



Multi-domain physical fitness profiles and machine learning classification by sport category across 13 competitive sports

Sahabuddin*

Universitas Negeri Makassar,
INDONESIA

Syahrudin

Universitas Negeri Makassar,
INDONESIA

Muhammad Ishak

Universitas Negeri Makassar,
INDONESIA

Muflih Wahid Hamid

Universitas Negeri Makassar,
INDONESIA

M Rachmat Kasmad

Universitas Negeri Makassar,
INDONESIA

Article Info

Article history:

Received: June 8, 2026

Revised: July 28, 2026

Accepted: September 17, 2026

Keywords:

Athletic profiling;
Fitness battery;
Machine learning;
Multi-sport classification;
Physical fitness.

Abstract

Background: Comparative multi-sport fitness profiling using a common institutional assessment framework remains limited in Southeast Asian competitive sport.

Aims: To compare multi-domain fitness profiles across 13 sports grouped into four sport categories and to evaluate machine-learning classification of sport category.

Methods: This cross-sectional STROBE-reported study included 197 KONI athletes (137 males, 60 females; age 20.6 ± 5.8 years). Standardized procedures were used within applicable sport groups, although several tests were category-dependent. Athletes were grouped as Combat ($n=69$), Team ($n=72$), Individual Technical ($n=37$), and Precision/Shooting ($n=19$). Between-category analyses and Random Forest (RF) and Linear Discriminant Analysis (LDA) with 5-fold out-of-fold cross-validation were performed.

Results: The largest between-category effects were observed for vertical jump ($\eta^2=0.438$, $p<.001$), the Bleep Composite/estimated VO_2max signal ($\eta^2=0.217$, $p<.001$), and sit-up endurance ($\eta^2=0.199$, $p<.001$). RF achieved 79.7% accuracy and macro-AUC=0.907; the three largest Gini importances in the final nine-variable model were vertical jump (0.211), Bleep Composite (0.162), and push-up endurance (0.144).

Conclusion: Multi-domain fitness patterns differed across the four administrative sport categories, but age, maturation, sex composition, and category-linked test availability constrain interpretation. The classifier is exploratory and requires external validation; the findings describe already-trained athletes and do not establish prospective talent-identification validity.

To cite this article: Sahabuddin., Syahrudin., Ishak, M., Hamid, M. W., & Kasmad, M. R. (2026). Multi-domain physical fitness profiles and machine learning classification by sport category across 13 competitive sports. *Journal of Coaching and Sports Science*, 5(3), 137-150. <https://doi.org/10.58524/jcss.v5i3.1332>

This article is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/) ©2026 by author/s

INTRODUCTION

Physical fitness profiling integrates multiple biomotor domains to describe athletes' functional capacities and support evidence-based monitoring. Its value is greatest when measurements are obtained under comparable protocols, because cross-sport interpretation becomes difficult when samples, instruments, or test procedures differ (Currell & Jeukendrup, 2008; Wang et al., 2023). National sport systems routinely collect large field-test datasets, yet the published literature remains dominated by single-sport studies. In Indonesia, KONI conducts multi-sport assessments across regional athlete pools, but these data have rarely been examined within a single comparative analytical framework (Republic of Indonesia, 2022; Syahputra et al., 2022; Sukamti et al., 2020).

A multi-domain perspective is especially important in competitive settings because no single field test can represent the complete functional demands placed on an athlete. Strength, explosive power, muscular endurance, cardiorespiratory capacity, flexibility, coordination, and related attributes interact within a broader performance system, and the relative importance of each domain can differ substantially between sports. Examining these qualities together therefore provides a more informative description of athlete phenotype than interpreting isolated test scores. At the same

* Corresponding author:

Sahabuddin., Universitas Negeri Makassar, INDONESIA. ✉sahabuddin@unm.ac.id

time, cross-sport profiling requires restraint: an apparent difference between categories may reflect sport-specific preparation, selection into a discipline, demographic composition, or differences in which tests were administered. For this reason, the value of a comparative battery lies not only in detecting contrasts but also in identifying overlap, heterogeneity, and the boundaries of what the available measurements can legitimately support.

Cross-sport comparison is relevant because physiological demands differ substantially across sport families. Team sports combine intermittent high-intensity actions with aerobic recovery; combat sports combine explosive, anaerobic, isometric, and recovery demands; individual technical sports vary widely in coordination, power, flexibility, and metabolic requirements; and shooting emphasizes postural, visual, and fine-motor control more than gross locomotor capacity (Modric et al., 2019; Chaabene et al., 2018; Phomsoupha & Laffaye, 2015; Mononen et al., 2007). These categories are therefore useful administrative groupings rather than assumptions of physiological homogeneity. This distinction is particularly important here because the Precision category contains only Shooting.

The use of broad sport categories also creates an important analytical tension. Categories simplify a complex sporting landscape and make system-level comparison feasible, yet they inevitably compress meaningful discipline-level variation. Two sports assigned to the same administrative family may differ in movement structure, work-to-rest pattern, technical constraints, competitive duration, and the developmental pathway through which athletes reach competitive status. Consequently, a category-level profile should be understood as an aggregate pattern rather than a physiological template that every athlete within that category is expected to match. This distinction is central to the present study because the analytical objective is to determine whether broad groupings organize fitness data to a useful degree, while simultaneously testing how much individual-level variation remains after those labels are imposed.

Machine learning (ML) offers a complementary way to examine whether multivariate fitness signatures discriminate sport categories. Random Forest can model non-linear interactions without multivariate-normality assumptions, whereas Linear Discriminant Analysis (LDA) provides a transparent linear benchmark. Recent sport applications demonstrate the growing use of supervised and unsupervised ML for athlete profiling and classification, but most remain focused on a single sport, playing position, or performance subgroup rather than broad cross-sport discrimination (Figueiredo et al., 2021; Sneha et al., 2024; de Haan et al., 2025; Cattle et al., 2026). A targeted update of this recent literature therefore supports, rather than proves, the continuing relevance of the present cross-sport question. Model performance must nevertheless be interpreted cautiously when classes are imbalanced, predictors are correlated, or measurement availability is linked to the target label.

From this perspective, classification accuracy is not treated as an end in itself. A model that separates existing sport categories can reveal how strongly the measured fitness domains jointly encode current sport membership. Still, successful discrimination does not establish that the same variables caused athletes to enter those sports or that the model would correctly allocate unspecialized individuals. The comparison between RF and LDA is therefore useful because it contrasts a flexible ensemble approach with a simpler linear decision structure under the same data constraints. Feature rankings can then be read alongside conventional effect sizes and multivariate clustering to determine whether different analytical approaches converge on similar domains or whether a more complex combination of variables drives apparent predictive performance. Such triangulation is particularly valuable when the dataset contains structural irregularities that would be easy to overlook if only one analytical framework were used.

The present dataset provides an opportunity to examine these issues within one institutional assessment system covering 13 sports. A key methodological limitation is that the KONI battery was standardized within applicable sport groups but not every test was administered to every sport. Consequently, category-linked structural missingness and demographic imbalance must be treated as design constraints rather than ordinary random missing data. The study therefore emphasizes descriptive cross-sport profiling and internally validated classification, not normative cut-offs, causal training effects, or prospective talent identification.

