

## Research Article

# Design and Optimization of a Decision Support System for Sports Training Based on Data Mining Technology

Lijing Cao 

*Department of Sports, Tianjin Vocational Institute, Tianjin 300410, China*

Correspondence should be addressed to Lijing Cao; 000697@tjtc.edu.cn

Received 14 March 2022; Revised 6 April 2022; Accepted 26 April 2022; Published 16 May 2022

Academic Editor: Sheng Bin

Copyright © 2022 Lijing Cao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents an in-depth study and analysis of sports training decision support systems through data mining techniques, and designs a corresponding sports training decision support system for application in practical situations. It investigates the application of data mining techniques in solving the problem of optimal analysis of sports training indicators. The Apriori algorithm is used to find out the hidden correlations behind the data, while the correlation study applicable to college students' physical fitness indicators is proposed and practiced, and the potential hidden relationships between the indicator data are found. The analysis results are further verified by the improved AP algorithm, FP algorithm, which provides important opinions for college physical education curriculum reform and youth physical culture development. Then, the Shi-Tomasi algorithm is used for feature point tracking and the support vector machine approach for human motion posture recognition. Considering that the collected action data contain a large amount of stationary action data, a threshold-based action segmentation algorithm is designed to extract useful action segments; after extracting the action segments, common action features are extracted from both statistical features, such as mean, variance, skewness, kurtosis, and physical features, such as acceleration and plantar pressure, to give a specific description of human daily actions. Thus, the fatigue of different postures is judged according to their number of occurrences and a preliminary assessment result is obtained. Then, the surface EMG data collected by the sensors are pre-processed to reduce noise, and the surface EMG signals are fused at the feature level based on the wavelet change algorithm, and the feature vectors are trained with BP neural networks to evaluate the fatigue of human muscle movements and obtain a preliminary evaluation result. Finally, by studying the application of information fusion technology at the decision level, the assessment results of the comprehensive human movement analysis are obtained using D-S evidence theory. In addition, a web platform is built on this basis to display and manage human exercise data and assessment results, and to facilitate the storage and query of historical exercise records.

## 1. Introduction

Data mining and machine learning are branches of artificial intelligence research and application, and classification algorithms are one of the important directions of research in these two fields and have an important position in practical applications such as prediction, decision-making, image recognition and processing, communication, and intelligent speech recognition [1]. Classification algorithms simulate the human ability to distinguish matters through computer programs and classify data into different results based on the characteristic attributes of the data, which can be divided into dichotomous and multiclassification problems according to

the number of classification target labels [2]. The continuous development and advancement of classification algorithms have a crucial position in the era of artificial intelligence and the era of Internet big data. Due to the highly open nature of modern sports systems, this change in the way decisions are made in the information society also profoundly affects the decision-making process in the field of sports. How to find suitable decision support tools to grasp the essence of the problem and make reasonable and correct decisions in the vastness of sports information is an issue that is being closely followed from macro decision-makers to coaches at all levels [3]. In addition to the introduction of various methods such as comprehensive evaluation and matrix decision-making,

sports researchers have been making unremitting efforts to develop modern decision support systems for sports applications based on the theoretical basis of computer science and the characteristics of sports itself.

It is also common for students to approach the physical test in the form of a test. The reason for this phenomenon is that, on the one hand, schools spend a lot of time each year to conduct physical test data, which are only saved in the computer as a resource, and the test results are seldom fed back to students. So, most students cannot correctly judge their physical health; on the other hand, because the physical test data are huge and complicated, it is difficult for teachers to make research and judgment, and it is even more difficult to give reasonable suggestions for individual students through reports [4]. It is also difficult for students to follow up and give feedback after training based on the test results. The main shortcomings of these tools are the high work intensity of manual input, the complexity and inefficiency of form stacking, the weak analytical ability, and the poor readability of data. For administrators and teachers, it not only increases the daily workload but also makes it difficult to process and analyze data [5]. For students, it is impossible to get real-time and effective feedback from daily teaching activities and physical fitness tests. The analysis methods used by the above tools remain on top of simple variance, mean, and reliability calculations, resulting in conclusions obtained from the analysis staying on the surface and failing to bring the full value of a large amount of data into play. But what administrators, teachers, and students care about is the hidden information that is not so easy to find.

Data mining was created in response to cloud computing, and the analysis of indicators can help to improve students' physical fitness scientifically by further improving our ability to access data through the comparison of big data and the specific study of certain indicators that people can use in their daily academic life. Data mining technology is further developing at a rapid pace worldwide [6]. As a pioneer in the field of data mining, First Capital Financial Group has combined the knowledge related to data mining with the traditional banking industry to further promote the traditional business of the group's bank and to a certain extent subvert traditional banking industry. There are still many uncharted areas of data mining that need to be explored and discovered to further promote the way of working of people's production life in the new era and to provide people's life more convenience.

*1.1. Related Works.* Data mining techniques can be divided into two categories, namely, descriptive data mining and predictive data mining. Descriptive data mining usually describes the data in an aggregated manner, which provides one with the general nature of the data, i.e., deriving patterns that can be aggregated for potential connections in the data. Predictive data mining usually involves building a model or a set of models to produce predictions about the data, i.e., predicting the value of a particular attribute given the values of other attributes [7]. With the improvement of collection management systems, collection of management

data has reached a provincial level of concentration, so there is a database to monitor the collection management process. The goal of further emancipating the mind and implementing scientific, fine, and professional management requirements is to comprehensively track and control tax sources through a series of legal means, analyze and predict the development and change trends of tax sources, strengthen the collection and management of tax sources, and effectively prevent tax loss in a series of tax management activities [8]. Using data mining technology, a law enforcement behavior monitoring system is established to monitor the law enforcement behavior of tax authorities, the management process, and the suspicious points of tax law enforcement, so as to realize premonitoring and standardize law enforcement behavior after events. Its essence is to extract and filter data features, and summarize them according to the distribution characteristics of the data itself, not based on the experience of experts. Moreover, the membership function of the axiomatic fuzzy set is also determined according to the semantics contained in the fuzzy concept, and does not require any prior information.

