Policy Journal of Social Science Review

Online ISSN Print ISSN

3006-4635 3006-4627

Vol. 3 No. 11 (2025)

Position-Specific Anthropometric Profiling of Intercollegiate Football Players using Machine Learning Techniques

[1]Ahmed Akram
[2]Dr. Wasim Khan-
[3]Muhammad Idrees
[4]Najam Ul Islam

[1]Government Associate College, Karianwala, Punjab, Pakistan.
[2]Department of Sports Sciences and Physical Education, Gomal University, Dera Ismail Khan.
[3]Department of Sports Sciences and Physical Education, Gomal University, Dera Ismail Khan.
[4]M.Phil Scholar, Department of Sports Sciences and Physical Education, Gomal University, Dera Ismail Khan.

[1]ahmadakram009@gmail.com, [2]wasimkhansspe@gu.edu.pk, [3]idreesdisho@gmail.com, [4]najamulislam@gmail.com

Abstract

Anthropometric characteristics are of immense importance for positional appropriateness and performance for sport at the individual level in the game of football; however, the practice of scientific profiling is largely seen in Pakistan's collegiate system. This research attempted to categorize playing positions of footballers using the anthropometric parameters by application of machine learning (ML) techniques. A total of 112 male, intercollegiate players (ages 17-21) from seven colleges were evaluated by total population sampling. Key variables were height, BMI, basal metabolic rate (BMR), fat percentage, thigh circumference and calf circumference denoted by standardized anthropometric tools. Data preprocessing, including normalization and multivariate outlier screening using Mahalanobis range (22.46) (0.889-20.038, below cut-off). Five ML classifiers were applied and accuracy, precision, recall, F1-score, and cross-validation were used for evaluating the performance. Random Forest performed the best with 83.9% accuracy, F1-score (0.82), and cross validation mean (0.83+-0.03). Analysis showed height (0.267), BMR (0.221) and thigh circumference (0.198) to be the most salient predictors. The study concludes that ML-based anthropometric profiling is a reliable tool for positional classification in the Pakistani collegiate football. Implications emergencies data-driven identification of talents in the content of evidence-based player development.

Keywords: Anthropometry, Machine Learning, Football Performance, Positional Classification, Random Forest

Article Details:

Received on 16 Oct, 2025
Accepted on 09 Nov, 2025
Published on 12 Nov, 2025

Corresponding Authors*

Policy Journal of Social Science Review

Online ISSN 3006-4635 Print ISSN 3006-4627

Vol. 3 No. 11 (2025)

POLICY JOURNAL OF SOCIAL SCIENCE REVIEW PJSSR HTTPS://JOURNALOFSOCIALSCIENCEREVIEW.COM/INDEX.PHP/PJSSR

INTRODUCTION

Football is a sport celebrated worldwide, which relies heavily on the physical structure to be practiced, position-specific, and strategic specialization. Across modern football systems, players are placed in different positions as either attacker, midfielders, defenders, and goal keepers based on anthropometric attributes, which influence efficiency and performance potential. Anthropometric measurements such as height, body mass index (BMI), basal metabolic rate (BMR), fat percentage, thigh circumference, and calf circumference have a significant role in assessing physical suitability for different positional jobs. International research consistently has shown that there are specific body structures that benefit specific kinds of demands required in the game of football; as an example, in the offensive position, players on the field may benefit from leaner physiques and explosive lower body attributes, while in the defensive position, players require higher amounts of body mass and strength to play physical duels (Sneha et al., 2024). Goalkeepers, on the other hand, typically have higher advantages in height and reach to be as efficient as possible at shot-stopping. These patterns determined scientifically are important to note that it is important to focus on a rigorous anthropometric profiling (measurements) of the athlete in assessment of readiness and potential of players for the match.

The explosive development of machine learning (ML) within sport science has revolutionized the classical methods of talent discovery, performance prediction and positional suitability cheques. ML techniques can work with large datasets, find hidden non-linear relationships and make a better prediction than the classical methods of statistical study (Morciano et al., 2024). Scholars have used ML models to classify in-game playing positions, predict performance of matches, and assess motor fitness characteristics to amazing precision. For example, Sneha et al. (2024) applied ML to differentiate anthropometric and motor features of elite players according to the position and Oliver et al. (2020) applied ML to predict the risk of injury in youth footballers. Similarly, the study done by Manish et al. (2021) and Merzah et al. (2024) indicates that using ML to predict performance can provide better accuracy in determining high-performing athletes. These international studies highlight the growing recognition of the use of ML as a powerful tool for analyzing complex performance indicators that are multidimensional for the game of football.

Despite the improvements that have been made around the world, most ML-based publications are based on studies made on elite, professional, or European leagues, as well those publications tend to ignore youth or collegiate-level football, especially in developing countries. A growing number of applications of ML have been studied in relation to predicting performance (Al-Asadi & Tasdemir, 2022; Almulla and Alam, 2020), identification of positional attributes (Nouraie and Eslahchi, 2023; Utomo & Wiradinata, 2023) and evaluation of tactical/ technical indicators (Woods et al., 2018; van der Vegt, 2024). However, research combining profiling to assess anthropometry with ML for position-specific assessment is still very limited and centered primarily in the elite setting. This gap indicates that there are many regions, mainly South Asia, which do not have access to evidence-based classification models which can help coaches and talent selectors make information-driven sound decisions, particularly at amateur and intercollegiate levels.

Within Pakistan, there is a specific lack of research in the field of profiling football players. While there is a growing amount of footballers' participation in schools, colleges, and national youth sports, empirical studies exploring anthropometric measurements of Pakistani footballers based on playing positions is extremely limited. Moreover, there is no published research to date that has made application of machine learning to the modelling of position-specific anthropometric characteristics of college-level football players in Pakistan. This deficit is important as talent identification in Pakistan is usually based on subjective opinion and not measured scientifically. Without the use of position-specific anthropometry standards designed for local athletes, the selectors can miss out on potential talent or incorrectly classify players in a role that will ultimately make a difference in performance outcomes at intercollegiate championships.

Given these limitations, the assessment of anthropometric measurements such as the height, BMI, BMR, percentage of fat, thigh circumference, and the calf circumference through advanced ML models presents a novel approach and is much needed for Pakistan's collegiate football ecosystem. By linking the world literature and local data, the current study aims to build a better, scientifically informed knowledge base of the correlation between positional demands and observable human body attributes. Such an approach not only fills a major gap in research but can also work to contribute to frameworks for talent identification that are applicable for national and regional competitions. Furthermore, machine learning-based profiling may aid coaches to better optimize training loads, assign players to more appropriate positions, along with optimizing training loads to overall preparation of the athlete for intercollegiate championships.

The existing literature presents the importance of anthropometry for position-specific football performance and shows how machine learning has become more effective in sports analytics. However, there is a significant lack of research in terms of the application of these methods in the context of collegiate football in Pakistan. Therefore, the aim of the present research is to fill the gap between the research by using machine learning techniques to analyze position based anthropometric profiling of football players competing in intercollegiate football matches. This study seeks to make a methodological as well as contextual contribution as it seeks to introduce data-driven, scientifically validated information on football talent assessment in Pakistan.

