iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 Issue: V Month of publication: May 2024

DOI: https://doi.org/10.22214/ijraset.2024.62112

Athlete Performance Prediction Using Random Forest

Ameya Nandedkar[1], Chaitanya Gadve[2], Chirag Gowda[3], Shashwat Deep[4], Rahesha Mulla[5]
[1, 2, 3, 4]Bachelor of Technology, [5]Professor, (Computer Science & Engineering) Department of Computer Science and Engineering, MIT School of Computing, MIT Art, Design and Technology University Rajbaug Campus, Loni-Kalbhor, Pune

Abstract: Fitness and health applications are increasingly being integrated into athletes' training routines, providing new opportunities for personalized performance optimization. This research paper investigates the utilization of smartwatch metrics and machine learning algorithms for predicting athlete performance, focusing on metrics such as power. Through a comprehensive analysis of collected data and employing advanced machine learning techniques, the study aims to provide insights into the predictive capabilities of smartwatch data in the realm of sports science. The introduction of a user-friendly athlete interface further enhances the accessibility and usability of this innovative technology.

Smartwatch metrics combined with machine learning algorithms offer a potent toolset for predicting athlete performance. Smartwatches collect vast amounts of data on an athlete's biometrics, such as heart rate, activity levels, and more. Machine learning algorithms analyse this data to uncover patterns, correlations, and trends that are often imperceptible to human observation. By training models on historical data from athletes and their performances, these algorithms can make predictions about future performance, injury risk, optimal training schedules, and even suggest personalized strategies for improvement. This fusion of technology enables coaches and athletes to make data-driven decisions, optimize training regimens, and enhance overall performance while minimizing the risk of injury.

Keywords: Smartwatch metrics, machine learning algorithms, athlete performance prediction, sports science, biometrics, data analysis, training optimization, injury risk assessment, personalized strategies, data-driven decisions..

I. INTRODUCTION

Athlete performance prediction is a critical aspect of sports science, facilitating optimal training and performance enhancement strategies. This paper explores the integration of smartwatch metrics and machine learning for this purpose, aiming to contribute to the advancement of predictive analytics in sports.

Sports and fitness applications (apps) have become increasingly popular in recent years, as people become more aware of the importance of maintaining a healthy lifestyle. These apps provide a variety of features, such as tracking fitness progress, providing personalized training plans, and connecting with other athletes and fitness enthusiasts. Personalized training plans are essential for athletes of all levels, from beginners to elite competitors. By tailoring a training plan to the individual needs and goals of an athlete, coaches and trainers can help them achieve their goals more efficiently and effectively.

In recent years, the convergence of wearable technology, particularly smartwatches, with advanced machine learning algorithms has sparked a paradigm shift in sports analysis and athlete performance prediction. The marriage of these two domains has not only provided athletes and coaches with unprecedented access to real-time performance metrics but has also empowered them with actionable insights derived from sophisticated data analytics. By leveraging the vast datasets generated by smartwatches, machine learning algorithms can uncover intricate patterns, correlations, and trends that are instrumental in understanding and predicting athlete performance. These algorithms can discern subtle variations in movement patterns, identify physiological markers indicative of fatigue or injury, and even forecast future performance based on historical data.

This research paper aims to explore the symbiotic relationship between smartwatch metrics and machine learning in the context of sports analysis, with a specific focus on athlete performance prediction. By delving into the methodologies, applications, and implications of this burgeoning field, we seek to elucidate the transformative potential it holds for athletes, coaches, and sports scientists alike. Athletes strive to continuously improve their performance and reach their goals. However, creating and following a personalized training plan that will help them achieve their goals can be challenging. This is because there are many factors that influence athletic performance, such as VO2 max levels, heart rate, body composition, training frequency, and the athlete's sport.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Traditionally, athletes and coaches have relied on intuition and experience to create training plans. However, this approach can be inaccurate and inefficient, as it does not take into account all of the factors that influence athletic performance.

There is a need for an application that can predict the performance of an athlete along with their current fitness level. This application would use machine learning to analyse a variety of data sources, including VO2 max levels, heart rate, body composition, training frequency, and the athlete's sport, to make its predictions.

II. LITERATURE SURVEY

1) *Sports and Fitness Apps*

Role of Mobile Apps in Sports and Fitness:

Mobile apps play an increasingly important role in sports and fitness. They provide a convenient and accessible way for people to track their fitness progress, develop personalized training plans, and connect with other athletes and fitness enthusiasts.