Researchers in various fields have studied the application of information fusion technology in various scenarios, obtained many research results, and concluded on effective engineering implementation methods. The United States has been the world leader in the research of this technology. Awaysheh A has studied the hidden value of social media information by visualizing and analyzing big data, and Lindsey has proposed a way to predict students' knowledge of the skills necessary for their majors by mining information about their academic performance to compensate for their bias in the learning process [9]. Krtalić and Bajić propose a way to predict students' knowledge of the necessary skills in their majors to remedy the problem of bias in the learning process. Based on student achievement data, Hongyan Jiang adopted the decision tree ID3 algorithm and association rules Apriori algorithm for data mining analysis [10]. The ID3 algorithm was used to analyze what factors are related to students' good grades, and the Apriori algorithm was used to find out the influence of good grades in one course on other courses. The coach of the famous American national basketball team, the NBA, used the data mining technology provided by IBM to decide to replace players on the spot, which was once a good story in the database world [11]. In this way, people's application of data, from the low-level end of the query operation, provides favorable decision support for business decision-makers at all levels. Based on the Hadoop big data platform, Omair et al. mine the information of informational campus applications and recommends campus information for the corresponding characteristics of learning [12]. Based on the physical fitness test data of the universities, Yung et al. use the FP-Growth algorithm to study the student physical fitness test data from a deeper perspective. It is also important to point out that all discovered knowledge is relative, has specific prerequisites and constraints, is domain-oriented, and at the same time should be easily understood by users, preferably by expressing the discovery results in natural language [13].

Therefore, the results of data mining and knowledge discovery (DMKD) research are very pragmatic. The third split point selects the maximum value among the attributes. Its semantics can be roughly expressed as “big, medium, small.” After generating a simple concept, the weights of all samples need to be initialized, and the initial weights of all samples are all the same.

For sports training, which itself is a very skilful sports competition, the quantitative analysis of training methods and training competitions is very complicated. In daily sports training, a large amount of raw data such as basic information of athletes, performance information, training plan information, competition information, etc., are accumulated [14]. How to make better use of these data resources, discover the potential correlations and rules implied in them, provide a practical basis to improve the training level of China’s sports training for the decision-making of leaders and coaches of various departments are urgent problems for China’s sports teams to solve at present. The successful application of data mining technology in sports training decision support systems will surely provide a successful answer for the solution of this problem.

*1.2. Sports Training Data Mining Algorithm Design.* Data mining is essentially a knowledge discovery process that is largely based on statistics, artificial intelligence, machine learning, and other techniques that highly automate the process of analyzing data, generalizing inferences, and extracting potential models from it. It also analyzes and predicts future scenarios, helping managers and decision-makers to assess risks and make the right decisions [15]. After distinguishing the sample subsets, the stability is evaluated according to the size of the information gain value: if the information gain value is low, the instability is large; if the information gain value is high, the instability is small. At the same time, knowledge mining of data includes a range of methods aimed at finding useful but undiscovered templates from the data set. For the sake of accuracy, extracting potentially useful and unknown information, models, and trends from large amounts of data is one level of a deeper analysis of the data. The source data may be structured, such as data in a relational database, or semi-structured, such as text, graphical, and graphic data, or even various data published on the Web.

The methods can be mathematical or nonmathematical, deductive, or inductively obtained. The discovered knowledge can be used for information management, optimization queries, decision support, and process management, as well as for the data itself that support the application. Thus, knowledge analysis techniques for data are an interdisciplinary subject, which helps to enable knowledge to be derived from data by providing management and decision support using data provided by low-level simple queries. Therefore, there is a need for researchers in different fields, especially scientists and engineers in the field of database technology, artificial intelligence technology, mathematical statistics, which is a new area of research in knowledge analytics.

A decision tree is a machine learning algorithm that follows a partitioning strategy. For classification problems, if we need to decide which category a sample belongs to, instead of arriving at a classification result in one step, we divide it into a series of sub-decisions so that a tree-structured classification and the prediction model is generated, where each internal node represents a test of an attribute, each branch represents the result of the test, and each leaf node has a category label. The meaning of entropy in the field of data mining is a quantification of the degree of randomness and uncertainty of the information contained in the current sample. The higher the entropy, the greater the randomness of that information and the more difficult it is to conclude from that information.

$$H(x) = \sum_{i=1}^{n_i} p(x) \ln p(x_i). \quad (1)$$

Information gain can be understood as a measure of the magnitude of the change in the randomness of information or data after a change in a condition, which can be considered as the difference between the entropy of the parent node and the weighted average entropy of the children.

$$IG(S, A) = H(S) - \frac{H(A) + H(S)}{H(S, A)}. \quad (2)$$

If a sample is randomly classified according to the distribution of classification labels in the data set, Gini impurity is a measure of the likelihood of misclassifying a new sample of a random variable. Use the combination of DW + OLAP + DM to build a decision support system. Make full use of the advantages of the three to improve the auxiliary decision-making ability. Based on the above situation, the structure of the sports training decision support system is shown in Figure 1. The system is supported and managed through metadata.

$$Gini(E) = 1 + \sum_{j=1}^c p_j^2 \times \sum_{j=1}^c p_c^2. \quad (3)$$

The information gain is the difference between the information entropy of the current node and the conditional entropy of the divided sub-nodes, which indicates the degree of information uncertainty reduction. If the information gain of an attribute is larger, it means that using this attribute to classify the samples can better reduce the uncertainty of the divided samples, and of course, choosing this attribute can accomplish our classification goal faster and better. After obtaining the information gain of each attribute, the information gain rate is obtained by dividing the intrinsic information with the information gain. The one with the highest information gain rate is selected as the splitting attribute to divide the sub-nodes.