Within the Indonesian athlete-development context, this question has practical relevance beyond the performance of a single classifier. Institutional testing programs generate repeated streams of information that can potentially support monitoring, communication between coaches

and sport-science staff, and decisions about where additional assessment is needed. However, these uses depend on understanding what the tests actually distinguish and where category-level summaries become misleading. A comparative analysis of a shared institutional dataset can therefore serve two functions: it can describe the dominant fitness contrasts presently visible across the competitive cohort, and it can expose measurement-design issues that should be addressed before more ambitious predictive applications are attempted. In this sense, the present study treats the KONI dataset as both a source of empirical information and a case study in how heterogeneous multi-sport field data should be interpreted responsibly.

Accordingly, the study had two aims: (1) to characterize and compare multi-domain physical fitness profiles of athletes from 13 sports grouped as Combat, Team, Individual Technical, and Precision/Shooting; and (2) to evaluate sport-category discrimination using Random Forest against an LDA baseline, with explicit attention to feature attribution, cross-validation, structural missingness, and demographic confounding. The intended contribution is a bounded, hypothesis-generating account of cross-sport fitness structure within Indonesian competitive sport and a reproducible analytical framework for subsequent external validation.

By integrating univariate comparisons, unsupervised multivariate structure, and supervised classification within the same cohort, the study is positioned to evaluate consistency across analytical levels. If the same domains emerge as large between-category effects, contribute to recognizable multivariate patterns, and receive high model importance, the convergence strengthens their descriptive relevance within this dataset. Conversely, incomplete cluster alignment, broad within-category dispersion, or sensitivity to predictor availability would indicate that administrative sport labels capture only part of the underlying fitness structure. This integrated interpretation is intended to produce a more informative and appropriately bounded account than either conventional significance testing or machine-learning accuracy could provide in isolation.

METHOD

Figure 1 summarizes participant recruitment, fitness assessment, preprocessing, statistical comparison, and machine-learning analysis.

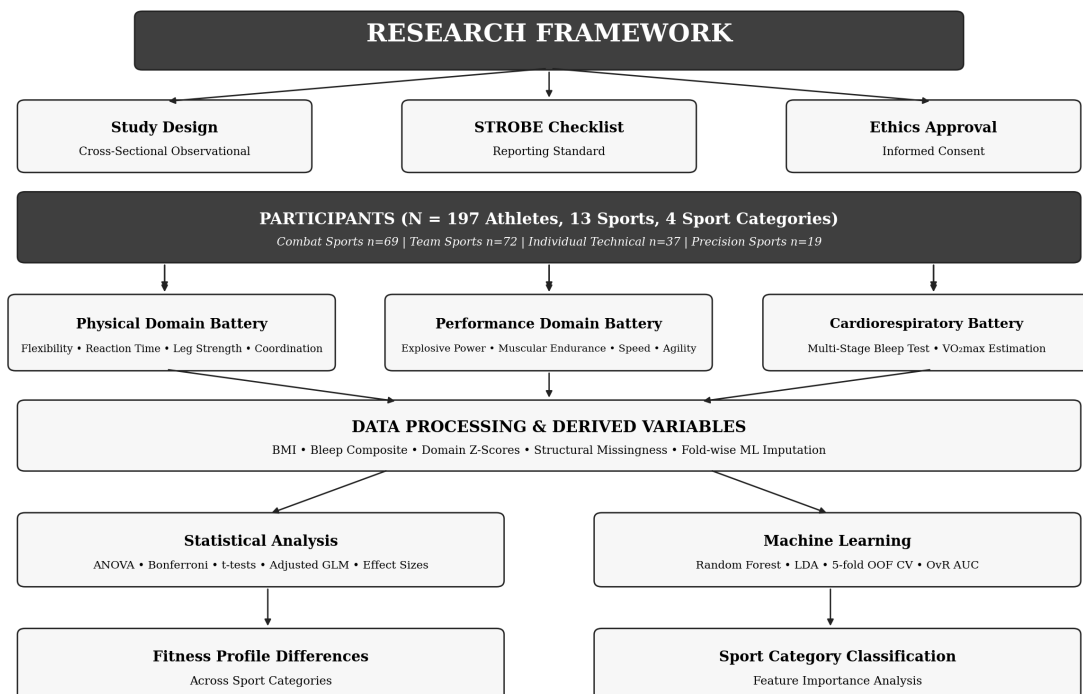


Figure 1. Research Framework and Analytical Workflow

Research Design

A cross-sectional observational design was used and reported with reference to STROBE (Levin, 2006; von Elm et al., 2007). Testing was conducted between January and April 2023 by

certified KONI assessors using standardized institutional procedures within each applicable test at the participating provincial assessment sites; this refers to protocolized assessment sessions rather than one literal session for the entire cohort.

Participants

Athletes were recruited from KONI competitive pools in Lampung and South Sulawesi between January 10 and April 30, 2023. Of 250 eligible athletes approached, 215 were assessed and 18 with incomplete data across more than three domains were excluded, yielding 197 athletes (137 males, 60 females; age 20.6 ± 5.8 years). Categories were Combat Sports (Karate, Kempo, Judo, Yongmodo; $n=69$), Team Sports (Football, Futsal, Sepaktakraw, Softball; $n=72$), Individual Technical Sports (Gymnastics, Badminton, Tennis, Fencing; $n=37$), and Precision Sports (Shooting; $n=19$). Category age/sex distributions were: Combat 20.8 ± 3.7 years, 55.1% male; Team 20.1 ± 2.9 years, 95.8% male; Individual Technical 17.7 ± 3.8 years, 45.9% male; and Shooting 27.2 ± 10.2 years, 68.4% male.

Inclusion required active KONI registration, injury- and illness-free status at testing, and at least three consecutive months of organized training. Mean training age was 5.2 ± 2.1 years and mean weekly training volume was 12.5 ± 3.4 hours—the three-month threshold defined minimum eligibility rather than total competitive experience. Of the 215 athletes assessed, 18 were excluded because data were incomplete across more than three domains; the exclusion pattern is reported descriptively, and no inferential claim about exclusion bias is made. Written informed consent was obtained from adults and from parents/legal guardians of minors in accordance with the Declaration of Helsinki ([World Medical Association, 2013](#)). Minors aged 8–12 years provided verbal assent and those aged 13–17 years signed a simplified assent form. Prospective ethical approval was obtained from the Health Research Ethics Committee of Universitas Negeri Makassar on December 15, 2022, before data collection.

Instrumentation and Fitness Battery

The field battery used standardized procedures within applicable sport groups. Common measures were flexibility (Stand and Reach), bilateral isometric leg strength, push-up, sit-up, half-squat endurance, and the Multi-Stage Fitness/Bleep Test ([Mayorga-Vega et al., 2014](#); [Bishop et al., 2017](#); [Santos et al., 2022](#); [Ramsbottom et al., 1988](#)). Whole-body reaction time was available for Combat and Shooting; coordination for Team and Individual Technical sports; vertical jump for Combat, Team, and Individual Technical sports; and speed/agility primarily for Combat and Team sports ([Engel et al., 2021](#); [Sekulic et al., 2017](#); [Markovic, 2007](#)). A medicine-ball throw recorded for Shooting was treated as a distinct upper-body measure and was not merged with vertical jump. Estimated VO_2max was calculated as $0.1428 \times \text{Bleep Composite} + 28.498$, where Bleep Composite = Level $\times 10 + \text{Shuttles}$ ([Léger et al., 1988](#); [Mayorga-Vega et al., 2015](#)). Speed, reaction time, and coordination were excluded from the primary cross-category ANOVA because of limited availability. Agility was retained in the final nine-variable ML feature set as a structurally missing predictor and was handled using fold-internal median imputation, as specified below. No site-specific test-retest or inter-rater reliability study was conducted for this KONI cohort, so no pooled reliability coefficient is claimed for the present dataset.

