machine learning with explainable AI provides a robust, transparent, sex-specific framework for stratifying adolescent fitness; cardiorespiratory capacity and abdominal endurance are

Academic Editor: Oliver Ramos
Álvarez

Received: 28 June 2026

Revised: 7 July 2026

Accepted: 15 July 2026

Published: 24 July 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\) license](#).

the components that most consistently distinguish the profiles, while lower-limb power and upper-body endurance differentiate the higher performance tiers.

Keywords: unsupervised learning; cluster analysis; explainable artificial intelligence; SHAP; physical fitness; cardiorespiratory fitness; sex differences; adolescents; population health surveillance

1. Introduction

Adolescence is a transitional, crucial life stage that plays a vital role in shaping long-term healthy behaviors [1,2]. During adolescence, physical fitness attributes are a key factor positively influencing not only physical health, but also cognitive [3], psychological [4], and social development [5]. However, in our current era, adolescents remain insufficiently active [1,6]. This is concerning because physical inactivity, sedentary behaviors, and low cardiorespiratory fitness are considered strong risk factors contributing to the onset of chronic diseases, leading to increased morbidity and mortality [7,8]. Emerging evidence suggests that these risks manifest differently across sexes, with girls exhibiting earlier metabolic consequences of inactivity [9,10] and boys showing more pronounced declines in musculoskeletal fitness [11]. Critically, this sex-specific divergence is rarely captured by the single, pooled cut-off points that dominate current fitness surveillance, which can mask clinically meaningful subgroups and misclassify adolescents located at the extremes of the performance distribution. This gap motivates analytical approaches that recover the latent structure of fitness data and its sex-specific organization—without imposing predetermined thresholds.

Traditional methods of classifying physical fitness often rely on predefined cutoffs or normative data, which may not capture the full spectrum of fitness levels within a population [12,13]. Moreover, many of the batteries used may be limited in their ability to identify distinct subgroups or phenotypes of physical fitness, further complicating comparisons and the development of targeted interventions [14]. Machine learning (ML) methods are being increasingly adopted for modeling epidemiological and health-related data in recent years [15], as these methodologies address limitations in classical statistical approaches by providing enhanced analytical capabilities in terms of physical activity, sedentary, and sleep data, facilitating innovative hypothesis generation [16]. Unsupervised ML algorithms offer a promising alternative by allowing the data to “speak for itself,” identifying natural groupings or clusters within the data without prior assumptions about the population’s structure [17,18]. Given the multivariate character of physical fitness, in which several correlated capacities co-vary within the same individual, these techniques are well-suited to efficiently process and analyze complex, high-dimensional datasets [15,16,19]. Indeed, data-driven clustering is now well established in the wider health and data-science literature as a means of recovering subtypes that rigid, threshold-based scoring systems overlook: comparable unsupervised approaches have been used to derive frailty phenotypes in community-dwelling adults through latent class analysis [20] and to identify clinically meaningful chronic-disease subtypes from large electronic health-record cohorts [21]. By contrast, their application to nationally representative pediatric fitness data, and to the sex-specific organization of that data, remains scarce.

Chile presents a unique case study due to its high rates of adolescent obesity, with 25% of children and adolescents being obese and 29% overweight in 2020 [22], coupled with declining cardiorespiratory fitness levels [23]. Unsupervised ML algorithms offer a solution by detecting natural clustering patterns in high-dimensional fitness data without a priori assumptions [24], while accounting for sex-specific physiological differences in

pubertal development [25]. Therefore, the objective of this study was to derive data-driven physical fitness profiles in Chilean adolescents using unsupervised ML operationally labeled as low, medium, and high performance, and to incorporate interpretability with a surrogate model and SHapley Additive exPlanations (SHAP) values to identify the determinants of stratification by sex.

The contribution of this work is therefore applied and translational rather than algorithmic. We do not propose a new clustering algorithm; instead, we (i) benchmark six unsupervised algorithms against three complementary internal validity indices combined in a single, explicitly defined composite criterion, with bootstrap-based stability assessment, so that model selection is transparent and reproducible rather than arbitrary; (ii) apply this framework to a large, nationally representative pediatric sample and stratify the analysis by sex, a dimension that pooled cut-off approaches typically collapse; and (iii) couple the unsupervised solution with an explainable artificial intelligence (AI) layer (a surrogate model and SHAP values) that translates the resulting profiles into interpretable, component-level information usable by teachers, coaches, and public-health practitioners. To our knowledge, this combination has not previously been reported for nationally representative adolescent fitness data.

2. Methods

This section describes the analytical workflow in the order in which it was implemented. We first present the study design, the source and structure of the dataset, and the ethical framework (Sections 2.1 and 2.2); we then describe the four physical fitness indicators and their measurement (Section 2.3), the descriptive procedures (Section 2.4), and the computational environment (Section 2.5). The core analysis comprises data preprocessing and dimensionality reduction (Section 2.6), the six unsupervised clustering algorithms and their configuration, internal validity indices, the composite selection criterion, and bootstrap stability assessment (Section 2.7), cluster interpretation and labeling (Section 2.8), and the explainable-AI layer based on a surrogate model and SHAP values (Section 2.9). The analysis was conducted separately for boys and girls throughout. The full, seeded analysis code is openly available (see Data Availability Statement).

2.1. Study Design and Participants

This cross-sectional study included a nationally representative sample of 25,912 Chilean schoolchildren (14,184 boys and 11,728 girls) aged 13 to 17 years. For the purposes of this study, adolescence was operationalized in line with the World Health Organization's definition of adolescence as the second decade of life (10–19 years); the 13–17-year span analyzed here reflects the age structure of the national school-based dataset rather than a different conceptual definition [26]. The data derive from the National Study of Physical Education administered by the Chilean Education Quality Agency (Agencia de Calidad de la Educación), in which the Sistema de Medición de la Calidad de la Educación en Educación Física (SIMCE EF) standardized fitness battery is applied under nationally regulated, supervised testing conditions by trained examiners [27]. The present analysis pools de-identified microdata from this national assessment, comprising 25,912 students with complete data on the four fitness indicators, which are openly available (see Data Availability Statement).

2.2. Ethical Considerations

The data used in this study were fully anonymized and openly available through a government-authorized education-quality platform of the Chilean Ministry of Education (MINEDUC). Because this work is a retrospective analysis of openly available, de-identified public data, ethical approval was not required, in accordance with national guidelines

for secondary data analysis. All procedures adhered to the principles of the Declaration of Helsinki [28].

2.3. Physical Fitness Assessments

Four core indicators of physical fitness were assessed under the standardized SIMCE EF protocol, administered by trained examiners following the national testing manual [27]:

- Cardiorespiratory fitness: maximal oxygen uptake ($\text{VO}_{2\text{max}}$, $\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) estimated from the 20-m shuttle-run test (Course-Navette), with $\text{VO}_{2\text{max}}$ derived from the last completed stage using a validated prediction equation.
- Core (abdominal) endurance: number of sit-ups (curl-ups) completed in a fixed time interval, performed with a standardized cadence and movement criterion.
- Upper-body muscular endurance: maximal number of push-ups executed with a standardized technique and cadence until volitional failure or loss of form.
- Lower-limb explosive power: standing horizontal (long) jump distance (cm), recorded as the best of the permitted attempts measured from the take-off line to the nearest heel contact.

The four fitness tests were applied with the standardized SIMCE EF equipment, cadences, and scoring criteria defined in the national testing manual, and $\text{VO}_{2\text{max}}$ was obtained from the shuttle-run performance using the protocol-specified prediction equation [27].

2.4. Descriptive Statistics

The mean and standard deviation for each physical fitness variable were calculated using conventional parametric methods. The mean (M) represents the arithmetic average of all individual scores within each performance cluster, while the standard deviation (SD) quantifies the degree of variation or dispersion from the mean. These variables were pre-processed to ensure quality: missing data were removed using listwise deletion, and outliers were excluded based on a z-score threshold $> \pm 3.5$.

2.5. Computational Environment

All statistical and ML analyzes were conducted using Python (version 3.10; Python Software Foundation, Beaverton, OR, USA), executed in Jupyter notebooks on a local workstation using libraries from the Python Scientific Stack: scikit-learn 1.3.2, pandas 2.1.4, NumPy 1.26.4, Matplotlib 3.8.2, and seaborn 0.13.2. All stochastic steps were initialized with a fixed random seed (`random_state = 42`), and the complete analysis code is openly available (see Data Availability Statement). This environment enabled reproducible, transparent, and flexible exploration of clustering models and data visualizations.