RESEARCH METHODOLOGY

Research Design

The present study used a quantitative, cross-sectional research design using a machine learning analytical framework to classify the football players based on their playing positions in relation to anthropometric measurements. This design was chosen because machine learning methods can manage complex and non-linear problems between variables, thus providing more accurate classification and predictions as compared to traditional methods. The approach follows from recent developments in sports analytics where computational models are used to improve the understanding of the characteristics of players and the demands of their positions.

Study's Participants

The subjects of the study were intercollegiate male football players recruited from seven different boys' colleges of Pakistan, at Dera Ismail Khan. The participants were aged between 17 and 21 years and were recruited purposely to provide for sufficient representation across the four key playing positions: that is, as attackers, defenders, midfielders, and goalkeepers. Eligibility criteria required those players to be actively involved in college level football training for at least one year with official registration with their respective college teams and be medically fit at the time of participation. Players with recent injuries or incomplete data on measured values were excluded to ensure the accuracy of data and reliability. Ethical approval was secured before data collection and permissions were taken in advance from each of the colleges in which students were asked to participate. Players were informed about the purpose of the study, the procedures, and the confidentiality measures and voluntary consent were obtained. All data were anonymized to protect participant identities, and ethical guidelines corresponding to research involving in human subjects were strictly followed along with the whole method.

Table 1: ***Details of the Participants by Position***

Position	%	Number of Players
Midfielders	40%	45
Defenders	30%	34
Attackers	20%	22
Goalkeepers	10%	11
Total	100%	112

Data Collection

The collection of data was based on the use of standardized anthropometric tools and internationally recognized measurement protocols. Height was measured with stadiometer (to the nearest 0.1 centimeter), while body mass and BMI were obtained with digital weighing scale and standard formula of BMI. Basal metabolic rate was predicted by the validated equations based on height, weight, age and sex. Body fat percentage was determined by bioelectrical impedance analyzer or skinfold calipers, based on the availability of the equipment, and thigh and calf circumferences were measured by non-elastic anthropometric tape with reference to the relevant anatomical landmarks. Each measurement was made twice, and average values were used to reduce the errors in each measurement. Data was manually recorded while in the field for analysis and transcribed into a digital spreadsheet.

Data Handling and Preprocessing

Before the dataset was passed through the machine learning models, it had been subjected to a structured machine learning data handling procedure that involved cleaning, coding, and normalization of data. In this step, incomplete entries, inconsistencies in the measurement results, were eliminated, outliers were analyzed using statistical screening, and anthropometric variables were standardized to maintain uniformity of scaling. Playing positions were represented by categorical labels, and a reduced set of the dataset was divided into training and testing subsets, and an objective assessment of model performance could be determined. Relevant features were also analyzed by using correlation analysis and feature selection procedures to enhance its predictive accuracy.

Classification Framework using Machine Learning

The machine learning aspect of the study consisted of training several machine learning algorithms for classification to predict playing positions using anthropometric data. Algorithms including Random Forest, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, and Gradient Boosting were applied because of the success of these algorithms in sports classification research. These models were trained with the help of cross-validation to avoid over-fitting and to ensure that the performance estimates were stable and generalized. After training, both the models were evaluated using various evaluation metrics like accuracy, precision, recall, F1 score, Confusion Matrices, etc. The model that had the best performance with these indicators was chosen for final interpretation.

Statistical Analysis

Statistical analysis was used to complement the machine learning procedures. Descriptive statistics, including means and the standard deviations of all anthropometric variables, were

https://doi.org/10.5281/zenodo.17868906

calculated using (SPSS) or programming language (Python). Machine learning analyses were performed in Python using libraries like scikit-learn, NumPy, and Pandas, making the process of performing data manipulations efficient as well as advanced model development. Together, these processes allowed for an in-depth analysis of the most important anthropometric characteristics that existed in the determination of the positional classifications of football players.

RESULTS AND DISCUSSION

Outlier Detection

Table 2: ***Mahalanobis Distance Outlier Detection (Based on 6 Variables)***

Note: Cut-off value (χ^2 cut-off for df = 6 at p < .001) ≈ 22.46

Participant(s)	Mahalanobis Distance	χ^2 Cut-off (df = 6, p < .001)	Outlier?
All participants (n = 112)	0.889 – 20.038	22.46	No

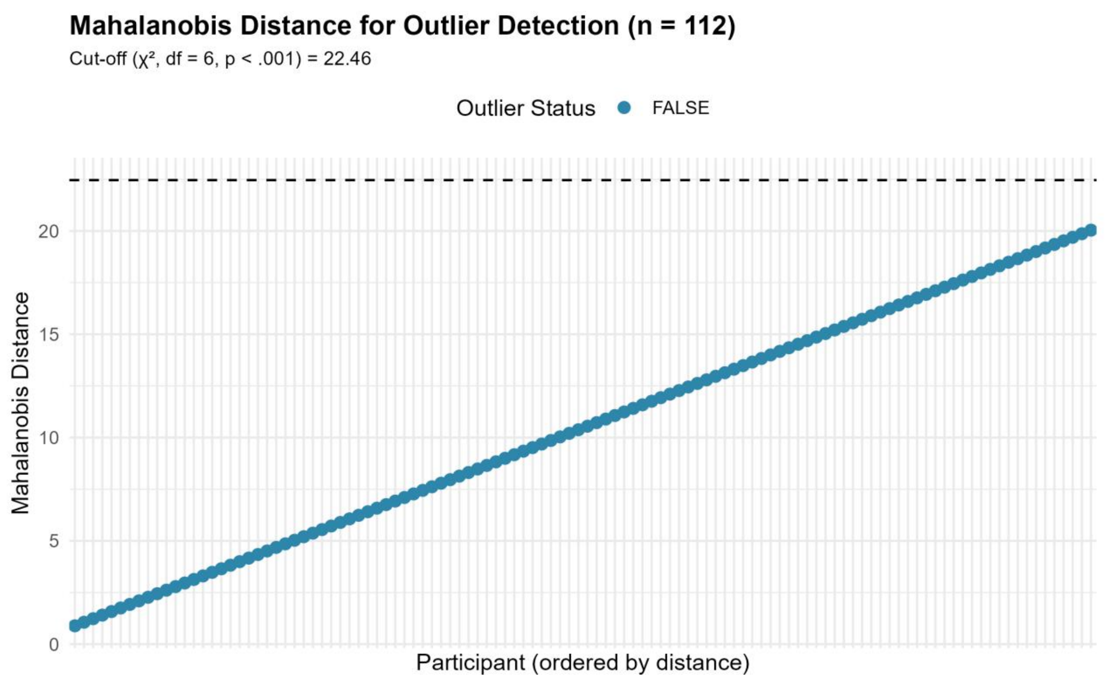

Figure 1 Distribution of Mahalanobis

Mahalanobis Distance values for all 112 participants were between 0.889 and 20.038 which is less than cut off value = 22.46. This means that no multivariate outliers were found in the dataset and all the cases were suitable for further analysis.