Data Processing and Derived Variables

Data were processed in Python 3.12 (scikit-learn 1.4; pandas 2.2). Missingness was structurally related to the KONI protocol and therefore to sport category. Descriptive and ANOVA results used available observations without imputation. For ML only, median imputation was fitted separately within each training fold and then applied to the corresponding held-out fold through a scikit-learn Pipeline. This procedure reduced preprocessing-related data leakage by preventing evaluation-fold information from entering the imputation step, while not eliminating the broader risks associated with category-linked structural missingness or same-sport athletes appearing across folds. A sensitivity analysis re-fitted RF and LDA using only predictors observed across the full cohort. The restricted-feature RF achieved 76.5% cross-validated accuracy and macro-AUC=0.880, 3.2 percentage points below the primary RF accuracy. This analysis was used to gauge, not eliminate, the influence of structurally unavailable predictors ([van Buuren, 2018](#)). BMI was calculated as kg/m^2 . For multivariate visualization, selected fitness outcomes were standardized to full-sample Z-scores.

Estimated VO₂max was retained for descriptive interpretation but excluded from ML because it is a deterministic transformation of the Bleep Composite.

Statistical Analysis

Descriptive statistics are reported as mean $\pm$ SD. Between-category differences were tested with one-way ANOVA and Bonferroni post hoc comparisons; η^2 was interpreted as small $\geq .01$, medium $\geq .06$, and large $\geq .14$ (Cohen, 1988; Field, 2024). Bonferroni adjustment controlled pairwise comparisons within each outcome but not the family of separate outcomes, so the analyses remain exploratory. Alpha was .05. Sex-stratified comparisons used independent-samples t-tests and Cohen's d and are presented descriptively because sex composition differed sharply by sport category. Exploratory adjusted general linear models used age as a continuous covariate, sport category as a categorical fixed factor, and sex as the focal fixed factor for selected outcomes. A sex $\times$ sport-category interaction was considered but not retained because several category-by-sex cells were too sparse for stable estimation. Distributional assumptions were screened with Shapiro–Wilk and Levene tests, with Welch/Games–Howell checks used where unequal variance was evident.

Machine Learning Classification

Two classifiers were evaluated: LDA as a linear baseline and Random Forest (RF) as a non-parametric ensemble (Breiman, 2001). RF settings were fixed a priori at 300 trees, maximum depth 10, and minimum samples per leaf 2; no data-driven hyperparameter search was conducted. The initial feature pool contained 10 variables after preprocessing. Estimated VO₂max was excluded to avoid duplicating Bleep Composite information, leaving the definitive nine-variable RF/LDA feature set: flexibility, leg strength, vertical jump, push-up, sit-up, half-squat, Bleep Composite, BMI, and agility. This same nine-variable specification is used consistently in the Results, Table 4, Figure 4, Discussion, and Limitations. Performance was estimated with stratified 5-fold cross-validation and out-of-fold (OOF) predictions. Primary metrics were accuracy and one-vs-rest ROC-AUC by class and macro-average (Hastie et al., 2022). Because athletes from the same sport could occur in training and test folds, this design estimates generalization to new athletes from sports already represented rather than to entirely new sports. Gini importance was calculated from RF refitted on the full analysis dataset and treated as descriptive attribution, not out-of-sample effect estimation. Cross-validated permutation importance was used as a sensitivity check because impurity-based rankings can be biased by predictor properties and correlation (Strobl et al., 2007; Pedregosa et al., 2011). The two approaches were interpreted as partially concordant rather than interchangeable; agreement at the domain level was used to assess robustness without claiming identical predictor rankings. Ward hierarchical clustering with Euclidean distance was applied to selected domain Z-scores as an unsupervised descriptive analysis. A four-cluster cut was examined to enable direct comparison with the four administrative sport categories and yielded an average silhouette score of 0.44. Cophenetic correlation assessed preservation of pairwise distances, while Adjusted Rand Index (ARI=0.48) and Normalized Mutual Information (NMI=0.55) quantified partial agreement with the administrative labels. The clustering metrics are therefore reported descriptively and are not used to claim that k=4 is a uniquely optimal natural partition.

RESULTS AND DISCUSSION

Results

Participant Characteristics

Table 1 shows marked demographic imbalance across categories. Team Sports were 95.8% male, Individual Technical Sports contained the youngest athletes (17.7 ± 3.8 years), and Shooting athletes were the oldest (27.2 ± 10.2 years). These differences are important confounders for all subsequent category comparisons.

Table 1. Participant Demographics and Body Composition By Sport Category

Sport Category	n	Male / Female	Age (yr)	Weight (kg)	Height (cm)	BMI (kg/m ²)
Combat Sports	69	38 / 31	20.8 $\pm$ 3.7	61.5 $\pm$ 11.2	161.4 $\pm$ 8.3	23.5 $\pm$ 3.7
Team Sports	72	69 / 3	20.1 $\pm$ 2.9	63.2 $\pm$ 8.1	165.4 $\pm$ 6.2	23.3 $\pm$ 3.4
Individual Technical	37	17 / 20	17.7 $\pm$ 3.8	59.2 $\pm$ 13.4	160.2 $\pm$ 11.1	22.7 $\pm$ 4.1
Precision Sports	19	13 / 6	27.2 $\pm$ 10.2	68.4 $\pm$ 15.7	164.1 $\pm$ 8.5	25.2 $\pm$ 4.1
Total	197	137 / 60	20.6 $\pm$ 5.8	62.3 $\pm$ 10.8	163.0 $\pm$ 8.6	23.6 $\pm$ 3.8

Note. Values are mean $\pm$ SD unless stated.

Between-Category Fitness Profile Differences

Table 2 shows significant between-category differences for all listed outcomes except flexibility. Vertical jump had the largest effect ($\eta^2=0.438$, $F=67.94$, $p<.001$), followed by the Bleep Composite/estimated VO_2max signal ($\eta^2=0.217$, $p<.001$) and sit-up endurance ($\eta^2=0.199$, $p<.001$). Push-up endurance showed a medium effect ($\eta^2=0.080$), leg strength a small effect ($\eta^2=0.052$), and flexibility a small non-significant effect ($\eta^2=0.013$). Because estimated VO_2max is a positive linear transformation of Bleep Composite, both rows represent the same cardiorespiratory signal in different units. Team Sports recorded the highest mean vertical jump (55.3 ± 8.1 cm), whereas Individual Technical Sports showed substantially lower and more dispersed values (23.9 ± 21.1 cm). Shooting recorded the lowest mean Bleep Composite (44.8 ± 18.8) and estimated VO_2max (34.9 ± 2.7 $\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$). Figure 2 illustrates these contrasts and the substantial within-category overlap in leg strength.