Post-pruning is the process of pruning after the decision tree has been constructed. In the pruning process, some subtrees are removed and replaced by leaf nodes, and the category identified by this leaf node is identified by the category to which most of the training samples in this subtree belong. The pruning process can also reduce the tree size

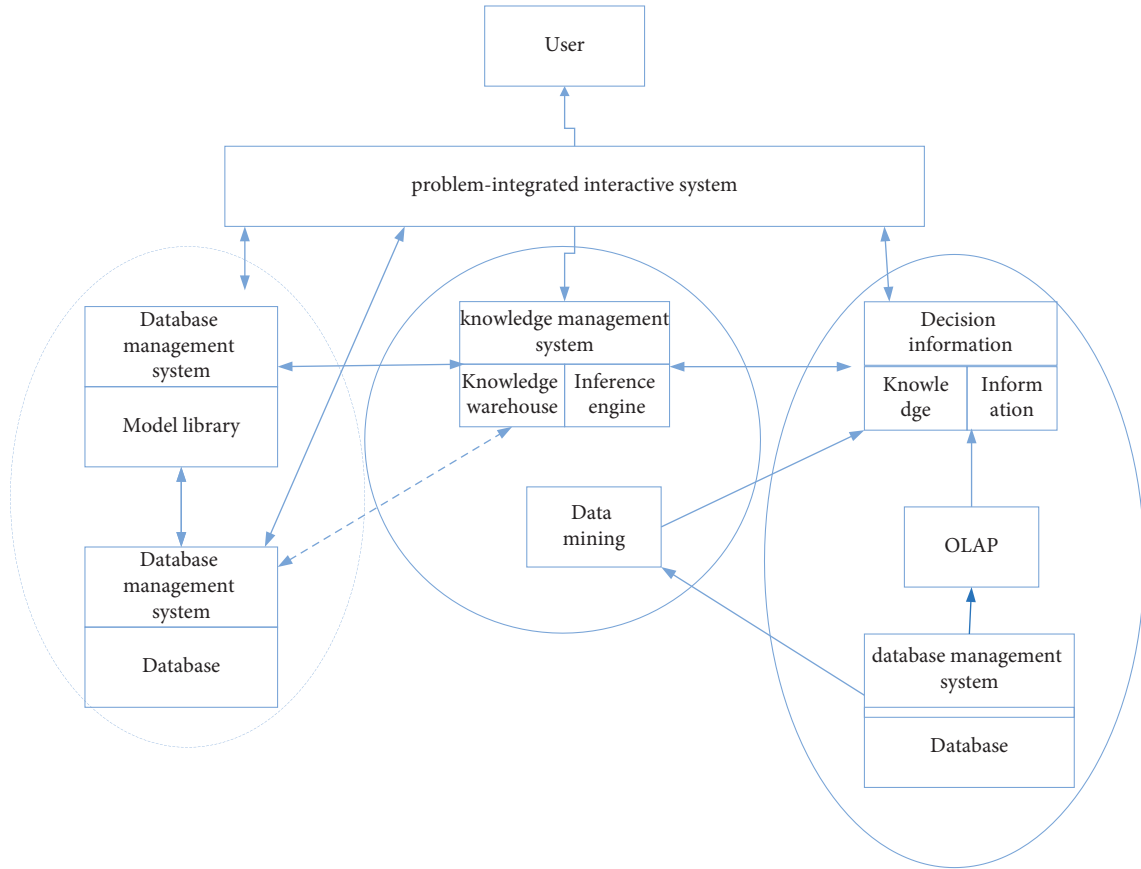


FIGURE 1: Structure of sports training decision support system.

by merging some similar nodes, selecting two child nodes for merging, and calculating the information gain for the new node; if the change in information gain is within an acceptable range, the merge is effective, and the node label after the merge is determined by the category with the largest proportion of all samples in the original two nodes. Post-pruning is currently the most common pruning processing method, as shown in Figure 2.

In the decision tree algorithm, the information entropy of a node is used to describe the chaotic degree of sample distribution among nodes, and the information gain is used to measure the change of information chaos relative to the child nodes after dividing the child nodes from the current node, and the larger the change, the more obvious it is to divide the data based on this attribute. The information gain rate is used to reduce the loss.

In the ABAFS decision tree, let the node being constructed be  $N$ , the training samples be  $X$ , the total number of samples contained in node  $N$  be denoted  $E_N$ , the degree of samples containing class  $m$  be denoted  $E_{m|N}$ , and the sample weights be set to  $W = \{w_1, w_2, \dots, w_N\}$ .

$$E^N = E_M^N - \sum_{i=1}^{|x|} w_{jk} u_{\epsilon}^{\delta}. \quad (4)$$

After combining axiomatic fuzzy sets and decision trees, discrete or continuous attributes can be fuzzified by generating simple concepts, the essence of which is to extract

and filter data features and generalize them according to the distribution characteristics of the data itself, not based on the experience of experts to build. Moreover, the affiliation function of the axiomatic fuzzy set is also determined based on the semantics contained in the fuzzy concept and does not require any a priori information. In general, three partition points are selected on the data set, the first partition point selects the minimum value of the current attribute, the second partition point selects the value after averaging all the data, and the third partition point selects the maximum value of the attribute.