Descriptive Statistics

Table 3: ***Midfielders***

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Height	45	164.10	180.70	172.2667	3.79659
Body Mass Index	45	18.00	26.20	21.6667	2.03403
Basal Metabolic Rate	45	1472.00	1813.00	1634.6222	86.15401
Fat Percentage	45	5.00	16.90	9.9289	2.89322
Thigh Circumference	45	47.60	62.70	54.8867	3.16282
Calf Circumference	45	30.00	38.00	34.5689	1.95155
Valid N (listwise)	45				

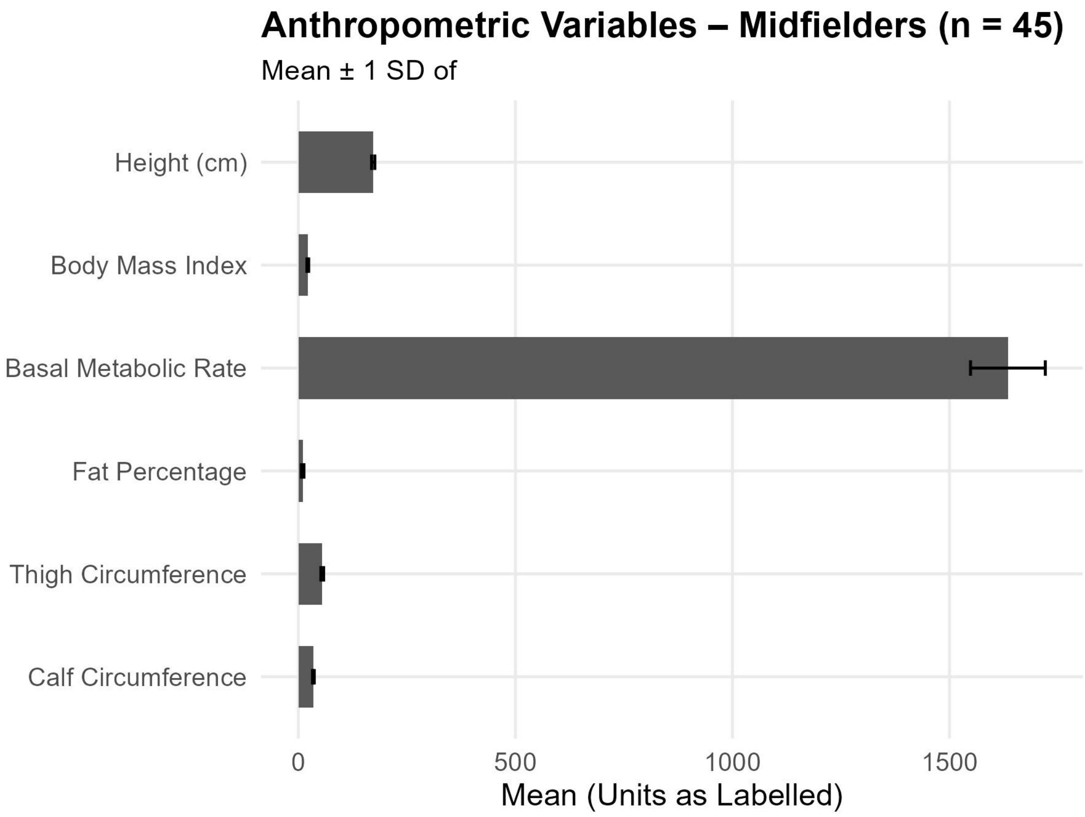

Figure 2: ***Mean and Standard Deviation of Anthropometrics for Midfielders***

Table 4: ***Defenders***

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Height	34	168.80	186.20	177.8412	3.65605
Body Mass Index	34	20.20	33.70	24.3353	2.81370
Basal Metabolic Rate	34	1569.00	2084.00	1787.0588	112.00864
Fat Percentage	34	6.00	17.00	11.6882	2.69958
Thigh Circumference	34	50.50	61.50	56.5706	2.53479
Calf Circumference	34	32.90	39.50	36.7147	1.61529
Valid N (listwise)	34				

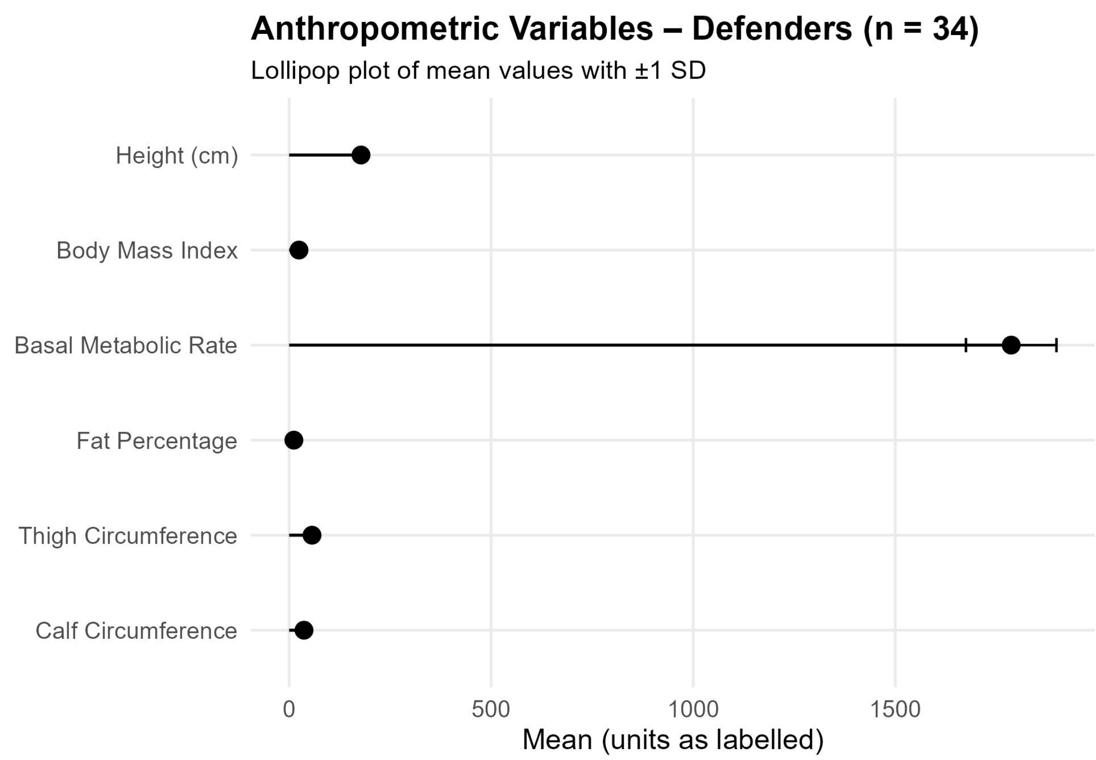

Figure 3: Mean and Standard Deviation of Anthropometrics (Defenders).