2.6. Data Processing and Dimensionality Reduction

All performance metrics were standardized using z-score normalization:

$$z = (x - \mu) / \sigma \quad (1)$$

where x is the raw score of a given fitness variable, μ is the variable mean, and σ is its standard deviation, all computed within each sex.

All subsequent clustering analyzes were performed in the resulting four-dimensional standardized feature space. Principal component analysis (PCA) was applied separately for visualization only: the first two principal components were used to construct the two-dimensional cluster maps and were not used as input to the clustering. Because the feature space comprised four standardized variables, PCA yielded four principal components;

the first two were displayed because a two-dimensional scatter plot can represent only two axes, and together they captured the majority of the total variance (66.4% in boys and 64.7% in girls), whereas the third and fourth components carried the remaining variance (33.6% and 35.3%, respectively). Because the clustering was performed in the full four-dimensional standardized space rather than on the components, this projection is a visualization device only and does not discard information; the explained and cumulative variance of all four components is shown in Supplementary Figure S2.

2.7. Unsupervised Clustering Algorithms

Six unsupervised machine-learning algorithms were applied separately to the male and female subgroups: K-means clustering, agglomerative (hierarchical) clustering, Gaussian mixture models (GMM), spectral clustering, density-based spatial clustering of applications with noise (DBSCAN), and hierarchical density-based spatial clustering of applications with noise (HDBSCAN) [29].

Each algorithm was configured to partition participants into three clusters ($k = 3$) representing low, medium, and high physical performance levels. The principal configuration was as follows: K-means was run with ten random initializations ($n_init = 10$) and a fixed random seed ($random_state = 42$); the Gaussian Mixture Model used three components with full covariance and the same seed; Agglomerative (hierarchical) clustering used Ward linkage on Euclidean distances; Spectral clustering used a nearest-neighbour affinity graph with ten neighbours ($n_neighbors = 10$) and K-means label assignment; DBSCAN was evaluated over a grid of $eps \in \{0.5, 1.0, 1.5\}$ with $min_samples = 10$; and HDBSCAN was evaluated over $min_cluster_size \in \{100, 500\}$. The quality of clustering solutions was assessed using three complementary internal validity indices: the Silhouette Score, which quantifies cohesion and separation of the clusters; the Calinski–Harabasz Index, which evaluates the ratio of between-cluster to within-cluster variance; and the Davies–Bouldin Index, which penalizes overlapping clusters. To obtain a single, transparent ranking criterion, the three indices were combined into a composite score. Each index was first min–max normalized within each sex:

$$X_{norm} = (X - X_{min}) / (X_{max} - X_{min}) \quad (2)$$

yielding S_{norm} for the Silhouette Score and C_{norm} for the Calinski–Harabasz Index. Because lower Davies–Bouldin (D) values indicate better partitions, the normalized Davies–Bouldin index was inverted ($D_{inv} = 1 - D_{norm}$) so that all three components share the same direction (higher = better). The composite score was then computed as the unweighted (equal-weight) mean of the three normalized components:

$$Score = (S_{norm} + C_{norm} + D_{inv}) / 3 \quad (3)$$

The stability of the selected solution was assessed with 500 age-stratified bootstrap resamples: in each resample, the selected algorithm was refitted and its partition compared with the full-sample solution using the adjusted Rand index (ARI). The selected K-means solution was highly stable in both sexes (mean ARI 0.973 in boys and 0.967 in girls; 95% bootstrap interval 0.939–0.994 in boys and 0.929–0.990 in girls). The behavior of the three internal validity indices across candidate numbers of clusters ($k = 2$ –8) for K-means in each sex is reported in Supplementary Figure S1. For DBSCAN and HDBSCAN, the grids tested ($eps \in \{0.5, 1.0, 1.5\}$ with $min_samples = 10$; $min_cluster_size \in \{100, 500\}$) produced unstable, over-fragmented solutions: DBSCAN yielded six clusters in boys and five in girls with approximately 10% of observations labeled as noise at $eps = 0.5$ and collapsed to a single cluster at $eps \geq 1.0$, whereas HDBSCAN classified all observations as noise (no clusters recovered). These density-based methods were therefore retained for

benchmarking but not selected, and the final models were prioritized on the basis of multi-metric performance and parsimony.

2.8. Cluster Interpretation and Labeling

Clusters were interpreted based on mean values of the four fitness variables. Labels (“Low”, “Medium”, and “High”) were assigned by ranking cluster centroids, confirming coherence across other metrics. Final distributions were visualized using 2D scatter plots generated from PCA projections.

2.9. SHAP-Based Explainable ML for Sex-Specific Fitness Stratification

In this study, after applying multiple unsupervised clustering algorithms (K-means, DBSCAN, Agglomerative, Gaussian Mixture, Spectral and HDBSCAN) and selecting the optimal model for each sex according to internal validity metrics (Silhouette, Calinski–Harabasz, and Davies–Bouldin), a supervised surrogate model was implemented using a Random-Forest Classifier to reproduce cluster assignments and provide an interpretable framework. The surrogate was trained on the four standardized fitness variables to predict the cluster label assigned by the selected unsupervised model, using a fixed random seed. The surrogate reproduced the cluster labels with high fidelity in both sexes (boys: macro-averaged F1 (macro-F1) 0.984, balanced accuracy 0.986, Cohen’s κ 0.973; girls: macro-F1 0.984, balanced accuracy 0.985, Cohen’s κ 0.975), evaluated on a stratified 20% hold-out set (random_state = 42; 11,347/2837 train/test observations in boys and 9382/2346 in girls), so that the SHAP attributions can be interpreted as reflecting the structure of the original clusters rather than artifacts of the surrogate. Subsequently, SHAP, grounded in cooperative game theory, was employed to quantify the individual contribution of each physiological variable to the differentiation of physical fitness levels. This methodological combination enabled the estimation of the relative importance of VO₂max, sit-ups, push-ups, and horizontal jump in a transparent and reproducible manner, ensuring comparability across sexes and fitness strata.

3. Results

In our clustering analysis of 25,912 Chilean schoolchildren using six unsupervised learning algorithms, K-means was identified as the most robust model for both boys and girls. This determination was based on a composite scoring method that integrated the Silhouette, Calinski–Harabasz, and Davies–Bouldin indices to balance cluster cohesion, separation, and compactness. Among boys, K-means achieved the highest composite score (0.959), combining the highest Silhouette value (0.290), the strongest Calinski–Harabasz index (5757.7) and a low Davies–Bouldin value (1.203); Agglomerative clustering was the next-best solution (0.756). Although DBSCAN attained a comparable Silhouette value on its non-noise points, it fragmented the data into six clusters and was therefore not competitive on the composite criterion. Among girls, K-means likewise outperformed all other methods (composite score 0.976; Silhouette 0.255, Calinski–Harabasz 4156.1, Davies–Bouldin 1.349), followed by Spectral (0.778) and Agglomerative (0.715) clustering (Table 1). For K-means, Agglomerative, Gaussian Mixture, and Spectral clustering, $k = 3$; DBSCAN returned six clusters in boys, and five in girls (~10% of observations as noise), and HDBSCAN returned no valid clusters (all observations classified as noise), so their indices are computed on non-noise observations where defined.

Table 1. Comparative clustering metrics by sex and algorithm.

