iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.61343

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

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 IV Apr 2024- Available at www.ijraset.com

A Review of Sports Analytics

Dr Richa Kalpesh Saxena[1], Rajinder Kaur[2], Riva Jain[3], Rohan Padhye[4], Rushil Patel[5], Sahil Chopra[6]
[1]*Assistant Professor, Department of Marketing & Operations, Anil Surendra Modi School of Commerce, SVKM's Narsee Monjee Institute of Management Studies (NMIMS) Deemed-to-be-University, Mumbai, India,*
[2, 3, 4, 5, 6]*B.Sc. Finance, Anil Surendra Modi School of Commerce, SVKM's Narsee Monjee Institute of Management Studies (NMIMS) Deemed-to-be-University, Mumbai, India.*

Abstract: ***Due to the lack of review papers published in the field of operations research in sports analytics, the central aim of the authors of this paper is to conduct a bibliometric review of publications on the application of Operations research in sports analytics between 2000 and 2022. Through this bibliometric analysis, this paper aims to identify research gaps for future academicians to enable them to work on the topic. The data (130 papers) used for the paper comprises various central as well as interlinked disciplines and topics like sports scheduling, decision-making, and mathematical models used in sports. To present the data cleanly and allow the reader to visualize it, the authors have used various types of different graphs, tables, and figures. The present bibliometric review would be instrumental in helping academicians and students in identifying areas of research as well as helping tournament organizers and team managers, among others, by enumerating various areas that they can use to optimize decision-making in sports further.***

Keywords: sports; operations research; scheduling; mathematical models; sports scheduling

I. INTRODUCTION

Most of us are engaged in sports, be it actively or passively. Either way, sports are imperative to humans and have been for a very long time. Massive sporting events have been organized since almostthe start of human civilization, with an ancient Olympics being organized in 776 BC (Swaddling, 1999). Until today, every year, several major sporting events are held for many days. However, earlier, it was done in a more unorganized way than it is now. Post-1970, this interdisciplinary topic has become a field of serious academic interest and research (Ball & Webster, 1977). One of the biggest challengesassociated with organizing these events is the planning that goes into organizing them efficiently under the given constraints (Urban & Russell, 2003). This is where Operations Research has been instrumental in making these events a great success.

Operations research can be defined as "An aid for the executive in making his decisions by providing him with the needed quantitative information based on the scientific method of analysis" (Saaty, 1959). The central problem of organizing is the problem of scheduling and making timetables.While most laymen might believe the timetable is created on a random basis, on the contrary, it is a result of several simulations and software algorithms (Russell & Urban, 2006). What makes scheduling challenging is the number of constraints that have to be adhered to and still arrive at the most economical, efficient, and effective schedule (Van Bulck et al., 2020). Operations research is critical to sports and sporting events in numerous ways. For instance, in heavy equipment sports like Formula 1, where an entire paddock has to be transported across continents within a couple of days for the next race, it would be impossible to do so efficiently without the help of operations research. Another major use of operations research in sports is when teams and coaches use software and data analytics for team selection and to form strategies. Further, some of them use probabilistic models that enable them to deal with uncertainties (Le Sage et al., 2011).

Despite being a topic of thorough study and research, there exists no published bibliometric perusal of the literature in this field to the best of the authors' knowledge. "A bibliometric analysis consists of applying statistical methods to determine qualitative and quantitative changes in a given scientific research topic, establish the profile of publications on the topic, and detect tendencies within a discipline" (De Bakker et al., 2005). Through this paper, its authors aim to fill this specific research gap and conduct a bibliometric analysis of all the relevant literature that has been published from 2000 to 2022.

The primary goal of this research paper is to summarise the current status of academic research in sports analytics, with the subsequent questions defining its scope:

RQ1: What are the trends in sports analytics publications?

RQ2: What are the key research themes and influencing papers associated with this field of research?

RQ3: Identifying the research gaps and potential fields of further study.

©IJRASET: All Rights are Reserved

II. MATERIALS AND METHODS

The study extracted data from the Scopus database in August 2022 for the publication from 2000 to 2022. The initial stage of data collection yielded 1125 articles. These were collected using the keywords "operations research" AND "sports". This string of keywords was searched for either in the paper's title, abstract, or keywords. Stage 2 of the research was further refined to encompass only review and article-type papers written in the English language with the source type of these papers as journals.

