



Special Issue

A Fuzzy Framework to Evaluate Players' Performance in Handball

Francisco P. Romero^{1,*}, Eusebio Angulo², Jesús Serrano-Guerrero¹, José A. Olivas¹

¹Department of Information Systems and Technologies, University of Castilla La Mancha, Ciudad Real, Castilla-La Mancha, Spain

²Department of Mathematics, University of Castilla La Mancha, Ciudad Real, Castilla-La Mancha, Spain

ARTICLE INFO

Article History

Received 22 Nov 2019

Accepted 13 Apr 2020

Keywords

Group decision-making
Handball player evaluation
Rating system
Aggregation operators
Linguistic model

ABSTRACT

The evaluation of the players' performance in sports teams is commonly based on the opinion of experts who do not always agree on the importance of the chosen indicators. This paper presents a novel approach based on fuzzy multi-criteria group decision-making tools for selecting those criteria that best represent the handball player's performance in a match and for setting their relevance weights. Our approach consists of a fuzzy model to aggregate expert judgments. This methodology overcomes some drawbacks of classical systems, including the definition of the relevance of each criteria using linguistic labels. A preliminary evaluation analyzes handball players' performance indicators and their application to a short tournament. Considering the obtained results, we can conclude that the proposal is relevant and provides useful insights regarding player performance in different matches. The proposed methodology has also been compared with a basic plus-minus rating methodology. This comparison illustrates the feasibility of our approach. Results suggest that plus-minus rating is not the best solution to represent the performance of specialized players who only play when their team attack or defense. Our approach demonstrates being more appropriate for sports such as handball because it includes the valuation of a full set of positive actions in defense and attack.

© 2020 The Authors. Published by Atlantis Press SARL.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. INTRODUCTION

We are currently witnessing an increasing interest in evaluating the performance of sports teams. Any attempt to assess the process, as well as the success of a sports team, requires some objective method for assessing the players. Sports statistics offer a reasonable and recognizable approach to this. Nevertheless, the problem lies in deciding which statistics should be used. There are popular measures such as goals or assists, but they fail to capture some critical indicators. Therefore, one of the most important questions for today's coaches is to know both the variables that influence player performance and the relative contribution of these variables.

Handball is an important popular sport in Europe. The first German League (Handball Bundesliga) has an annual budget of approximately 80 million Euros and about 750.000 people practice this sport in multiple leagues [1]. Furthermore, top teams in France have a yearly budget of million Euros [2]. Today, we are witnessing a growing interest in evaluating the performance of handball teams, which involves establishing an objective method of assessing players. These assessments are often based on the opinion of various experts who do not always agree with the importance of the selected indicators. In the worst case, evaluations are merely subjective. The aim is to be able to make an objective assessment of the players based on their statistics on the essential actions in each match.

In this way, handball is far from the advances and research existing in other sports; the existing approaches [3] to represent the effectiveness of a handball team or a single handball player is now insufficient for evaluating the performance of handball player in a match.

On the other hand, plus-minus (PM) methodologies [4] are applied to different collective sports. Using this top-down rating strategies, the whole team in court is rewarded when positive action is achieved, whether they participate in the action or not. On the other hand, bottom-up methodologies the ratings of the players that do not participate in the action are not affected. A PM methodology is complicated to apply to handball because of its characteristics and rules. For example, some players are defensive or offensive specialists. Statistically, a defensive specialist will concede more goals and adverse actions than a player who always attacks. Therefore the evaluation would generally be negative. The opposite would happen with players who only attack and are usually on the court when their teams score goals. Consequently, there is no tool for understanding the reason player performances vary from one game to another, nor for knowing which variables are affecting performance.

Multi-criteria group decision-making (MCGDM) is very popular in the recent literature, mainly due to its usefulness in a wide range of solutions and fields of application [5], for example, engineering [6], sustainability [7], and biology [8]. Its primary purpose is to allow a set of experts to rank a set of elements according to a specific set of criteria. A performance assessment problem needs expert judgments, and thus, the complexity of this decision-making

* Corresponding author. Email: franciscop.romero@uclm.es

problem increases in the form of subjectivity and vagueness. Due to the shortcomings created by the subjectivity of human judgments and the vagueness of data, fuzzy set theory can be used in performance-assessment-related decision-making processes [9]. The use of a fuzzy approach provides approximate reasoning methods that can handle the inherent subjectivity.

On the other hand, handball, like other sports, constitutes a dynamic environment; in this context, it is quite common that not all coaches have experience corresponding to all the criteria. The experts have the means to provide each of their preferences. Criteria weights can be modified at any time during the process. Therefore, it would be desirable to apply a MCGDM method that allows for the consideration of all aspects.

This paper addresses the problem of identifying variables that may affect the performance of a handball player from one game to another, with a group of experts. Our purpose is to allow a set of experts (handball coaches) to rank a set of performance criteria according to their preferences, and in turn, obtain a ranking of players according to their performance. Thus, coaches can use an organized framework to make rational decisions about their players.

The proposal is based on a fuzzy model for analyzing the opinion of several experts and creating definitions by aggregating the specifications of the linguistic labels given by them. In order to verify the feasibility of the proposal, we present a preliminary experimental evaluation of complete application of the approach to a very short tournament: the 2019 Spanish Women's Promotion Playoffs.¹ Our results provide relevant insights into the features that may distinguish an outstanding performance in a very short tournament from the rest of the players. Consequently, the main contributions of this article are the following:

- Designing of a fuzzy-based mechanism to compute expert judgments via linguistic values instead of numerical values.
- Implementing a fuzzy-based aggregation mechanism to compute evaluation of handball players, depending on individual facet linguistic opinions that adapt to the specific problem.
- Putting forward a real-world application and evaluation for the sport of handball.

The novelty of our work lies in that we handle two difficult issues not previously addressed in the literature: (i) to provide a framework for a set of experts, for example, the team coaching staff, to evaluate the performance of the players in each match. (ii) The proposed framework allows the creation of an objective metric which compares and takes into account all the important actions (positive and negative), establishing a player rating at each moment of the match.