Group	Model	Silhouette	Calinski–Harabasz	Davies–Bouldin	Sil_{norm}	Cal_{norm}	Dav_{norm}	Score
Boys	K-means	0.290	5757.7	1.203	1.000	1.000	0.123	0.959
	Agglomerative	0.220	3999.8	1.415	0.749	0.691	0.171	0.756
	DBSCAN	0.289	69.4	0.660	0.997	0.000	0.000	0.666
	Spectral	0.127	3597.5	1.646	0.411	0.620	0.223	0.603
	GMM	0.013	1581.9	5.075	0.000	0.266	1.000	0.089
	HDBSCAN	—	—	—	—	—	—	—
Girls	K-means	0.255	4156.1	1.349	1.000	1.000	0.073	0.976
	Spectral	0.183	3243.7	1.560	0.658	0.779	0.103	0.778
	Agglomerative	0.168	2770.4	1.590	0.588	0.665	0.107	0.715
	DBSCAN	0.135	22.1	0.832	0.435	0.000	0.000	0.478
	GMM	0.042	1131.2	7.903	0.000	0.268	1.000	0.089
	HDBSCAN	—	—	—	—	—	—	—

Note. Sil_{norm} , Cal_{norm} and Dav_{norm} are the within-sex min–max-normalized Silhouette, Calinski–Harabasz and Davies–Bouldin indices, respectively (Equation (2)); *Score* is the composite internal-validity criterion, $Score = (Sil_{norm} + Cal_{norm} + (1 - Dav_{norm}))/3$ (Equation (3)), with higher values indicating better-defined clusters. Em dashes (—) indicate indices that could not be computed because the algorithm returned no valid clusters. GMM, Gaussian mixture model; DBSCAN, density-based spatial clustering of applications with noise; HDBSCAN, hierarchical DBSCAN.

Based on the distribution of the 25,912 schoolchildren across the three data-driven performance clusters, the medium-performance group was the largest (44.9% of the sample; 24.8% boys and 20.1% girls), closely followed by the high-performance group (41.0%; 24.1% boys and 16.9% girls). The low-performance cluster comprised 14.1% of students (5.8% boys and 8.3% girls). Girls were proportionally more represented in the low-performance cluster (8.3% vs. 5.8%) and boys in the high-performance cluster (24.1% vs. 16.9%), indicating a sex asymmetry at both ends of the performance distribution (Table 2).

Table 2. Sample distribution by physical performance level.

Cluster	Boys	Boys %	Girls	Girls %	Total	Total %
Low	1509	5.82%	2141	8.26%	3650	14.09%
Medium	6421	24.78%	5216	20.13%	11,637	44.91%
High	6254	24.14%	4371	16.87%	10,625	41.00%
Total	14,184	54.74%	11,728	45.26%	25,912	100.00%

Note. “Low”, “Medium”, and “High” denote the three data-driven physical-performance clusters obtained from the selected unsupervised model in each sex (not predefined cut-off categories); labels were assigned by ranking the cluster centroids on the four fitness variables (see Section 2.8). Percentages are expressed relative to the total sample ($N = 25,912$).

Cluster-based stratification in boys revealed that the low-performance cluster was set apart chiefly by markedly lower abdominal endurance (sit-ups 15.7 reps vs. ~24.5–24.7 in the other clusters), whereas VO_{2max} , push-ups, and horizontal jump most strongly separated the high-performance cluster (VO_{2max} 51.2 $mL \cdot kg^{-1} \cdot min^{-1}$, push-ups 18.8 reps, horizontal jump 186.8 cm). VO_{2max} was not strictly monotonic across the ordered clusters—being comparable in the low and medium clusters—because the sit-ups test, which primarily defined the low cluster, approached its ceiling in the medium and high clusters (Table 3).

Table 3. Metrics obtained in boys.

Level	VO ₂ max (mL·kg ⁻¹ ·min ⁻¹)	Sit-Ups (reps)	Push-Ups (reps)	Horizontal Jump (cm)
Low	45.0 ± 5.7	15.7 ± 3.2	10.1 ± 7.1	159.8 ± 26.1
Medium	43.7 ± 4.8	24.5 ± 1.2	9.1 ± 6.2	153.5 ± 22.4
High	51.2 ± 4.5	24.7 ± 1.1	18.8 ± 7.9	186.8 ± 20.6

Note. Values are presented as $M \pm SD$ (mean $\pm$ standard deviation). “Low”, “Medium” and “High” denote the three data-driven clusters obtained from the selected K-means solution in boys, labeled by ranking the cluster centroids on the four fitness variables (see Section 2.8). VO₂max, maximal oxygen uptake.

Figure 1 shows the two-dimensional PCA projection of the boys’ K-means solution ($k = 3$; composite score 0.959; PC1 and PC2 explaining 43.2% and 23.2% of the variance): the high-performance cluster separates clearly, whereas the medium- and low-performance clusters partially overlap.

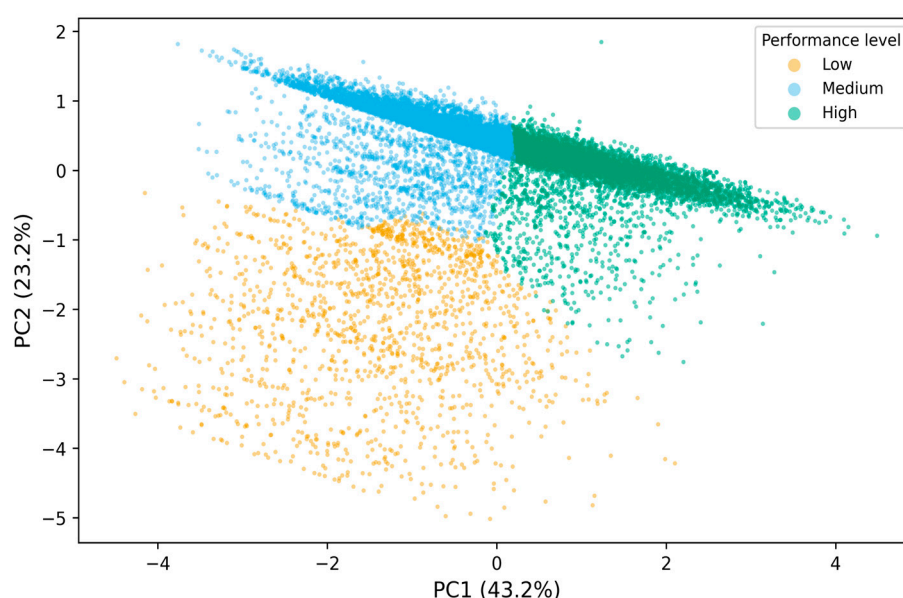


Figure 1. Two-dimensional PCA projection of the K-means clustering solution for Chilean adolescent boys ($k = 3$). Axes are the first two principal components (PC1, 43.2%; PC2, 23.2%; 66.4% cumulative) of the four standardized fitness variables; points are color-coded by performance level (Low, Medium, High).

In girls (Table 4), the low-performance cluster was likewise distinguished mainly by lower abdominal endurance (sit-ups 9.9 reps vs. ~23.5–23.6 in the other clusters), while the high-performance cluster showed the highest VO₂max (43.7 mL·kg⁻¹·min⁻¹), horizontal jump (148.9 cm), and push-ups (16.5 reps). As in boys, VO₂max was not strictly monotonic across the ordered clusters, being marginally higher in the low than in the medium cluster, owing to the sit-up ceiling that primarily defined the low cluster.

Table 4. Metrics obtained in girls.

Level	VO ₂ max (mL·kg ⁻¹ ·min ⁻¹)	Sit-Ups (reps)	Push-Ups (reps)	Horizontal Jump (cm)
Low	38.2 ± 4.2	9.9 ± 4.5	9.9 ± 6.3	122.4 ± 20.7
Medium	37.1 ± 3.2	23.5 ± 2.4	11.0 ± 6.6	119.6 ± 17.9
High	43.7 ± 3.9	23.6 ± 2.9	16.5 ± 7.0	148.9 ± 19.6

Note. Values are presented as $M \pm SD$ (mean $\pm$ standard deviation). “Low”, “Medium” and “High” denote the three data-driven clusters obtained from the selected K-means solution in girls, labeled by ranking the cluster centroids on the four fitness variables (see Section 2.8). VO_{2max} , maximal oxygen uptake.

Figure 2 shows the two-dimensional PCA projection of the girls’ K-means solution ($k = 3$; composite score 0.976; PC1 and PC2 explaining 40.8% and 23.9% of the variance): the high-performance cluster separates in the upper region, the medium cluster is densest centrally, and the low cluster concentrates separately.