The semantics can be roughly expressed as “large, medium, small.” After generating the simple concept, the weights of all samples need to be initialized, and the initial weights of all samples are the same, as follows:

$$\omega_{ki} = \frac{1}{N}, \quad i = 1, 2 \dots N. \quad (5)$$

All samples of data that form equivalence classes are indistinguishable, i.e., they are equivalent concerning the attributes describing the data. Given real-world data, there are usually classes that cannot be distinguished by the available attributes. Rough sets can be used to approximate or “roughly” define such classes. The definition of a rough set for a given class  $C$  is approximated by two sets: the lower approximation of  $C$  and the upper approximation of  $C$ . The lower approximation of  $C$  consists of several such data samples that, based on knowledge of the attributes, are

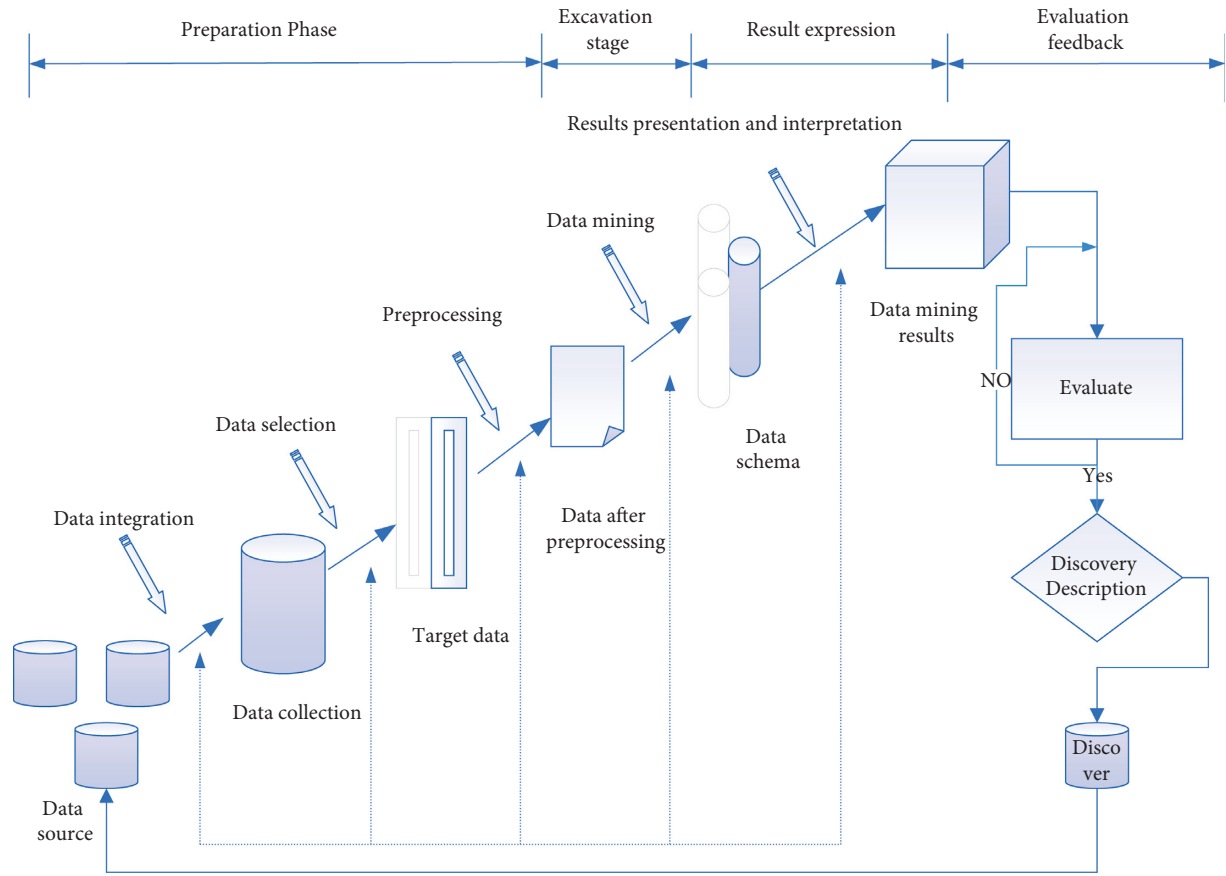


FIGURE 2: Data mining process.

unquestionably part of  $C$ . The upper approximation of  $C$  consists of all such samples that, based on knowledge of the attributes, cannot be considered not part of  $C$ . The lower approximation of  $C$  consists of several such samples that, based on knowledge of the attributes, are unquestionably part of  $C$ .

The ID3 algorithm uses the information gain of an attribute as the node selection criterion and considers the attribute with high information gain as a good attribute. By calculating the information gain of each attribute, the attribute with the highest information gain is selected as the division criterion for each division, and the process is repeated several times until a decision tree with perfectly classified training samples is generated. In the actual operation, the test attribute is determined by the attribute with the highest information gain, and the different test attributes can distinguish several different subsets of samples [16]. After distinguishing the subsets of samples, their stability is evaluated based on the magnitude of the information gain value: if the information gain value is low, the instability is large; if the information gain value is high, the instability is small.

