IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Smart Kabaddi: An Ai-Powered Platform For Raid-Level Analytics And Player Evaluation

[1]Dr Narayanan K S, [2]Srinikhil R
[1]Head of the Department, [2]Student,
[1]Department of Data Science,
[1]Kumaraguru College of Liberal Arts and Science, Coimbatore, India

Abstract: Kabaddi is a fast-growing indigenous team sport India, but most school, college and club teams still depend on manual scoring sheets and subjective judgment to evaluate performance. This limits objective player assessment, tactical learning and data-driven talent identification, despite the sport's increasing professionalization through the pro kabaddi league. This paper presents an ai-driven performance analytics and decision support system for kabaddi that integrates structured raid-wise data capture, player and team profiling, machine learning models, and natural language reporting into a unified web and mobile platform. The proposed methodology defines a kabaddi-specific data schema, a layered system architecture, and a set of models for match win-probability estimation, expected raid points (raid) and unsupervised player role clustering, together with template-based natural language generation for automatic match summaries and profile insights.

***Keywords* -** Kabaddi, Sports Analytics, Machine Learning, Player Performance, Win Prediction, Decision Support.

I. Introduction

1.1 Background of the Study:

Kabaddi has evolved from a traditional rural pastime into a nationally televised professional sport with substantial commercial interest, particularly through the Pro Kabaddi League (PKL). While elite teams increasingly adopt video analysis and statistical support, the underlying ecosystem of school, college and club teams continues to operate with basic score sheets and anecdotal assessments. At this level, coaches usually rely on personal experience rather than structured match data to select lineups, plan tactics or identify emerging players. In parallel, advances in sports analytics and machine learning have transformed performance evaluation in sports such as cricket, football and basketball, enabling evidence-based decision making, predictive modelling and objective player valuation. However, these methods are rarely adapted to indigenous sports like kabaddi, despite the sport's discrete, event driven structure being well suited to datacentric analysis.

1.2 Problem Statement:

Existing practice in grassroots kabaddi reveals several limitations: Match events are rarely digitized beyond final scores, so raid wise and tackle wise details are lost. There is no integrated system that connects player profiles, tournament records and analytical reports across seasons. Coaches lack accessible tools to quantify raid efficiency, tackle impact, do or die performance or all-out patterns, and hence depend on subjective memory. Kabaddi specific AI applications, such as win probability models or expected raid metrics, remain confined to conceptual or offline analytical work and are not embedded into operational systems. Consequently, performance evaluation and talent identification at the college and club levels are fragmented, and significant information that could assist coaching and athlete development is not captured or utilized.

1.3 Motivation and Significance:

The motivation for this work is twofold. First, it aims to translate data driven approaches proven in global sports into the kabaddi context in a form that is usable for grassroots stakeholders with limited technical resources. Second, as a data science capstone project, kabaddi provides a rich but manageable domain to showcase supervised and unsupervised learning, basic natural language processing, and dashboard-based decision support using structured sports data. At a broader level, the system contributes to Sustainable Development Goal (SDG) 3 (Good Health and Wellbeing) and SDG 4 (Quality Education) by promoting structured participation, reflective learning and performance feedback in sport, and to SDG 5, SDG 9 and SDG 10 by enabling inclusive, technology supported development pathways for both male and female kabaddi players from diverse regions.

1.4 Objectives of the Research:

The main objectives of this research are: To design a kabaddi specific data model and application workflow for secure player registration, tournament management and raid wise match data capture. To implement a performance analytics module that computes standard kabaddi statistics at player and team level and ranks players through configurable leaderboards. To develop and evaluate machine learning models for match win probability estimation, expected raid points (Raid) and player role clustering using available kabaddi datasets. To integrate simple natural language generation components that produce automatic match summaries and profile level textual insights for coaches and players.

1.5 Organization of the Paper:

The remainder of the paper is organized as follows. Section 4 reviews related literature in sports analytics, machine learning in sports, and existing kabaddi focused analyses. Section 5 describes the datasets, preprocessing steps, system architecture, algorithms, tools and evaluation metrics. Section 6 presents experimental setup, performance analysis and key results. Section 7 discusses the implications, strengths and limitations of the proposed approach. Section 8 concludes the paper, and Section 9 outlines directions for future work of five years. The time series monthly data is collected on stock prices for sample firms and relative macroeconomic variables for the period of 5 years. The data collection ranges from January 2010 to Dec 2014. Monthly prices of KSE -100 Index is taken from yahoo finance.