Table 2. One-Way ANOVA Results for Physical Fitness Variables across Four Sport Categories

Variable	Combat (n=69)	Team (n=72)	Individual (n=37)	Precision (n=19)	df	F	p	η^2	Effect
Flexibility (cm)	15.4 $\pm$ 5.7	15.0 $\pm$ 8.1	14.5 $\pm$ 7.7	12.6 $\pm$ 5.7	3,188	0.81	.488	.013	Small
Leg Strength (kg)	146.6 $\pm$ 55.3	155.9 $\pm$ 53.3	121.5 $\pm$ 50.4	140.7 $\pm$ 46.8	3,187	3.52	.016*	.052	Small
Vertical Jump (cm)	45.8 $\pm$ 12.4	55.3 $\pm$ 8.1	23.9 $\pm$ 21.1	—	2,175	67.94	<.001***	.438	Large
Push-Up (reps)	30.4 $\pm$ 20.1	22.8 $\pm$ 9.9	31.1 $\pm$ 18.9	17.7 $\pm$ 11.5	3,190	5.59	.001**	.080	Medium
Sit-Up (reps)	37.0 $\pm$ 13.6	33.6 $\pm$ 10.7	51.1 $\pm$ 32.5	18.3 $\pm$ 8.5	3,190	15.96	<.001***	.199	Large
Bleep Composite	69.0 $\pm$ 21.0	88.3 $\pm$ 23.1	91.5 $\pm$ 44.4	44.8 $\pm$ 18.8	3,192	17.29	<.001***	.217	Large
VO_2max est. ($\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$)	38.3 $\pm$ 3.0	41.1 $\pm$ 3.3	41.6 $\pm$ 6.3	34.9 $\pm$ 2.7	3,192	17.29	<.001***	.217	Large

Note. * $p<.05$; ** $p<.01$; *** $p<.001$. η^2 : eta-squared effect size (Cohen, 1988). Descriptive statistics are presented as mean $\pm$ standard deviation. The reported degrees of freedom (df) reflect pairwise-available data for each specific test. Vertical jump was not assessed in Precision Sports (—); the vertical jump ANOVA therefore compares only three groups (Combat, Team, Individual Technical; $n=178$, $df=2$, 175). Estimated VO_2max was calculated via the equation: $\text{VO}_2\text{max} = 0.1428 \times \text{Bleep_Composite} + 28.498$ (adapted from Ramsbottom et al., 1988); as a deterministic transformation of the Bleep Composite, its statistical sample size and df strictly mirror those of the source variable.

Figure 2 visualizes the strongest power and cardiorespiratory contrasts together with leg strength, which showed considerable overlap despite a statistically significant small effect.

Distribution of Key Fitness Variables Across Sport Categories

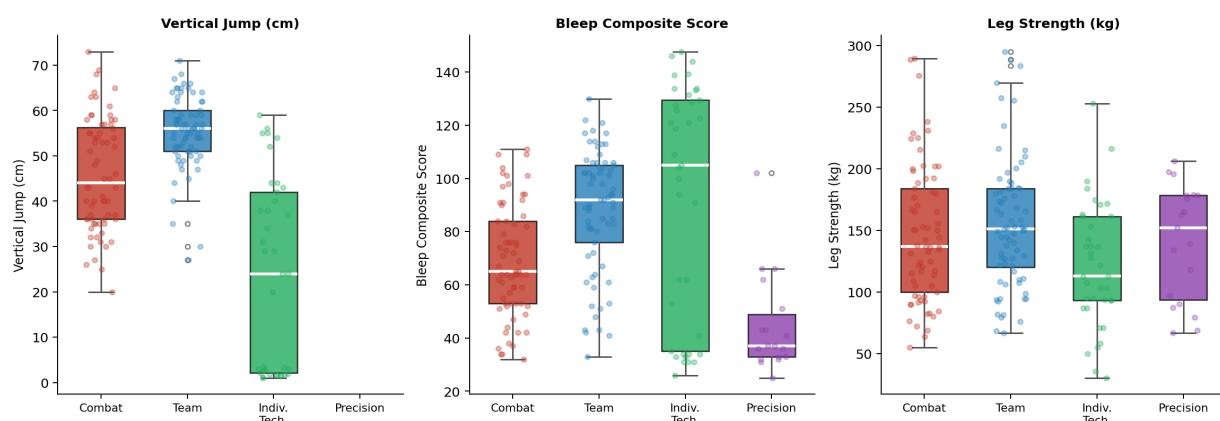


Figure 2. Box plots with individual data-point overlays (jitter) for three key fitness variables across four sport categories: vertical jump (cm; left), Bleep Composite score (center), and leg strength (kg; right). Boxes span the interquartile range; horizontal lines indicate medians; whiskers extend to $1.5 \times \text{IQR}$. Individual data points are overlaid with horizontal jitter to reduce overplotting. Absent boxes indicate the variable was not assessed in that category.

Sex-Stratified Fitness Comparisons

Pooled sex-stratified comparisons (Table 3) showed large differences in leg strength ($d=1.44$) and vertical jump ($d=1.30$), and medium differences in push-up endurance ($d=0.66$) and the Bleep Composite/estimated VO_2max signal ($d=0.60$); flexibility and sit-up endurance were small

and non-significant. These estimates remain confounded by unequal sex composition across sport categories (Sandbakk et al., 2018; Konrad et al., 2021).

Table 3. Sex-Stratified Fitness Comparison: Independent-Samples t-test (N=197)

Variable	Male	Female	t	df	p	d
	Mean $\pm$ SD	Mean $\pm$ SD				
Flexibility (cm)	14.4 $\pm$ 7.6	15.7 $\pm$ 5.4	-1.23	190	.219	0.19
Leg Strength (kg)	162.9 $\pm$ 50.8	103.2 $\pm$ 34.8	8.29	189	<.001***	1.44
Vertical Jump (cm)	51.1 $\pm$ 15.2	31.3 $\pm$ 15.3	8.01	176	<.001***	1.30
Push-Up (reps)	28.2 $\pm$ 15.6	18.3 $\pm$ 14.5	4.16	192	<.001***	0.66
Sit-Up (reps)	35.6 $\pm$ 14.8	33.1 $\pm$ 18.0	0.99	192	.323	0.15
Bleep Composite	84.9 $\pm$ 28.2	68.1 $\pm$ 27.6	3.75	194	<.001***	0.60
VO₂max est. (ml·kg⁻¹·min⁻¹)	40.6 $\pm$ 4.0	38.2 $\pm$ 3.9	3.75	194	<.001***	0.60

Note. *** $p < .001$. d = Cohen's d : small ≥ 0.2 , medium ≥ 0.5 , large ≥ 0.8 . Outcome-specific available-case counts varied because of structural missingness. Vertical jump was unavailable for the 19 Shooting athletes, yielding $n=178$ for that comparison ($df=176$). Estimated VO₂max is a positive linear transformation of Bleep Composite; therefore, it uses the same observations and has the same t statistic, df , p value, and standardized effect size as the Bleep Composite, while its means and SDs are expressed in ml·kg⁻¹·min⁻¹.

In exploratory models adjusting for age and sport category, the sex main effect remained significant for push-up endurance ($F(1,192)=15.42$, $p < .001$, partial $\eta^2=0.08$) and Bleep Composite ($F(1,192)=12.18$, $p < .001$, partial $\eta^2=0.06$). Sparse cells and the absence of a stable interaction model limit stronger inference.

Multivariate Fitness Structure

To complement the univariate comparisons reported above, hierarchical cluster analysis was used to examine whether athletes exhibited coherent multivariate fitness configurations when the six standardized fitness domains were considered simultaneously. This unsupervised analysis was intended to characterize the natural organization of the observed fitness profiles rather than to reproduce the predefined sport-category labels.

