International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

Determining Essential Fitness and Motor Skill Parameters for Talent Identification in 12-Year-Old Girls: A Comparison of Machine Learning-Based Feature Extraction Techniques

Rabiu Muazu Musa[1], Siti Musliha Mat-Rasid*[2], Jeffrey Fook Lee Low[3], Gunathevan Elumalai[4], Mohd. Izwan Shahril[5], Mohamad Azri Ismail Ahmad[6], Norlaila Azura Kosni[7]

Submitted: 28/01/2024 **Revised**: 06/03/2024 **Accepted**: 14/03/2024

Abstract: A successful talent identification program relies on the assessment and mapping of relevant parameters for the evaluation of an athlete's ability. Physical fitness and motor-skill parameters are performance variables that contribute to success in a variety of sports. Although these parameters are prerequisites for any sporting event, their requirements may differ depending on the athletes' level and age. The present investigation endeavors to identify essential physical fitness and motor skill-related parameters for talent identification in 12-year-old girls. A total of 236 girls were recruited from different schools in Malaysia and completed a standard test for physical fitness and motor skills that constituted of 30-meter run, step test, 1-minute curl up, hand grip, agility T-test, stork balance stand test, and standing broad jump test. Two different feature extraction techniques viz. Symmetrical Uncert Attribute Evaluation (SymmU) and Recursive Feature Elimination (RFE) were employed to determine the important parameters that could be considered for talent identification in 12-year-old girls. Both the SymmU and RFE demonstrated four sets of parameters worthy of consideration. There was agreement in the selection of three parameters across the two techniques (sit and reach, agility, and stork balance stance test). However, SymmU revealed the standing broad jump as the fourth essential parameter, while RFE indicated a 1-minute curl up. Overall, these tests are shown to be essential for consideration during talent identification programs for 12-year-old girls. The identified parameters provide actionable knowledge for coaches and sports organizations, fostering the development of the next generation of elite female athletes.

Keywords: *Fitness, Machine Learning, Motor Skill, Talent Identification*

1. Introduction

Talent identification (TId) in sports is a multifaceted process crucial for the effective development of athletes and the optimization of sports performance [1]. The ability to accurately identify and nurture athletic potential is fundamental for sports organizations aiming to cultivate a pool of elite athletes [2]. The significance of talent identification becomes more pronounced when considering the dynamic nature of physical fitness and motor skill parameters that contribute to success in various sports [3][4]. These parameters, encompassing a wide range of abilities from agility to muscular strength, serve as critical indicators of an athlete's potential and are vital for tailored training and development programs [5][6].

[1] *Centre for Fundamental and Continuing Education, Department of Credited Co-Curriculum, Universiti Malaysia Terengganu, Malaysia*
ORCID ID : 0000-0001-5332-1770
[2] *Faculty of Sports Science and Coaching, Sultan Idris Education University, Tanjong Malim, Perak, Malaysia*
ORCID ID : 0000-0003-3650-7391
[3] *Faculty of Sports Science and Coaching, Sultan Idris Education University, Tanjong Malim, Perak, Malaysia*
ORCID ID : 0000-0003-1573-5029
[4] *Faculty of Sports Science and Coaching, Sultan Idris Education University, Tanjong Malim, Perak, Malaysia*
ORCID ID : 0000-0002-7882-0741
[5] *Faculty of Sports Science and Coaching, Sultan Idris Education University, Tanjong Malim, Perak, Malaysia*
ORCID ID : 0000-0002-0910-5795
[6] *Faculty of Sports Science and Coaching, Sultan Idris Education University, Tanjong Malim, Perak, Malaysia*
ORCID ID : 0000-0002-7665-7637
[7] *Faculty of Sport Science & Recreation, Universiti Teknologi MARA, Pahang Branch, Pahang, Malaysia*
ORCID ID : 0000-0002-1729-8398
* *Corresponding Author Email: sitimusliha@fsskj.upsi.edu.my*

Moreover, age-specific considerations in talent identification cannot be overstated. Researchers emphasized the importance of recognizing developmental differences among athletes of various ages [6][7]. In the context of youth sports, understanding the unique physical and physiological characteristics of 12-year-old girls is essential for crafting targeted talent identification strategies. Not only does this aid in recognizing potential future elite athletes, but it also ensures that the training regimens are developmentally appropriate and align with the specific needs of this age group [3]. Identifying talent in sports requires one to have a scientific approach as talent can be challenging, biased, and difficult to predict, the application of a scientific approach has broadly been discussed to replace the traditional methods in TId.