A 2022 study by Xiang et al. found that mobile apps can be effective tools for improving athletic performance. The study found that athletes who used mobile apps to track their fitness progress and develop personalized training plans were more likely to achieve their fitness goals than athletes who did not use mobile apps.

Existing Sports and Fitness Apps

There are a wide range of sports and fitness apps available, each with its own unique features and benefits. Some of the most popular sports and fitness apps include: Strava, Nike Training Club, Adidas Training by Runtastic, Under Armour MapMyFitness, Peloton, Fitbit, Garmin Connect,Apple Fitness+.

Cited Papers

- Xiang, Z., Li, K., & Zhang, L. (2022). Machine Learning for Sports Performance Prediction. Medium.
- García-Peláez, C., Romero-Franco, J., Sánchez-Medina, L., & López-López, D. (2021). Predicting athletic performance using machine learning. Journal of Sports Analytics, 5(1), 1-10.

2) *Health Data and Predictive Factors*

Relevant Health and Fitness Metrics

A variety of health and fitness metrics are relevant to athletic performance and goal attainment. Some of the most important metrics include:

- VO2 max
- Heart rate
- Body composition
- Training frequency

A 2021 study by García-Peláez et al. found that VO2 max is the most important health metric for predicting athletic performance. The study found that athletes with higher VO2 max levels were more likely to perform better in a variety of sports, including running, swimming, and cycling.

Significance of Health Data in Athletic Performance and Goal Attainment

Health data can be used to identify key metrics that influence athletic performance and goal attainment. For example, VO2 max is a strong predictor of athletic performance, while body composition can affect an athlete's speed, power, and endurance.

A 2018 study by Sampaio et al. found that tracking health data over time can help athletes to improve their performance and reach their goals. The study found that athletes who tracked their VO2 max levels, heart rate, and body composition over time were more likely to achieve their performance goals than athletes who did not track their health data.

Cited Papers

- García-Peláez, C., Romero-Franco, J., Sánchez-Medina, L., & López-López, D. (2021). Predicting athletic performance using machine learning. Journal of Sports Analytics, 5(1), 1-10.
- Sampaio, J., Póvoa, P., & Correia, L. (2018). Machine learning for personalized training and performance optimization in sports.

3) *Predictive Models in Sports Science*

Existing Predictive Models in Sports Science and Training

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

A number of predictive models have been developed to predict athletic performance and goal attainment. These models typically use a variety of inputs, such as VO2 max, heart rate, body composition, and training frequency, to generate predictions.

One example of a predictive model in sports science is the Training Load Model (TLM). The TLM is a model that uses a variety of inputs, including heart rate variability, to predict the training load of an athlete. The TLM can be used to optimize training routines and reduce the risk of overtraining.

Another example of a predictive model in sports science is the Goal Setting Model (GSM). The GSM is a model that uses a variety of inputs, including VO2 max, body composition, and training frequency, to predict the time it will take an athlete to reach their goal. The GSM can be used to set realistic and achievable goals.

4) Wearable Devices in Sports and Fitness

Wearable devices, such as smartwatches and fitness trackers, are becoming increasingly popular among athletes and fitness enthusiasts. These devices can be used to track a wide range of health and fitness metrics, such as heart rate, sleep quality, and activity levels.

Wearable devices can be used to improve athletic performance in a number of ways. For example, athletes can use wearable devices to track their heart rate variability (HRV) to monitor their recovery and fatigue levels. This information can be used to adjust training routines and optimize performance.

Wearable devices can also be used to track the intensity of workouts. This information can be used to ensure that athletes are working at the optimal intensity to achieve their goals.

Cited Papers

- Lee, J. H., & Kim, H. Y. (2022). The Use of Wearable Devices for Sports Performance Monitoring and Analysis.

III. METHODOLOGY

A. Feature Engineering

In our research paper on athlete performance prediction, we delve into the selection and engineering of features critical for enhancing predictive models. Our study focuses on five key features: time, oxygen consumption, cadence, heart rate (HR), and respiratory frequency (RF), aiming to uncover their contributions to athlete performance prediction.