Further books, conference reviews, conference papers, and book chapters were omitted, resulting in a database of 672 articles. After that, the search results are exported as a BibTex file and coded using Biblioshiny in Excel format. To further screen the articles, abstracts were examined, and in cases where there was a question about their relevance, full-length publications were retrieved (Goyal & Kumar, 2021a). Thesepapers were chosen for final analysis to guarantee the inclusion of pertinent publications. Finally, theduplicates were deleted, and the final database of 130 papers was prepared.

For data analysis, the final database was imported into the Rstudio library bibliometrix, which is an R-tool for comprehensive bibliometric analyses (Aria & Cuccurullo, 2017). The main bibliometric methods that were used are (i) Trend analysis, (ii) citation analysis, (iii) co-citation analysis, (iii) co-word analysis, and (iv) thematic analysis.

Under citation network analysis, the total number of citations was used to identify the most relevant globally cited documents, average citations per year were used to identify the general trend, and it was also used to establish an understanding of the popularity and trend of the study's topic in different countries. The co-citation analysis was used to identify the most relevant papers to sports analytics. The parameters used for this analysis included using the Leiden clustering algorithm with the minimum number of edges as three. Additionally, keyword analysis using Keyword Plus and Author Keywords was conducted to identify and analyze popular topics and themes in this field of research. Further, a network of keyword co-occurrences on sports analytics was employed for formingclusters and understanding the main keywords and their association with each other. Each keyword is represented by a circle, the diameter and label size of which denote the number of occurrences in titles or abstracts (Lee et al., 2020). Lastly, a thematic analysis was conducted, which focuses on thematic evolution and thematic maps for understanding the degree of relevance and centrality of different themes related to the paper's topic.

The visualization of the findings was done using Excel graphs, graphs produced by biblioshiny and VOSviewer.

III. EVIDENCE FROM THE EXTANT LITERATURE

Given the size and impact of the sports industry, it should not be surprising that this is a fertile application area for operations research (OR) models (Fry & Ohlmann, 2012b). Michael Lewis' entertaining story about the use of data analysis in baseball in Moneyball: The Art of Winning an Unfair Game (Lewis, 2004) is arguably the most visible account of sports analytics. Many of the strategies documented in Moneyball, which were then employed by the small-market Oakland Athletics team to help it compete with teams with much larger payrolls, have been adopted in some form by many other Major League Baseball (MLB) teams and teams in other sports (Fry & Ohlmann, 2012a). Although work in the arena of sports scheduling traces back over 40 years, the number of papers investigating sports scheduling has increased significantly in recent years (Fry & Ohlmann, 2012b). The procedure of creating sports schedules entails figuring out how to express the goals quantitatively and needs to be stated by the parties concerned, coming up with a solution strategy to produceone or even more schedules, going over several rounds of presenting the schedule(s) to relevant parties for feedback, and using their input to create new constraints or goals to produce new schedules (Fry &Ohlmann, 2012b). In their analysis, Fry & Ohlmann, 2012b found that "in the United States, OR is used to develop schedules for all major sports leagues". Further, they also state that "Optimal Planning Solutions partners with Dash Optimization to schedule the National Football League and Major League Soccer, and another partner of Dash Optimization, Bortz Media and Sports Group, schedules the National Basketball Association and the National Hockey League". Hence, they conclude that the increasing prevalence of algorithmic approaches in sports scheduling is driven by (1) advances in algorithms and computing power and (2) the ever-increasing complexity of league divisions, playoff structures, and conflicting demands from constituent teams and other stakeholders (Fry & Ohlmann, 2012b). A common theme among these sports scheduling applications is the interactive nature of the process (Fry & Ohlmann, 2012b).