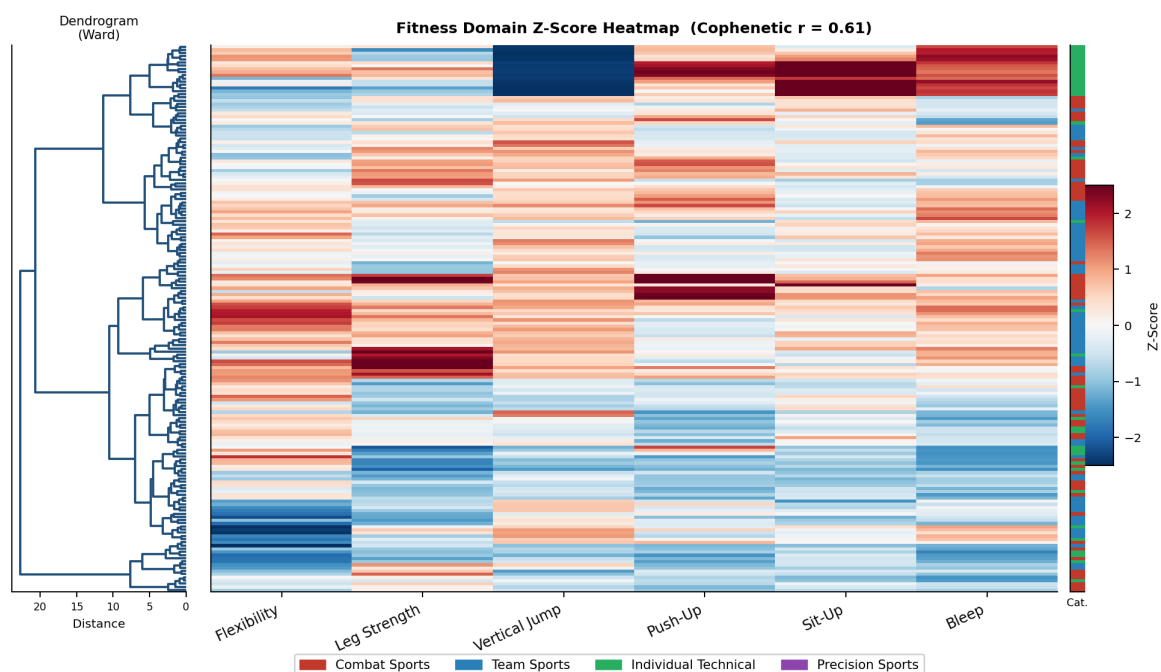


Figure 3. Hierarchical clustering of multi-domain fitness profiles (Ward linkage, Euclidean distance). Left: dendrogram; center: heatmap of six Z-scored fitness domains; right: sport-category membership. Cophenetic $r=0.611$; the four-cluster solution is descriptive rather than a one-to-one reconstruction of sport categories.

Ward linkage with Euclidean distance was therefore applied to the domain Z-scores, allowing athletes with similar combinations of flexibility, leg strength, vertical-jump performance, push-up endurance, sit-up endurance, and Bleep-test performance to be grouped according to overall profile similarity. The resulting dendrogram and heatmap provide a visual representation of both between-

athlete heterogeneity and the extent to which sport categories coincide with the multivariate fitness structure.

As shown in Figure 3, the multivariate structure exhibited partial, but not complete, concordance with sport-category membership. At $k = 4$, agreement between the data-driven clusters and the predefined categories was moderate ($ARI = 0.48$; $NMI = 0.55$), indicating that sport category accounted for a meaningful proportion of the observed fitness organization while substantial within-category heterogeneity remained. Team Sports athletes occurred more frequently in branches characterized by comparatively higher vertical-jump and Bleep-test Z-scores, consistent with a profile emphasizing explosive lower-body power and cardiorespiratory capacity. In contrast, athletes from the Precision Sports category, represented by Shooting in the present sample, were concentrated more often in branches with relatively lower scores on these domains. Combat Sports and Individual Technical Sports athletes were distributed across several branches, suggesting greater internal diversity in their combinations of strength, endurance, power, and flexibility.

The cophenetic correlation of $r = 0.611$ further indicates that the hierarchical representation preserved the original pairwise distance structure to a moderate degree. Accordingly, the clustering results should be interpreted as evidence of overlapping multivariate fitness phenotypes rather than discrete physiological signatures that map uniquely onto sport categories. This distinction is important because the observed clusters may reflect not only sport-specific training demands, but also the substantial variation in age, sex composition, maturation, training history, and discipline-specific characteristics within each category. Taken together, the analysis extends the preceding univariate findings by showing that between-category differences in individual fitness variables coexist with considerable athlete-level overlap when physical fitness is considered as a multidimensional construct.

Machine Learning Classification Performance

Table 4 summarizes the RF and LDA results; Figure 4 shows RF out-of-fold ROC curves and the three largest Gini importances from the final nine-variable model.

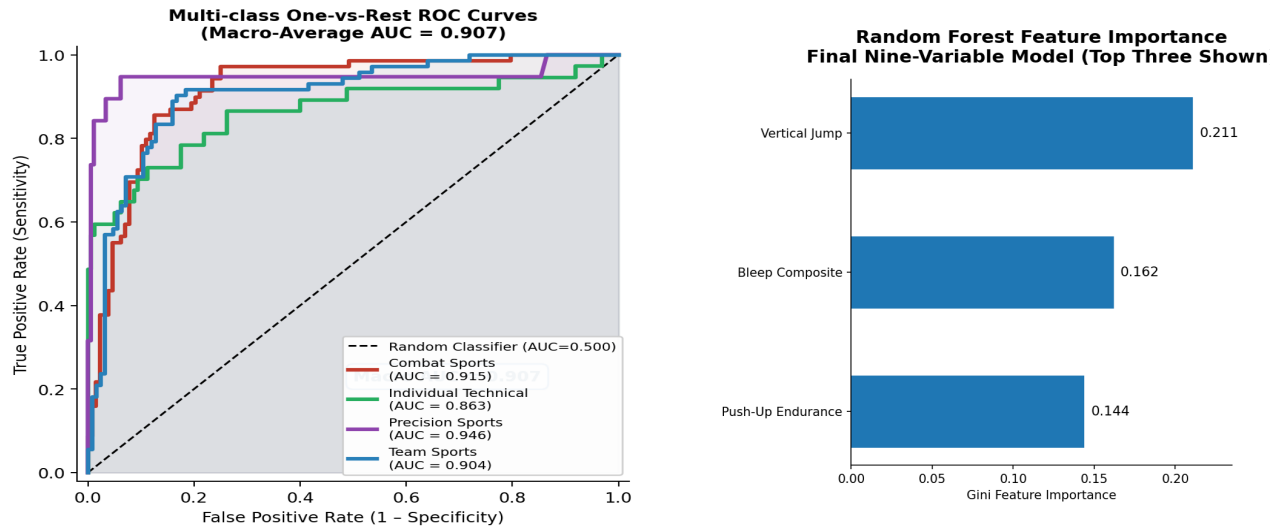


Figure 4. Random Forest classification results. Left: one-vs-rest ROC curves from 5-fold out-of-fold predictions. Right: the three largest Gini importances in the final nine-variable model. AUC = area under the receiver operating characteristic curve.

