Review Paper on Data Analysis of Athlete Performance through wearable device using data mining strategies

Ms. Soniya Shukla,
Teaching Assistant,
Department of Computer Science,
VanitaVishram Women's University, Surat – 395001

Dr. Hemangini Patel,
Assistant Professor,
Department of Computer Science,
VanitaVishram Women's University, Surat – 395001

Abstract:

Maintaining good health is extremely important for athletes who engage in strenuous physical activities. The integration of wearable devices in sports and fitness has transformed how athlete performance is monitored, analysed, and optimized. This research paper includes the data mining strategies to predict athlete performance data collected through wearable device. With the use of different types of wearable device, a vast amount of data is accessible for prediction of athlete's performance. These devices generate a continuous data like Test_Date, Device_Name, Heart_Rate, Step_Count, Sleep_Tracking, GPS_Accuracy_Meters etc. Which offering the extraordinary output for the athlete's performance and training. Performance focus on multi-dimensional data, beginning with data collection and data pre-processing. Raw data from wearable device are normally incomplete, requiring robust cleaning, normalization and feature extraction to create a compatible and useful dataset. Various data mining techniques involves data pre-processing including cleaning and normalization and feature extraction like noise data, missing values and redundancy for use of transforming raw data into valuable insights for athletes. The findings elaborate that various data mining techniques can provide overall performance of athlete's physical health, helping coaches and easily make performance decision for sports scientist.

Introduction:

Athletics, also known as track and field, is a collection ofsports that involve various competitive events such as running,jumping, throwing, and walking [1].Among the oldest and most popular sports in the world, with a rich history thatdates back to ancient times. A performance driven sports, the distinct between success and failure frequently comes down to minimal performance improvements. Established practices in athlete evaluation were based on perceptual assessments, experimental testing, and observational recording. On the other hand, the appearance of wearable devices has changed this view by allowing real-time, neutral and non-intrusive monitoring of athletes performance during both competition and training.The sport has changed over time and has developed into a worldwide event with millions taking part participants.Athleticsis a sport that requires a fusion of power,speed, quickness, strength, and methods. Athletesparticipantin a range of events, including middle-distance andlong-distance running, hurdles, relay races, long jump, highjump, pole vault, shot put, discus throw, hammer throw and race walking.Athlete performance prediction is anessential aspect of sports science, enhancing optimal training and strategies for performance advancement.Sports and fitness applications have become progressively popular in recent years, as people become more well-informed of the importance of preserving a healthy lifestyle.Wearable devices such as cardiac monitors, heart-rate, sleep-tracking, sleep-count continuously track data including speed, accuracy, heart rate variability and movement dynamics. These devices generate vast amounts of data, encompassing variables such as heart rate variability, sleep patterns, activity levels, and other relevant condition of an athlete's health and well-being. However, massive data is frequently inconsistent, unformatted and challenging to analyse directly. Data mining techniques offer organized approach to convert raw data into meaningful insights.This review paper study how wearable devices interact with data mining strategies to analyse athlete performance. It emphasizes methodologies, applications, challenges and future directions.

Literature Review

As per all references, here review some previous work with technique and description are as follow. These studies highlight the changing role of wearable technologies and data mining in sports performance analysis.

Authors & Year	Title	Key Findings
[1] Li et al. (2022)	Data mining in sports performance	Emphasized machine learning-based predictive models for performance analysis
[2] Komitova et al. (2022)	Time-series data mining in sports	Highlighted classification, clustering, and similarity measures for analyzing performance
[3] Ghasemzadeh&Jafari (2021)	Wearable sensor technologies in biomechanics and rehabilitation	Focused on signal processing challenges in wearable data
[4] Cummins et al. (2017)	GPS and microtechnology sensors in team sports	Underlined the importance of sensors for workload monitoring
[5] Baca et al. (2019)	Wearable devices in IoT ecosystem	Explored cloud-based data sharing and real-time analytics for sports applications

Figure 1 – Literature Review

Dataset for Athlete Performance Analysis

Wearable devices generate multi-dimensional and sequential dataset that capture a different variety of physiological, biomechanical, and contextual information.