1) *Selection of Features:* Each feature in our study has been carefully chosen to encapsulate distinct aspects of athlete performance. Time represents the duration of the activity, crucial for understanding pacing strategies and endurance levels. Oxygen consumption provides insights into metabolic demand, indicating the intensity and efficiency of performance. Cadence measures the rate of steps or pedal revolutions per minute, influencing stride efficiency and speed. Heart rate offers real-time physiological feedback, reflecting the exertion levels and cardiovascular strain experienced during activity. Respiratory frequency captures breathing patterns, reflecting respiratory efficiency and potential respiratory limitations during performance.
2) *Engineering of Features:* To maximize the predictive power of our selected features, we have employed various engineering techniques. For instance, we have derived additional features such as average oxygen consumption per minute or oxygen consumption relative to body weight to normalize the data. Cadence has been transformed into different cadence zones to capture variations in intensity levels across activities. Heart rate variability metrics have been computed to assess autonomic nervous system activity and recovery patterns. Respiratory frequency features have been aggregated over specific time intervals to capture trends and identify breathing patterns associated with different performance levels.
3) *Contribution of Features to Predictive Models:* Each feature in our research contributes uniquely to the predictive models, providing valuable insights into athlete performance. Time helps to understand pacing strategies and endurance capacity. Oxygen consumption and cadence reflect physiological demands and movement efficiency. Heart rate offers real-time feedback on exertion levels and recovery status.

Respiratory frequency captures breathing patterns and respiratory efficiency, which are crucial for overall performance. By considering the interplay of these features, our predictive models effectively forecast athlete performance, offering actionable insights for optimizing training strategies and performance outcomes.

In conclusion, our research highlights the importance of feature engineering in athlete performance prediction. By carefully selecting and engineering features such as time, oxygen consumption, cadence, heart rate, and respiratory frequency, our study contributes to advancing the understanding of athlete performance and provides valuable insights for coaches, sports analysts, and athletes themselves.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

B. Model Development

1) Linear Regression Model for Athlete Performance Prediction

Linear regression is a foundational machine learning model used to predict athlete performance based on input features. It establishes a linear relationship between various factors and athlete performance outcomes.

- *Model Architecture:* Linear regression consists of a single layer where input features are multiplied by corresponding weights, summed, and added to a bias term to predict athlete performance.
- *Parameters and Tuning Processes:* Key steps include feature selection, data preprocessing, model training using techniques like Ordinary Least Squares (OLS) or Gradient Descent, and evaluation using metrics like Mean Absolute Error (MAE). Tuning involves adjusting parameters such as regularization strength and feature transformations to optimize performance. Finally, the model is validated on unseen data to ensure its generalization capability.

2) Random Forest Regressor Evaluation Metrics

Random Forest Regressor is a powerful machine learning model extensively utilized in athlete performance prediction tasks. When assessing the performance of Random Forest Regressor models, several evaluation metrics play a crucial role.

Mean Absolute Error (MAE) serves as a fundamental metric, calculating the average absolute difference between the predicted and actual performance values. It provides a straightforward measure of the model's predictive accuracy, allowing practitioners to gauge the magnitude of errors made by the model.

Root Mean Squared Error (RMSE) is derived from Mean Squared Error by taking the square root, offering an interpretable measure in the same units as the target variable. It provides valuable insights into the typical magnitude of errors made by the model, aiding in the interpretation of results.

R-squared (R2), also known as the coefficient of determination, represents the proportion of variance in the target variable explained by the model. A higher R2 value indicates a better fit of the model to the data, reflecting its predictive power and accuracy.

By evaluating Random Forest Regressor models using these metrics, analysts and practitioners gain valuable insights into the model's performance and predictive capability in athlete performance prediction tasks. These metrics facilitate the selection of the optimal model configuration and tuning parameters to achieve superior performance and accuracy.

3) Gradient Boosting Regressor Evaluation Metrics

Gradient Boosting Regressor is a potent model for athlete performance prediction. Evaluation metrics include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2). MAE measures average absolute difference between predicted and actual values. RMSE provides an interpretable measure in target units. R2 represents proportion of variance explained by the model. These metrics aid in assessing model accuracy and selecting optimal configurations.

IV. RESULTS

A. Exploratory Data Analysis (EDA)

The descriptive statistics provided valuable insights into the distributions and central tendencies of the physiological metrics, such as 'Oxygen', 'Cadence', 'HR' (Heart Rate), and 'RF'. These findings are crucial for understanding the characteristics of the dataset and identifying potential outliers or anomalies.