Both the ID3 and C4.5 algorithms use the concept of information first, and the C4.5 algorithm uses the information gain rate instead of information gain as a criterion to determine the test attributes. The branch nodes generated in this way are generated by different attributes, and each branch represents a subset of the samples divided.

The decision tree algorithm can be used to obtain a decision tree with less instability through a partitioning technique. Let there be  $s$  samples of data in the set  $S$ , while having different values of the category attributes  $S: C (i = 1, 2, \dots, n)$ , and then assume that the number of samples in class  $C$  is  $S_j$ , for which the total information of the sample is

$$I(S_1, S_2, \dots, S_n) = a + m \log_a(p_i). \tag{6}$$

Association rules are usually used to find frequent patterns and association rules that exist between things or objects with associated data and transaction data. The most typical example is the application of association rules in the famous shopping basket analysis. Supermarkets generate many consumption records every day, and association rules can dig out the hidden rules between each order and each product from these consumption records, and analyze these rules to get certain consumer consumption habits, to help supermarkets better develop corresponding marketing strategies, as shown in Figure 3.

In the scale decomposition of different levels, the wavelet decomposition coefficient with a larger absolute value has a larger energy value, which has a greater impact on the reconstruction of the signal. Therefore, the maximum value of the absolute value of the wavelet coefficient at each level is taken as the characteristic parameter, so as to obtain a 16-dimensional Feature vector. Frequent itemset

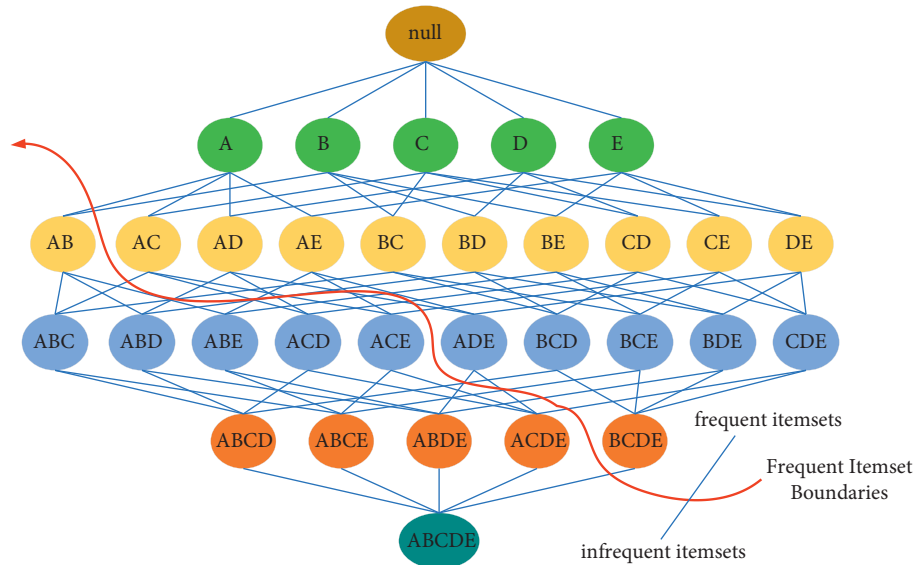


FIGURE 3: Core process of frequent itemset.

mining is the basis for many important data mining tasks such as association rules, correlation analysis, causality, sequential itemsets, local periodicity, and mapping segments. Apriori algorithm and FP-Tree algorithm are both association rule algorithms in data mining that deal with the simplest single-level, single-dimensional Boolean association rules. Data items are found by constructing datasets to find frequent itemsets. Through the discussion, it can be known that the above algorithms can be used to determine whether there is an association between data indicators; Apriori algorithm and FP-Growth algorithm can do the aggregation of data, and further optimize the algorithm by FP-Tree algorithm to reduce the redundant filtering of data; the combination of the two has the rationality to ensure that the model is feasible and improve the reliable analysis for the validation of the analysis results [17]. The combination of the two is reasonable and ensures the feasibility of the model, and improves the reliability of the analysis for the verification of the analysis results.

The interaction between the data layer and the data service layer is the data source. The data layer architecture has three layers in a comprehensive view. The bottom layer is the data source, the data source is the tax data of each province and of the Internet, and various data sources need to be used in the data demand handled by the data service group. Next is the data storage service layer. The data storage can be divided into offline data storage and real-time data storage. Real-time data are generally used by building a real-time monitoring platform, while the data stored in offline data storage is the data source commonly used by the data service group.

*1.3. Optimization Analysis of Sports Training Decision Support System Design.* After the data acquisition system is established, to recognize the human body's daily behavior, it is necessary to collect data on the human body's movements

in various daily situations. This paper introduces how to collect the data and the corresponding preprocessing process. Before data acquisition, the type of action to be captured is defined. After determining the action to be captured, the data are collected based on the action type, the acquisition environment, the location of the sensor, etc., to provide the data basis for subsequent action classification and recognition. After the data acquisition is completed, the original data need to be filtered, through wild point rejection, data segmentation, and other operations for subsequent action classification and recognition. It can be seen by observation that the error value continues to decrease with multiple trainings, and the mean square error reaches the target value when the iteration reaches 262 times.

In the actual data acquisition process, the information collected by the motion acquisition system will inevitably be mixed with noise due to the influence of the external environment and the state of the sensor itself. Moreover, not all signals are useful for the result in the process of signal propagation. The signal may be mixed with other noises or distorted by the environment in each process of generation, transformation, and transmission. Noise or aberrations mixed into the signal can interfere with the subsequent processing of the data and affect the final recognition result. Therefore, the original signal must be processed to reduce the effect of noise or distortion.