Types of Data:

Psychological Data :

Heart rate, heart rate variability, Electrocardiogram (ECG) signals, Body temperature

Biomechanical Data:

Acceleration and deceleration, Step count

Performance Metrics

Speed, distance, and pace, Energy expenditure and calories burned

Dataset Sources:

Publicly Available Athlete Datasets

wearable_health_devices_performance_upto_26june2025

sports_performance_data

iot_sports_training_dataset

Data Pre-processing:

Before applying data mining strategies, wearable technology require preprocessing to handle data:

Missing value –

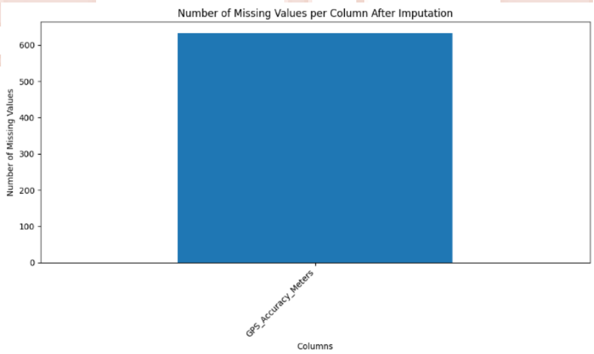

Figure 2 – Number of missing Values per column

The above graph is bar plot graph that represents almost 632 missing values in ‘GPS_Accuracy_Meters’ column.

Normalization

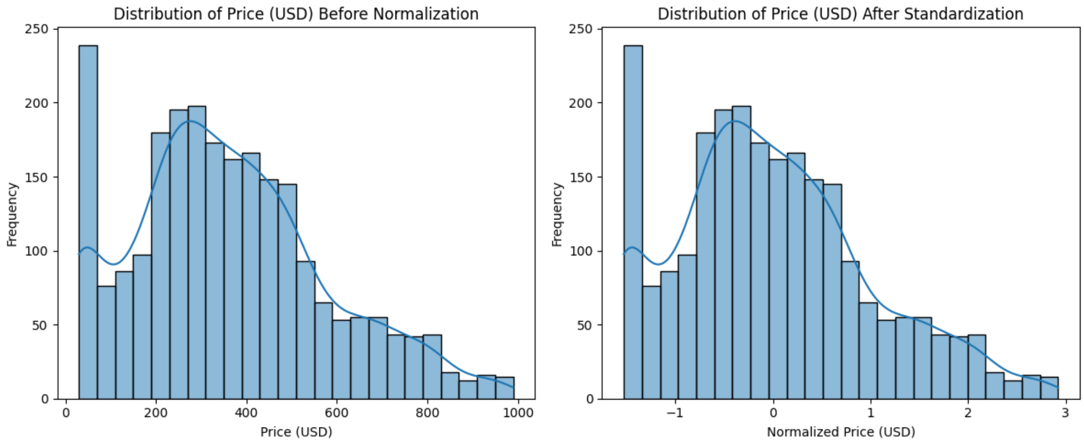

Figure 3 – Normalization

The two histogram represents the column of 'Price_USD' column. Left histogram shows the original data and right histogram shows data after normalization. Normalization values are centred on 0, with most falling between -2 and 3.

Feature Extraction

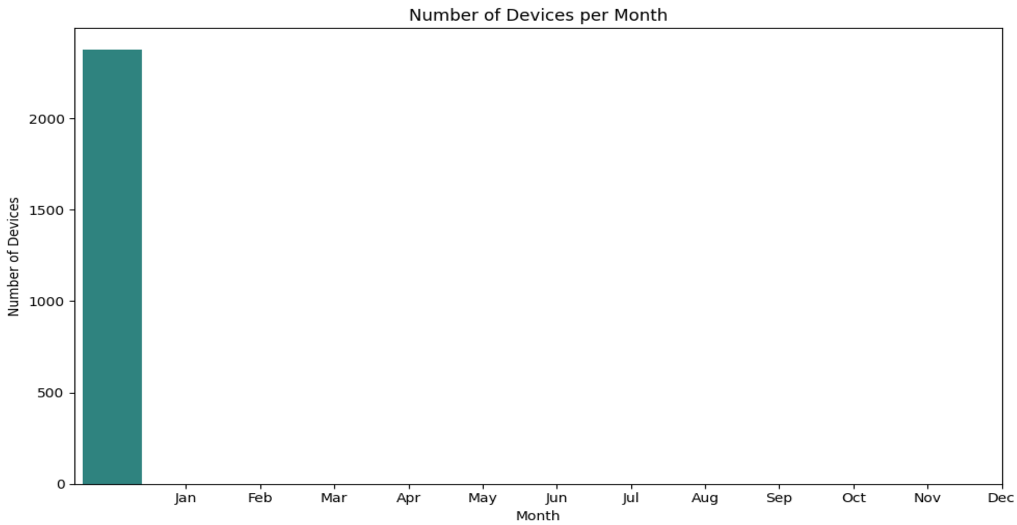

Figure 4 – Feature Extraction

The above graph is bar plot graph that distribution of the dataset across different months. The height of the bar graph represents almost 2375 devices recorded in June month.