Parallel to the development of technology, the assessment in TId is also advancing and developing. Application of advanced statistical analysis has been used broadly in many disciplines in TId because of the common interest in understanding the individual characteristics that lead to long-term success [8][9][10][11][12]. Such approaches tend to replace subjective assessments with objective approaches by quantifying performance into number scores. Indeed, as the advancement of knowledge across multiple disciplines allows for a greater depth of understanding about TId, applying machine learning has become one of the most popular techniques in sports-related analyses. There are numerous research papers in which machine learning is applied to sports [13][14][15][16]. They emphasized that the application of machine learning in sports is not limited to topics of sports performance, athlete talent identification, or technical analysis of games. Sports organizations' economic operations, both on- and off-field, are quickly transitioning to a data-driven culture.

The primary objective of this study was to identify the essential parameters for talent identification in 12-year-old girls, recognizing the distinctive demands and characteristics of this age group using machine learning. The study also seeks to contribute to the ongoing discourse on effective talent identification methodologies by comparing two distinct feature extraction techniques: Symmetrical Uncert Attribute Evaluation (SymmU) and Recursive Feature Elimination (RFE).

2. Methodology

2.1. Participants

The study engaged a diverse cohort of 236 girls aged 12 years recruited from various schools across Malaysia. The inclusion criteria ensured the representative sample of 12-year-old girls actively participated in physical education and extracurricular sports activities. Before participation, informed consent was obtained from both the parents and participants, ensuring compliance with ethical standards and guidelines.

2.2. Procedures

The assessment encompassed a battery of standardized tests that were carefully chosen to evaluate a comprehensive spectrum of physical fitness and motor skill parameters. These assessments were meticulously designed to align with specific developmental characteristics of 12-year-old girls.

2.2.1. 30-Meter Run

This test gauged the participants' speed and agility, providing insights into their anaerobic capacity and lower limb strength.

2.2.2. Step Test

A reliable measure of cardiovascular endurance, the step test required participants to perform a set number of steps within a specified time frame.

2.2.3. 1-Minute Curl-Up

Focused on core strength and muscular endurance, this test involved performing as many curl-ups as possible in a minute, providing valuable information on abdominal muscle function.

2.2.4. Hand Grip

A fundamental measure of upper body strength, the hand grip test involved using a dynamometer, providing a quantifiable metric for participants' grip strength.

2.2.5. Agility T-Test

This test evaluated the participants' agility and quickness by measuring the time taken to complete a T-shaped course.

2.2.6. Stork Balance Stand Test

Assessing the participants' static balance, this test requires maintaining balance on one leg for a specified duration.

2.2.7. Standing Broad Jump

Measuring explosive lower body power, the standing broad jump provided insights into the participants' lower limb strength and coordination.

2.2.8. Sit and Reach

Sit and reach test could be a sort of adaptability test that assesses the extend of movement. This sit and reach test measures the adaptability of low back muscles and hamstring muscles.

2.3. Data Analysis

To identify the essential parameters for talent identification, two distinct feature extraction techniques were employed: SymmU and RFE.

2.3.1. Symmetrical Uncert Attribute Evaluation (SymmU)

This technique evaluates the relevance and importance of each feature concerning the target variable, providing a symmetrical assessment of attribute relevance [18].

2.3.2. (SymmU) Recursive Feature Elimination (RFE)

RFE systematically removes the least significant features, iteratively refining the set of parameters to those that were most critical for talent identification [19]. The data analysis process involved rigorous statistical examinations, considering both the overall performance observed during the battery of tests and individual characteristics identified using the SymmU and RFE. The final set of parameters considered essential for talent identification emerged from the convergence of results from both techniques.

3. Results

Table 1 shows the descriptive statistics of fitness tests for

12-year-old girls involved in this study.

Table 1. Descriptive of fitness data for girls aged 12 years.