This contribution is set out as follows: Section 2 presents the foundations of our proposal and briefly reviews the most relevant literature on MCGDM and player evaluation in sport teams. Section 3 describes the consensus processes based on fuzzy aggregation functions and the details the proposal put forward in this article in full.

Section 4 presents an illustrative example of the application of this method. Sections 5 and 6 explain the study case, showing and discussing the corresponding results. Finally, Section 7 sets out some conclusions and future research work.

2. RELATED WORK

In this section, the work related to the issues of evaluating the performance of handball players is presented. In Section 2.1, the associated techniques of MCGDM methods are reported. Section 2.2 reviews the literature related to player evaluation methods.

2.1. Multicriteria Group Decision-Making

MCGDM methods can be considered to be an extension of traditional group decision-making methods [2]. In a traditional group decision-making method, experts need to rank a set of alternatives based on their preferences. On the other hand, in MCGDM methods experts are asked to rank the alternatives considering a pre-specified set of criteria. Introducing criteria to group decision-making processes helps experts carry out the decision in a less subjective way. Moreover, the criteria will not have the same importance. Therefore, it is possible to associate a weighting schema that indicates the weight that should be assigned to each criteria when calculating the final results.

The choice and weighting of the criteria derive from a dynamic procedure where it is necessary to manage situations in which alternatives are added or removed throughout the process [10]. Some approaches propose the use of a multi-granular fuzzy linguistic modeling process in order to allow the modification of experts, criteria, and alternatives at any time during the process [11].

One of the major issues restraining the ability to determine the right method for choosing the best player in a match is the absence of the ability to handle uncertain and inadequate information, which mostly happens in match conditions. In the practical problems of player performance evaluation, several criteria are affected by uncertainty. Experts' judgments are usually uncertain and difficult to measure by exact numerical values, so the application of fuzzy logic has become the best alternative for solving the shortcomings characterized by vagueness and imprecision. Recently, several studies have applied the typical MCGDM methods to a range of fuzzy environments [12].

At the end of the process, an aggregation method has to be applied to evaluate the player performance. For this purpose, different aggregation operators can be found in the literature [13]. Fuzzy logic aggregations operators are useful for managing uncertainty, vagueness, and subjectivity. These operators go beyond combining numerical values on [0, 1] and provide mechanisms for combining fuzzy information components [14]. These approaches have been applied with success in a wide range of applications [15–17].

2.2. Player Evaluation Methods

The evaluation of a player's performance is one of the most critical aspects of the application of advanced analytics and statistics in sport.

¹ <https://bit.ly/2KvZqbr>

It seeks to achieve a more objective view of productivity, efficiency, effectiveness, and value of the components of the game at the individual level. In addition, this aspect can play an important role in making better decisions both within a club and at the level of the team.

In many sports, there is a great deal of literature related to the evaluation of player performance [18]. Various approaches have been proposed to rate players for specific tasks, for example, a Bayesian regression model to obtain an estimation of the number of goals scored of any soccer player [19]; and a soccer player ranking system based completely on the value of the passes completed [20]. The final value is computed based on the relationship between pass locations and shot opportunities generated. Data-driven frameworks that offers a multi-dimensional and role-aware evaluation of the performance of soccer players [21] are useful tools to address some limitations of the valuation approaches, for example, the mono-dimensional focus, the specificity of each player's role on the field, and the lack of a gold standard.

The player ratings can be labeled as bottom-up and top-down. In bottom-up ratings, a player's scoring increases if he performs an action that is associated with a positive result for the team. Player's scoring is not affected by players who did not participate in the action. On the other hand, in top-down ratings, the main idea is to reward the team as a whole for each positive action. The credit for this performance is distributed among individual players involved in the match, independently of the actions they performed.

One of the most popular top-down productivity metrics is the PM rating model [22]. The PM rating and its variations have been used extensively in ice-hockey [23], basketball [24], and soccer [25]. However, there are some doubts about its usefulness when it comes to comparing individual players, as the rating is established by many factors beyond the control of the player being evaluated. Moreover, in sports where player-specific functions (NFL or handball) stand out, the effectiveness of this metric is limited and requires attention beyond this study. For this reason, we apply a bottom-up rating method instead of a top-down rating like PM.

Handball experts such as coaches and sports journalists have long estimated player performance. Although the academic and professional attention to quantitative analysis of data has exponentially grown over the past years, there is no much work concerning evaluating a handball player in a game-by-game scenario [26].

According to several studies [27], variables like shots saved by the opposing goalkeeper, technical fouls, steals, and the goalkeeper's saves are the key indicators of team performance.

In this context, we aim to provide a comprehensive framework for assessing handball player performance on a match-by-match basis.

3. A FUZZY FRAMEWORK TO EVALUATE PERFORMANCE IN TEAM SPORTS

In order to perform the necessary computation, the following steps are carried out (Figure 1):

1. *Initial setting*: The initial parameters of the MCGDM framework are established (experts, criteria, linguistic labels,...).
 2. *Expert opinions*: Each expert, depending on his or her own experience, must express an opinion regarding the different criteria. They can do this by using the linguistic label set they prefer.
 3. *Aggregation*: First, the opinions that have been provided by the experts are standardized and then, the aggregation method is applied to generate the fuzzy result that represents the overall opinion of all the experts.
 4. *Information gathering*: The necessary information on the statistics of each match is compiled either from the official statistics of the game or through its direct visualization. The proposed approach is applied to obtain a ranking of players.

The following subsections will explain in detail the way these steps are carried out.