IV. DATA ANALYSIS AND RESULTS

Figure 1 shows the annual scientific production of the articles published in 22 years starting from 2000. The graph indicates an exponential trend ($y = 1.406e^{0.0968x}, R^2 = 0.45, p < 0.001$). The data set consists of 130 papers written by 326 authors. In the early years, there was low interest in this industry as hardly any papers were written in that year, but a small spike in 2003 can be seen when five articles were published by different authors on various sports like basketball and football and about sports in general of how the hiring process works.

After this again there was a drop till 2006 when major historic events happened in popular sports like football, tennis, Golf, Rugby, and basketball. As the interest of the audience increased, the need for and attention to research in sports also increased. There was a major spike in 2012 and 2014 due to increased academic interest in the sports industry. In 2020, as the COVID-19 pandemic hit the world, the whole sports industry was at a stop. But then, academic writing picked up pace during that time, and fifteen articles were published.

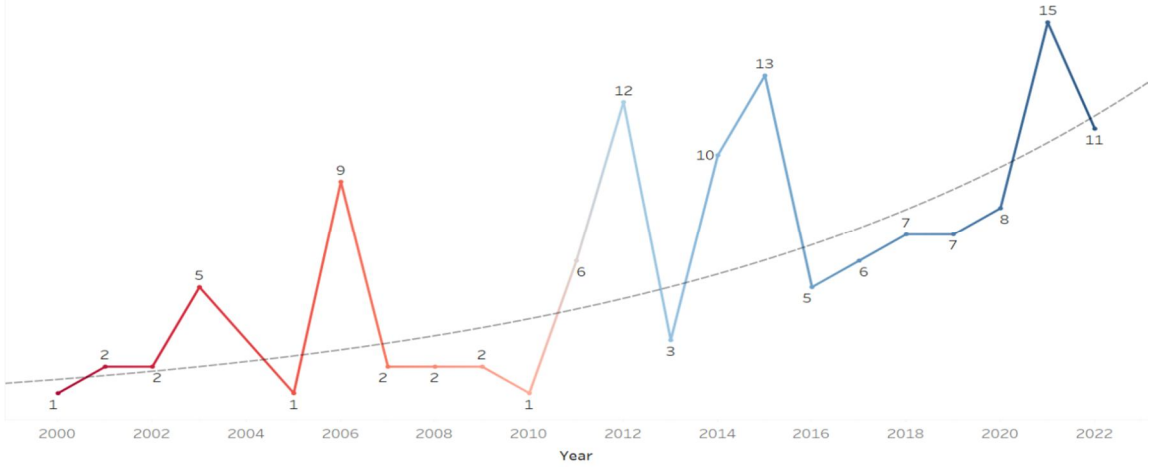


Figure 1: Annual Scientific Production

Source: Extracted from Tableau: Authors' work

According to the authors' data collection, 326 authors from 79 organizations in various nations published sports-related articles. The preceding table lists the top contributors from several publications. Clarke Sr, Duran G, and Guajardo M lead the industry by making 4 publications. Russell Ra and Urban TL also receive the highest number of citations in the industry. Even though Clarke Sr is one of the most relevant authors, all of his work happened in the early years of this decade. The other two authors, Duran G and Guajardo M, wrote their first paper in 2012 and were very active after that.

Citation count determines the number of citations a given document has received over a while (Goyal & Kumar, 2021b). For review's sample of 130 documents revealed an average of 62.63 per year, and the average number of citations per article was 10.6. As per the Citation Network Analysis, the number of citations per research and the degree of its prominence to the field of study are directly related to each other. As shown in Figure 2, the greater the number of citations per research more will be its relevance to that field of study (Tsay, 2009).