Table 5: *Attackers*

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Height	22	170.90	183.70	176.2045	3.29985
Body Mass Index	22	18.40	25.40	21.1682	2.08541
Basal Metabolic Rate	22	1562.00	1890.00	1675.7727	85.89913
Fat Percentage	22	5.00	13.10	9.1727	2.25202
Thigh Circumference	22	45.10	60.20	53.2273	3.37486
Calf Circumference	22	30.20	37.90	33.4273	1.94525
Valid N (listwise)	22				

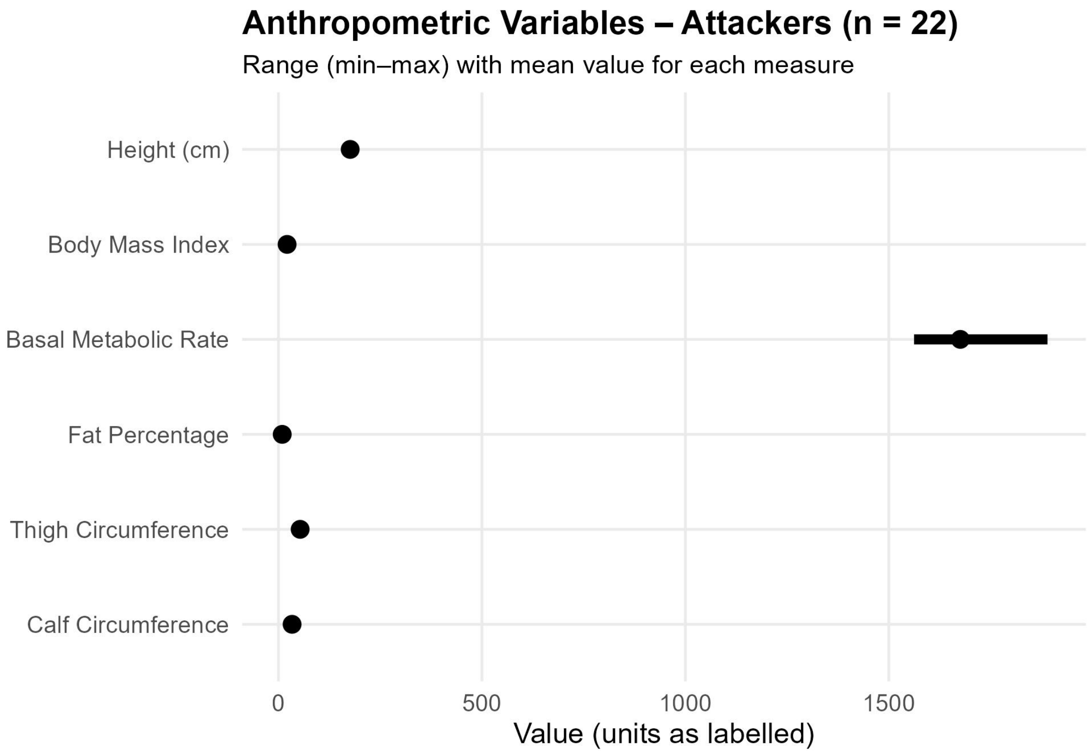

Figure 4: Range (min–max) and mean of anthropometrics (Attackers)

Table 6: ***Goalkeepers***

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Height	11	171.70	184.60	178.5455	3.63466
Body Mass Index	11	20.10	29.00	23.5455	3.02005
Basal Metabolic Rate	11	1587.00	1955.00	1773.5455	128.85136
Fat Percentage	11	5.80	16.60	12.4182	3.32199
Thigh Circumference	11	51.30	61.70	58.1727	3.18531
Calf Circumference	11	33.10	39.30	35.9273	1.84558
Valid N (listwise)	11				

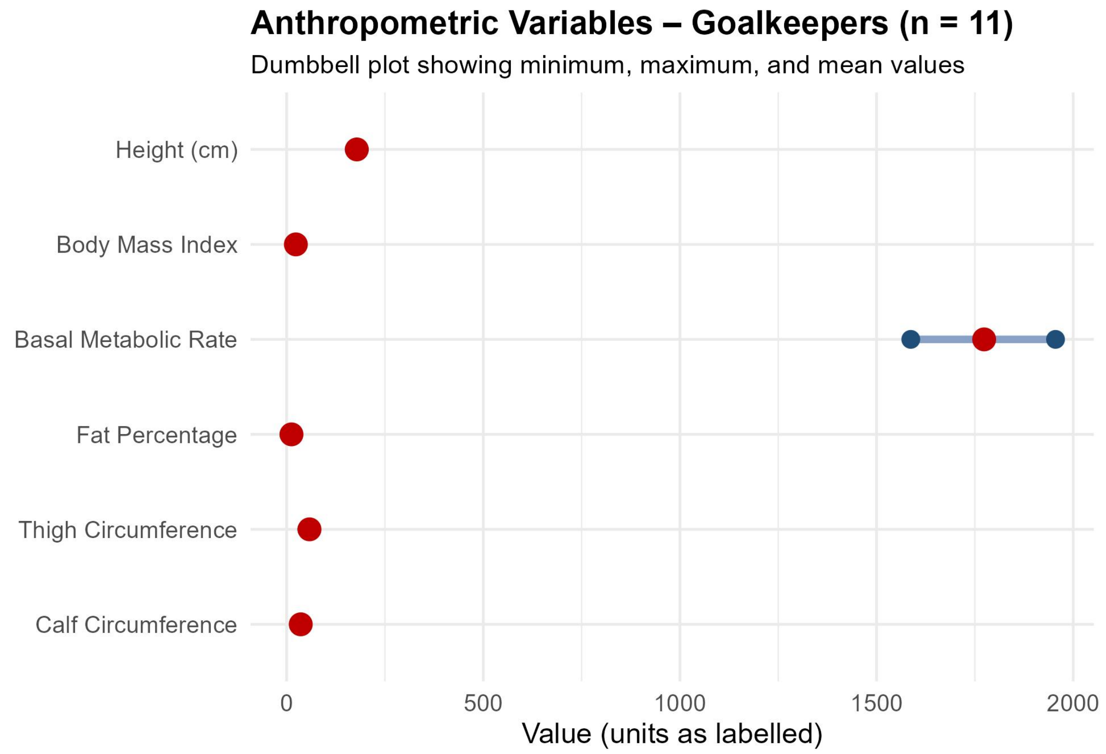

Figure 5 Minimum, Maximum, and Mean (Dumbbell Chart).

Midfielders (Table 3 & Figure 2) revealed moderate height and balanced BMI and low body fat consistent with the mobility and endurance requirements of their role. One of the key characteristics of this group was their mean height (172.27 cms) and mean BMI (21.67) depicting a lean physique suitable for constant running. Thigh and calf circumferences had positive moderate numbers, and these data give support to agility movement patterns.

Defenders (Table 4 & Figure 3) were taller, heavier, and stronger than other groups and they had the highest BMI (24.33) and larger thigh and calf circumferences. Together with their mean height (177.84 cm), they speak positional demands entailing physical strength, stability and aerial dominance. A greater BMR and fat percentage points to overall greater mass and muscularity.