3.1. Initial Settings

Before starting the process, the initial set of parameters for carrying out the MCGDM

- *Initial set of experts:* The initial set of experts who will participate in the discussion is established.

$$E \equiv \{e_1, e_2, \dots, e_n\}$$

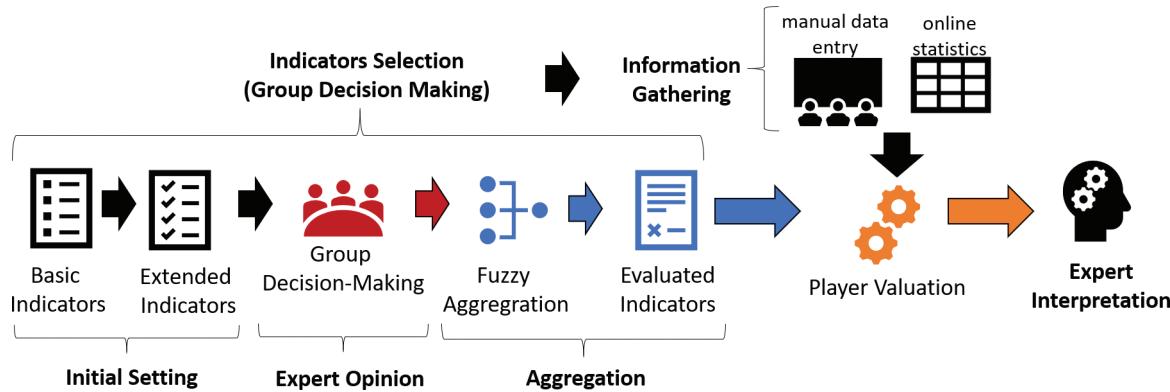


Figure 1 | Overview of the proposal.

It is possible to assign weights to the experts. Usually, experts that have more experience are assigned higher weighting values.

$$W_e = \{w_{e_1}, \dots, w_{e_n}\}$$

- *Initial set of criteria:* The initial set of criteria values is defined.

$$C = \{c_1, c_2, \dots, c_l\}$$

that refer to the performance of the set of players X based on the preferences they provided.

$$X = \{x_1, x_2, \dots, x_m\}$$

Criteria can make a positive or negative contribution to performance. Thus, a sign vector is associated with the set of criteria:

$$\sigma_c = \{\sigma_{c_1}, \sigma_{c_2}, \dots, \sigma_{c_l}\}$$

where $\sigma_{c_i} \in \{+, -\}$

- *Initial linguistic label set:* Experts can express their views on the criteria using one of the pre-defined sets of labels, each with a different discriminatory power.

$$S = \{S^3, S^5, S^7\}$$

where the superindex represents the number of available labels. For instance, the experts can use the following linguistic label set for providing preferences:

$$S^5 = \{Very Low, Low, Medium, High, Very High\}$$

Once all these parameters have been established, experts can start the selection process.

In traditional group decision-making problems, experts are required to rank a set of alternatives. However, in this study, the issue to solve consists of the difference in importance of each criterion. Then, experts must evaluate the set of criteria. Therefore, it is necessary to obtain a weighting vector,

$$W_c = \{w_{c_1}, w_{c_2}, \dots, w_{c_l}\}$$

which indicates the weight that should be assigned to each criterion while calculating the evaluation of a player x_i in one match.

3.2. Expert Opinion

Every expert, depending on their own experience, must express an opinion regarding the selected criteria. In this process, experts have the following degrees of flexibility:

- Each expert chooses the linguistic label set (S^n) that they want to use.
- Each expert selects the criteria that they want to provide evaluations on.
- Each expert can add new criteria that were not considered at the beginning of the process.

Once all the experts have provided the required information, if the expert applied different sets of linguistic labels, the information is standardized (represented by the same scale) to be used later in the aggregation processes. For this purpose, multi-granular fuzzy linguistic modeling methods [28] can be applied. Our study adopts a model based on the concept of 2-tuple fuzzy linguistic representation [29]. This approach has produced excellent results in practical applications, especially in terms of efficiency [30].

A high level of consensus concerning the expert opinions is mandatory to assure the viability of the aggregation process, that is, we require that the expert estimates have a common intersection at some α -level cut. For this purpose, we compute a global consensus degree that is compared against a consensus threshold provided by the set of experts prior to the application of the consensus. If this degree is greater than or equal to this consensus threshold, then the consensus-reaching process is considered successful and hence it should end. Otherwise, the experts need to discuss it further.

Once the expert opinions are validated, a fuzzy aggregation process is applied to summarize the results as explained in the following subsection.

3.3. Aggregation

We use fuzzy techniques for finding a compact and synthesized representation of expert opinions. The aggregation process is carried out in the following steps:

1. Each expert defines the linguistic vector $V^e = \{v_1^e, v_2^e, \dots, v_l^e\}$ indicating the importance that they give to each of the chosen criteria. The union of all these linguistic vectors generates the global evaluation matrix containing all the valuations from all the experts.
2. *Individual Aggregation:* The linguistic vector $V^e = \{v_1^e, v_2^e, \dots, v_l^e\}$ representing the expert evaluations of all indicators S_e is aggregated. The weights assigned to the experts W_e are used. This is done through the following steps:
 - (a) *Baseline:* A fuzzy definition of each label is used (balanced or unbalanced)
 - (b) *Aggregated Valuation Matrix:* We obtain the weight of each label in each criterion according to the occurrence frequency and the expert weighting w_{e_n} . The weighted mean operator is used with this aim.
 - (c) *Fuzzification:* We use a fuzzy operator in order to obtain a fuzzy set for each criterion, for example, the minimum t-norm, which truncates each baseline membership function according to the previous computed weighting of each label.

Then, the aggregated output will be the fuzzy set representing how relevant each indicator would be in measuring player performance.

3. *Defuzzification:* The center of area (COA) defuzzification method computes the center of mass of the membership function of the fuzzy set (the centroid). The COA method

maintains the underlying semantic ranking relation within the set of linguistic labels, that is, given two linguistic labels $s_i, s_j \in S$ such that $s_i < s_j$ then $u_{COA}(s_i) < u_{COA}(s_j)$. Thus, the centroid of a type-1 fuzzy set A in a continuous domain X is calculated as follows:

$$U_{COA}(A) = \frac{\int_x x * \mu_A(x) dx}{\int_x \mu_A(x) dx}$$

and it will be the method adopted here to obtain a numeric output.

4. *Stratification*: In order to reduce the sparsity of the weighting distribution, we carry out a stratification process. The purpose is to group criteria with similar evaluations using clustering. As a result, we obtain the criteria segmented into groups according to this importance.
5. *Collective Aggregation*: The aggregation process obtains the final solution according to the opinions given by the experts. These results allow us to define a weighting scheme to evaluate the performance in accordance with this set of criteria. For this purpose, we aggregate this value for each group of indicators and normalize the final result.