Fitness Test	Mean	SD	Min	Max
30 Meter Run (s)	6.20	0.88	4.00	8.26
Step Test (60 s pulse count)	156	22	99	204
1-minute Curl Up (count)	22	6	5	38
Handgrip Test (kg)	20.1	4.0	9.0	34.7
Agility T-Test (s)	16.55	2.35	12.10	23.41
Stork Balance Stand Test (s)	4.24	1.97	1.11	10.57
Standing Broad Jump (cm)	113.5	22.4	62.0	177.0
Sit and Reach (cm)	25.8	5.9	10.5	45.0

SD: Standard Deviation

Figure 1 demonstrates the features/parameters selected using the RFE technique. It can be observed from the figure that four physical fitness and motor skill-related parameters, namely the 1-minute curl, sit and reach, agility, and stork balance tests, are indispensable in the consideration of talent identification among 12-year-old girls. The figure further depicts the cross-validation accuracy of the selected parameters with a total of 29%, revealing that the selected parameters could explain up to 29% of the total variability within the datasets.

Figure 2 reveals the feature/parameters selected using the SymmU technique. It can be detected from the figure that four physical fitness and motor skill-related parameters, including the sit and reach, agility, stork balance tests, and standing broad jump, are highlighted as essential when considering talent identification among 12-year-old girls. Moreover, the figure displays the scores of the selected parameters, with each parameter having a score greater than 0.03 revealing that the selected parameters are worthy of consideration during the talent identification-related tasks.

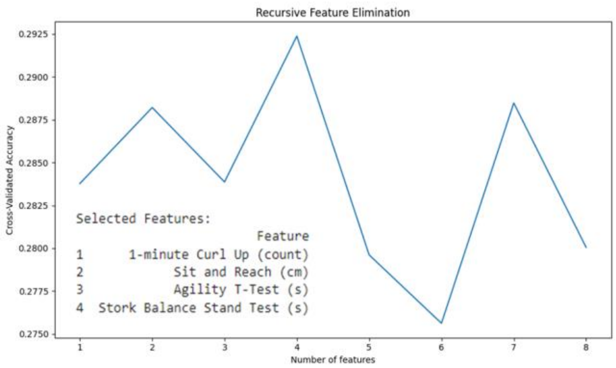

Fig. 1. Recursive Feature Elimination scores for the essential parameters identified.

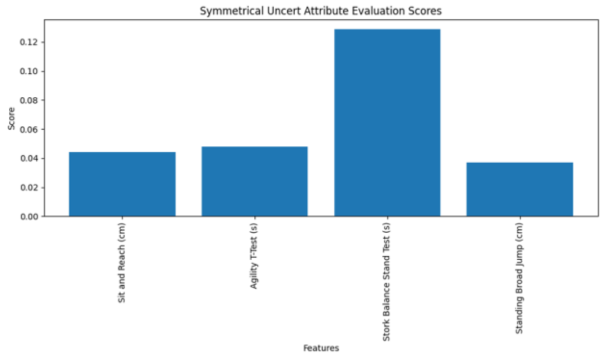

Fig. 2. Recursive Feature Elimination scores for the essential parameters identified.

4. Discussion

The identification of essential parameters for talent identification in 12-year-old girls holds significant implications for both sports science and talent development programs. The consistent selection of three parameters sit and reach, agility, and stork balance stand tests across both the SymmU and RFE reinforces their importance in assessing the potential of young female athletes. These findings align with the broader literature emphasizing the role of flexibility, agility, and balance in predicting athletic success, particularly in youth populations [5,7]. The discrepancy in the fourth identified parameter between the SymmU (standing broad jump) and RFE (1-minute curl-up) introduces an interesting dimension to the discussion. The standing broad jump, reflecting lower limb power and coordination, may be indicative of explosive strength that is crucial in various sports. On the other hand, the 1-minute curl-up, focusing on core strength and muscular endurance, is equally pivotal. This disparity highlights the multifaceted nature of talent identification and suggests that a combination of explosive power and muscular endurance should be considered when evaluating the potential of 12-year-old female athletes. This is in line with previous investigations where these parameters are shown to influence the performance of young athletes [20,21].