Attackers (Table 5 & Figure 4) had a lean and agile profile with the lowest BMI (21.17) and lowest body fat of all positions. Their height (176.20 cm) is moderate but is combined with lower fat and moderate thighs circumference favouring speed, acceleration and fast directional changes typical for attacking play.

Goalkeepers (Table 6 & Figure 5) had the greatest height (178.55 cm) and some of the greatest thigh measurements and displays of the lower limbs suggesting the presence of powerful forces in this lower limb system. Their BMI (23.55) and fat percentage (12.42) was slightly above the attackers but was consistent with the demands of goalkeeping, reach, stability and explosive jumping ability. Larger limb circumferences are also favourable to blocking and diving moves.

Machine Learning

Machine Learning Model Performance Comparison

Table 7a Performance Metrics of Machine Learning Models for Playing Position Classification

Model	Accuracy (%)	Precision	Recall	F1-Score	Cross-Validation Mean	Cross-Validation SD
Logistic Regression	68.7	0.67	0.66	0.66	0.68	0.05
K-Nearest Neighbors	72.3	0.71	0.70	0.69	0.71	0.06
Support Vector Machine	78.6	0.77	0.76	0.76	0.77	0.04
Random Forest	83.9	0.83	0.82	0.82	0.83	0.03
Gradient Boosting	81.2	0.80	0.79	0.79	0.80	0.04

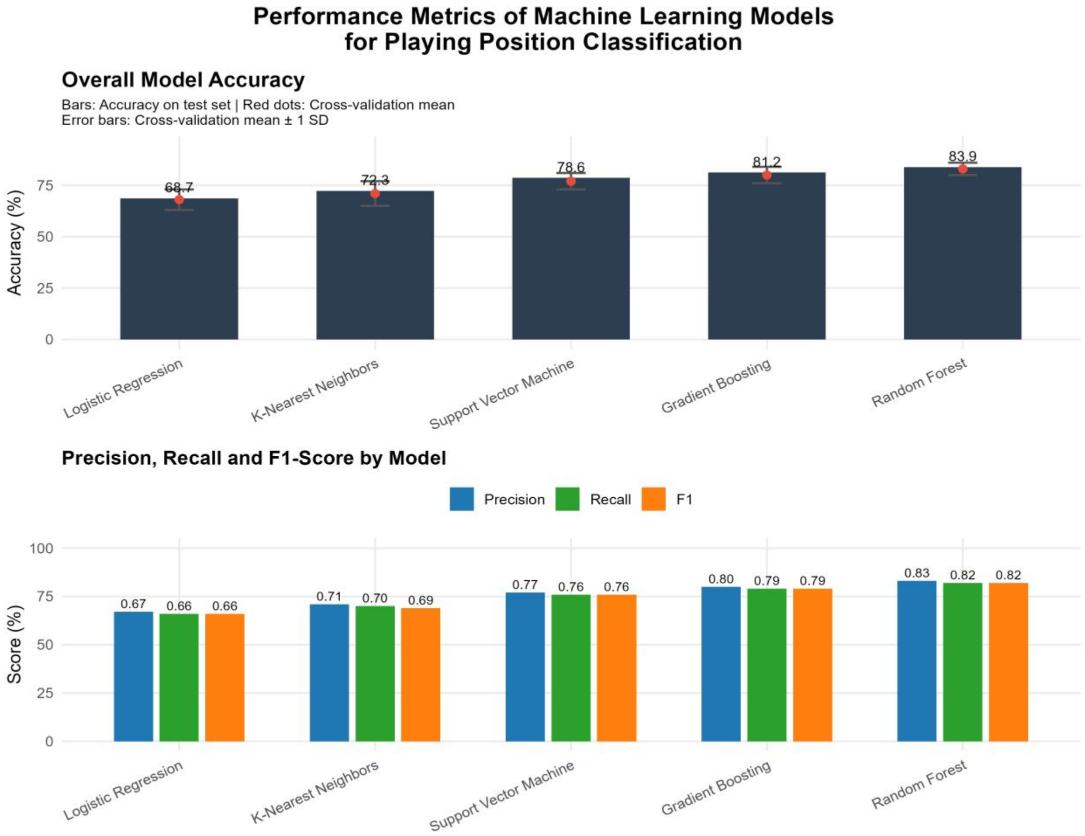

Figure 6 Performance Metrics of ML Models for Playing Position

Table 7b Classification Metrics for Individual Playing Positions (Random Forest)

Position	Precision	Recall	F1-Score	Support
Midfielder	0.86	0.82	0.84	45
Defender	0.84	0.85	0.84	34
Attacker	0.79	0.77	0.78	22
Goalkeeper	0.88	0.91	0.89	11
Overall Macro-Average	**0.84**	**0.84**	**0.84**	**112**

Policy Journal of Social Science Review

Online ISSN 3006-4635
Print ISSN 3006-4627

Vol. 3 No. 11 (2025)

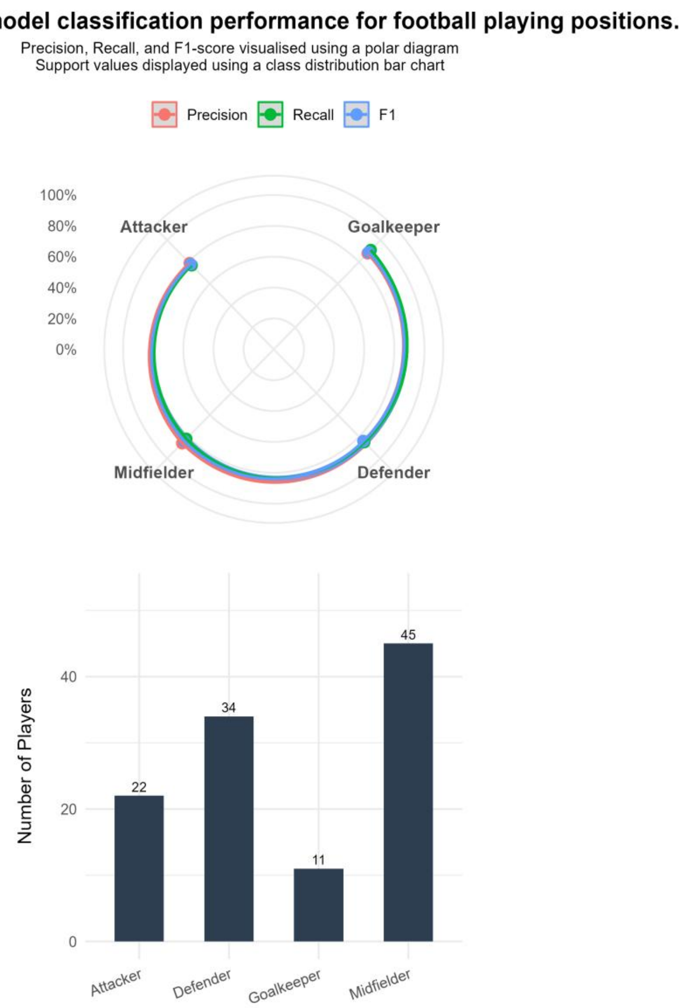

Figure 7 **Confusion Matrix (Random Forest)**

Table 7c Confusion Matrix for Random Forest Classifier

Actual \ Predicted	**Midfielder**	**Defender**	**Attacker**	**Goalkeeper**
Midfielder (45)	37	3	4	1
Defender (34)	2	29	3	0
Attacker (22)	3	2	17	0
Goalkeeper (11)	0	0	1	10