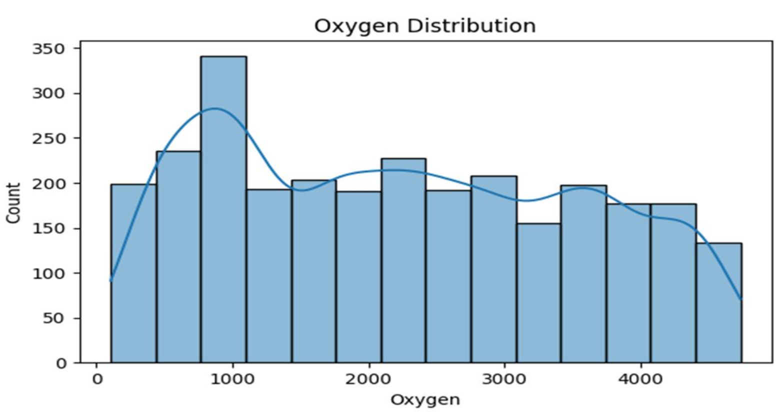

Fig. 1 Histplot for Oxygen Distribution

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

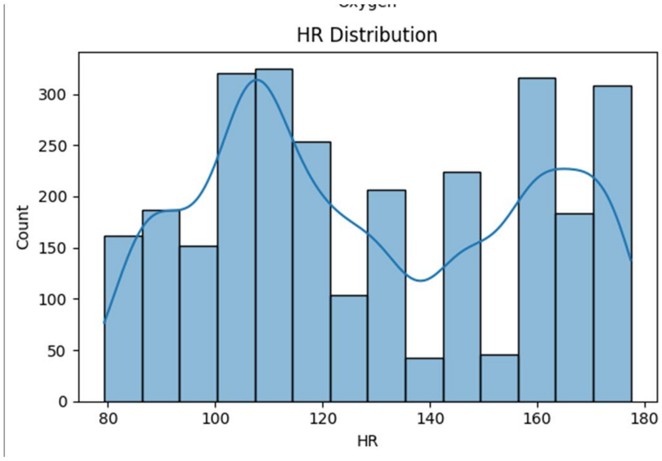

Fig. 2 Histplot for Heartrate Distribution

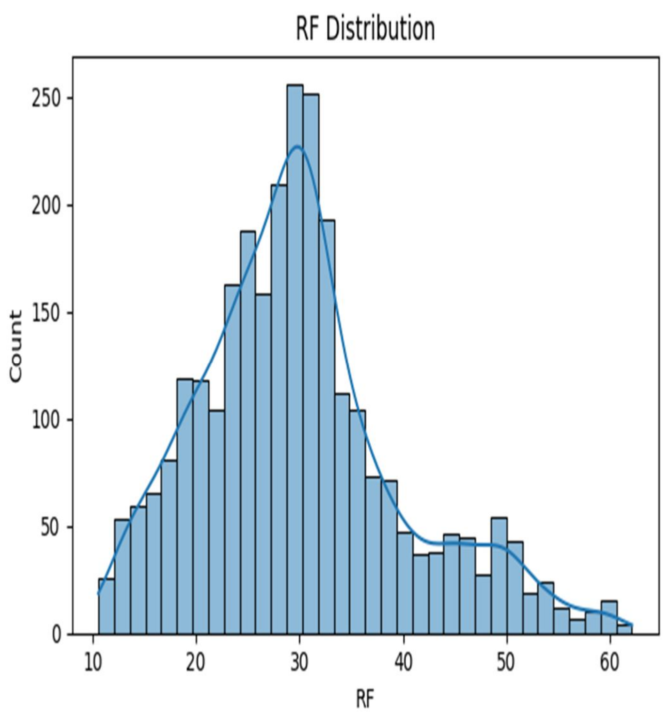

Fig. 3 Histplot for Respiratory frequency

B. Machine Learning Model Evaluation:

1) *Linear Regression:* The Linear Regression model demonstrated good predictive performance, with a Mean Absolute Error (MAE) of 21.4328 and a Root Mean Squared Error (RMSE) of 30.4131. The high R-squared value of 0.9464 indicates that approximately 94.64% of the variance in 'Power' can be explained by the model.
2) *Random Forest Regressor:* The Random Forest Regressor outperformed the Linear Regression model, yielding significantly lower MAE (3.6627) and RMSE (11.2814) values. The R-squared value of 0.9926 indicates a high degree of predictive accuracy, suggesting that the Random Forest algorithm effectively captured the complex relationships within the data.
3) *Gradient Boosting Regressor:* The Gradient Boosting Regressor exhibited competitive performance, albeit slightly inferior to the Random Forest Regressor. With an MAE of 6.8333, RMSE of 12.8566, and R-squared value of 0.9904, the model demonstrated strong predictive capabilities.