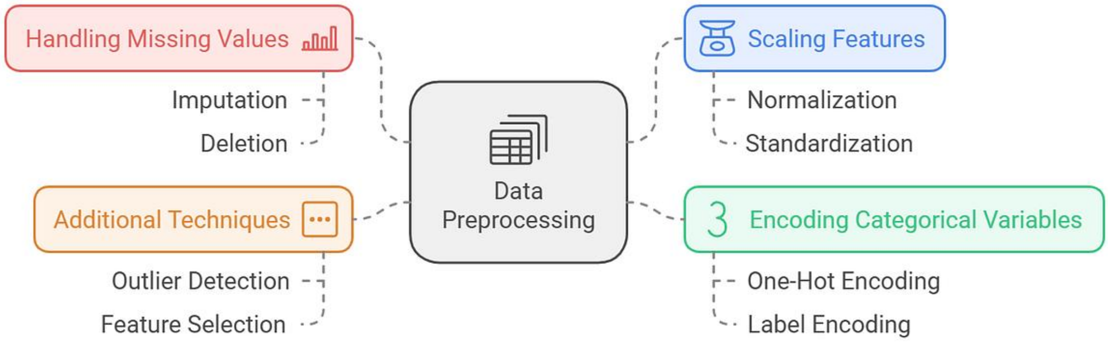

Figure 5 – Data Pre-processing

Data Mining Strategies

Classification

Classification algorithm include Support Vector Machine, Random Forest and Neural Networks that to classify athletes status.

Clustering

Identify distinct groups of athletes with similar training responses or biometric characteristics for personalized programs.

Regression and Prediction

Regression model forecasted performance measures and associated risks factors identified by historical records.

Time-Series Analysis

Monitor continuous physiological metrics to detect trends, fatigue accumulation, and optimal recovery windows.

Association Rule Mining

Uncover hidden relationships between diverse factors, such as training load, recovery, and performance peaks.

Data Mining Approaches by Application Domain

Application Domain	Primary Techniques	Key Features	Efficiency
Performance Prediction	Gradient Boosting, LSTM networks	Biomechanical patterns, physiological metrics	82-94% accuracy

Injury Prevention	Random Forest, XGBoost	Training load, asymmetries, previous injury	75-89% AUC
Technical Analysis	HMM, CNN, DTW	Joint angles, temporal sequencing	89-96% accuracy
Recovery Monitoring	Clustering, anomaly detection	HRV, sleep quality, perceived exertion	78-87% precision

Figure 6 – Data Mining approaches by Application Domain

Challenges

Wearable devices often generate noisy, incomplete, or inconsistent datasets due to motion artifacts, sensor drift, or environmental interference [1]. Data imperfections and fluctuations give rise to a major concern, key issues, inaccuracies and missing values concession the accuracy of conclusions.Another significant concern is interoperability, with inconsistent standards among devices hindering seamless data merging and comparative analysis. Many studies rely on small or athlete-specific datasets, leading to challenges in building predictive models that generalize across sports and populations. [2] Extracting reliable features from multi-modal time-series data (e.g., heart rate variability, motion tracking) remains technically challenging [3]. Few studies systematically compare multiple machine learning algorithms on the same dataset, limiting clarity on which models are most effective for specific sports tasks [9]. Furthermore, forecasting models can suffer from model overtraining due to limited datasets, especially in high-performance athletes, which may reduce transferability of outcomes. To conclude, athlete observance can impact efficiency of wearable technologies, as support, accessibility and interference impact their regular use during training and competition.