3.4. Information Gathering

For each player (x_i) within each match, the values that each player obtains, associated with each rating criterion ($[x_i, c_i]$), are collected. Subsequently, using the weighting scheme previously obtained, the valuation of each player and the rankings of the players are calculated.

4. ILLUSTRATIVE EXAMPLE

In this section, in order to clarify the way this approach works, a simple example is shown. Imagine that a set of five experts $E = \{e_1, e_2, e_3, e_4, e_5\}$ need to rank six performance criteria:

$$C = \{c_1, c_2, c_3, c_4, c_5, c_6\}$$

In this example, every criterion contributes in a positive way to the final valuation.

Each expert can decide individually on the level of importance that should be given to each criteria value. For this purpose, they used the following linguistic label set:

$$S^3 = \{s_1^3 = H - High, s_2^3 = M - Medium, s_3^3 = L - Low\}$$

The valuations provided by the experts are specified below (see Table 1):

In order to carry out the criteria ranking, the aggregation matrix must be calculated. This is done by aggregating the evaluations of the experts. Table 2 shows the weights (W_e) and the normalized weights (W_e^N) assigned to each expert according to their level of experience. The results obtained are shown in Table 3.

Using the fuzzy definition of the linguistic labels (see Figure 2), the fuzzification of each of the criteria evaluation can be made. The fuzzification results are shown in Figure 3.

Table 1 | Example of global valuation matrix.

	e_1	e_2	e_3	e_4	e_5
c_1	H	H	M	H	H
c_2	M	L	L	M	L
c_3	H	H	H	H	H
c_4	M	L	H	L	L
c_5	M	L	H	L	M
c_6	L	L	L	M	L

Table 2 | Expert weighting scheme.

	W_e	W_e^N
e_1	1.00	0.29
e_2	0.75	0.21
e_3	0.50	0.14
e_4	0.25	0.07
e_5	1.00	0.29

Table 3 | Example of aggregated valuation matrix.

	L - Low	M - Medium	H - High
c_1	0.00	0.14	0.86
c_2	0.64	0.36	0.00
c_3	0.00	0.00	1.00
c_4	0.57	0.29	0.20
c_5	0.28	0.58	0.20
c_6	0.93	0.07	0.00

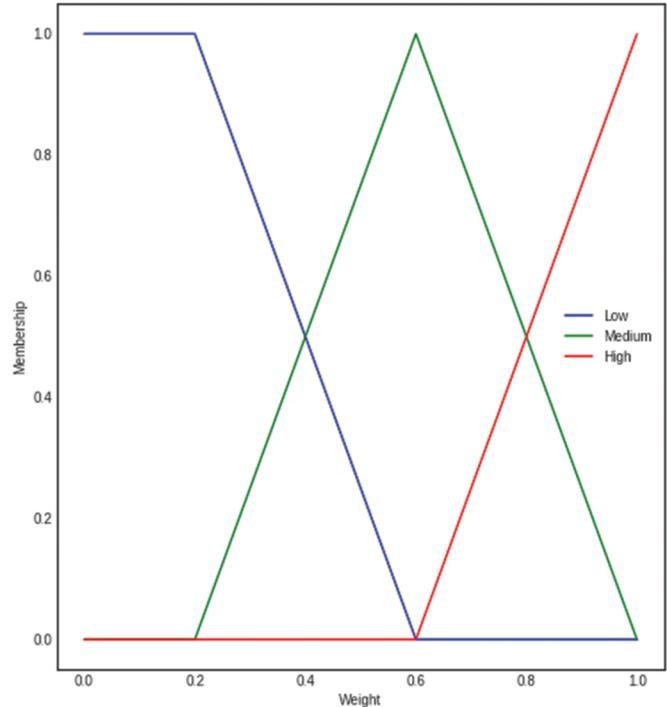


Figure 2 | Linguistic labels definitions as fuzzy sets.

After a defuzzification process, we obtain a weighting for each criterion represented as a weight vector:

$$W = \{0.80, 0.36, 0.86, 0.38, 0.50, 0.24\}$$

In order to reduce the sparsity of the weightings obtained, we carry out a clustering/stratification process (see Figure 4).

Consequently, different groups of indicators

$$C_{groups} = \{c_{high}, c_{med}, c_{low}\}$$

with their corresponding weightings are found:

$$W_{groups} = \{w_{high} = 0.83, w_{med} = 0.44, w_{low} = 0.30\}$$

Thus, the weighting scheme obtained can be used to evaluate and rank a set of five players

$$X = \{x_1, x_2, x_3, x_4, x_5\}$$

according to the values achieved for each criterion (see Table 4).

After the information gathering, the aggregation operator (linear combination of each element according to the weight of the criteria) is applied, and the ranking results are shown in Table 5.

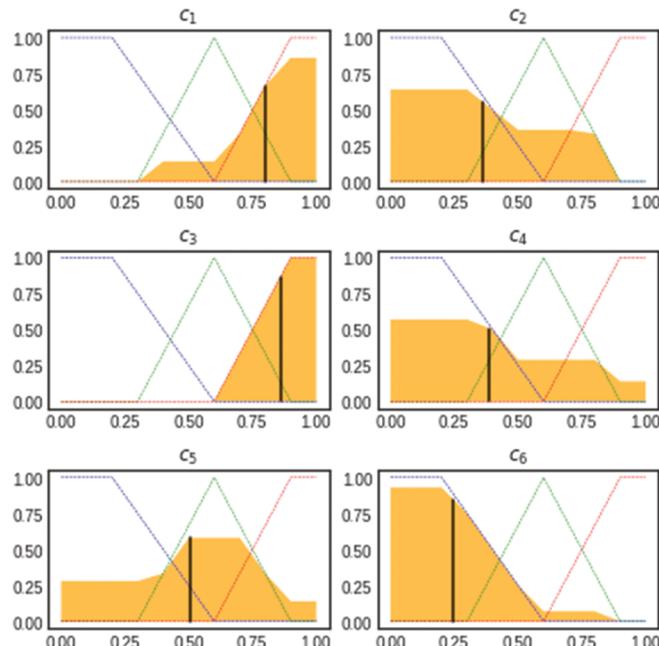


Figure 3 | Fuzzification of the criteria valuation.