RF achieved 79.7% accuracy and macro-AUC=0.907, compared with 75.6% and 0.815 for LDA. RF per-class AUC ranged from 0.863 (Individual Technical) to 0.946 (Shooting). The three largest Gini importances were vertical jump (0.211), Bleep Composite (0.162), and push-up endurance (0.144). Gini rankings are descriptive and were interpreted with the partially concordant permutation-importance sensitivity analysis.

Table 4. Machine Learning Classification Performance: Random Forest vs. Linear Discriminant Analysis (5-fold out-of-fold cross-validation)

Metric	Random Forest	Linear Discriminant Analysis
Cross-validated accuracy	79.7%	75.6%
AUC — Combat Sports (OvR)	0.915	0.813

Metric	Random Forest	Linear Discriminant Analysis
AUC — Team Sports (OvR)	0.904	0.882
AUC — Individual Technical (OvR)	0.863	0.791
AUC — Precision Sports / Shooting (OvR)	0.946	0.774
Macro-average AUC	0.907	0.815
RF-LDA accuracy gap	4.1 pp	—
#1 Gini feature importance	Vertical Jump (0.211)	N/A
#2 Gini feature importance	Bleep Composite (0.162)	N/A
#3 Gini feature importance	Push-Up Endurance (0.144)	N/A

Note. OOF = out-of-fold; AUC = one-vs-rest ROC-AUC; pp = percentage points. Both classifiers were evaluated under the same stratified 5-fold cross-validation framework. The RF-LDA difference is descriptive; no paired inferential test was performed. Random Forest feature importances are model-based descriptive attributions from the final nine-variable feature set and should not be interpreted as out-of-sample effect estimates.

Discussion

The results show that cross-sport separation in this cohort was driven most strongly by explosive power and cardiorespiratory fitness, while multivariate classification was useful but clearly exploratory. These patterns describe already-trained athletes and cannot distinguish self-selection into a sport from training-induced adaptation (Gil & Andacao, 2025). Viewed across the three analytical layers, the results are notable for both convergence and discordance. The univariate analyses identified clear category contrasts in selected domains, the hierarchical clustering revealed recognizable but overlapping multivariate patterns, and RF achieved useful within-dataset discrimination. Together, these findings indicate that sport membership is associated with a structured fitness signal, but that the signal is neither uniform within categories nor sufficiently isolated from demographic and measurement characteristics to justify deterministic interpretation. This combined pattern is more informative than a high or low classification score alone because it shows that the four administrative categories organize part of the fitness space while leaving substantial athlete-level variability unexplained.

Vertical jump was the strongest univariate discriminator and the largest Gini feature. This is physiologically plausible because jumping integrates rapid force production and stretch-shortening-cycle function, qualities emphasized in many intermittent and explosive sports (Suchomel et al., 2016; Chaabene et al., 2018). Team Sports had the highest mean jump, but the wide dispersion in Individual Technical Sports and the broad age range of that category indicate that maturation and discipline mix likely contribute to the observed effect. The category contrast should therefore not be interpreted as a direct effect of training.

The magnitude of the vertical-jump contrast is especially important when considered together with the distributional plots and the clustering solution. A large category effect does not mean that vertical-jump ability uniquely defines any one sport family; rather, it indicates that the distributions were sufficiently separated in this cohort to make explosive lower-body performance a strong organizing variable. Team Sports showed a comparatively high and compact profile, whereas Individual Technical Sports exhibited much greater dispersion. That dispersion is analytically meaningful because it demonstrates how a category mean can conceal athletes whose demands and developmental status differ markedly from one another. Thus, vertical jump is best interpreted here as a high-information descriptor of current category structure rather than as a universal marker of sport suitability or future competitive potential.

Cardiorespiratory fitness provided the second major signal. Team and Individual Technical categories showed higher Bleep values, whereas Shooting was lowest. The VO₂max row is a transformed expression of the same Bleep signal, not an independent replication. The Shooting result is specific to this sampled discipline: postural control, visual function, fine-motor stability, and autonomic regulation were not measured, so the present battery cannot characterize the full performance architecture of precision sport (Sadowska et al., 2019).

The cardiorespiratory findings reinforce the same principle. The Bleep Composite produced substantial category differentiation and contributed strongly to the RF model, yet its interpretation depends on the sporting context represented in each group. Higher scores among Team and Individual Technical athletes are compatible with the locomotor and repeated-effort demands represented in several of those disciplines. In contrast, the lower values in Shooting should not be read as evidence of globally inferior fitness. Instead, they indicate lower expression of one measured

capacity within a battery that does not directly quantify several performance qualities central to shooting. The deterministic relation between the Bleep Composite and estimated VO_2max also illustrates why statistical duplication must be avoided: reporting both units can aid interpretation, but they constitute a single underlying cardiorespiratory signal rather than independent evidence.

RF numerically exceeded LDA by 4.1 percentage points in accuracy, but no paired inferential comparison was performed, and the difference does not establish practical superiority. The fully observed-feature sensitivity model retained useful discrimination (76.5% accuracy; macro-AUC=0.880), suggesting that performance was not wholly dependent on structurally missing predictors. Nevertheless, category-linked missingness, demographic imbalance, and the absence of external validation remain fundamental threats to generalization. Contemporary sport-ML literature similarly emphasizes that internally accurate models require careful validation before operational use (Sneha et al., 2024; de Haan et al., 2025; Cattle et al., 2026).

The multivariate and machine-learning results further qualify the apparent category separation. ARI and NMI showed only partial correspondence between the four-cluster solution and the four administrative labels, demonstrating that athletes with similar fitness profiles do not necessarily belong to the same category. This finding is consistent with the heterogeneous composition of Combat and Individual Technical Sports and argues against assuming that administrative classification is equivalent to physiological classification. In practical terms, the cluster structure suggests that some athletes share cross-category fitness phenotypes, while others occupy less typical positions relative to their assigned group. Such overlap is expected in a multidimensional performance system and supports the use of profile-based interpretation at the individual level rather than reliance on category averages alone.

The pooled sex comparisons were directionally consistent with established strength, power, and aerobic differences, but they are not isolated sex effects because category composition differs sharply by sex. The selected adjusted models reduce but do not remove this concern. Likewise, small category effects for leg strength and non-significant flexibility do not imply that these qualities are unimportant for performance; they indicate only limited discrimination among the four administrative categories in this dataset (Konrad et al., 2021).

Demographic structure is therefore not a peripheral limitation but part of the explanation for the observed patterns. The extreme male predominance in Team Sports, the younger composition of Individual Technical Sports, and the older Shooting group mean that category differences partly embed sex- and age-related variation. The adjusted analyses provide some evidence that selected fitness differences persist after accounting for age and category, but sparse cells prevent a comprehensive interaction model and leave residual confounding. This is particularly relevant for machine learning because an algorithm can exploit any reproducible information that helps separate labels, regardless of whether that information represents sport-specific adaptation, demographic composition, or protocol design. Interpretation of model performance must therefore remain linked to the structure of the sample from which the model was learned.