The application of SymmU and RFE provided a comprehensive analysis of the importance of individual parameters in talent identification in the current investigation. SymmU, with its symmetrical attribute evaluation approach, and RFE, with its recursive elimination strategy, demonstrated a remarkable degree of agreement in selecting the three primary parameters. This concordance enhanced the robustness of the identified parameters and underscored their significance for talent identification process. However, the divergence in the fourth parameter underscores the rigorous nature of the feature extraction techniques. SymmU, by emphasizing the standing broad jump, accentuates the importance of explosive lower-body power, while RFE, with its emphasis on the 1-minute curl-up, underscores the relevance of core strength and endurance. Integrating the insights from both techniques can provide a more comprehensive framework for talent identification, acknowledging the multifaceted nature of athletic potential.

It is worth reiterating that while this study sheds light on essential parameters for talent identification in 12-year-old girls, it also opens avenues for future research. Investigating the long-term predictive validity of these parameters, especially for actual athletic performance and specialization, could provide valuable insights. Additionally, refining and expanding feature extraction techniques and exploring other innovative approaches can further enhance the precision of talent identification methodologies.

5. Conclusion

The findings from the present study show that certain physical fitness- and motor skill-related parameters that constitute the sit and reach, agility, and stork balance stand tests and additional insights from the standing broad jump and 1-minute curl-up tests could offer tangible insights for coaches, sports organizations, and talent development experts. Tailoring training regimens to enhance these specific attributes in 12-year-old female athletes could prove instrumental in optimizing their athletic potential. Furthermore, these parameters can be incorporated into talent identification protocols, providing a systematic and evidence-based approach to identifying future elite athletes. Coaches and sports organizations should consider the individual nuances of these parameters and recognize that a holistic approach to talent identification is essential. Acknowledging developmental differences among 12-year-old girls and tailoring interventions accordingly can foster a supportive and effective environment for talent development.

Acknowledgments

The authors thank the Ministry of Higher Education for providing the Fundamental Research Grant Scheme (FRGS/1/2020/SS02/UPSI/02/1) and Sultan Idris Education University for managing this grant. The authors would like to thank the teachers and students involved in this study for their support.

Author contributions

Rabiu Muazu Musa: Manuscript preparation and revising article, Final approval of the version to be published **Siti Musliha Mat-Rasid:** Study design, Data collection, Statistical analysis; Manuscript preparation and revising article **Jeffrey Fook Lee Low:** Study design, Data collection, Statistical analysis **Gunathevan Elumalai:** Study design, Data collection, Statistical analysis **Mohd Izwan Shahril:** Data collection, Statistical analysis **Mohamad Azri Ismail Ahmad:** Data collection, Statistical analysis **Norlaila Azura Kosni:** Manuscript preparation and revising article.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] S. M. Mat-Rasid et al., "The Relative Age Effect on Physical Fitness of School Children," *International Journal of Human Movement and Sports Sciences*, vol. 10, no. 5, pp. 1098–1106, Oct. 2022, doi: https://doi.org/10.13189/saj.2022.100528.

[2] A. M. Williams and T. Reilly, "Talent identification and development in soccer," *Journal of sports sciences*, vol. 18, no. 9, pp. 657–67, 2000, doi: https://doi.org/10.1080/02640410050120041.

[3] I. Balyi, R. Way, and C. Higgs, Long-Term Athlete Development. Human Kinetics, pp. 27–46, 2013.

[4] R. Vaeyens, M. Lenoir, A. M. Williams, and R. M. Philippaerts, "Talent Identification and Development Programmes in Sport," *Sports Medicine*, vol. 38, no. 9, pp. 703–714, 2008, doi: https://doi.org/10.2165/00007256-200838090-00001.

[5] T. A. Haugen, E. Tønnessen, and S. K. Seiler, "The Difference Is in the Start: Impact of Timing and Start Procedure on Sprint Running Performance," *Journal of Strength and Conditioning Research*, vol. 26, no. 2, pp. 473–479, Feb. 2012, doi: https://doi.org/10.1519/jsc.0b013e318226030b.

[6] R. M. R., A. N., M. A. B. H. M., M. R. M., K. L. A., and J. H., "Unsupervised Pattern Recognition of Physical Fitness Related Performance Parameters among Terengganu Youth Female Field Hockey Players," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, no. 1, p. 100, Feb. 2017, doi: https://doi.org/10.18517/ijaseit.7.1.1145.