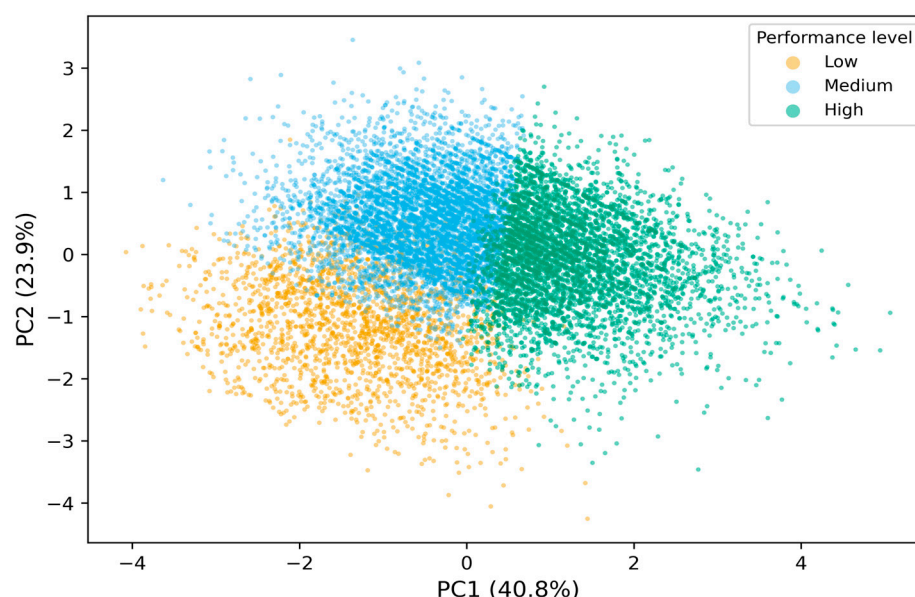


Figure 2. Two-dimensional PCA projection of the K-means clustering solution for Chilean adolescent girls ($k = 3$). Axes are the first two principal components (PC1, 40.8%; PC2, 23.9%; 64.7% cumulative) of the four standardized fitness variables; points are color-coded by performance level (Low, Medium, High).

Figure 3 shows the SHAP-based analysis of physical fitness stratification in boys, computed from the Random-Forest surrogate of the K-means clusters. Globally (Panel A), the four components contributed comparably to the surrogate’s class predictions, with VO_{2max} showing the largest mean $|SHAP|$ value (0.134), followed by sit-ups (0.123), horizontal jump (0.123), and push-ups (0.108). The cluster-specific beeswarm plots show how each component drives membership. For the low-performance cluster (Panel B), sit-ups dominated: low sit-up counts produced large positive SHAP values that pushed participants toward this cluster, whereas the other variables contributed little—consistent with this cluster being defined by below-ceiling abdominal-endurance scores. For the medium cluster (Panel C), VO_{2max} , horizontal jump, and push-ups were the principal drivers, with sit-ups largely inactive. For the high-performance cluster (Panel D), high values of VO_{2max} and horizontal jump produced the strongest positive SHAP values, identifying superior cardiorespiratory capacity and lower-limb power as the markers of high performance.

In girls (Figure 4), the SHAP analysis revealed analogous patterns. Globally (Panel A), sit-ups showed the largest mean $|SHAP|$ value (0.189), followed by VO_{2max} (0.153), horizontal jump (0.119), and push-ups (0.070). In the low-performance cluster (Panel B), low sit-up counts again produced large positive SHAP values and dominated membership, with minimal input from the other variables. In the medium cluster (Panel C),

VO₂max, sit-ups, and horizontal jump were the strongest drivers. In the high-performance cluster (Panel D), high VO₂max and horizontal-jump values produced the largest positive SHAP contributions, again identifying cardiorespiratory capacity and lower-limb power as the principal markers of high fitness.

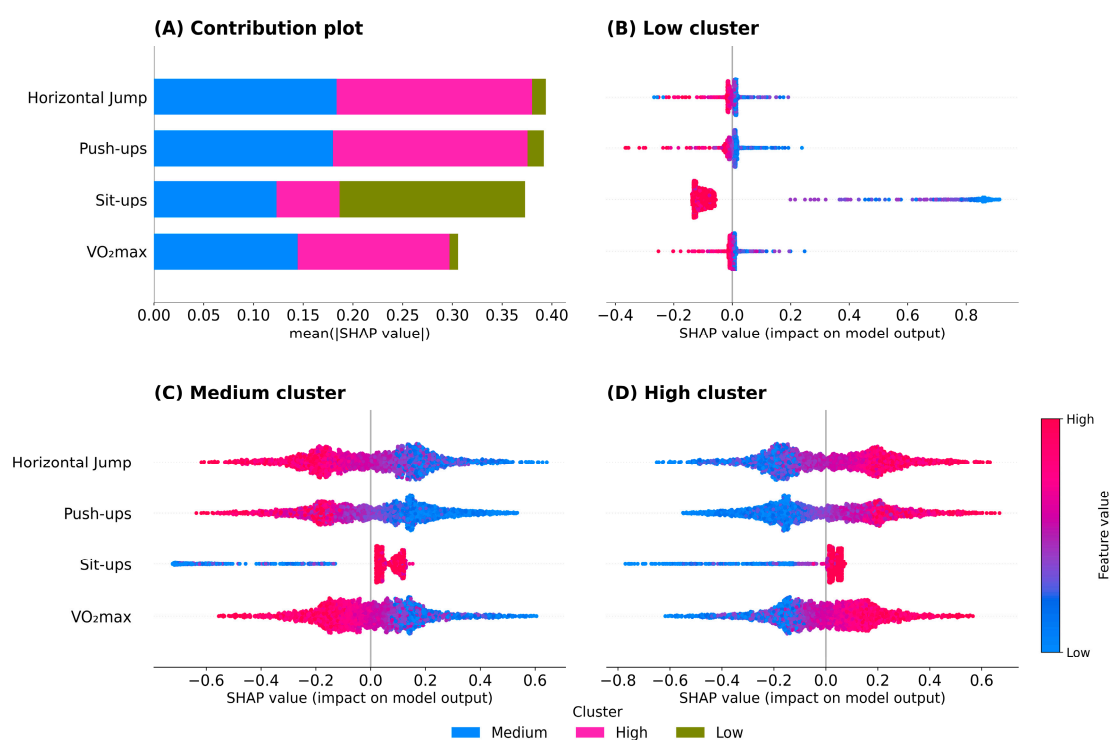


Figure 3. SHAP analysis of physical-fitness stratification in boys, computed from the Random-For-est surrogate of the K-means clusters. (A) Contribution plot: global mean absolute SHAP value per feature, stacked by cluster (Low, Medium, and High). (B–D) SHAP beeswarm summary plots for the Low (B), Medium (C), and High (D) clusters; each point is a participant, the horizontal position is the SHAP value (impact on the predicted cluster membership), and color encodes the feature value (blue = Low and red = High).

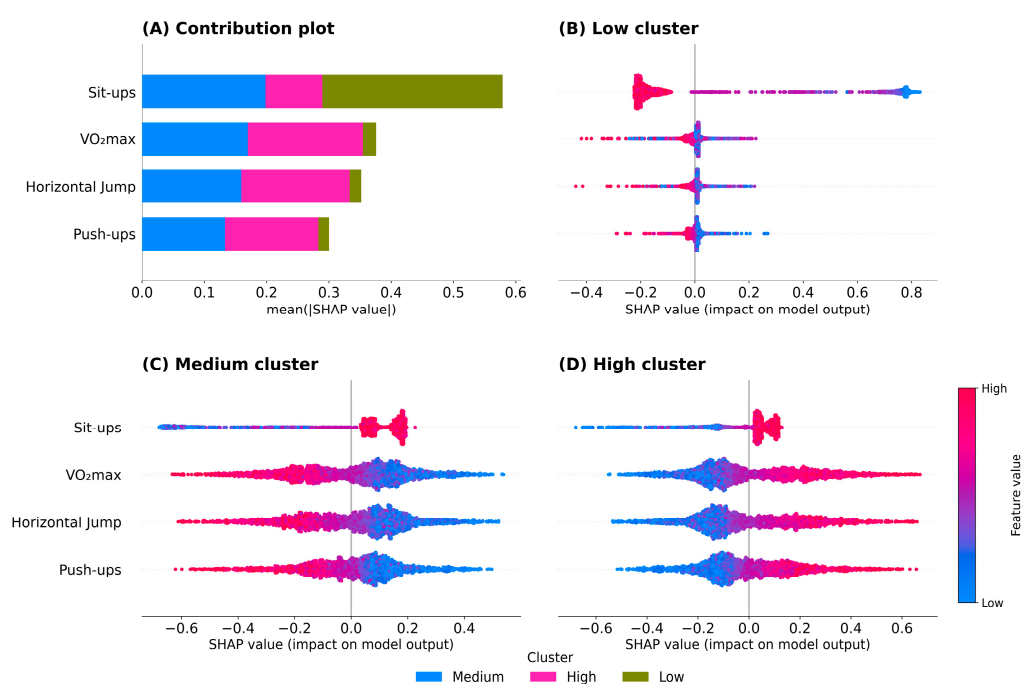


Figure 4. SHAP analysis of physical-fitness stratification in girls, computed from the Random-Forest surrogate of the K-means clusters. (A) Contribution plot: global mean absolute SHAP value per feature, stacked by cluster (Low, Medium, and High). (B–D) SHAP beeswarm summary plots for the Low (B), Medium (C), and High (D) clusters; each point is a participant, the horizontal position is the SHAP value, and color encodes the feature value (blue = Low and red = High).