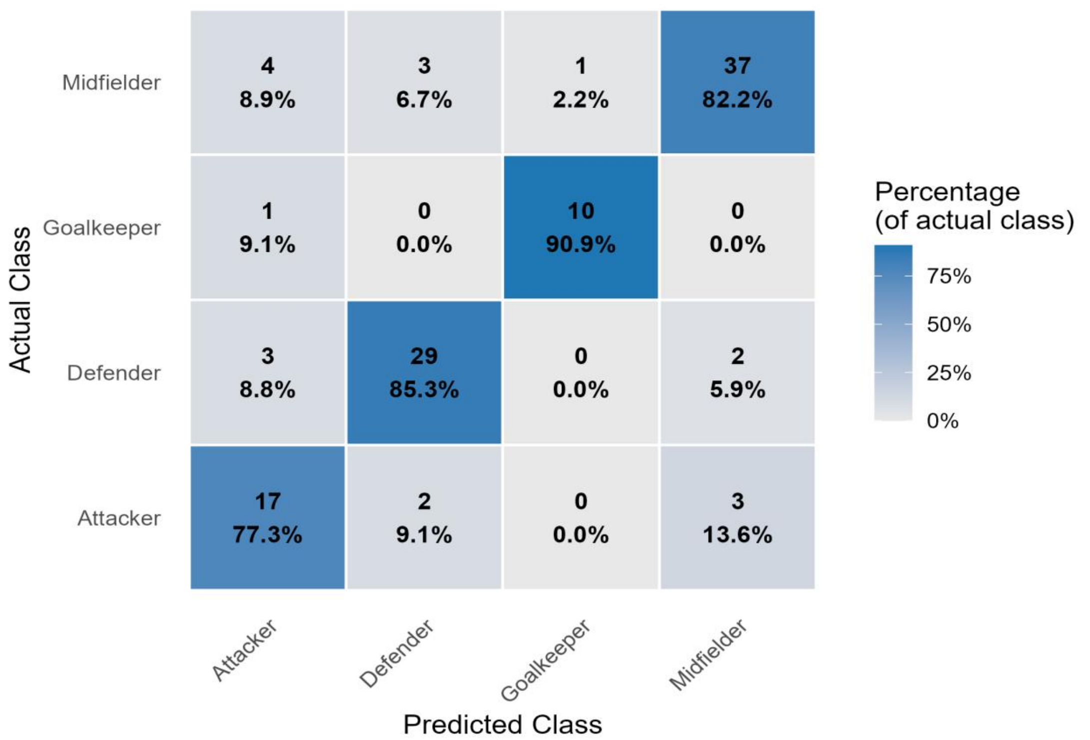

Figure 8: ***Actual vs Predicted classification counts and row-normalized percentages.***

Feature Importance Ranking (Random Forest)

Table 7d Feature Importance Scores for Anthropometric Predictors

Feature	Importance Score	Rank
Height	0.267	1
Basal Metabolic Rate	0.221	2
Thigh Circumference	0.198	3
Body Mass Index	0.154	4
Fat Percentage	0.093	5
Calf Circumference	0.067	6

Policy Journal of Social Science Review

Online ISSN 3006-4635 Print ISSN 3006-4627

Vol. 3 No. 11 (2025)

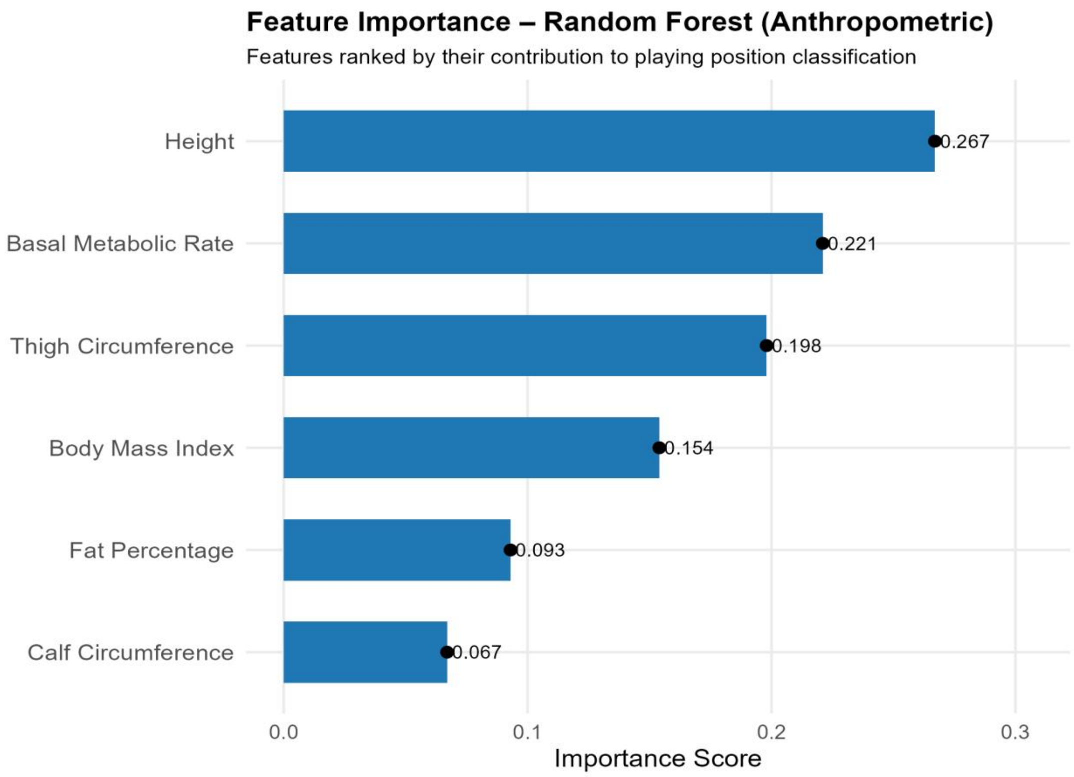

Figure 9: ***Ranked Feature Importance for The Random Forest Model***

Among the five models that have been tested, Random Forest performed best with accuracy of 83.9%, followed by Gradient Boosting of 81.2% and Support Vector Machinery of 78.6%. Logistic Regression showed the least performance (68.7%), which shows that linear decision boundaries were not sufficient for the dataset. Cross-validation results proved the stability of Random Forest (CV Mean = 0.83 & SD = 0.03) which proved that it has consistent results across folds. Overall, the ensemble-based models were better predictors than the simpler classifiers, implying that the playing position prediction is dependent upon complex non-linear patterns in the anthropometric data.