[7] R. M. Malina, "Early Sport Specialization," *Current Sports Medicine Reports*, vol. 9, no. 6, pp. 364–371, Nov. 2010. doi: https://doi.org/10.1249/JSR.0b013e3181fe3166

[8] A. Nadzmi et al., "Development of Motor Performance Index: A Preliminary Study Among 7 Years Old Malaysian Kids," Lecture notes in bioengineering, pp. 143–156, Jan. 2023, doi: https://doi.org/10.1007/978-981-99-0297-2_11.

[9] A. F. M. Rozi et al., "The Development of Malaysian Body Somatotype Model using Comprehensive Multivariate Techniques," International Journal of Recent Technology and Engineering, vol. 8, no. 2S7, pp. 264–267, Sep. 2019, doi:

https://doi.org/10.35940/ijrte.b1053.0782s719.

[10] J. Pion, A. Hohmann, T. Liu, M. Lenoir, and V. Segers, “Predictive models reduce talent development costs in female gymnastics,” *Journal of Sports Sciences*, vol. 35, no. 8, pp. 806–811, Jun. 2016, doi: https://doi.org/10.1080/02640414.2016.1192669.

[11] F. Louzada, A. C. Maiorano, and A. Ara, “iSports: A web-oriented expert system for talent identification in soccer,” *Expert Systems with Applications*, vol. 44, pp. 400–412, Feb. 2016, doi: https://doi.org/10.1016/j.eswa.2015.09.007.

[12] K. Till et al., “Identifying Talent in Youth Sport: A Novel Methodology Using Higher-Dimensional Analysis,” *PLOS ONE*, vol. 11, no. 5, p. e0155047, May 2016, doi: https://doi.org/10.1371/journal.pone.0155047.

[13] G. A. Whitaker, R. Silva, D. Edwards, and I. Kosmidis, “A Bayesian approach for determining player abilities in football,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 70, no. 1, pp. 174–201, Nov. 2020, doi: https://doi.org/10.1111/rssc.12454.

[14] D. Barron, G. Ball, M. Robins, and C. Sunderland, “Identifying playing talent in professional football using artificial neural networks,” *Journal of Sports Sciences*, vol. 38, no. 11–12, pp. 1211–1220, Jan. 2020, doi: https://doi.org/10.1080/02640414.2019.1708036.

[15] N. Chmait, S. Robertson, H. Westerbeek, R. Eime, C. Sellitto, and M. Reid, “Tennis superstars: The relationship between star status and demand for tickets,” *Sport Management Review*, vol. 23, no. 2, Apr. 2019, doi: https://doi.org/10.1016/j.smr.2019.03.006.

[16] J. G. Claudino, D. de O. Capanema, T. V. de Souza, J. C. Serrão, A. C. Machado Pereira, and G. P. Nassis, “Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a Systematic Review,” Sports Medicine - Open, vol. 5, no. 1, Jul. 2019, doi: https://doi.org/10.1186/s40798-019-0202-3.

[17] R. S. Lloyd et al., “Long-Term Athletic Development, Part 2,” *Journal of Strength and Conditioning Research*, vol. 29, no. 5, pp. 1451–1464, May 2015, doi: https://doi.org/10.1519/01.jsc.0000465424.75389.56.

[18] I. Kononenko, E. Šimec, and M. Robnik-Šikonja, “Overcoming the myopia of inductive learning algorithms with RELIEFF.” *Applied Intelligence*, vol. 7, no. 1, pp. 39–55, 1997, doi: https://doi.org/10.1023/a:1008280620621.

[19] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene Selection for Cancer Classification Using Support Vector Machines,” *Machine Learning*, vol. 46, no. 1/3, pp. 389–422, 2002, doi: https://doi.org/10.1023/a:1012487302797.

[20] Rabiu Muazu Musa, A. Majeed, Norlaila Azura Kosni, and Mohamad Razali Abdullah, Machine Learning in Team Sports. 2020. doi: https://doi.org/10.1007/978-981-15-3219-1.

[21] Mat-Rasid, S. M. et al., “Applied Multidimensional Analysis for Assessing Youth Performance in Sports Talent Identification Program,” *International Journal of Recent Technology and Engineering*, vol. 8, no. 2S7, pp. 207–211, Sep. 2019, doi: https://doi.org/10.35940/ijrte.b1051.0782s719.