4. Discussion

The results of this study show that unsupervised ML algorithms combined with explainable AI (SHAP) are effective in classifying fitness levels in Chilean adolescents without predefined labels. This ability to group individuals according to their physiological and performance characteristics represents a significant methodological advance for school settings where resources for personalized assessments are often limited, providing an operational tool for exercise and health professionals.

According to our results, K-means provided the best clustering solution for both boys and girls, indicating a similar global geometry of the fitness data in the two sexes; the sex-specific signal emerged not in the choice of algorithm but in the composition of the clusters and in the SHAP attributions. This is consistent with findings [30] from a systematic review, which identified variations in physical activity patterns and sedentary behavior between boys and girls, suggesting the value of sex-disaggregated analysis. In addition, the Cabanas-Sánchez [31] study found that lifestyle profiles and their relationship to physical fitness vary between sexes, with girls showing greater vulnerability to sedentary behaviors and boys to low levels of physical activity.

Moreover, the use of a composite metric that integrates the Silhouette, Calinski–Harabasz, and Davies–Bouldin indices in our study provides a robust assessment of cluster quality. This multidimensional approach is supported by research such as the HEL-ENA study [32], which combined hierarchical and non-hierarchical methods to identify patterns of healthy behavior in European adolescents, highlighting the importance of multiple metrics to validate the stability and consistency of clusters. Positioned against the broader literature, our findings extend a growing body of work that uses unsupervised learning to derive health phenotypes without predefined thresholds: Bonilla et al. [33] profiled the physical fitness of physical-education students and recovered distinct profiles differing in VO_2max and agility, while in clinical populations latent-class and density-based approaches have yielded reproducible frailty phenotypes in older adults [20] and clinically meaningful chronic-disease subtypes from electronic health records [21]. Compared with these studies, the present work is distinguished by its nationally representative pediatric sample, its explicit sex-stratification, and the addition of an explainable-AI layer that attaches interpretable, component-level meaning to each profile—features that, to our knowledge, have not previously been combined for adolescent fitness data.

Based on the results obtained in boys, the fact that K-means with $k = 3$ achieved a high composite score (0.959) and a highly stable partition across bootstrap resamples (mean ARI 0.973) suggests that this algorithm is particularly effective in detecting consistent patterns of physical performance in Chilean adolescent boys. The use of PCA provided a two-dimensional projection that aided the visualization and interpretation of the group structure, while the clustering itself was performed on the full set of standardized fitness variables. Our study highlights the ability of K-means to clearly distinguish a high physical performance group, characterized by higher values in VO_2max , muscular endurance, and lower-limb power.

The clear separation of the high-performance group in the PCA space indicates the existence of a subset of adolescents with well-defined physical characteristics, including higher VO_2max , greater strength-endurance (as measured by sit-ups and push-ups), and greater lower-body power (as demonstrated by the horizontal jump). In contrast, the par-

tial overlap between the middle and lower groups reflects the natural continuity of physical performance, particularly at this stage of pubertal development, where biological maturation can vary significantly among individuals of the same chronological age. This finding is consistent with the literature, which indicates that fitness levels during adolescence tend to be distributed on a continuous rather than a discrete curve, highlighting the value of unsupervised learning in capturing these smooth transitions, rather than imposing arbitrary cutoffs [34]. The use of PCA in our study to visualize the cluster structure in two dimensions is consistent with a broader movement-science literature that pairs data-driven, unsupervised methods with the analysis of movement and fitness data. In this vein, Thornton et al. [35] applied an unsupervised machine-learning approach to raw accelerometry in children and obtained a more sensitive, data-driven classification of physical-activity intensity than conventional cut-point methods.

In the female subgroup, the high-performance cluster showed the highest cardiorespiratory capacity (VO_2max 43.7 $\text{mL}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$), exceeding both the medium (37.1) and low (38.2) clusters; VO_2max was marginally higher in the low than in the medium cluster, rather than increasing monotonically, because the low cluster was defined chiefly by lower abdominal endurance rather than by aerobic capacity. This pattern is consistent with interventions combining aerobic and strength training that report gains in cardiorespiratory fitness in adolescents [36,37]. In girls, horizontal jump also separated the high-performance cluster clearly, which may be linked to neuromuscular and hormonal differences characteristic of adolescence that modulate the response to training [38]. Overall, the results underscore the need for integrated assessments that consider multiple components of fitness [39,40] and their usefulness in school surveillance to detect areas for improvement early and implement targeted interventions [41,42]. Unsupervised segmentation therefore has direct implications for the personalization of physical education and school health strategies; this is consistent with the evidence from Xie et al. [43] who identified subgroups with different responses to cardiorespiratory training. In contrast to our study, that research analyzed responses to an aerobic endurance intervention program.

A key contribution of this work is explainability: a Random-Forest surrogate—a widely used ML method for prediction modeling [44–47]—allowed us to estimate SHAP importance by variable and by cluster. Rather than a single dominant predictor, the four components contributed comparably at the global level, but with a clear division of labor across clusters: abdominal endurance (sit-ups) almost entirely defined membership in the lowest-performing cluster, whereas VO_2max , lower-limb power, and upper-body endurance jointly distinguished the medium and high clusters. This level of traceability improves the plausible understanding of the profiles, promotes responsible use in public and clinical health, and enables pedagogical feedback to coaches and teachers (e.g., which component to improve to move a student to a higher profile) [48]. This emphasis on explainability aligns with very recent applications in pediatric populations: explainable machine learning with SHAP has been used to build a nationally representative obesity-risk model in children and adolescents [49], and AI-based pipelines have been proposed for analyzing health-related fitness in youth [50]; unlike these supervised, single-outcome models, our framework applies post hoc explainability to an unsupervised, multi-component fitness phenotype. Because the SHAP attributions are derived from a surrogate model, their interpretability depends on how faithfully that model reproduces the unsupervised cluster labels; here, the surrogate achieved high fidelity in both sexes (macro-F1 0.984; Cohen's κ 0.973 in boys and 0.975 in girls), indicating that the reported SHAP values reflect the structure of the original clusters rather than artifacts of the surrogate. Integrating this framework into school screening would enable early detection of those who require support and optimize the allocation of resources toward interventions with greater health returns.

Limitations

This study presents several limitations that should be considered when interpreting the findings. First, although a large and representative sample of Chilean adolescents was used, the data derive from independent cross-sectional samples of students, which precludes longitudinal inference about within-individual change over time. Second, participant classification into performance levels (Low, Medium, and High) was based solely on objective physical fitness metrics, without accounting for contextual factors such as habitual physical activity, nutrition, or mental health, which may significantly influence performance. Third, the abdominal-endurance (sit-ups) test approached its scoring ceiling in the medium and high clusters, so that the lowest-performing cluster was defined largely by below-ceiling sit-up scores; one consequence is that VO_2max was not strictly monotonic across the ordered clusters, and the labels should be read as an overall ordering rather than as a uniform gradient on every component. Fourth, because the explainable-AI layer used a Random-Forest surrogate trained on the same four variables that defined the clusters, its high fidelity is expected by construction and should be interpreted as confirming that the SHAP attributions faithfully describe the cluster geometry, not as independent validation of the clusters themselves.