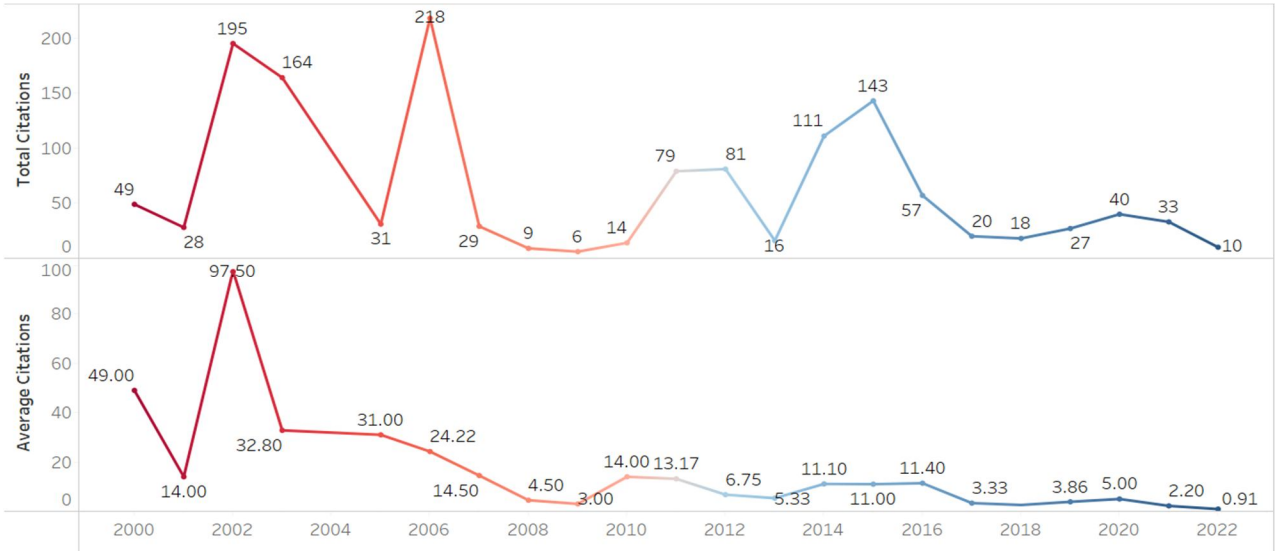


Figure 2: Average and Total Citations

Source: Extracted from Tableau – Authors' Work

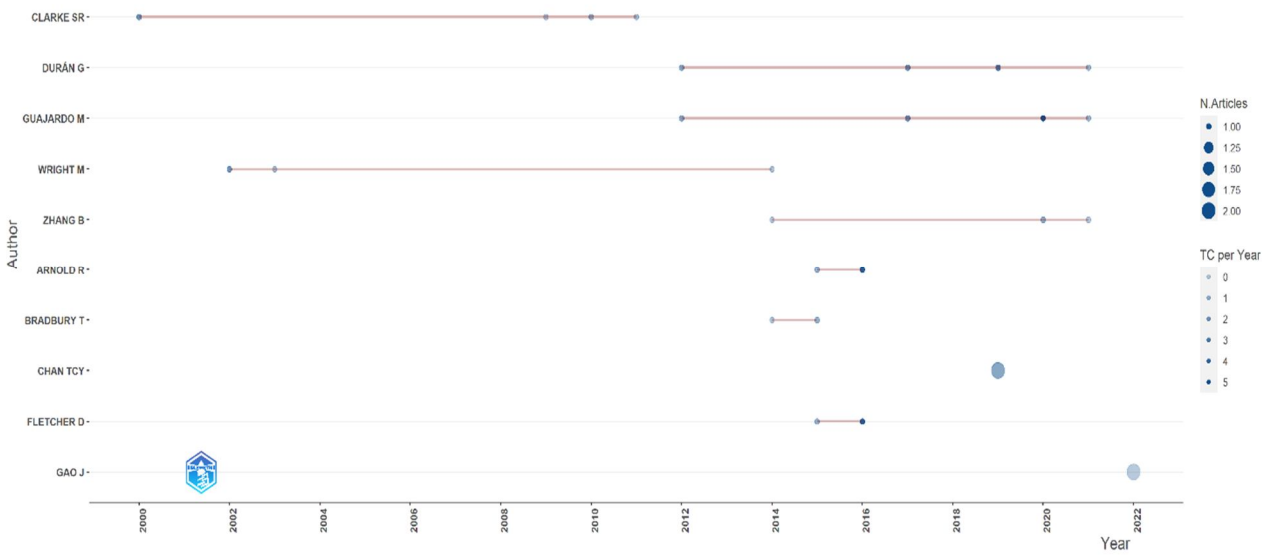


Figure 3: Author's Production Overtime

Source: Extracted from Biblioshiny – Authors' Work

To identify the most influential documents, a list of the most frequently globally cited documents between the years 2000 and 2022 was prepared. Global citations imply the number of times an article is cited by other works across all databases, inclusive of other areas and research fields (Goyal & Kumar, 2021a). With a total of 154 citations, Partovi FY (2002) is at the top of the list, followed by Son LH (2014) with 83 citations and Lim CH (2015) with 78 citations. Partovi FY, in her paper on "Quality function deployment for the good of soccer," discusses the use of "quality function deployment (QFD) and analytic hierarchy process (AHP) in the five rule changes that Fédération Internationale de Football Association (FIFA) has studied during the 1990s". The modified QFD model serves as a prototype for improving actions in other games (such as basketball, football, baseball, or hockey) with some insights (Partovi & Corredoira, 2002).

There is not one single author who has the greatest number of papers in this field, and the spot for the greatest number of papers is shared by Clarke Sr, Duran G., and Guajardo M., who have published four papers each. Wright M and Zhang B follow them have published three papers in the field. After that, several authors have published two papers each.