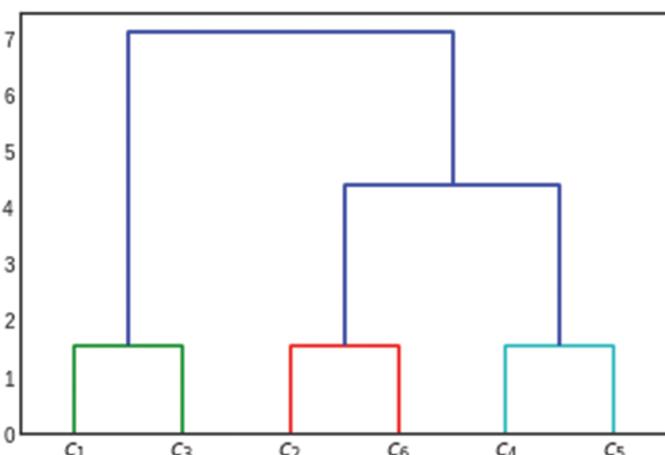


Figure 4 | Clustering results.

5. HANDBALL PLAYER VALUATION PROCESS

This section presents the application of the previous proposal to obtain a set of indicators for evaluating handball player performance, and its importance according to a set of experts.

5.1. Initial Setting

A collection of basic criteria is selected as a baseline of the decision process to determine the indicators that must be considered to model the player performance. Table 6 shows those indicators that can be obtained from online statistics of handball games.

This basic set is extended using the criteria of a national handball coach, and the indicators used for the evaluation of junior players by the Spanish Handball Federation. This extension achieves a higher level of granularity than the Basic Indicator Set, for example, missed shots are broken down into several subcategories to better capture the cause of the error.

Table 7 shows the criteria collection for this extended baseline.

5.2. Expert Opinion

The obtained set of criteria is given to the 15 experts (Spanish handball coaches) who are weighted according to their coach experience. Fuzzy linguistic terms, such as most important, important,

Table 4 | Example of player results.

	High		Medium		Low	
	c_1	c_3	c_4	c_5	c_2	c_6
x_1	3	2	0	3	4	5
x_2	2	1	2	3	7	4
x_3	2	3	2	6	5	0
x_4	6	0	7	8	2	1
x_5	1	3	1	2	3	1

Table 5 | Players final score.

Final	Score
x_3	14.93
x_1	14.10
x_4	10.69
x_2	5.79
x_5	4.90

Table 6 | Basic indicators set.

Metric	Description
7MGoals	7-metre Goals
7mMS	7-metre Missed Shots
6MGoals	6-metre Goals
6mMS	6-metre Missed Shots
9MGoals	9-metre Goals
9mMS	9-metre Missed Shots
FBGoals	Fast Break Goals
FBMS	Fast Break Missed Shots
Additionals	Positive actions (steals, etc.)
Punishments	Exclusions, mistakes, etc.

and normal, were used to determine the experts' influence weightings, which were represented by fuzzy membership functions. For example, a National Handball coach was assigned the most important influence, while a Level 1 coach was assigned an important influence weighting [31].

Every expert, depending on his own experiences, must express an opinion regarding the chosen criterion. In this case, each criterion is evaluated by linguistic labels belonging to one of the following predefined label sets.

$$\begin{aligned} S^3 &= \{s_1^3 = H, s_2^3 = M, s_3^3 = L\} \\ S^5 &= \{s_1^5 = VH, s_2^5 = H, S_3^5 = M, s_4^5 = L, S_5^5 = VL\} \end{aligned}$$

These labels sets are represented by using Trapezoidal and Triangular Fuzzy Numbers as can be seen in Figure 5.

Once the expert opinions are known, the fuzzy techniques explained below are applied to aggregate the set of linguistic labels representing the evaluations of all criteria.

5.3. Aggregation

Then, the fuzzy aggregation process is applied. Table 8 shows the results of the importance of each chosen indicator.

These results allow us to define a weighting scheme to evaluate the performance of a handball player in a specific match (see Eq. 1).

Table 7 | Extended indicators set.

Positive	Negative
9mGoal	9mSaved, 9mPost, 9mOut, 9mBlock
6mGoal	6mSaved, 6mPost, 6mOut
FBDGoal, 7mGoal	FBDmissed, 7mMissed
Assist, Steal, Block	Turnover Handling
Received 7m Foul	Committed 7m Foul
Defensive Action	Unc. Defensive Withdrawal
Provoke 2min	Defensive Mistake
	Punishments

Table 8 | Final results.

Rating	Criteria
Positive	
High	9mGoal, 6mGoal, FBDGoal, 7mGoal
Medium	Assist, Steal, Block, Provoke 2min
Low	Received 7m Foul, Defensive Action
Negative	
High	9mOut, 9mBlock, 6mOut, FBDmissed, 7mMissed, Punishments
Medium	9mSaved, 6mSaved, Turnover Handling, Defensive Mistake
Low	Committed Penalty Foul 9mPost, 6mPost, Uncomplete Defensive Withdrawal

$$\begin{aligned} P_{score} = & +0.95 * \{\text{High}(+)\} \\ & -0.80 * \{\text{High}(-)\} \\ & +0.85 * \{\text{Medium}(+)\} \\ & -0.70 * \{\text{Medium}(-)\} \\ & +0.65 * \{\text{Low}(+)\} \\ & -0.60 * \{\text{Low}(-)\} \end{aligned} \quad (1)$$