Additionally, while unsupervised ML algorithms are powerful, they can be sensitive to data distribution, noise, and parameter selection, particularly methods such as DBSCAN and HDBSCAN, which here failed to recover stable clusters. Although a robust model-selection approach was implemented using multiple clustering-evaluation metrics and bootstrap stability assessment, the biological interpretation of clusters remains indirect and requires external validation. Finally, the clusters were evaluated only with internal indices and were not validated against external criteria (e.g., adiposity, cardiometabolic markers, objectively measured physical activity, or academic or health outcomes); such external validation is needed before the derived profiles are used for screening or risk stratification.

In addition, the cross-sectional design of the study prevents the establishment of causal relationships between physical fitness levels and their potential determinants. Future research should incorporate longitudinal designs, integrate multifactorial data (environmental, socioeconomic, and genetic), and pursue clinical validations to ensure these findings can inform sustainable and equitable public health policies.

Beyond these methodological considerations, a natural extension of the present framework is the transition from a fully unsupervised approach toward semi-supervised learning. In real-world settings, reference labels are available for at least a subset of participants—for example, from expert clinical or pedagogical judgment, or from established criterion-referenced fitness standards—so that a limited amount of labeled information can be combined with the much larger pool of unlabeled records. Semi-supervised methods exploit these partial labels to guide and stabilize cluster formation and to anchor the resulting profiles to recognized external categories, rather than relying on internal data structure alone. Recent biomedical work illustrates the value of this strategy: a semi-supervised non-negative matrix-factorization model that incorporates partial annotations as prior information improved the accuracy and robustness of high-dimensional clustering [51], and a self-supervised, pseudo-label-driven method that iteratively refines K-means assignments yielded clinically distinct phenotypes that outperformed a purely unsupervised baseline on the same internal validity indices used here—silhouette, Calinski–Harabasz, and Davies–Bouldin [52]. Applying such schemes to adolescent fitness data, by seeding the model with a small set of expertly labeled or criterion-referenced cases, could strengthen the clinical and educational interpretability of the derived profiles, improve their alignment with normative performance categories, and support their eventual use for individualized screening and targeted intervention.

5. Conclusions

This study demonstrates that the combination of unsupervised ML algorithms and explainable artificial intelligence approaches, such as SHAP, provides a robust, transparent, and sex-specific framework for stratifying adolescent physical fitness at scale. By identifying clear profiles of low, medium, and high performance in both sexes, this work transcends traditional cut-off-based methods and enables more precise and personalized population surveillance. The evidence indicates that abdominal endurance most strongly distinguishes the lowest-performing adolescents, whereas cardiorespiratory capacity, lower-limb power, and upper-body endurance jointly differentiate the higher-performing tiers. These findings reinforce the need for preventive interventions tailored to sex and fitness level and validate the potential of ML and game-theory-based explainability as next-generation tools to guide school health and sports policies. Integrating this framework into educational systems could transform the early identification of at-risk populations and optimize the allocation of resources toward evidence-based interventions, positioning adolescent physical fitness at the core of twenty-first-century public health strategies.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/data11080184/s1>. Figure S1: Internal cluster-validity indices (Silhouette Score, Calinski–Harabasz Index, and Davies–Bouldin Index) for K-means clustering as a function of the number of clusters ($k = 2–8$), shown separately for boys and girls. Higher Silhouette and Calinski–Harabasz values and lower Davies–Bouldin values indicate better-defined clusters. The dashed line marks the selected solution ($k = 3$). Figure S2: Scree plot of the four principal components obtained from the four standardized fitness variables, showing the individual and cumulative explained variance for boys and girls. The first two components (highlighted), which are displayed in Figures 1 and 2, jointly account for 66.4% (boys) and 64.7% (girls) of the total variance; the third and fourth components carry the remaining variance and do not reveal additional cluster structure.

Author Contributions: Conceptualization: R.Y.-S., J.P.Z.-C., C.H.-T., R.O., P.O., E.G.-C., G.C.-R., J.P.-H., J.O.-A., B.A.B.-P., J.D.P.-U., E.G.-M. and J.F.L.-G.; Methodology: R.Y.-S., J.P.Z.-C., C.H.-T., R.O., P.O., E.G.-C., G.C.-R. and J.P.-H.; Formal analysis: R.Y.-S., J.P.Z.-C., C.H.-T., R.O. and P.O.; Investigation: R.Y.-S., J.P.Z.-C., C.H.-T., R.O., P.O., E.G.-C., G.C.-R., J.P.-H., J.O.-A., B.A.B.-P., J.D.P.-U., E.G.-M. and J.F.L.-G.; Software and validation: R.Y.-S., J.P.Z.-C., C.H.-T., R.O. and P.O.; Writing—original draft preparation: R.Y.-S., E.G.-C., G.C.-R., J.P.-H., J.O.-A., B.A.B.-P., J.D.P.-U., E.G.-M. and J.F.L.-G.; Writing—review and editing: R.Y.-S., J.P.Z.-C., C.H.-T., R.O., P.O., E.G.-C., G.C.-R., J.P.-H., J.O.-A., B.A.B.-P., J.D.P.-U., E.G.-M. and J.F.L.-G.; Visualization: R.Y.-S., J.P.Z.-C., C.H.-T., R.O., P.O., E.G.-C., G.C.-R., J.P.-H., J.O.-A., B.A.B.-P., J.D.P.-U., E.G.-M. and J.F.L.-G.; Supervision: R.Y.-S. and J.F.L.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study because it consisted exclusively of a retrospective analysis of anonymized, openly available public microdata released by the Chilean Ministry of Education, in accordance with national guidelines for secondary analysis of public data (see Section 2.2). All procedures adhered to the principles of the Declaration of Helsinki [28].

Informed Consent Statement: Participant consent was waived because the study analyzed anonymized, openly available public microdata from which no individually identifiable information was accessible to the authors.

Data Availability Statement: The data analyzed in this study are openly and publicly available from the Chilean Ministry of Education (MINEDUC) through the public platform of the Agencia de Calidad de la Educación (SIMCE EF—Estudio Nacional de Educación Física). Because the dataset is

already openly accessible from this official source, no additional data deposit by the authors was required. A copy of the analyzed dataset is also available at https://figshare.com/articles/dataset/EFI_SIMCE/30214864 (accessed on 14 July 2026).