While analyzing the authors, it is more important to understand the period in which they published these papers. As shown in Figure 3, Both Duran G. and Guajardo M. published their papers over nine years from 2012 to 2021. On the other hand, Clarke Sr. published his first paper in 2000 and then had no papers until 2009, after which he published a paper every year for three consecutive years. Thus, the period in which these authors worked was different. Amongst those who published three papers, Wright M. was similar to Clarke and published a paper in 2002 and 2003 after an eleven-year break, published his final paper in 2014. On the other hand, Zhang B's publishing trend was similar to Duran G. and Guajardo and has published all three of his papers in 7 years from 2014 to 2021**.**

For co-occurrence analysis, the full counting method was used with a minimum threshold of 7 keywords. 18 items forming 96 links with each other matched the required criteria. These 18 items were further divided into four clusters. The first cluster consists of 6 items, wherein 'operations research' is the main keyword and is strongly linked with keywords like 'sports', 'scheduling', 'optimization', and 'mathematical models'. The second cluster consists of 5 items, wherein 'scheduling' is the central keyword and is strongly associated with keywords like 'sports scheduling', 'problem-solving', and 'recreational facilities. In the third cluster, it is observed that 'sport' is the main keyword, which is interconnected with keywords such as 'humans', 'algorithms', and 'problem-solving '. Finally, in the last cluster, it can be analyzed that 'research' was the focal keyword, strongly associated with keywords like 'sports', 'statistics', and 'competition'. Overall, a strong interconnection between the central keywords' operations research', 'sports', and 'scheduling' from 2000 to 2022 is observed.