Practically, the profiles are best used as preliminary descriptive information for KONI rather than normative cut-offs. The data argue against treating broad administrative categories as physiologically homogeneous and against using the classifier for sport assignment or talent identification. A common prospective battery, larger discipline-level samples, and external cohorts are needed before model-assisted athlete evaluation can be considered. For applied sport-science practice, the most defensible implication is to use the present profiles as a diagnostic map of the assessment system rather than as fixed standards. Domains that show strong separation can help practitioners identify which capacities currently differentiate groups and can guide more focused follow-up assessment. In contrast, domains with broad overlap may still be highly important within individual sports even if they contribute little to cross-category discrimination. A future monitoring framework would benefit from a common core battery administered to all athletes, supplemented by discipline-specific modules where physiological or technical requirements justify additional tests. This would preserve comparability for system-level analysis while avoiding the assumption that one identical battery can fully capture performance demands across every sport.

The sensitivity analysis also has practical methodological significance. The modest reduction in RF performance when analysis was restricted to fully observed predictors indicates that useful discriminative information remained even after category-dependent test availability was removed. At the same time, the decrease suggests that structurally unavailable variables contributed some

information to the primary model. Rather than treating this as evidence for or against imputation in absolute terms, it highlights the importance of reporting sensitivity analyses whenever measurement availability is tied to the prediction target. For institutional datasets assembled for operational assessment rather than originally designed for machine learning, this form of transparency is essential to distinguish genuine multivariate signal from artifacts of the data-collection process.

More broadly, the study demonstrates why statistical significance, clustering, and predictive performance should be interpreted together in multi-sport profiling. Large univariate effects identify variables with substantial mean separation, clustering describes whether those variables combine into naturally recurring athlete profiles, and supervised classification quantifies how well the measured information reproduces known labels under internal validation. None of these approaches independently establishes causality, normative status, or future performance. Their combined use, however, provides a richer description of how fitness information is organized and where the evidence is sufficiently stable to motivate prospective testing. This analytical distinction is important for preventing exploratory machine-learning results from being translated prematurely into selection or talent-identification decisions.

Implications

The findings suggest that multi-domain fitness assessment should be used primarily to support individualized athlete profiling rather than to establish rigid standards for broad sport categories. Although vertical jump and cardiorespiratory fitness emerged as prominent dimensions differentiating the four sport categories, substantial within-category heterogeneity indicates that athletes within the same administrative group may possess markedly different physical profiles. Accordingly, sport organizations and practitioners should consider implementing a standardized core fitness battery across sports to improve comparability, supplemented by discipline-specific assessments that capture unique physiological, technical, and perceptual demands. The machine-learning model may assist practitioners in identifying informative fitness patterns and prioritizing follow-up assessment, but it should not yet be used for sport assignment or talent identification because the current findings are based on already-trained athletes, are affected by demographic and measurement-related confounding, and lack external validation.

Research Contributions

This study contributes to the sport-science literature by providing a multi-sport analytical framework that integrates conventional between-group comparisons, hierarchical clustering, and supervised machine learning to examine physical-fitness structure across 13 competitive sports. Rather than relying solely on statistical significance or classification accuracy, the study triangulates effect sizes, multivariate profile structure, and predictive discrimination to show that administrative sport categories capture meaningful but incomplete information about athletes' multidimensional fitness characteristics. Methodologically, the study also contributes by explicitly addressing category-linked structural missingness through fold-specific preprocessing, out-of-fold cross-validation, and sensitivity analysis using commonly observed predictors, thereby offering a more transparent approach for analysing heterogeneous institutional sport datasets. Collectively, these contributions provide both empirical evidence from an underrepresented multi-sport setting and a reproducible analytical approach that can be further tested through longitudinal designs and external validation.

Limitations

Several limitations constrain interpretation. The cross-sectional design cannot separate athlete self-selection from training adaptation; the age range (8–46 years), unequal sex distribution, and likely maturation differences create substantial confounding; and the nine-variable ML model includes predictors with category-linked structural missingness. Fold-specific imputation prevents direct preprocessing leakage but cannot remove that design feature. Estimated VO_2max is regression-derived rather than laboratory-measured, site-specific test-retest reliability was not established, and no independent external cohort was available. Sport-level samples were also small for several disciplines, and the Precision category consisted only of Shooting ($n=19$). These limitations preclude normative, deployment-readiness, and prospective talent-identification claims.

Suggestions

Future studies should use repeated assessments and external cohorts, maximize common test availability across sports, recruit sufficient sex- and maturity-stratified samples, and test whether fitness profiles predict sport-specific performance rather than current category membership alone.

CONCLUSION

Athletes from 13 sports showed distinct but demographically confounded multi-domain fitness patterns. Vertical jump had the largest category effect ($\eta^2=0.438$), and the Bleep Composite/estimated VO_2max signal also showed a large effect ($\eta^2=0.217$). The final nine-variable RF achieved 79.7% out-of-fold accuracy and macro-AUC=0.907, numerically above LDA (75.6%; 0.815), with vertical jump, Bleep Composite, and push-up endurance as the three largest Gini importances. These findings demonstrate within-dataset discriminability only. External validation, harmonized measurement availability, and demographically stratified prospective samples are required before operational athlete-evaluation use.

ACKNOWLEDGMENT

The authors thank KONI for facilitating data access and the participating athletes and assessors for their cooperation.

FUNDING INFORMATION

This research received no external funding.

AUTHOR CONTRIBUTION STATEMENT

S Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Project administration. Sy & MWH Methodology, Investigation, Validation, Writing – review & editing. MI & MRK Supervision, Resources, Writing – review & editing. All authors read and approved the final manuscript.

AI DISCLOSURE STATEMENT

AI-assisted writing tools were used for language refinement and structural suggestions. All outputs were reviewed and edited by the authors, who retain full responsibility for the scientific content; AI was not used as an autonomous data analyst or author.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest. The funder statement is not applicable because this research received no external funding.

ETHICAL APPROVAL

De-identified data and the analysis code are available from the corresponding author upon reasonable request, subject to KONI institutional approval, the conditions of the ethics approval (Number: 045/UNM/OR/12/2022), and applicable data-governance requirements.

REFERENCES

- Bishop, C., Turner, A., Jarvis, P., Chavda, S., & Read, P. (2017). Considerations for selecting field-based strength and power fitness tests to measure asymmetries. *Journal of Strength and Conditioning Research*, 31(9), 2635–2644. <https://doi.org/10.1519/JSC.0000000000002023>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cattle, A., Johnston, K., McAuley, A. B. T., Kelly, A., & Baker, J. (2026). Machine learning applications in sport: A scoping review. *Frontiers in Psychology*, 17, 1802549. <https://doi.org/10.3389/fpsyg.2026.1802549>
- Chaabene, H., Negra, Y., Bouguezzi, R., Capranica, L., Franchini, E., Prieske, O., Hbacha, H., & Granacher, U. (2018). Tests for the assessment of sport-specific performance in Olympic combat sports: A systematic review with practical recommendations. *Frontiers in Physiology*, 9, 386. <https://doi.org/10.3389/fphys.2018.00386>