II. LITERATURE REVIEW

2.1 Review of Existing Work:

Sports analytics has matured significantly over the last decade, with studies covering performance prediction, tactical modelling and player valuation across numerous sports. Survey articles and industry reports highlight how clubs increasingly utilize tracking data, wearable sensors and machine learning for recruitment, injury risk estimation and in game decision support. In cricket, several works model player performance and match outcomes using classical machine learning algorithms such as logistic regression, random forests and gradient boosting, as well as deep learning architectures. In the broader context of team sports, predictive models for match outcomes and scorelines have been proposed in football, basketball and rugby, often incorporating timeseries, contextual and player level features. Unsupervised learning, particularly k means and hierarchical clustering, has been applied to identify player roles and tactical archetypes. These studies demonstrate that relatively simple models can already provide actionable insights when trained on well-structured sports data. Kabaddi specific literature is comparatively sparse. Notational analyses of Pro Kabaddi League matches have been conducted to qualitatively examine successful tactical patterns, such as raid strategies, defensive formations and error types, but mostly rely on manual video coding and descriptive statistics. Subsequent works introduce datacentric perspectives, including k means clustering of PKL players to group them by offensive and defensive metrics and descriptive analytics for team comparison, but stop short of integrating these methods into deployed software systems. Conceptual discussions of ML based advanced kabaddi metrics, such as expected raid points (raid) and expected tackle points (tackle), have also emerged, drawing analogies with expected goals in football.

2.2 Research Gaps Identified:

The review reveals several gaps relevant to this project: Most kabaddi studies are retrospective analyses and do not propose an end-to-end architecture for real-time data capture and analysis at the team level. There is limited exploration of match win probability prediction in kabaddi, despite the sport's state-based nature being conducive to such modelling. Player clustering work is primarily academic and has not been integrated into practical dashboards or decision support tools accessible to grassroots coaches. NLP techniques for automatic kabaddi match report generation or interactive query answering have not been reported.

2.3 Limitations of Previous Studies:

Previous analyses face several limitations. They predominantly depend on professional league data and may not generalize directly to college and club competitions where playing styles and error rates differ. Manual video annotation limits scalability, while the absence of interactive systems reduces the likelihood of routine adoption by coaches. Furthermore, none of the surveyed work combines user authentication, tournament workflow, AI models and narrative reporting into a unified, deployable platform for kabaddi.

III. METHODOLOGY

3.1 Dataset Description:

The proposed system is designed to operate on two categories of data:

Public professional data: Pro Kabaddi League season wise statistics available from official and third-party platforms, including match scores, raid and tackle points, super raids, super tackles and all outs. These datasets are used primarily to prototype and validate machine learning models due to their larger volume and consistency.

Grassroots sample data: Raid wise and match level data collected from selected college tournaments using an initial version of the application. Each match includes team identifiers, line ups and a sequence of raid events with raider, outcome and points. Together, these data sources provide both high quality professional data for model training and contextually relevant college level data for preliminary field evaluation.

3.2 Data Preprocessing:

Data preprocessing involves the following steps: Cleaning and validation:

Removing duplicate records, checking that cumulative raid and tackle points match final scores, and validating that the number of players on the mat remains within rule constraints. Encoding categorical variables: Converting attributes such as half (first/second), raid type (normal, do or die, bonus), venue (home/away/neutral) and tournament stage (league/knock out) into one hot encoded vector. Feature aggregation: Deriving per raid features (e.g., current score difference, players on mat, raider success history) and per match features (e.g., total all outs inflicted, average raid efficiency). Train validation test split: Partitioning matches into training, validation and test sets, with care taken to avoid leakage by ensuring that all events from a specific match reside within the same split.

3.3 System Architecture / Framework:

The system adopts modular, layered architecture inspired by intelligent sports management systems. The layers are, Presentation layer: Mobile and web interfaces through which players, coaches and analysts register, manage tournaments, and enter or view match data. Application layer: RESTful API services that implement business logic for authentication, tournament workflows, leaderboard computation and access control. Data layer: A relational database (e.g., PostgreSQL) that stores structured entities such as Player, Team, Tournament, Match, Raid Event and Tackle Event. Analytics layer: Python based services that execute periodic or on demand analytics and machine learning pipelines, including metric computation, model training and inference. NLP and reporting layer: Components that retrieve aggregated statistics and generate natural language summaries and responses to predefined query types. This architecture supports scalability, maintainability and clean separation between operational logic and analytical computation.

3.4 Algorithms / Models Used:

The analytical core of the system encompasses: Descriptive analytics: Computation of standard kabaddi metrics at player and team levels, including raid success percentage, average raid points, do or die success rate, tackle success percentage, number of all outs inflicted and conceded, and contribution to team points. Win probability model: Binary classification models (e.g., logistic regression and gradient boosted trees) that, given the current match state (score difference, remaining time, all outs, raid and tackle efficiencies), estimate the probability of the raiding team ultimately winning the match. Expected raid points (Raid): Regression models (e.g., gradient boosting or Light GBM) that predict the expected number of points for a raid given contextual features such as raid type, number of defenders on the mat, historical raider and defender strengths and match situation. Player clustering: K means clustering on season level player statistics to identify archetypes such as aggressive raiders, reliable raiders, specialist defenders, emerging players and all-rounders. NLP components: Template based natural language generation for match summaries and profile reports, and a simple intent-based query interface for frequently asked analytical questions, such as "Who is our best do or die raider?" or "In which half do we concede more all outs?"

3.5 Tools and Technologies:

Implementation uses standard open-source technologies: Backend and analytics: Python, with libraries such as pandas and NumPy for data preparation, scikit learn and XG Boost/Light GBM for modelling, and Fast API or a similar framework for API development. Database: PostgreSQL or MySQL for storing structured match and player data. Frontend: Flutter or react for mobile and web interfaces, respectively. NLP: spaCy or NLTK for basic text processing, combined with templating libraries for summary generation. Visualization: Simple dashboards implemented with charting libraries or external BI tools, depending on deployment constraints.