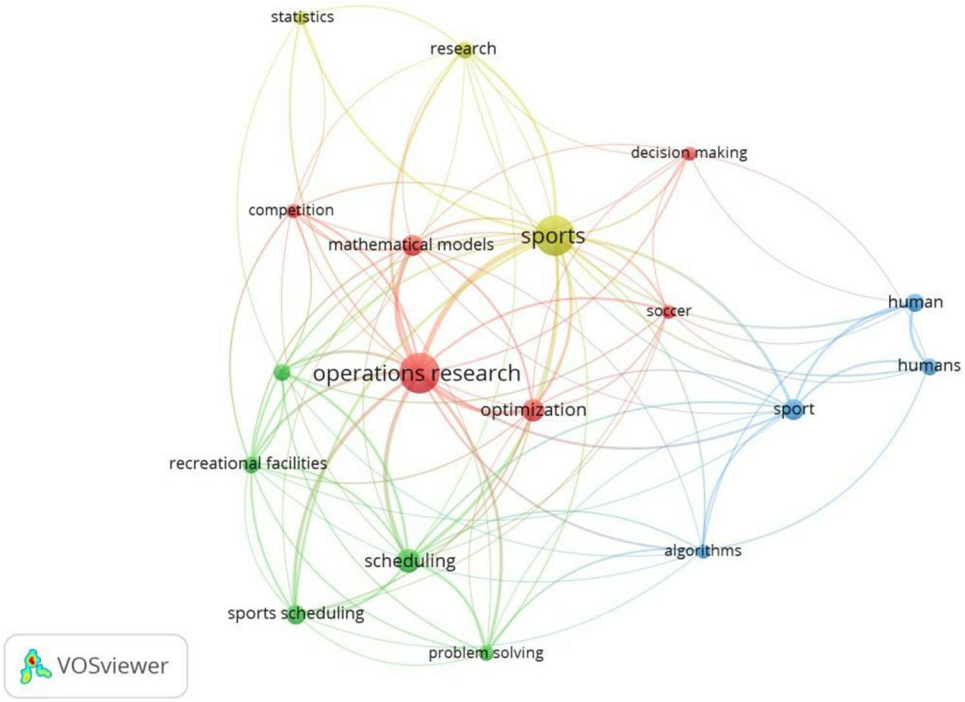


Figure 4: Co-occurrence of Keywords

Source: Vos Viewer – Authors' Work

V. DISCUSSIONS

The review has been able to meet its primary objective of understanding the publications in the field of sports analytics. The publication trends show an exponential increase in the publication, indicating increasing academic interest in the field. However, the field is still nascent, and therefore, so far, no review article has been published.

Based on the authors' findings, themes like mathematical models, computational methods, scheduling, and optimization are new and emerging topics over the past decade. However, a more in-depth study, post-COVID-19 has shown that big data analytics in sports management, game outcome prediction models by integrating data mining methods, intelligent processing, and analysis of images from real-time sports using the internet of things, risk prediction, fuzzy analytics, multimodal neural network models, etc. are the topics that authors have majorly focused on during the last two years. The bibliometric analysis allowed the authors to estimate the value of the area's influential authors, affiliated nations, resourceful journals, keywords, and interrelationships between works.

The article has provided comprehensive insight into the use of operations research in athletics while also shedding light on the dearth of research pertaining to this subject, as well as providing the first systematic review. Carrying out a thorough study on this subject would lead to much higher productivity for the parties that are impacted by this and would also assist in boosting the productivity of other organizations, such as sports teams. Some proposals for future research include including an ever-increasing number of teams in the relevant projects and using an increased number of mathematical models.

This bibliometric review would be extremely helpful in assisting academicians and students in determining areas of research. Additionally, it would help assist tournament organizers and team managers, amongst others, by enumerating various areas that they can use to optimize decision-making in sports further. This bibliometric review demonstrates not only how efficient OR is in thearea of sports but also why it needs to be used more often and widely by sports teams. In addition, considerable study ought to be undertaken on this subject for the authors to make more discoveries.