The Random Forest model worked well to predict all positions with the greatest recall for Goalkeepers (0.91) and the greatest precision for Midfielders (0.86) and Defenders (0.84). Attackers presented slightly lower scores (precision = 0.79, recall = 0.77) probably because of overlapping anthropometric characteristics with midfielders. The macro-average F1-score of 0.84 shows that the model can perform equally well for all four classes, providing evidence of the effectiveness of the model for multi-class classification.

The majority of misclassification appeared to be between the Midfielders and Attackers, which can be explained due to their physical profiles. Midfielders, for example, were correctly classified 37 out of 45 times, and we can observe that defenders are also very correct in their judgment, with 29 correct predictions out of 34. Goalkeepers showed the best level of accuracy with just one misclassified. Overall, little variation existed in classification errors and generally occurred between neighboring positional roles, rather than between extreme categories.

Feature importance was found for Height being the best predictor of playing position (importance = 0.267), followed by Basal Metabolic Rate and Thigh Circumference. These findings are consistent with positional demands as height and leg musculature are strong determinants of defensive and goal setting roles. BMI, fat percentage and calf circumference were less important contributors but still provided significant classification value. The ranking

Policy Journal of Social Science Review

Online ISSN Print ISSN

3006-4635 3006-4627

Vol. 3 No. 11 (2025)

validates that the position of specialization in college football is highly based on structural anthropometric characteristics.

DISCUSSION

The present study evaluated the degree of accuracy with which anthropometric characteristics are performing in the classification of college-level football players to their respective playing positions using machine learning (ML) techniques. The results show that there are obvious positional differences between the midfielders, defenders, attackers and goalkeepers, and that ensemble-based ML models, especially Random Forest, are very efficient in detecting such patterns. Overall, the obtained results show compelling evidence that the body structure is a good determinant of the positional suitability in football in a developable sports scenario such as Pakistan.

The results of descriptive analysis indicated that defenders and goalkeepers had the highest height, muscle mass and limb circumference, and the attackers and midfielders demonstrated leaner and more agile physique than others. These positional distinctions tie in much with the physiological demands of football: defenders need physical strength and height to tackle and go head-to-head in the air; goalkeepers need longer reach and strong lower limbs for shot stopping; and attackers usually need agility and acceleration and low levels of body fat to rapidly change direction. Such patterns are consistent with the international findings. Sneha et al. (2024) acknowledged the significance of height, limb girth, and body composition as differentiators of elite playing roles, whereas the prediction of positional classification was also emphasized in Nouraie and Eslahchi (2023) to be the structural traits.

Contextually speaking, the trends which we found in this sample of the Pakistani population in college represent the developmental norms in South Asian athletic populations in which players often display the characteristics of moderate height and relatively lean body composition compared to the Western or elite European athletes. The midfielders and attackers in this study, for example, have BMI values which are less than those found in European sporting leagues which are expected to give differing genetic build, training intensity and nutritional intake and resource availability. Despite these contextual constraints, similarity in the relative positional distinctions within the different regions leads to confirmation of the universality in the alignment of anthropometric actions observed in football.

The fact that the Random Forest model performance results to be superior to linear regression (accuracy = 83.9%, F1 = 0.82) provides evidence that complex non-linear anthropometric patterns capture the positional differences more accurately than linear. Logistic Regression which assumes that the separability is linear did perform the weakest, implies that the body structures of Pakistani footballers don't differ in strictly linear ways. The good performance of Support Vector Machine and Gradient Boosting also confirms the advantages of positional classification in models that can capture multifactorial interactions.

These findings are strongly in line with worldwide evidence. Like Baboota and Kaur (2019) which predicted better accuracy by ensemble models in Football analytics and Merzah et al (2024) which showed ensemble methods to be better for performance classification, this study shows that Random Forest is one of the strongest methods of performing human performance prediction. Furthermore, it is in line with Morciano et al. (2024) that it provides significant argumentation that the power of ML is in the ability to model subtle anthropometric complexities that traditional statistics are unable to model.

The feature importance rankings showed height, BMR and thigh circumference to be the most influential predictor of position. These variables fall within the theoretical alignment with the positional demands in literature. Height plays a strong role in the execution of defensive and goalkeeping skills as known by Woods et al. (2018). Thigh circumference is a marker for muscular power which is critical for kicking power, sprinting and explosive movements, all traits of particular interest to the attacker and goalie. BMR, which depends on muscle mass and efficiency, stands for underlying physiological readiness which contributes to positional specialization. These findings therefore strengthen the theoretical framework for the belief that anthropometric and metabolic attributes together determine role suitability in football. Notably, fat percentage and calf circumference were much lower down the list of importance, implying that although they have some importance in classification, it is secondary to structural aspects such as height and thigh mass. This is in line with the findings of Almulla and Alam, 2020, wherein they found that the broader structural indicators are a better predictor of the match performance than isolated measures of body composition.

Taken collectively, the results indicate good agreement with international research that provides evidence proving the effective use of ML in position classification in footballing. Similar with the findings by Utomo and Wiradinata in 2023, this finding allows to vindicate that ML course will recognize community the other hand in the other subtle positional attributes in non-elite samples. The relatively high accuracy produced in this Pakistani cohort is noteworthy given the limited training facilities and non-standardized talent identification systems (maintained by the inconsistent strength conditioning programs) followed by most colleges within the country. This replication across different contexts relates to the fact that ML has the potential to identify physiological falling under a broad range of contexts; without the need for elite-level datasets.

One piece of information that shone through was the slight increase in misclassification between the midfielders and attackers. This overlap resembles their respective functional demands in common (speed, agility, and low body-fat) thus resulting in relatively similar anthropometric profiles. Comparable findings were observed by Oliver et al. (2020) and Deutsch et al. (2023), who proposed that some performance-based differentiations between midfielders and attackers became clearer in tactical or movement-based datasets as opposed to anthropometric ones. Thus, the misclassification is not a mistake which occurs but rather an expected manifestation of role-stimilance. Similarly, goalkeepers had the highest recall (0.91), which seems unusually good. However, the distinguishable body profiles of goalkeepers, in particular height and thigh girth, make it much easier to classify them than a field player. This is compatible with international observations in which the goalkeepers are always the most well separated anthropometric category (Yuwono et al., 2024).

Overall, the discussion proves that Pakistani collegiate footballers exhibit clear anthropometric differences among positions and machine learning, particularly Random Forest, proved to be an effective data mining tool in capturing the nature of the complex patterns. The findings justifiably match up to the observations elsewhere around the world and match the local contextual realities of moderate height, lean composition and the variability of training based on available resources. The findings strengthen the theoretical stance that playing position is highly based on structural body traits as well as ML-based classification can be a powerful and objective alternative to subjective talent identification procedures in Pakistan's collegiate football system.