Figure 4 shows the change curve of the linear acceleration of one of the axes when the volunteer is performing the upward movement. In the beginning, the linear acceleration is stationary for a while, so the value of linear acceleration is around 0 at this time, and then when the upward movement starts, the acceleration value starts to change above and below the value of 0, corresponding to the change of acceleration and deceleration, and finally, at the stop, the acceleration value is fixed at the value of 0 again. It is obvious from the curve that it contains noise. The noise is mainly reflected in the following aspects. One is the noise inherent

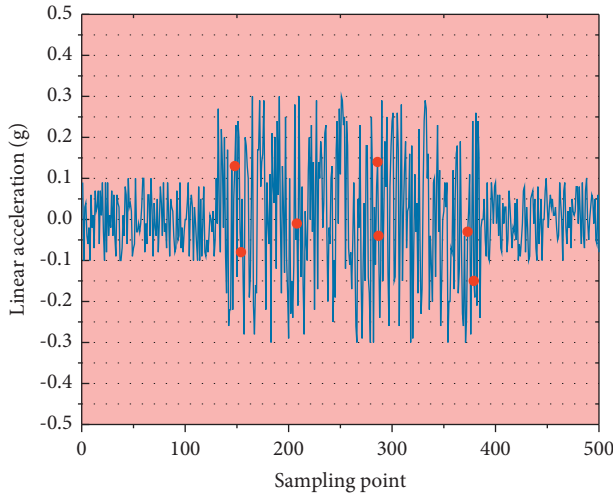


FIGURE 4: Linear acceleration signal during action.

in the inertial sensor itself, the second is the noise randomly generated by the device itself, and the third is the noise generated by jitter. As described in the previous section, the human action frequency is less than 20 Hz, for low frequency. High-frequency signals can be considered as noise. And, random noise has always existed. Since the object of our study is human action information, the remaining two types of signals can be considered as noise signals.

Through analysis, we know that this information is not independent and irrelevant, so this paper uses its internal connection to design a reasonable table structure, so as to make the database use the most reasonable data and not produce data redundancy. There are many common smoothing filtering methods, and the common one is the moving average filtering method. The moving average method is essentially a low-pass filter that exactly meets our requirements for suppressing high-frequency signals. The moving average method uses the average of  $N$  consecutive sampling points to replace the current sampling value, which can effectively suppress random interference signals. Suppose a sensor acquires a sequence of actions as

$$x = (x_1, x_2, \dots, x_n). \quad (7)$$

DW is a data organization and storage technology for decision-making that emerged from the development of database technology. DW consists of basic data, historical data, comprehensive data, and metadata, and can provide comprehensive analysis, time trend analysis, and other decision-making information. OLPA is a technology for analyzing multi-dimensional data. Since a large amount of data is concentrated in a multidimensional space, OLPA technology provides analytical data from multiple perspectives analysis paths to ascertain what users need to assist in decision-making. DM mines and analyzes data in a database or DW using a series of methods to identify and extract implicit and potentially useful information from it, and uses these techniques to assist in decision-making. According to the number of occurrences of the walking state in the whole exercise process, a probability value of exercise

fatigue can be given, which can better reflect the fatigue state of the athlete, so as to obtain the initial result of exercise evaluation.

DW is the basis and OLPA and DM are two different analytical tools. Based on the characteristics of sports training itself, a combination of all three is used to build a decision support system. The advantages of the three are fully utilized to improve the decision support capability. Based on the above, the structure of the decision support system for sports training is shown in Figure 1, which supports and manages the system through metadata.

This system proposes to build the decision support system based on data warehouse and data mining, which can make up for the lack of data organization and data inconsistency problems in DSS, and give full play to the characteristics of subject oriented data warehouse suitable for analysis. The data warehouse extracts, purifies, and transforms internal and external data sources, regroups data into a global-oriented data view, provides the basis for data storage and organization in DSS, and solves the previous data inconsistency problem. Through the information mining of students' academic performance, Lindsey proposes a kind of necessary skills and knowledge to predict students' majors, which can make up for the partial problems of students in the learning process. Data mining and online analytical processing (OLPA) play a key role in the complete solution of the whole DSS. OLPA provides an analysis-oriented multidimensional data model from integrated data in the data warehouse and uses multidimensional analysis methods to analyze and compare multidimensional data from multiple perspectives, sides, and levels, enabling users to analyze data more naturally. DM takes large amounts of data in data warehouses and multidimensional databases [18]. As shown in Figure 1, the DSS architecture based on data warehousing and data mining automatically discovers potential patterns in the data and makes predictions based on these patterns. The knowledge found in data mining can be used directly to guide the analytical processing of online analytical processing (OLPA), and new knowledge derived from data mining and online analytical processing can be immediately added to the system's knowledge base.

Make better use of these data resources and discover the potential associations and rules hidden in them so as to provide practical and feasible basis for improving the training level of sports training in our country for the decision making of leaders of various departments and coaches. The architecture of the sports training decision support system is shown in Figure 5. Firstly, the current data and historical database data of each MIS system such as sports athletes' information management system, sports training plan, quality monitoring system, sports athletes' selection system, and some other data are extracted, purified, converted, and loaded into the data warehouse. Then, several data marts such as scientific research department, center management department and coaching department are divided according to several big functional departments.