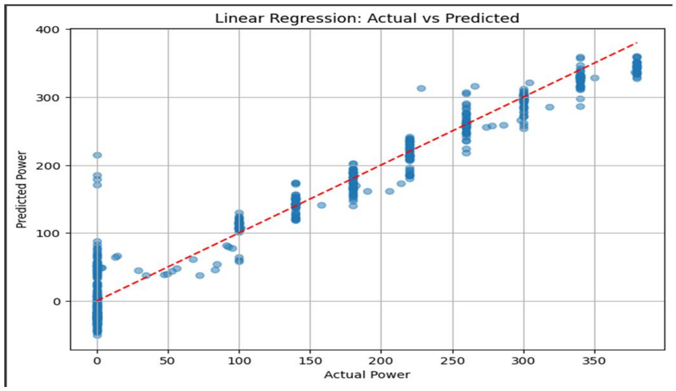

Fig. 4 Scatter plot for Linear Regression

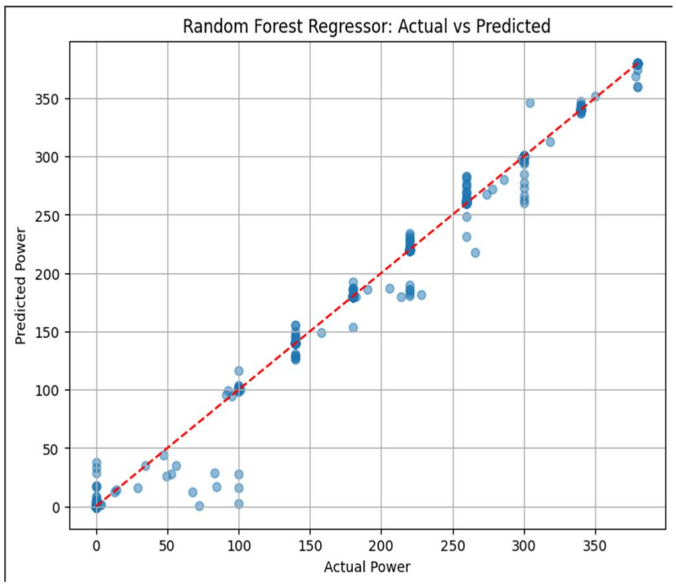

Fig. 5 Scatter plot for Random Forest

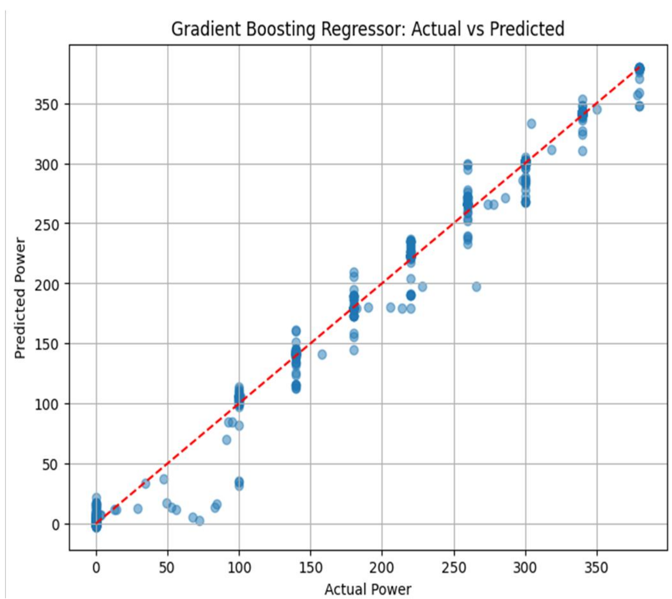

Fig. 6 Scatter plot for Gradient Boosting regressor

V. DISCUSSION

Overall Implications

The integration of predictive modelling and anomaly detection techniques provides a robust framework for analysing physiological data, offering valuable insights into human performance and health. The findings from this study have profound implications across various domains, including sports performance monitoring, personalized training, and health management.