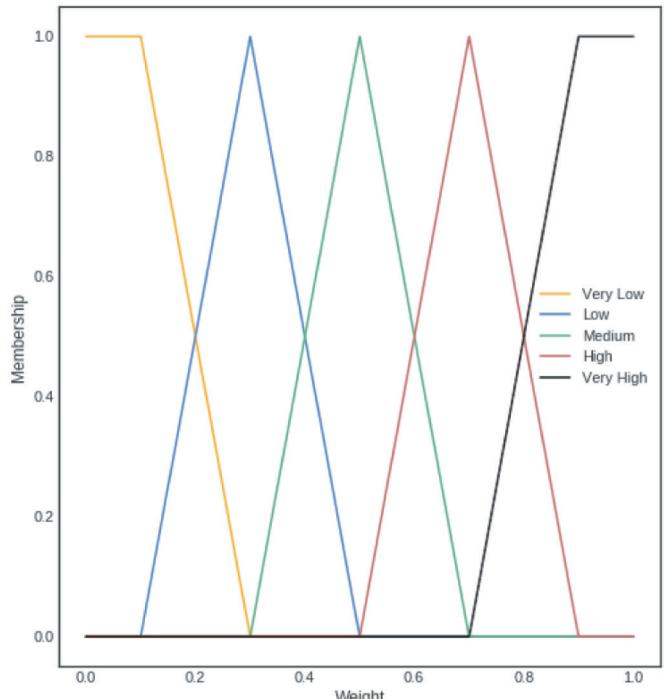
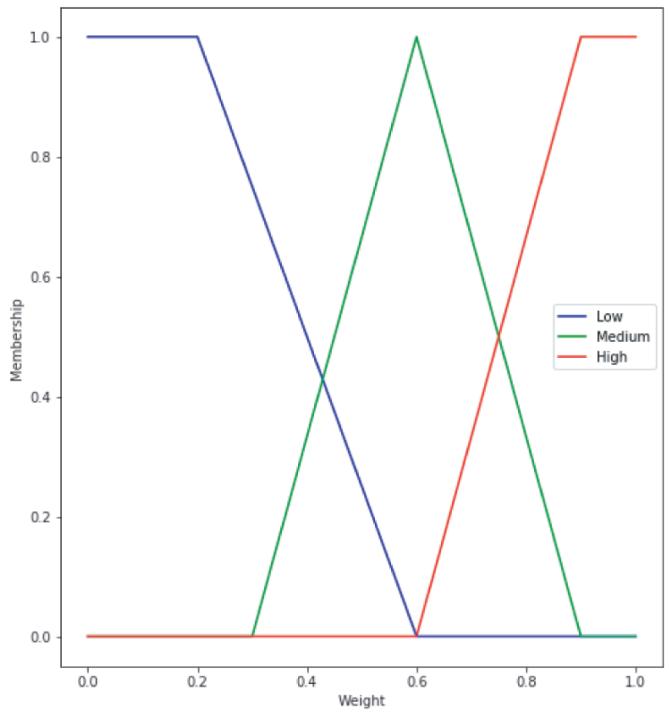


Figure 5 | Linguistic labels definitions as fuzzy sets.

Table 9 | Top players complete ratings (most valuable players in bold).

Player No.	Team	Player Rating						Acc.
		Match1	Match2	Match3	Match4	Match5	Match6	
70	Tenerife	–	4.7	–	1.0	–	–0.1	5.60
13	Tenerife	–	4.1	–	7.8	–	4.8	16.70
49	Tenerife	–	5.3	–	–1.2	–	10.2	14.30
8	Tenerife	–	–0.7	–	6.0	–	6.3	11.60
20	Tenerife	–	1.2	–	–0.5	–	7.1	7.8
8	Bolaños	–	1.8	2.8	–	0.9	–	5.5
97	Bolaños	–	–0.9	5.5	–	12.1	–	16.7
15	Bolaños	–	1.4	3.7	–	3.3	–	8.4
6	Bolaños	–	0.6	2.9	–	9.7	–	13.2
13	Bolaños	–	–1.6	4.0	–	6.9	–	9.3
4	S.Quirze	2.3	–	3.7	–	–	1.0	7.00
36	S.Quirze	2.2	–	–2.9	–	–	–0.1	–0.8
55	S.Quirze	4.5	–	3.5	–	–	–0.6	7.4
34	S.Quirze	0.2	–	4.6	–	–	0.0	4.8
2	La Rioja	2.9	–	–	9.2	17.2	–	29.3
23	La Rioja	3.8	–	–	4.0	3.1	–	10.9
89	La Rioja	1.8	–	–	7.0	5.3	–	14.1
44	La Rioja	–2.0	–	–	4.0	0.9	–	2.9

where the sign of each group is chosen according to the characteristics of the criteria.

6. EXPERIMENTS

A case study was carried out with real data. Data was collected from a very short tournament: the 2019 Spanish Women's Promotion Playoff to the Liga Guerreras Iberdrola (the top division). This was a four-team round robin tournament (data was collected for the 6 matches) in which the teams (Logroño Sporting La Rioja, Handbol Sant Quirze, Salud Tenerife, and Vino Doña Berenguela Bolaños), qualified from prior matches, fought for only one place that gave a place for the next season in the top category of Spanish women's handball.²

The statistical information, comprising the full set of indicators shown in Table 8, was gathered manually by direct viewing of matches. Subsequently, the aggregation formula presented in Eq. (1) has been used to obtain a valuation for each player in each game as can be seen in Table 9. Each "Player of the Match" is highlighted in bold.

The evaluation model is adjusted for all players. The results obtained with this model are compared with the best player of each match, as chosen by the coaches and sports journalists. Thus, we can check whether the results of the evaluation model are reasonable.

In the first match, the best player was *player No. 1* of Handbol Sant Quirze, who plays as goalkeeper a position which has no general valuation model, whereas the model proposed gave the best player *player No. 55* of Handbol Sant Quirze. In the second match, the best player of the match was *player No. 70* of Salud Tenerife, which is the player with the second highest valuation in our model.

Table 9 shows that in all other matches the choice of best player coincides fully with the highest evaluation provided by this methodology. In the third game, the best player was *player No. 97*

of Vino Doña Berenguela Bolaños; in the fourth and fifth, the best player was *player No. 2* of Logroño Sporting La Rioja, and in the sixth, *player No. 49* of Salud Tenerife.

To conclude, the valuations of the matches reflect what happened in the games. In four out of six games, the chosen "Player of the Match" by the coaches and the sports journalists was the one with most points in the valuation system. In the matches where the predicted result did not coincide, the chosen player is a goalkeeper in the first match, a position for which we do not have a model. The player chosen in the second match had the second most points.