The system has good scalability, as the application server, database server, software functions, etc., can be smoothly

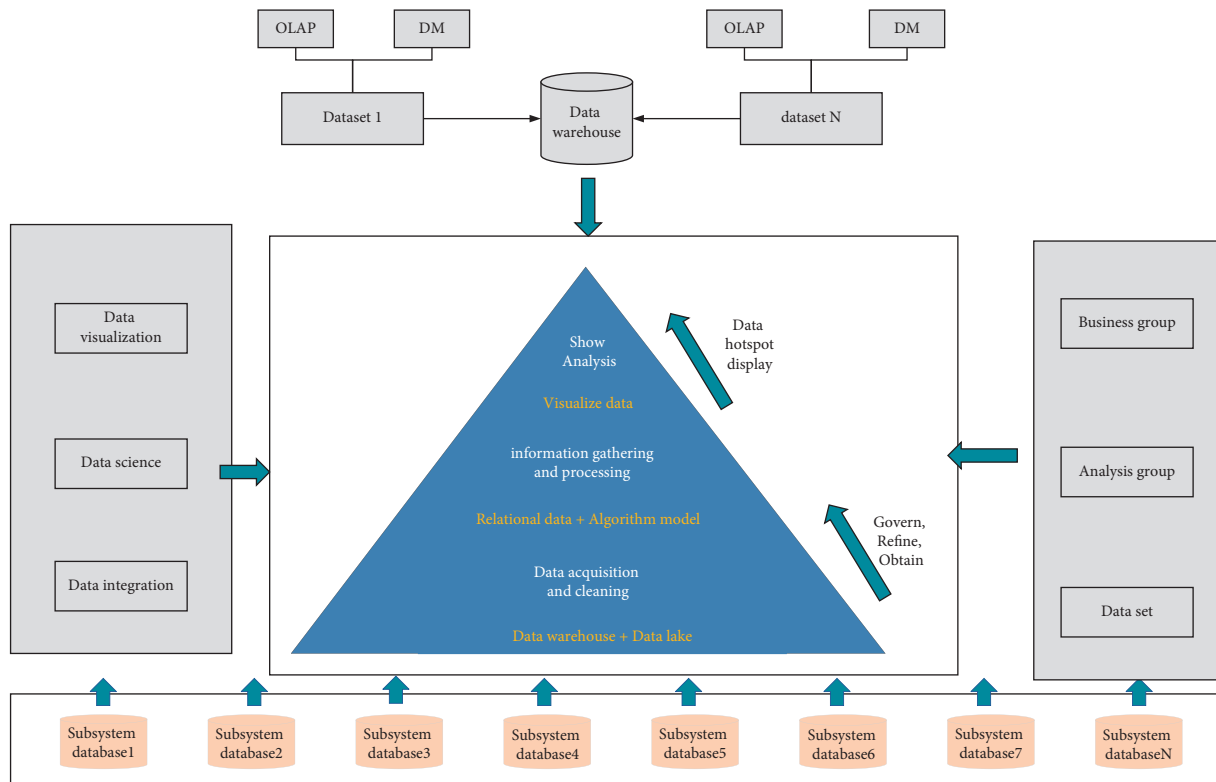


FIGURE 5: System structure of sports training decision support system.

expanded. The decision support system adopts an advanced C/S structure model to ensure the easy expandability of application servers and clients; the multi-node configuration ensures that the database system is easy to expand; the whole software system has good scalability and flexibility. The construction of a model library is necessary for the fencing training system because the basis of superb fencing athlete training methods is the predictive method, and it is necessary to establish and apply targeted predictive models: individual routine models of good athletes, teaching training models, training load models, physical fitness training levels, and competition ability models. In addition to the above models, there are also time-sensitive evaluation models: physical fitness evaluation models for all levels of athletes, competitive ability evaluation models, and other models dedicated to fencing training monitoring. It can also analyze and predict future conditions, helping managers and decision-makers assess risks and make the right decisions.

#### 1.4. Data Mining Algorithm Performance Results Analysis.

In this paper, we will use multi-scale wavelet decomposition to feature extraction of surface EMG signals, taking 4-channel surface EMG signals as input and performing 4 levels of scale decomposition using wavelet transform. In the different levels of scale decomposition, the wavelet decomposition coefficients with larger absolute values have larger energy values and have a greater impact on the reconstruction of the signal, so the maximum value of the absolute value of the wavelet coefficients at each level is taken

as the feature parameter to obtain the 16-dimensional feature vector [19].

And, it is observed that the data of sensor11 and sensor15 are not intercepted because their surface EMG signals are continuously transformed within one second and there is no redundant data. In this paper, the energy values of several groups of data were calculated, and the energy average of several groups of data was used as the threshold value for interception. When the energy value of the data does not exceed the threshold value of fatigue or nonfatigued state during exercise in Figure 6, this set of data is removed; when the threshold value is met, it is retained and used as the data set. After the signal is intercepted according to the energy value, accurate data of fatigue and nonfatigued states can be obtained. Figure 6 shows the average energy magnitude of the data collected by the sensor in 1 s for each state. If the information gain of an attribute is larger, it means that using this attribute for sample division can better reduce the uncertainty of the divided samples. Of course, selecting this attribute can complete our classification goal faster and better.