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Lawrence Erlbaum. <https://doi.org/10.4324/9780203771587>
- Currell, K., & Jeukendrup, A. E. (2008). Validity, reliability and sensitivity of measures of sporting performance. *Sports Medicine*, 38(4), 297–316. <https://doi.org/10.2165/00007256-200838040-00003>
- de Haan, M., van der Zwaard, S., Sanders, J., Beek, P. J., & Jaspers, R. T. (2025). Beyond playing positions: Categorizing soccer players based on match-specific running performance using machine learning. *Journal of Sports Science and Medicine*, 24(3), 565–577. <https://doi.org/10.52082/jssm.2025.565>
- Engel, F. A., Sperlich, B., & Holmberg, H. C. (2021). Reaction time assessment in elite sports: A systematic review. *Sports Medicine*, 51(7), 1471–1483. <https://doi.org/10.1007/s40279-021-01445-0>
- Field, A. (2024). *Discovering statistics using IBM SPSS statistics (6th ed.)*. SAGE Publications. <https://uk.sagepub.com/en-gb/eur/discovering-statistics-using-ibm-spss-statistics/book279152>
- Figueiredo, A. J., Coelho-e-Silva, M. J., Sarmiento, H., & Malina, R. M. (2021). Predictors of performance in youth male football players: A systematic review. *Journal of Sports Sciences*, 39(10), 1095–1113. <https://doi.org/10.1080/02640414.2020.1858602>
- Gil, B. B. T. P., & Andacao, A. (2025). Spirituality and mental training as predictors of athletic resilience. *Journal of Sports and Physical Activity*, 1(2), 108–17. <https://doi.org/10.64268/jospa.v1i2.19>
- Hastie, T., Tibshirani, R., & Friedman, J. (2022). *The elements of statistical learning (12th printing)*. Springer. <https://hastie.su.domains/ElemStatLearn/>
- Konrad, A., Gádányi, A., Müller, E., Reiter, K., & Tilp, M. (2021). The effects of a stretching program in physical education on flexibility and jump performance. *Scandinavian Journal of Medicine & Science in Sports*, 31(10), 1965–1974. <https://doi.org/10.1111/sms.14036>
- Levin, K. A. (2006). Study design III: Cross-sectional studies. *Evidence-Based Dentistry*, 7(1), 24–25. <https://doi.org/10.1038/sj.ebd.6400375>
- Léger, L. A., Mercier, D., Gadoury, C., & Lambert, J. (1988). The multistage 20 metre shuttle run test for aerobic fitness. *Journal of Sports Sciences*, 6(2), 93–101. <https://doi.org/10.1080/02640418808729800>
- Markovic, G. (2007). Does plyometric training improve vertical jump height? A meta-analytical review. *British Journal of Sports Medicine*, 41(6), 349–355. <https://doi.org/10.1136/bjism.2007.033738>
- Mayorga-Vega, D., Aguilar-Soto, P., & Viciano, J. (2015). Criterion-related validity of the 20-m shuttle run test for estimating cardiorespiratory fitness: A meta-analysis. *Journal of Sports Science and Medicine*, 14(3), 536–547. <https://www.jssm.org/jssm-14-536.xml>
- Mayorga-Vega, D., Merino-Marban, R., & Viciano, J. (2014). Criterion-related validity of sit-and-reach tests for estimating hamstring and lumbar extensibility: A meta-analysis. *Journal of Sports Science and Medicine*, 13(1), 1–14. <https://doi.org/10.4100/jhse.2014.91.18>
- Modric, T., Versic, S., Sekulic, D., & Liposek, S. (2019). Analysis of the association between running performance and game performance indicators in professional soccer players. *International Journal of Environmental Research and Public Health*, 16(20), 4032. <https://doi.org/10.3390/ijerph16204032>
- Mononen, K., Kontinen, N., Viitasalo, J., & Era, P. (2007). Relationships between postural balance, rifle stability and shooting accuracy among novice rifle shooters. *Scandinavian Journal of Medicine & Science in Sports*, 17(2), 180–185. <https://doi.org/10.1111/j.1600-0838.2006.00549.x>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://www.jmlr.org/papers/v12/pedregosa11a.html>
- Phomsoupha, M., & Laffaye, G. (2015). The science of badminton: Game characteristics, anthropometry, physiology, visual fitness and biomechanics. *Sports Medicine*, 45(4), 473–495. <https://doi.org/10.1007/s40279-014-0287-2>

- Ramsbottom, R., Brewer, J., & Williams, C. (1988). A progressive shuttle run test to estimate maximal oxygen uptake. *British Journal of Sports Medicine*, 22(4), 141–144. <https://doi.org/10.1136/bjsm.22.4.141>
- Republic of Indonesia. (2022). Law of the Republic of Indonesia Number 11 of 2022 on Sports (Undang-Undang Nomor 11 Tahun 2022 tentang Keolahragaan). <https://peraturan.bpk.go.id/Home/Details/203148/uu-no-11-tahun-2022>
- Sadowska, D., Krzepota, J., & Klusiewicz, A. (2019). Postural balance and rifle stability in a standing shooting position after specific physical effort in biathletes. *Journal of Sports Sciences*, 37(16), 1892–1898. <https://doi.org/10.1080/02640414.2019.1603136>
- Sandbakk, Ø., Solli, G. S., & Holmberg, H. C. (2018). Sex differences in world-record performance. *International Journal of Sports Physiology and Performance*, 13(10), 1325–1335. <https://doi.org/10.1123/ijssp.2018-0102>
- Santos, R., Coelho-e-Silva, M. J., Vaz, V., Figueiredo, A. J., & Castanheira, J. (2022). Fitness assessment in youth athletes: Muscular endurance reliability. *International Journal of Environmental Research and Public Health*, 19(6), 3284. <https://doi.org/10.3390/ijerph19063284>
- Sekulic, D., Pojskic, H., Uljevic, O., Pehar, M., Spasic, M., & Sattler, T. (2017). Side-step cutting technique and athletic performance. *Journal of Functional Morphology and Kinesiology*, 2(2), 11. <https://doi.org/10.3390/jfmk2020011>
- Sneha, S., Prithvi, B. S., Niranjanamurthy, M., Kiran Kumar, H. K., & Dayananda, P. (2024). Machine learning based assessment of elite football players based on anthropometric and motor fitness parameters with regard to their playing positions. *SN Computer Science*, 5, 974. <https://doi.org/10.1007/s42979-024-03261-x>
- Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 25. <https://doi.org/10.1186/1471-2105-8-25>
- Suchomel, T. J., Nimphius, S., & Stone, M. H. (2016). The importance of muscular strength in athletic performance. *Sports Medicine*, 46(10), 1419–1449. <https://doi.org/10.1007/s40279-016-0486-0>
- Sukamti, E. R., Saputra, J., & Sugiharto. (2020). Study of sprint and vertical jump for Indonesian athletes: Multi-sectorial testing. *Utopía y Praxis Latinoamericana*, 25(Extra 10), 419–428. <https://www.redalyc.org/journal/279/27964799040/html/>
- Syahputra, R., Oktarifaldi, Putri, L. P., & Bakhtiar, S. (2022). Physical performance characteristic of West Sumatera athletes in combat and game sports. *International Journal of Human Movement and Sports Sciences*, 10(4), 732–747. <https://doi.org/10.13189/saj.2022.100414>
- van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). <https://stefvanbuuren.name/fimd/>
- von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement. *Annals of Internal Medicine*, 147(8), 573–577. <https://doi.org/10.7326/0003-4819-147-8-200710160-00010>
- Wang, X., Soh, K. G., Samsudin, S., Deng, N., Liu, X., Zhao, Y., & Akbar, S. (2023). Effects of high-intensity functional training on physical fitness and sport-specific performance among the athletes: A systematic review with meta-analysis. *PLOS ONE*, 18(12), e0295531. <https://doi.org/10.1371/journal.pone.0295531>
- World Medical Association. (2013). WMA Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, 310(20), 2191–2194. <https://doi.org/10.1001/jama.2013.281053>