6.1. Basic Plus-Minus Comparison

In this section, the evaluation of the handball match is carried out with a basic PM rating, and the results obtained are compared with the proposed approach rating.

The main idea is to apply the basic PM rating method [22] in handball: (i) Consider a player from a handball team and a given match. (ii) Count the number of goals scored by the player's team while the player is in the game, and then subtract the number of points scored by the opposing team during the same time intervals. (iii) The resulting number will here after be referred to as the basic PM rating of the player.

Table 10 shows the results using both rating processes in a match from the second division of handball in 2019, between Handball Málaga and the Handball Pozuelo de Calatrava.³ The table is simplified by showing only the two best players and the two worst of each team according to both methodologies.

The best player according to the basic PM rating was *player No. 21* of Handball Pozuelo de Calatrava and according to our approach, it was *player No. 4* of Handball Málaga, who was chosen as Most Valuable Player. The worst player according to basic PM rating was *player No. 14* of Handball Málaga, and according to our approach, it was *player No. 16* of Handball Pozuelo de Calatrava.

² See https://www.youtube.com/playlist?list=PLWU4nHhhA7wGS6km9Qvj7msB_6r4wnakU

³ See <https://www.youtube.com/watch?v=nnc3zLP5PH0>

Table 10 | Basic plus/minus comparison (in bold best and worst players).

Player No.	Team	Basic Plus-Minus Rating	Proposed Approach Rating
Top Player Rating			
21	Pozuelo	15	1
3	Pozuelo	5	3.5
17	Pozuelo	-1	3.5
4	Málaga	14	9
5	Málaga	13	4
7	Málaga	12	8.5
Worst Player Rating			
22	Pozuelo	-11	-0.5
20	Pozuelo	-9	-1.5
16	Pozuelo	-3	-6
13	Pozuelo	-2	-2
14	Málaga	-12	1
65	Málaga	-9	4
17	Málaga	-4	-2
11	Málaga	3	-1

These results highlight the main problem of PM in handball: the evaluation of players who are defensive specialists (players who only defend) or offensive specialists (players who only attack). Defensive players have great difficulty performing positive actions, and will lose points as they only play when their team will probably concede goals, except goals in fast breaks. On the other hand, the player who only attacks will have positive ratings unless the opponent team scores a goal with a fast break.

The results obtained show that the basic PM rating generates false MVP's. This is the case with *player No. 21* offensive specialist of Handball Pozuelo de Calatrava, who only participated in the attacking actions of her team but was not decisive. The chosen MVP in this match was *player No. 4* of Handball Málaga, who was found to be the most valuable player with our approach.

In the same way, the defensive players get low PM ratings, for example, *players Nos. 20 and 22* of Pozuelo de Calatrava and *players Nos. 14 and 65* of Handball Málaga. They obtain the worst values according to the PM rating, being *player No. 14* of Handball Málaga, who only defended and accumulated fewer errors and negative actions than *player No. 16* of Handball Pozuelo de Calatrava, who had the minimum score with the proposed approach rating in this study.

In the case of handball, where changes are normally made without stopping the clock, the basic PM may offer false high and low values that do not reflect the outcome of the match. For example, specialized players who only play when it is more likely they will score goals (positive actions) or concede them (negative actions). Moreover, the basic PM rating system, as opposed to the approach proposed here, does not value many positive actions in defense and attack, nor the actions and errors of each player.

7. CONCLUSIONS AND FUTURE WORK

This study takes a fuzzy approach to choosing the indicators that best represent the contribution of the handball players to the results. The key idea is to combine the numerous evaluations provided by a group of experts into a single one. This methodology overcomes some shortcomings of classical systems, including the definition of the relevance of each indicator using linguistic labels.

The novelty of this work is two-fold: (i) we provide a framework to technical coaches to evaluate teams with the same method and (ii) we provide an objective measure of the performance of handball players.

Preliminary experiments have been carried out to verify the feasibility of the proposal. A very short tournament (2019 Spanish Women's Promotion Playoff) was intensely studied to extract and analyze the performance indicators. Considering our results, we can conclude that the proposal is relevant and provides useful insights regarding player performance in different matches.

The proposed methodology has been compared with a basic PM rating methodology. The results obtained suggest the definition of a PM metric for handball as a promising future line of work, since no reference has been found in the literature. Therefore, it will be possible to adapt the rating process to singularities and the nature of handball.

There are some possibilities for improvement in the process of assigning weights to the performance criteria; for example, the combination of experts provided opinions with results acquired by machine-learning techniques. The use of machine-learning techniques to automate the process of weighting performance features has been successfully tested in several studies to evaluate soccer players [32].

Finally, some limitations of our proposal should be noted and studied. First, the proposed approach rating does not allow for evaluation of the goalkeeper (essential in handball), which requires a specific form of evaluation different from that of other players. Furthermore, it is very complicated to consider so many aspects of the game in real time; it would therefore be recommendable to simplify the process of capturing data from each match, that is, to find a balance between quality of information and the effort required to gathering it.

CONFLICT OF INTEREST

Authors have no conflict of interest.

ACKNOWLEDGMENTS

This work has been partially supported by the Spanish Government under grant MERINET: TIN2016-76843-C4-2-R (AEI/FEDER, UE).

REFERENCES

- [1] DHB, Unser Markenleitbild, 2019. <https://dhb.de/der-dhb-verband.html>.
- [2] DPA, Start der EHF-Champions League: THQ Kiel und SG Flensburg müssen Vollgas geben, 2019. <https://www.shz.de/14920136>.
- [3] F. Schwenkreis, A graded concept of an information model for evaluating performance in team handball, in Conference: 8th International Conference on Data Science, Technology and Applications, Prague, Czech Republic, 2019, vol. 7, pp. 196–202.
- [4] T. Kharrat, J. Peña, I.G. McHale, Plus-minus player ratings for soccer, Eur. J. Oper. Res. 283 (2017), 726–736.
- [5] M. Romaniello, U. Fiore, Special issue on recent advances in soft set decision making: theories and applications, Appl. Soft Comput. 54 (2017), 364–365.