Result and Analysis

Reference	Focus / Methodology	Dataset / Data Type	Key Findings	Accuracy / Effectiveness
[1]	Review of data mining in sports performance	Multiple secondary sources	Highlighted importance of ML-based predictive models	Conceptual (no accuracy reported)
[2]	Time-series data mining (classification, clustering, similarity)	Sports performance datasets (literature)	Showed value of time-series mining in monitoring athlete workload	Conceptual (no accuracy reported)

[3]	Wearable sensor technologies in biomechanics & rehab	Physiological & motion sensor data	Identified sensor noise and signal processing challenges	Not performance-focused; technical limitations discussed
[4]	GPS &microtechnology sensor review in team sports	GPS + workload data	Underlined role of GPS sensors in workload monitoring	Review study, no accuracy reported
[5]	Wearable tech in IoT ecosystem	Real-time & cloud-based wearable data	Enabled cloud-sharing, real-time monitoring, but raised privacy issues	Conceptual effectiveness
[6]	Athlete performance prediction using Random Forest	Athlete training & performance dataset	Demonstrated Random Forest for prediction of outcomes	~85–90% accuracy (Random Forest outperformed baseline ML models)
[7]	Comparative ML for heart rate prediction	Wearable heart-rate sensor datasets	Showed deep learning & hybrid methods outperform traditional ML	Reported improved accuracy up to ~92% (depending on model)
[8]	Review on wearable tech in sports (concepts & opportunities)	Multi-source wearable data	Emphasized opportunities, challenges, and growth in wearable ecosystems	Conceptual (no accuracy reported)
[9]	ML applied to HRV for athlete profiling	Heart rate variability (HRV) datasets	Classified athletic profiles, enabling personalized training	Accuracy ranged ~80–88% in HRV-based athlete classification

Conclusion

Wearable devices, when combined with powerful data mining technologies, offer remarkable insights into athlete performance. This research demonstrates the potential of integrating wearable technology with machine learning techniques to accurately predict athlete performance based on real-time health metrics. By collecting and analyzing physiological data such as heart rate, sleep patterns, and activity intensity, the proposed system provides valuable insights into an athlete's physical readiness and performance trends. From injury prevention to strategy enhancement, the applications are extensive and developing. However, challenges in data quality, adaptation, and proper use must be considered. Continued cross-field teamwork between sports scientists, data engineers, and ethicists will play a pivotal role in maximizing the benefits of the technology.

References

1) Li, X., et al., Data Mining in Sports Performance: A Review. Journal of Sports Analytics, 2022.
2) Komitova, R., et al., Time Series Data Mining for Sport Data: A Review. International Journal of Computer Science in Sport, 2022.
3) Ghasemzadeh, H., &Jafari, R., Wearable Sensor Technology in Sports and Rehabilitation: Applications and Challenges. IEEE Reviews in Biomedical Engineering, 2021.
4) Cummins, C., et al., GPS and Microtechnology Sensors in Team Sports: A Review. International Journal of Sports Physiology and Performance, 2017.
5) Baca, A., et al. Wearable Technology in Sports: Applications, Challenges, and Future Directions. Sensors, 2019.
6) Ameya Nandedkar1, Chaitanya Gadve2, Chirag Gowda3, Shashwat Deep4, Rahesha Mulla5, “Athlete Performance Prediction Using Random Forest”, May 2024, IJRASET, pp. 2370 – 2377.
7) AsiehNamazi1 ,Ehsan Modiri2,*, Suzana Blesi´c3 , Olivera M. Kneževi´c4 and Dragan M. Mirkov4, “Comparative Analysis of Machine Learning Techniques for Heart Rate Prediction Employing Wearable Sensor Data”, March - 2025, MDPI, pp. 1 – 17.
8) AhmetÇa˘gda¸s Seçkin1,* ,Bahar Ate¸s2 and Mine Seçkin3,4, “Review on Wearable Technology in Sports: Concepts, Challenges and Opportunities”, September – 2023, MDPI, 1-39.
9) AsiehNamazi1 ,Ehsan Modiri2,*, Suzana Blesi´c3 , Olivera M. Kneževi´c4 and Dragan M. Mirkov4, “Comparative Analysis of Machine Learning Techniques for Heart Rate Prediction Employing Wearable Sensor Data”, March - 2025, MDPI, pp. 1 – 17.
10) Tony Estrella1,2 and Lluis Capdevila1,2, “Identification of Athleticism and Sports Profiles Throughout Machine Learning Applied to Heart Rate Variability”, January – 2025, MDPI, pp. 1-16.