Acknowledgments: The authors thank the Agencia de Calidad de la Educación de Chile (Education Quality Agency of Chile) for their valuable contribution to the development of this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. van Sluijs, E.M.F.; Ekelund, U.; Crochemore-Silva, I.; Guthold, R.; Ha, A.; Lubans, D.; Oyeyemi, A.L.; Ding, D.; Katzmarzyk, P.T. Physical activity behaviours in adolescence: Current evidence and opportunities for intervention. *Lancet* **2021**, *398*, 429–442. [https://doi.org/10.1016/S0140-6736\(21\)01259-9](https://doi.org/10.1016/S0140-6736(21)01259-9).
2. Mastorci, F.; Lazzeri, M.F.L.; Vassalle, C.; Pingitore, A. The Transition from Childhood to Adolescence: Between Health and Vulnerability. *Children* **2024**, *11*, 989. <https://doi.org/10.3390/children11080989>.
3. Ruiz-Ariza, A.; Alberto, G.-C.; de Loureiro, N.E.; Martínez-López, E.J. Influence of physical fitness on cognitive and academic performance in adolescents: A systematic review from 2005–2015. *Int. Rev. Sport Exerc. Psychol.* **2017**, *10*, 108–133. <https://doi.org/10.1080/1750984X.2016.1184699>.
4. Li, N.; Wang, D.; Zhao, X.; Li, Z.; Zhang, L. The association between physical exercise behavior and psychological resilience of teenagers: An examination of the chain mediating effect. *Sci. Rep.* **2024**, *14*, 9372. <https://doi.org/10.1038/s41598-024-60038-1>.
5. Lemes, V.; Gaya, A.R.; Sadarangani, K.P.; Aguilar-Farias, N.; Rodriguez-Rodriguez, F.; Martins, C.M.d.L.; Fochesatto, C.; Cristi-Montero, C. Physical Fitness Plays a Crucial Mediator Role in Relationships Among Personal, Social, and Lifestyle Factors With Adolescents' Cognitive Performance in a Structural Equation Model. The Cogni-Action Project. *Front. Pediatr.* **2021**, *9*, 656916. <https://doi.org/10.3389/fped.2021.656916>.
6. Guthold, R.; Stevens, G.A.; Riley, L.M.; Bull, F.C. Global trends in insufficient physical activity among adolescents: A pooled analysis of 298 population-based surveys with 1.6 million participants. *Lancet Child Adolesc. Health* **2020**, *4*, 23–35. [https://doi.org/10.1016/S2352-4642\(19\)30323-2](https://doi.org/10.1016/S2352-4642(19)30323-2).
7. Kumar, B.; Robinson, R.; Till, S. Physical activity and health in adolescence. *Clin. Med.* **2015**, *15*, 267–272. <https://doi.org/10.7861/clinmedicine.15-3-267>.
8. Lavie, C.J.; Ozemek, C.; Carbone, S.; Katzmarzyk, P.T.; Blair, S.N. Sedentary Behavior, Exercise, and Cardiovascular Health. *Circ. Res.* **2019**, *124*, 799–815. <https://doi.org/10.1161/CIRCRESAHA.118.312669>.
9. Bucciarelli, V.; Mattioli, A.V.; Sciomer, S.; Moscucci, F.; Renda, G.; Gallina, S. The Impact of Physical Activity and Inactivity on Cardiovascular Risk across Women's Lifespan: An Updated Review. *J. Clin. Med.* **2023**, *12*, 4347. <https://doi.org/10.3390/jcm12134347>.
10. Tsai, Y.-J.; Wu, M.-P.; Hsu, Y.-W. Emerging health problems among women: Inactivity, obesity, and metabolic syndrome. *Gynecol. Minim. Invasive Ther.* **2014**, *3*, 12–14. <https://doi.org/10.1016/j.gmit.2013.07.005>.
11. Đurić, S.; Sember, V.; Starc, G.; Sorić, M.; Kovač, M.; Jurak, G. Secular trends in muscular fitness from 1983 to 2014 among Slovenian children and adolescents. *Scand. J. Med. Sci. Sports* **2021**, *31*, 1853–1861. <https://doi.org/10.1111/sms.13981>.
12. Kokkinos, P. Physical Fitness Evaluation. *Am. J. Lifestyle Med.* **2015**, *9*, 308–317. <https://doi.org/10.1177/1559827613520128>.
13. Wilder, R.P.; Greene, J.A.; Winters, K.L.; Long, W.B., III; Gubler, K.D.; Edlich, R. Physical fitness assessment: An update. *J. Long-Term Eff. Med. Implants* **2006**, *16*, 193–204. <https://doi.org/10.1615/JLongTermEffMedImplants.v16.i2.90>.
14. Marques, A.; Henriques-Neto, D.; Peralta, M.; Martins, J.; Gomes, F.; Popovic, S.; Masanovic, B.; Demetriou, Y.; Schlund, A.; Ihle, A. Field-Based Health-Related Physical Fitness Tests in Children and Adolescents: A Systematic Review. *Front. Pediatr.* **2021**, *9*, 640028. <https://doi.org/10.3389/fped.2021.640028>.
15. Wiemken, T.L.; Kelley, R.R. Machine Learning in Epidemiology and Health Outcomes Research. *Annu. Rev. Public Health* **2020**, *41*, 21–36. <https://doi.org/10.1146/annurev-publhealth-040119-094437>.
16. Farrahi, V.; Rostami, M. Machine learning in physical activity, sedentary, and sleep behavior research. *J. Act. Sedent. Sleep Behav.* **2024**, *3*, 5. <https://doi.org/10.1186/s44167-024-00045-9>.
17. Trezza, A.; Visibelli, A.; Roncaglia, B.; Spiga, O.; Santucci, A. Unsupervised Learning in Precision Medicine: Unlocking Personalized Healthcare through AI. *Appl. Sci.* **2024**, *14*, 9305. <https://doi.org/10.3390/app14209305>.
18. Jiang, T.; Gradus, J.L.; Rosellini, A.J. Supervised Machine Learning: A Brief Primer. *Behav. Ther.* **2020**, *51*, 675–687. <https://doi.org/10.1016/j.beth.2020.05.002>.

19. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 160. <https://doi.org/10.1007/s42979-021-00592-x>.
20. Segaux, L.; Oubaya, N.; Broussier, A.; Baude, M.; Canoui-Poitaine, F.; Naga, H.; Laurent, M.; Leissing-Desprez, C.; Audureau, E.; Ferrat, E.; et al. Identification of five frailty profiles in community-dwelling individuals aged 50–75: A latent class analysis of the SUCCEED survey data. *Maturitas* **2019**, *127*, 1–11. <https://doi.org/10.1016/j.maturitas.2019.05.007>.
21. Pikoula, M.; Quint, J.K.; Nissen, F.; Hemingway, H.; Smeeth, L.; Denaxas, S. Identifying clinically important COPD sub-types using data-driven approaches in primary care population based electronic health records. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 86. <https://doi.org/10.1186/s12911-019-0805-0>.
22. Junta Nacional de Auxilio Escolar y Becas (JUNAEB). *Mapa Nutricional 2020*; JUNAEB, Ministerio de Educación: Santiago, Chile. Available online: <https://www.junaeb.cl/mapa-nutricional> (accessed on 14 July 2026).
23. Yáñez-Sepúlveda, R.; Clemente-Suárez, V.J.; de Souza-Lima, J.; Ferrari, G.; Hurtado-Almonacid, J.; Zavala-Crichton, J.P.; Hinojosa-Torres, C.; Reyes-Amigo, T.; Páez-Herrera, J.; Cortés-Roco, G.; et al. Cardiorespiratory fitness and socioeconomic influences in Chilean schoolchildren: A cross-sectional study. *J. Pediatr.* **2024**, *100*, 660–666. <https://doi.org/10.1016/j.jpeds.2024.06.010>.
24. El Abbassi, M.; Overbeck, J.; Braun, O.; Calame, M.; van der Zant, H.S.J.; Perrin, M.L. Benchmark and application of unsupervised classification approaches for univariate data. *Commun. Phys.* **2021**, *4*, 50. <https://doi.org/10.1038/s42005-021-00549-9>.
25. Adeli, E.; Zhao, Q.; Zahr, N.M.; Goldstone, A.; Pfefferbaum, A.; Sullivan, E.V.; Pohl, K.M. Deep learning identifies morphological determinants of sex differences in the pre-adolescent brain. *NeuroImage* **2020**, *223*, 117293. <https://doi.org/10.1016/j.neuroimage.2020.117293>.
26. Sawyer, S.M.; Azzopardi, P.S.; Wickremarathne, D.; Patton, G.C. The age of adolescence. *Lancet Child Adolesc. Health* **2018**, *2*, 223–228. [https://doi.org/10.1016/S2352-4642\(18\)30022-1](https://doi.org/10.1016/S2352-4642(18)30022-1).
27. Agencia de Calidad de la Educación. *Estudio Nacional de Educación Física (SIMCE EF): Informes de Resultados*; Agencia de Calidad de la Educación: Santiago, Chile. Available online: <https://www.agenciaeducacion.cl/orientar/estudios/estudios-nacionales/> (accessed on 14 July 2026).
28. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants. *JAMA* **2025**, *333*, 71–74. <https://doi.org/10.1001/jama.2024.21972>.
29. Zareba, M.; Danek, T.; Stefaniuk, M. Unsupervised Machine Learning Techniques for Improving Reservoir Interpretation Using Walkaway VSP and Sonic Log Data. *Energies* **2023**, *16*, 493. <https://doi.org/10.3390/en16010493>.
30. Mello, G.T.; Bertuol, C.; Minatto, G.; Filho, V.C.B.; Oldenburg, B.; Leech, R.M.; Silva, K.S. A systematic review of the clustering and correlates of physical activity and sedentary behavior among boys and girls. *BMC Public Health* **2023**, *23*, 372. <https://doi.org/10.1186/s12889-022-14869-0>.
31. Cabanas-Sánchez, V.; Martínez-Gómez, D.; Izquierdo-Gómez, R.; Segura-Jiménez, V.; Castro-Piñero, J.; Veiga, O.L. Association between Clustering of Lifestyle Behaviors and Health-Related Physical Fitness in Youth: The UP&DOWN Study. *J. Pediatr.* **2018**, *199*, 41–48.e1. <https://doi.org/10.1016/j.jpeds.2018.03.075>.
32. Ottevaere, C.; Huybrechts, I.; Benser, J.; De Bourdeaudhuij, I.; Cuenca-Garcia, M.; Dallongeville, J.; Zaccaria, M.; Gottrand, F.; Kersting, M.; Rey-López, J.P.; et al. Clustering patterns of physical activity, sedentary and dietary behavior among European adolescents: The HELENA study. *BMC Public Health* **2011**, *11*, 328. <https://doi.org/10.1186/1471-2458-11-328>.
33. Bonilla, D.A.; Sánchez-Rojas, I.A.; Mendoza-Romero, D.; Moreno, Y.; Kočí, J.; Gómez-Miranda, L.M.; Rojas-Valverde, D.; Petro, J.L.; Kreider, R.B. Profiling Physical Fitness of Physical Education Majors Using Unsupervised Machine Learning. *Int. J. Environ. Res. Public Health* **2023**, *20*, 146. <https://doi.org/10.3390/ijerph20010146>.
34. Tingelstad, L.M.; Raastad, T.; Till, K.; Luteberget, L.S. The development of physical characteristics in adolescent team sport athletes: A systematic review. *PLoS ONE* **2023**, *18*, e0296181. <https://doi.org/10.1371/journal.pone.0296181>.
35. Thornton, C.B.; Kolehmainen, N.; Nazarpour, K. Using unsupervised machine learning to quantify physical activity from accelerometer in a diverse and rapidly changing population. *PLoS Digit. Health* **2023**, *2*, e0000220. <https://doi.org/10.1371/journal.pdig.0000220>.
36. Lubans, D.R.; Smith, J.J.; Eather, N.; Leahy, A.A.; Morgan, P.J.; Lonsdale, C.; Plotnikoff, R.C.; Nilsson, M.; Kennedy, S.G.; Holliday, E.G.; et al. Time-efficient intervention to improve older adolescents' cardiorespiratory fitness: Findings from the 'Burn 2 Learn' cluster randomised controlled trial. *Br. J. Sports Med.* **2021**, *55*, 751–758. <https://doi.org/10.1136/bjsports-2020-103277>.
37. Leahy, A.A.; Eather, N.; Smith, J.J.; Hillman, C.; Morgan, P.J.; Nilsson, M.; Lonsdale, C.; Plotnikoff, R.C.; Noetel, M.; Holliday, E.; et al. School-based physical activity intervention for older adolescents: Rationale and study protocol for the Burn 2 Learn cluster randomised controlled trial. *BMJ Open* **2019**, *9*, e026029. <https://doi.org/10.1136/bmjopen-2018-026029>.