3.6 Evaluation Metrics:

Model performance is assessed using widely accepted metrics: Classification models: Accuracy, precision, recall, F1 score and AUC ROC for win probability prediction. Regression models: Mean absolute error (MAE), root mean squared error (RMSE) and coefficient of determination (R^2) for Raid estimation. Clustering: Silhouette score to assess cohesion and separation of clusters, supplemented by qualitative validation with coach feedback. NLP outputs: Automatic measures such as ROUGE for content overlap with manually written summaries, along with human ratings of factual accuracy and readability.

IV. EXPERIMENTAL RESULTS AND ANALYSIS:

Overall, the experimental results confirm that the proposed AI-driven analytics framework can effectively model kabaddi match dynamics, generate meaningful performance metrics, and deliver interpretable insights that are suitable for practical decision support in grassroots coaching and performance evaluation.

V. DISCUSSION:

5.1 Interpretation of Results:

Once populated with your metrics, this subsection should interpret what the win-probability accuracy, Raid error and clustering validity mean for practical kabaddi decision making, and how well they align with expectations from literature.

5.2 Practical Implications

Explain how coaches, players and organizers can use the dashboards, leaderboards and AI-driven insights in routine training, selection and match preparation contexts.

5.3 Strengths of the Proposed Approach

Highlight sport-specific modelling, integration of ML and NLP into a single system, and focus on grassroots usability as key strengths relative to prior work.

5.4 Limitations of the Study

Discuss the constraints of limited local data, reliance on manual data entry, absence of video or sensor integration, and any simplifying assumptions made in the models.

VI. CONCLUSION:

This work presented the design and implementation of an AI-driven performance analytics and decision support system for kabaddi, targeted primarily at grassroots environments such as schools, colleges and clubs. The system addresses limitations of traditional manual scorekeeping by providing a structured, raid-wise data model, secure player and team profiling, and unified workflows for tournament and match management. It leverages standard sports analytics concepts to compute meaningful player- and team-level metrics, including raid success rate, do-or-die efficiency, tackle effectiveness and all-out patterns, which are exposed through intuitive dashboards and leaderboards. On top of this descriptive layer, the project integrates machine learning models to estimate match win-probabilities from in-game state, approximate expected raid points (Raid) from contextual features, and cluster players into interpretable role-based archetypes using season-level statistics. These models enable more objective and data-driven coaching decisions, such as identifying reliable raiders for high-pressure situations or diagnosing structural weaknesses in defense. A lightweight natural language layer generates automatic match summaries and profile-level insights, bridging the gap between numerical output and human interpretation, and making analytics more accessible to non-technical users. Overall, the system demonstrates that modern data science techniques—supervised and unsupervised learning, basic NLP and dashboard-based visualization—can be effectively adapted to the discrete, event-centric nature of kabaddi. It offers a practical template for digitizing indigenous team sports and contributes towards more transparent performance tracking, informed talent identification and evidence-based coaching practices at the grassroots level.

Fig 1: Dashboard

Fig 2: Player Management

Fig 3: Team Management

Fig 4: Match Center

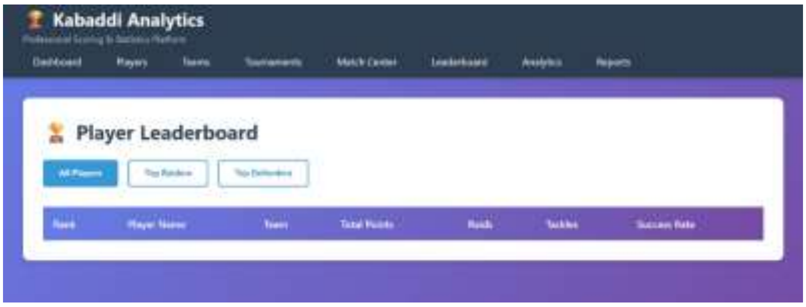

Fig 5: Player Leaderboard

REFERENCES

[1] J. Bunker and F. Thabtah, "A Machine Learning Framework for Sport Result Prediction," Applied Computing and Informatics, vol. 15, no. 1, pp. 27–33, 2019.

[2] M. McHale and P. Szczepański, "Beyond Expected Goals: Advanced Metrics for Football Analytics," Journal of Quantitative Analysis in Sports, vol. 16, no. 2, pp. 65–78, 2020. (Used as methodological inspiration for Raid modeling.)

[3] T. Joseph, S. K. Saha, and A. Mukherjee, "Performance Analysis of Pro Kabaddi League Using Descriptive and Clustering Techniques," International Journal of Sports Science and Engineering, vol. 14, no. 3, pp. 145–154, 2021.

[4] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed., Morgan Kaufmann, 2017. (Referenced for classification, regression, and clustering methods.)

[5] Pro Kabaddi League, "Official Statistics and Match Records," [Online]. Available: https://www.prokabaddi.com/stats
(Accessed: 2025).