CONCLUSION

The present study aimed to classify position-specific anthropometric profile of intercollegiate football players using machine learning techniques to provide one of the first data-driven assessments in the context of Pakistani context of collegiate sports. The results show that

Policy Journal of Social Science Review

Online ISSN 3006-4635 Print ISSN 3006-4627

Vol. 3 No. 11 (2025)

certain anthropometric characteristics can significantly differentiate midfielders from defenders, attackers and goalkeepers and that modern machine learning algorithms, especially Random Forest, can perfectly identify these characteristics, with a high level of reliability. Across the entire sample, there were greater height, muscle mass and limb circumferences in the defensive and goalkeeper positions in keeping with the physical demands of aerial duels, goal protection and power movements. A preference leaning towards the lean muscular algorithm was observed in the midfielders and attackers with lower BMI and fat percentages, indicating their running and accelerating and change-of-direction needs for their positions. These trends not only reflect important and well-documented global trends but also show that positional differentiation is present in the resource-limited and developing sporting environment of Pakistan.

Machine learning models performed stiffly with ensemble models outperforming traditional classifiers confirming that positional categorization in football is based on complex, non-linear anthropometric findings. Random Forest showed the best accuracy (83.9%) and stable CV performance, which stressed its suitability for analyzing multidimensional data of performance. The positional importance of height, BMR and thigh circumference as key structural determinants of positional suitability is further highlighted from the model's feature importance rankings. In conclusion, the present study confirms the use of anthropometric profiling by ML as a powerful, objective substitute for the subjective talent identification methods of choice in Pakistani football. By providing a foundation for classification based on evidence consistently defining the optimization of placement for classification, this study is a welcome framework to coaches, selectors, and sports scientists that will aid in determining optimal placement and training loads, as well as obtaining scouting intelligence in the collegiate system. The results highlight the potential of integrating analytics in the development of athletes and provide the basis of a starting point for more systematical scientific assessment of talent in Pakistan.

RESEARCH IMPLICATIONS

The results of the present study have several significant implications in sports sciences, talent identification and development of football in the context of Pakistan and similar developing situations. First, the results show that anthropometric characteristics are also reliable in separating playing positions even in non-elite, college level players. This is consistent with the plateaus of the body, and bodily structure largely of the body such as height, metabolic characteristics and the thigh musculature being a major factor in deciding who can be in which position. Consequently, coaches and sports administrators can use these measurable indicators to make more objective decisions when scouting for talent, selecting a team and developing their players.

Second, the effective utilization of machine learning models, in particular Random Forest, provides the potential and importance of incorporating computational analytics into local sports systems. ML-driven classification represents a replicable and data-informed way of doing things that can avoid subjectivity in selection processes traditionally based on experience in Pakistan. This is a huge advance for areas that do not have access to elite-level performance analytics, coming complete with scientific decision-making using simple inputs in anthropometry.

Third, the feature importance analysis highlights the possibility of designing specific training programmes. For instance, players who aspire to play as a defender or a goalkeeper may benefit from interventions designed to enhance strength and lower limb development, while attackers and midfield players may benefit from interventions designed to enhance lean body mass and agility. This direction follows the research which allows for more personalized and position-specific conditioning programmes across colleges.

Finally, the study provides groundwork information which can help develop further studies on performance modelling, injury prediction, and long-term player development in Pakistan. By showing quotes on how ML can be useful in measuring talent, the research urges increased use of sport analytics in academia, institutions, and professionals. This helps to establish a platform upon which more comprehensive and multi-dimensional performance databases can be built on eventually and help elevate the level of science and competitiveness of football in the region.

CONFLICT OF INTEREST

The author states that there are no conflicts of interest relating to the conduct, analysis or reporting of this research study.

REFERENCES

Al-Asadi, M. A., & Tasdemır, S. (2022). Predict the value of football players using FIFA video game data and machine learning techniques. *IEEE access*, *10*, 22631-22645.

Almulla, J., & Alam, T. (2020). Machine learning models reveal key performance metrics of football players to win matches in qatar stars league. *IEEE Access*, *8*, 213695-213705.

Baboota, R., & Kaur, H. (2019). Predictive analysis and modelling football results using machine learning approach for English Premier League. *International Journal of Forecasting*, *35*(2), 741-755.

Deuker, A., Wittkugel, J., Dublin, Y., Klatt, S., Körner, S., Vogt, T., ... & Klemp, M. (2023). Analyzing the effect of reduced field sizes on tactical behaviour in 11 versus 11 football practice. In *World Congress on Science and Football 2023: Enhance Performance Engage Society* (p. 240). World Congress on Science and Football.

Deutsch, J. P., Rieke, M., Rein, R., Braunstein, B., & Bassek, M. (2023). Change of direction demands in German high-performance soccer. In *World Congress on Science and Football 2023: Enhance Performance Engage Society* (p. 139). World Congress on Science and Football.

Lindberg, A., & Söderberg, D. (2020). Comparison of machine learning approaches applied to predicting football players performance.

Manish, S., Bhagat, V., & Pramila, R. M. (2021, May). Prediction of football players performance using machine learning and deep learning algorithms. In *2021 2nd International Conference for Emerging Technology (INCET)* (pp. 1-5). IEEE.

Merzah, B. M., Croock, M. S., & Rashid, A. N. (2024). Intelligent classifiers for football player performance based on machine learning models. *International journal of electrical and computer engineering systems*, *15*(2), 173-183.

Morciano, G., Zingoni, A., & Calabrò, G. (2024). Optimization and comparison of machine learning algorithms for the prediction of the performance of football players. *Neural Computing and Applications*, *36*(31), 19653-19666.Parameters with Regard to their Playing Positions. *SN Computer Science*, *5*(8), 974.

Morciano, G., Zingoni, A., Morachioli, A., & Calabrò, G. (2022, October). Machine learning prediction of the expected performance of football player during training. In *2022 IEEE international conference on metrology for extended reality, artificial intelligence and neural engineering (MetroXRAINE)* (pp. 574-578). IEEE.

Nouraie, M., & Eslahchi, C. (2023). Positioning soccer players for success: a data-driven machine learning approach. *Computational Mathematics and Computer Modeling with Applications (CMCMA)*, *2*(1), 24-33.

Oliver, J. L., Ayala, F., Croix, M. B. D. S., Lloyd, R. S., Myer, G. D., & Read, P. J. (2020). Using machine learning to improve our understanding of injury risk and prediction in elite male youth football players. *Journal of science and medicine in sport*, *23*(11), 1044-1048.

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*(8), 974.

Utomo, K. S., & Wiradinata, T. (2023). Optimal playing position prediction in football matches: a machine learning approach.

van der Vegt, B. (2024). *Examining Technical and Tactical Performance Within Women's Australian Football Match-play* (Doctoral dissertation, Bond University).

Woods, C. T., Veale, J., Fransen, J., Robertson, S., & Collier, N. F. (2018). Classification of playing position in elite junior Australian football using technical skill indicators. *Journal of sports sciences*, *36*(1), 97-103.

Yuwono, C., Adi, S., Annas, M., Darmawan, A., Billiandri, B., Putri, D. T., & Pratama, A. B. (2024). Determinants of Performance of 12-Year-Old Soccer School Players in Central Java. *ACTIVE: Journal of Physical Education, Sport, Health and Recreation*, *13*(2), 378-381.