VI. CONCLUSIONS AND LIMITATIONS

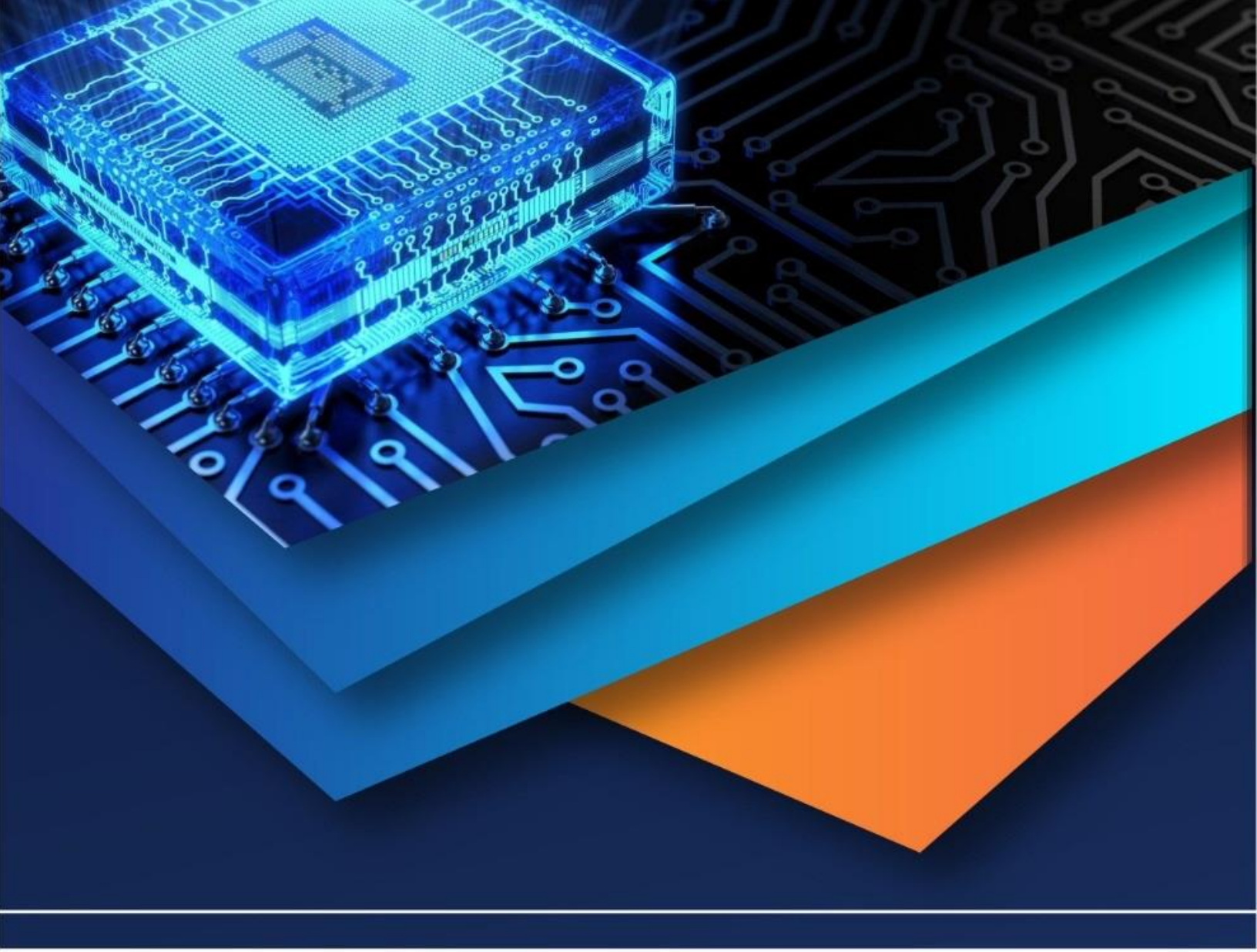
While it is true that operations research in sports analytics is employed in several events throughout the world, the authors' findings indicate that limited research has been done in the field between 2000and 2022. By carrying out a comprehensive analysis of 130 relevant papers from the Scopus database,the study presented a bibliometric review of sports analytics.

Overall, the authors find an individual growing focus on sports scheduling, operations research in sports, team selection strategies, and probabilistic models for team selection and game outcome predictions over the last decade, but have not been able to find any article that encompasses these topics in one paper. The lack of significant contributions made in the field was the most prominent research gap identified by the authors. The authors believe there is much scope for growth in the contributions to this field. The study advises that additional research be carried out through a combination of databases and papers published in other languages because the Scopus database, which served as the study's data source, only took into consideration documents written in English.

There is a lack of research articles that analyze the application of Operations Research and Analytics in sports for India and the types of sports practiced in India. During the first stage of data collection, several articles were related to sports medicine, and these had to be eliminated manually. This also significantly reduced the number of documents for analysis. Further, because the majority of the pertinent articles were centered on football, there was insufficient exposure to a variety of sports. Another limitation of this review was that the authors were unable to identify strong trends between author collaborations, i.e., the same authors co-authoring multiple papers. This is possibly a result of the number of papers published in this field being inadequate, leading to the article's sample size being reasonably low.