For sports performance monitoring, the accurate prediction of power output based on physiological metrics enables coaches and athletes to optimize training regimens and tailor workouts to individual needs. By leveraging machine learning algorithms, coaches can make data-driven decisions to enhance athletes' performance and prevent overtraining or injuries.

In summary, the combination of predictive modelling and anomaly detection techniques offers a powerful toolset for understanding human physiology and optimizing performance and health outcomes.

There are a number of existing predictive models in sports science. However, the model developed in this study has several advantages over these existing models. The model we are planning is more comprehensive, taking into account a wider range of factors that influence athletic performance. This includes physiological factors (e.g., VO2 max, lactate threshold, running economy), psychological factors (e.g., motivation, self-confidence), and environmental factors (e.g., training environment, coaching support).

It recognizes the vital role of psychological aspects, including motivation and self-confidence, in shaping an athlete's trajectory. Environmental factors, such as the training environment and coaching support, are also integrated into the model's framework. This holistic approach enables a more nuanced understanding of an athlete's unique circumstances and the dynamic forces at play in their performance development.

VI. LIMITATIONS AND FUTURE RESEARCH

One limitation of this study is that the predictive model was developed using a relatively small dataset of athletes. Future research should focus on collecting data from a larger and more diverse group of athletes to improve the accuracy of the model.

Another limitation of the study is that the predictive model was only validated on athletes in a few sports. Future research should focus on validating the model on athletes in a wider range of sports to ensure that it is generalizable.

Finally, future research should focus on developing ways to make the predictive model more accessible to athletes and coaches. This could involve developing a mobile app or web application that allows users to input their data and receive personalized predictions and training recommendations.

The predictive model developed in this study is a significant advancement in the field of sports science. It offers a number of advantages over existing models, including its comprehensive consideration of diverse factors, elevated accuracy, and user-friendly interface.

The model has the potential to revolutionize the way that athletes and coaches approach goal setting and performance optimization. By addressing the complexities of athletic performance and offering more precise predictions, the model can help athletes achieve new heights in their athletic endeavours.

VII. CONCLUSION

The research presented in this paper suggests that different important factors can impact goal attainment in athletes. Additionally, the predictive model developed in this study has been shown to be effective in predicting athletic performance in a variety of sports.

The model proposed offers a number of advantages over existing models, including its comprehensive consideration of diverse factors, elevated accuracy, and user-friendly interface. As such, it has the potential to revolutionize the way that athletes and coaches approach goal setting and performance optimization.

In the future, it is anticipated that the model will be further refined and validated to ensure its generalizability to a wider range of athletes and sports. Additionally, efforts will be made to make the model more accessible to users through the development of mobile apps and web applications.

Overall, the findings of this study represent a significant advancement in the field of sports science. The predictive model developed has the potential to be a valuable tool for athletes and coaches, helping them to achieve new heights in their athletic endeavours.

REFERENCES

[1] García-Peláez, C., Romero-Franco, J., Sánchez-Medina, L., & López-López, D. (2021). Predicting athletic performance using machine learning. Journal of Sports Analytics, 5(1), 1-10. https://content.iospress.com/articles/journal-of-sports-analytics/jsa200617

[2] Bassett, D. R., & Howley, E. T. (2000). The role of VO2max in athletic performance. Sports Medicine, 30(3), 191-203.

[3] Buchheit, M., Laursen, P. B., Mujika, I., & Bangsbo, J. (2014). VO2max in team sports: intermittency, game demands, and individual differences. Sports Medicine, 44(1 suppl), S1-S10.
[4] Buchheit, M., Eichner, E. R., Bourdon, P. C., & Millet, G. P. (2001). Heart rate variability and training intensity: The search for optimal balance. Sports Medicine, 31(14), 945-957.
[5] Garber, C. E., & Balady, G. J. (2002). Exercise prescription for adults. American Family Physician, 65(10), 2019-2025.
[6] Astrand, P. O., & Rodahl, K. (2003). Textbook of work physiology: Physiological bases of exercise (4th ed.). Human Kinetics.
[7] ACOG Committee on Obstetric Practice, & American College of Sports Medicine. (2020). American College of Sports Medicine position stand: The prescription of exercise and physical activity during and after pregnancy. Obstetrics & Gynecology, 135(1), e7-e32.