38. Greier, K.; Drenowatz, C.; Ruedl, G.; Kirschner, W.; Mitmannsgruber, P.; Greier, C. Physical Fitness across 11- to 17-Year-Old Adolescents: A Cross-Sectional Study in 2267 Austrian Middle- and High-School Students. *Adv. Phys. Educ.* **2019**, *9*, 258–269. <https://doi.org/10.4236/ape.2019.94018>.
39. Evaristo, S.; Moreira, C.; Lopes, L.; Oliveira, A.; Abreu, S.; Agostinis-Sobrinho, C.; Oliveira-Santos, J.; Póvoas, S.; Santos, R.; Mota, J. Muscular fitness and cardiorespiratory fitness are associated with health-related quality of life: Results from labmed physical activity study. *J. Exerc. Sci. Fit.* **2019**, *17*, 55–61. <https://doi.org/10.1016/j.jesf.2019.01.002>.
40. Ortega, F.B.; Zhang, K.; Cadenas-Sánchez, C.; Tremblay, M.S.; Jurak, G.; Tomkinson, G.R.; Ruiz, J.R.; Keller, K.; Nyström, C.D.; Sacke, J.M.; et al. The Youth Fitness International Test (YFIT) battery for monitoring and surveillance among children and adolescents: A modified Delphi consensus project with 169 experts from 50 countries and territories. *J. Sport Health Sci.* **2025**, *14*, 101012. <https://doi.org/10.1016/j.jshs.2024.101012>.
41. Pulimeno, M.; Piscitelli, P.; Colazzo, S.; Colao, A.; Miani, A. School as ideal setting to promote health and wellbeing among young people. *Health Promot. Perspect.* **2020**, *10*, 316–324. <https://doi.org/10.34172/hpp.2020.50>.
42. Neil-Sztramko, S.E.; Caldwell, H.; Dobbins, M. School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6 to 18. *Cochrane Database Syst. Rev.* **2021**, *9*, CD007651. <https://doi.org/10.1002/14651858.CD007651.pub3>.
43. Xie, L.; Gou, B.; Bai, S.; Yang, D.; Zhang, Z.; Di, X.; Su, C.; Wang, X.; Wang, K.; Zhang, J. Unsupervised cluster analysis reveals distinct subgroups in healthy population with different exercise responses of cardiorespiratory fitness. *J. Exerc. Sci. Fit.* **2023**, *21*, 147–156. <https://doi.org/10.1016/j.jesf.2022.12.005>.
44. Oehm, A.W.; Zablotzki, Y.; Campe, A.; Hoedemaker, M.; Strube, C.; Springer, A.; Jordan, D.; Knubben-Schweizer, G. Random forest classification as a tool in epidemiological modelling: Identification of farm-specific characteristics relevant for the occurrence of *Fasciola hepatica* on German dairy farms. *PLoS ONE* **2023**, *18*, e0296093. <https://doi.org/10.1371/journal.pone.0296093>.
45. Speiser, J.L.; Miller, M.E.; Tooze, J.; Ip, E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst. Appl.* **2019**, *134*, 93–101. <https://doi.org/10.1016/j.eswa.2019.05.028>.
46. Zale, A.D.; Abusamaan, M.S.; McGready, J.; Mathioudakis, N. Development and validation of a machine learning model for classification of next glucose measurement in hospitalized patients. *EClinicalMedicine* **2022**, *44*, 101290. <https://doi.org/10.1016/j.eclim.2022.101290>.
47. Loef, B.; Wong, A.; Janssen, N.A.; Strak, M.; Hoekstra, J.; Picavet, H.S.; Boshuizen, H.H.; Verschuren, W.M.; Herber, G.C. Using random forest to identify longitudinal predictors of health in a 30-year cohort study. *Sci. Rep.* **2022**, *12*, 10372. <https://doi.org/10.1038/s41598-022-14632-w>.
48. Yáñez-Sepúlveda, R.; Vásquez-Bonilla, A.; Olivares, R.; Olivares, P.; Zavala-Crichton, J.P.; Hinojosa-Torres, C.; Muñoz-Strale, C.; Giakoni-Ramírez, F.; de Souza-Lima, J.; Páez-Herrera, J.; et al. Supervised machine learning algorithms for the classification of obesity levels using anthropometric indices derived from bioelectrical impedance analysis. *Sci. Rep.* **2025**, *15*, 30681. <https://doi.org/10.1038/s41598-025-15264-6>.
49. Chen, Z.; Zhu, L.; Chen, P. Development and validation of an explainable machine learning-based risk prediction model for obesity in Chinese children and adolescents: A population-based study. *Front. Nutr.* **2026**, *13*, 1816724. <https://doi.org/10.3389/fnut.2026.1816724>.
50. Lv, M.; Wang, J.; Yang, Y.; Zhong, J.; Yang, J.; Dong, C. An AI-based algorithm for analyzing physical activity and health-related fitness in youth. *Sci. Rep.* **2026**, *16*, 5105. <https://doi.org/10.1038/s41598-026-35514-5>.
51. Qiu, Y.; Yan, C.; Zhao, P.; Zou, Q. SSNMDI: A novel joint learning model of semi-supervised non-negative matrix factorization and data imputation for clustering of single-cell RNA-seq data. *Brief. Bioinform.* **2023**, *24*, bbad149. <https://doi.org/10.1093/bib/bbad149>.
52. Ghaderi, H.; Foreman, B.; Nayeibi, A.; Tipirneni, S.; Reddy, C.K.; Subbian, V. A self-supervised learning-based approach to clustering multivariate time-series data with missing values (SLAC-Time): An application to TBI phenotyping. *J. Biomed. Inform.* **2023**, *143*, 104401. <https://doi.org/10.1016/j.jbi.2023.104401>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.