- [6] Y. Fu, M. Li, H. Luo, G.Q. Huang, Industrial robot selection using stochastic multicriteria acceptability analysis for group decision making, *Robot. Autonom. Syst.* 122 (2019), 103304.
- [7] L. Wang, G. Ma, F. Zhou, Y. Liu, T. Tian, Multicriteria decision-making approach for selecting ventilation heat recovery devices based on the attributes of buildings and the preferences of decision makers, *Sustain. Cities Soc.* 51 (2019), 101753.
- [8] U. Liu, S. Kenney, E. Breman, T.A. Cossu, A multicriteria decision making approach to prioritise vascular plants for species-based conservation, *Biol. Conserv.* 234 (2019), 221–240.
- [9] S. Aydin, C. Kahraman, Order fulfillment performance evaluation under intuitionistic fuzzy environment, in: 2019 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (EUSFLAT 2019), Prague, Czech Republic, Atlantis Press, 2019.
- [10] G. Campanella, R. A. Ribeiro, A framework for dynamic multiple-criteria decision making, *Decis. Support Syst.* 52 (2011), 52–60.
- [11] J.A. Morente-Molinera, X. Wu, A. Morfeq, R. Al-Hmouz, E. Herrera-Viedma, A novel multi-criteria group decision-making method for heterogeneous and dynamic contexts using multi-granular fuzzy linguistic modelling and consensus measures, *Inf. Fusion.* 53 (2020), 240–250.
- [12] I. Palomares, L. Martínez-López, F. Herrera, A consensus model to detect and manage noncooperative behaviors in large-scale group decision making, *IEEE Trans. Fuzzy Syst.* 22 (2014), 516–530.
- [13] A. Mardani, A. Jusoh, E.K. Zavadskas, Fuzzy multiple criteria decision-making techniques and applications-two decades review from 1994 to 2014, *Expert Syst. Appl.* 42 (2015), 4126–4148.
- [14] F. Blanco-Mesa, E. León-Castro, J.M. Merigó, A bibliometric analysis of aggregation operators, *Appl. Soft Comput.* 81 (2019), 105488.
- [15] J. Serrano-Guerrero, F. Chiclana, J.A. Olivas, F.P. Romero, E. Homapour, A T1OWA fuzzy linguistic aggregation methodology for searching feature-based opinions, *Knowl. Based Syst.* 189 (2019), 105–131.
- [16] A.M. Rosso-Cerón, V. Kafarov, G. Latorre-Bayona, R. Quijano-Hurtado, A novel hybrid approach based on fuzzy multi-criteria decision-making tools for assessing sustainable alternatives of power generation in San Andrés Island, *Renew. Sustain. Energy Rev.* 110 (2019), 159–173.
- [17] P. Aghapoor Khameneh, S. Miri Lavasani, R. Nabizadeh Nodehi, R. Arjmandi, Water distribution network failure analysis under uncertainty, *Int. J. Environ. Sci. Technol.* 17 (2020), 421–432.
- [18] D.J. Berri, Who is ‘most valuable’? Measuring the player’s production of wins in the National Basketball Association, *Managerial Decis. Econ.* 20 (2000), 411–427.
- [19] A. Saez Castillo, J. Rodriguez Avi, J.M. Perez Sanchez, Expected number of goals depending on intrinsic and extrinsic factors of a football player. An application to professional Spanish football league, *Eur. J. Sport Sci.* 13 (2013), 127–138.
- [20] J. Brooks, M. Kerr, J. Guttag, Developing a data-driven player ranking in soccer using predictive model weights, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, 2016, pp. 49–55.
- [21] L. Pappalardo, P. Cintia, P. Ferragina, E. Massucco, D. Pedreschi, F. Giannotti, PlayeRank: data-driven performance evaluation and player ranking in soccer via a machine learning approach, *ACM Trans. Intell. Syst. Technol.* 10 (2019), 1–27.
- [22] L.M. Hvattum, A comprehensive review of plus-minus ratings for evaluating individual players in team sports, *Int. J. Comput. Sci. Sport.* 18 (2019), 1–23.
- [23] N. Spagnola, The Complete Plus-Minus: a Case Study of the Columbus Blue Jackets, Master’s Thesis, University of South Carolina, 2013. <https://scholarcommons.sc.edu/etd/2708>.
- [24] P. Fearnhead, B. Taylor, On estimating the ability of NBA players, *J. Quant. Anal. Sports.* (2010), 11.
- [25] T. Decroos, J. Van Haaren, V. Dzyuba, J. Davis, STARSS: a spatio-temporal action rating system for soccer, in: J. Davis, M. Kaytoue, A. Zimmermann (Eds.), Proceedings of the 4th Workshop on Machine Learning and Data Mining for Sports Analytics, CEUR Workshop Proceedings, CEUR Workshops, Skopje, Macedonia, 2017, pp. 11–20.
- [26] J.M. Saavedra, Handball research: state of the art, *J. Hum. Kinet.* 63 (2018), 5–8.
- [27] G. Daza, A. Andrés, R. Tarragó, Match statistics as predictors of team’s performance in elite competitive handball, *RICYDE Rev. Int. de Ciencias del Deporte.* 48 (2017), 149–161.
- [28] J.A. Morente-Molinera, I.J. Pérez, M.R. Ureña, E. Herrera-Viedma, On multi-granular fuzzy linguistic modeling in group decision making problems: a systematic review and future trends, *Knowl. Based Syst.* 74 (2015), 49–60.
- [29] L. Martínez, R.M. Rodríguez, F. Herrera, 2-tuple linguistic model, in: The 2-tuple Linguistic Model, Springer, Cham, Switzerland, 2015, pp. 23–42.
- [30] J. Serrano-Guerrero, F.P. Romero, J.A. Olivas, Hiperion: A fuzzy approach for recommending educational activities based on the acquisition of competences, *Inf. Sci.* 248 (2013), 114–129.
- [31] P. Rezakhani, Fuzzy MCDM model for risk factor selection in construction projects, *Eng. J.* 16 (2012), 79–94.
- [32] T. Decroos, L. Bransen, J. Van Haaren, J. Davis, Actions speak louder than goals: valuing player actions in soccer, arXiv preprint arXiv:1802.07127, 2018.