REFERENCES

[1] Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics,11(4),959–975. https://doi.org/https://doi.org/10.1016/j.joi.2017.08.007
[2] Ball, B. C., & Webster, D. B. (1977). Optimal Scheduling for Even-Numbered Team Athletic Conferences. A I I E Transactions, 9(2), 161–169. https://doi.org/10.1080/05695557708975138
[3] Chen, X., Lun, Y., Yan, J., Hao, T., & Weng, H. (2019). Discovering thematic change and evolution of utilizing social media for healthcare research. BMC Medical Informatics and Decision Making,19(2), 50. https://doi.org/10.1186/s12911-019-0757-4
[4] De Bakker, F. G. A., Groenewegen, P., & Den Hond, F. (2005). A Bibliometric Analysis of 30 Years of Research and Theory on Corporate Social Responsibility and Corporate Social Performance. Business & Society, 44(3), 283–317. https://doi.org/10.1177/0007650305278086
[5] EGGHE, L. (2006). An improvement of the H-index: The G-index. ISSI Newsletter, 2. Fry, M. J., & Ohlmann, J. W. (2012a). General Sports Applications. April 2016, 104–108.
[6] Fry, M. J., & Ohlmann, J. W. (2012b). Introduction to the special issue on analytics in sports, part II: Sports scheduling applications. Interfaces, 42(3), 229–231.https://doi.org/10.1287/inte.1120.0632
[7] Goyal, K., & Kumar, S. (2021a). Financial literacy: A systematic review and bibliometric analysis. International Journal of Consumer Studies, 45(1), 80–105. https://doi.org/10.1111/ijcs.12605
[8] Goyal, K., & Kumar, S. (2021b). Financial literacy: A systematic review and bibliometric analysis. International Journal of Consumer Studies, 45(1), 80–105.https://doi.org/https://doi.org/10.1111/ijcs.12605
[9] Hodge, D. R., & Lacasse, J. R. (2010). Evaluating Journal Quality: Is the H-Index a Better Measure Than Impact Factors? Research on Social Work Practice, 21(2), 222–230. https://doi.org/10.1177/1049731510369141
[10] Istoan, R., Manea, L. D., Plesa, L., & Tintisan, M. L. (2022). Increasing the sustainability of construction sector by developing new products based on biomass and renewable polymers - bibliometric analysis. IOP Conference Series: Materials Science and Engineering, 1251(1), 12005. https://doi.org/10.1088/1757-899x/1251/1/012005
[11] Le Sage, T., Bindel, A., Conway, P. P., Justham, L. M., Slawson, S. E., & West, A. A. (2011). Embedded programming and real-time signal processing of swimming strokes. Sports Engineering, 14(1), 1–14. https://doi.org/10.1007/s12283-011-0070-7
[12] Lee, I.-S., Lee, H., Chen, Y.-H., & Chae, Y. (2020). Bibliometric Analysis of Research Assessing the Use of Acupuncture for Pain Treatment Over the Past 20 Years. Journal of Pain Research, Volume 13, 367–376. https://doi.org/10.2147/JPR.S235047
[13] Lewis, A. J. (2005). Towards fairer measures of player performance in one-day cricket. Journal of the Operational Research Society, 56(7), 804–815. https://doi.org/10.1057/palgrave.jors.2601876
[14] Partovi, F. Y., & Corredoira, R. A. (2002). Quality function deployment for the good of soccer. European Journal of Operational Research, 137(3), 642–656. https://doi.org/10.1016/S0377- 2217(01)00072-8
[15] Russell, R. A., & Urban, T. L. (2006). A constraint programming approach to the multiple-venue, sport-scheduling problem. Computers and Operations Research, 33(7), 1895–1906. https://doi.org/10.1016/j.cor.2004.09.029
[16] Saaty, T. L. (1959). Mathematical Methods of Operations Research.Swaddling, J. (1999). The Ancient Olympic Games.
[17] Tsay, M. Y. (2009). Citation analysis of Ted Nelson's works and his influence on hypertext concept.Scientometrics, 79(3), 451–472. https://doi.org/10.1007/s11192-008-1641-7
[18] Urban, T. L., & Russell, R. A. (2003). Scheduling sports competitions on multiple venues. European Journal of Operational Research, 148(2), 302–311. https://doi.org/10.1016/S0377- 2217(02)00686-0
[19] Van Bulck, D., Goossens, D., Schönberger, J., & Guajardo, M. (2020). RobinX: A three-field classification and unified data format for round-robin sports timetabling. European Journal of Operational Research, 280(2), 568–580. https://doi.org/10.1016/j.ejor.2019.07.023



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)