The error convergence curves after putting the data samples in different states into the model for training are shown in the figure below. By observation, the error value decreases with several training iterations, and the mean square difference reaches the target value when it reaches 262 iterations [20]. The output result of “fatigue” or “non-fatigue” obtained by the BP neural network is the initial evaluation result of the sensor data. After completing the BP neural network training, the two-dimensional feature vector obtained by inputting the test data is the discriminant value



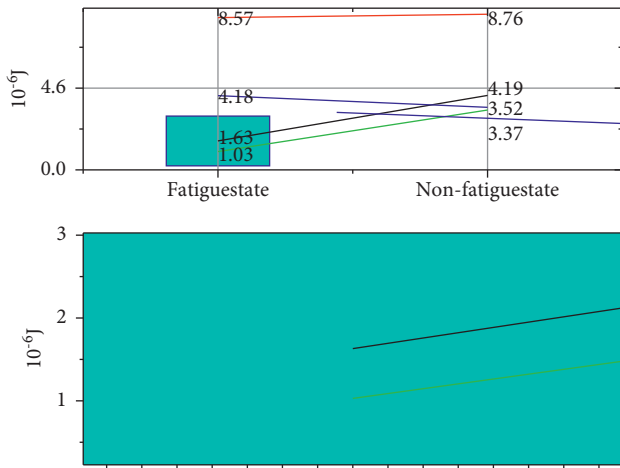


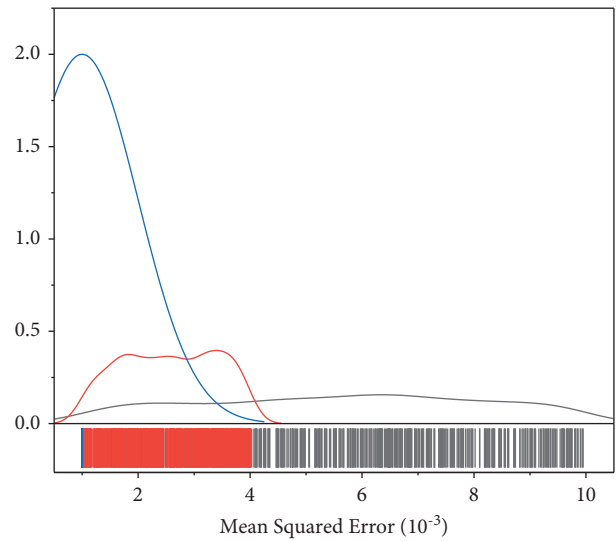
FIGURE 6: Energy values of surface EMG signals.

for the classification of the test data as “fatigue” or “non-fatigue.” For subsequent decision-level information fusion of the image and sensor data using D-S evidence theory, the results in the neural network need to be taken out to obtain the probability value of the corresponding classification, as shown in Figure 7.

The solid line shows the classification effect when the decision tree algorithm is applied to the training set, the dashed line shows the classification effect of the decision tree on the test set, and the horizontal axis is the number of nodes. From the figure, as the nodes of the tree gradually increase, the recognition rate becomes higher and higher on the training set, and the recognition on the test set increases first and then decreases. Therefore, the performance of the decision tree on the test set is related to the size of the tree, and more nodes of the tree are not good. When there are too many nodes in the tree, the performance is very good in the training set, but poor in the test set. To get the decision tree with strong generalization ability, the decision tree is often generated into a larger tree, and then pruned to get the decision tree with stronger generalization ability.

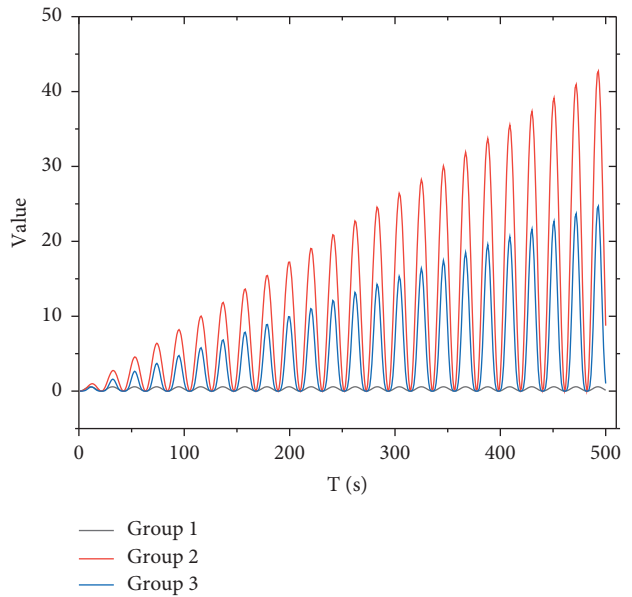
Pre-pruning is a pre-set threshold value that is pruned while the decision tree model is being built. The basic idea is to set a threshold value when determining whether a node needs to be generated downward, and stop the downward growth if the value of information gain is less than this threshold value, to limit the number of tree nodes generated. However, this method is less used in practice because the selection of the threshold value is very difficult. Post-pruning is the pruning of subtrees after the decision tree model is constructed. The basic idea is to start from the bottom up, i.e., from the leaf node of the tree. If the smallest subtree in which the node is located is merged and the loss function is smaller than the loss function before merging, the parent node corresponding to the node is set as the new leaf node, and the kind of this node is identified by the one with the largest number of samples of a certain category in that subtree.

1.5. Analysis of the Optimization Results of the Sports Training Decision Support System. The main data involved in the



— Fatigue  
 — Non-fatigue  
 — Objective

FIGURE 7: Error convergence curve.



— Group 1  
 — Group 2  
 — Group 3

FIGURE 8: Data information page.

intelligent motion evaluation system include basic user information data, general motion recording data, electromyographic data, sensor data as well as evaluation result data, and account management data. To manage the information efficiently and quickly, the above data need to be stored in a database for unified management and resource recall [21]. It is clear from the analysis that this information is not independent and unrelated, so in this paper, we need to design a reasonable table structure using their intrinsic connections so that the database can be utilized most reasonably without creating data redundancy. For example, the assessment result in the assessment information

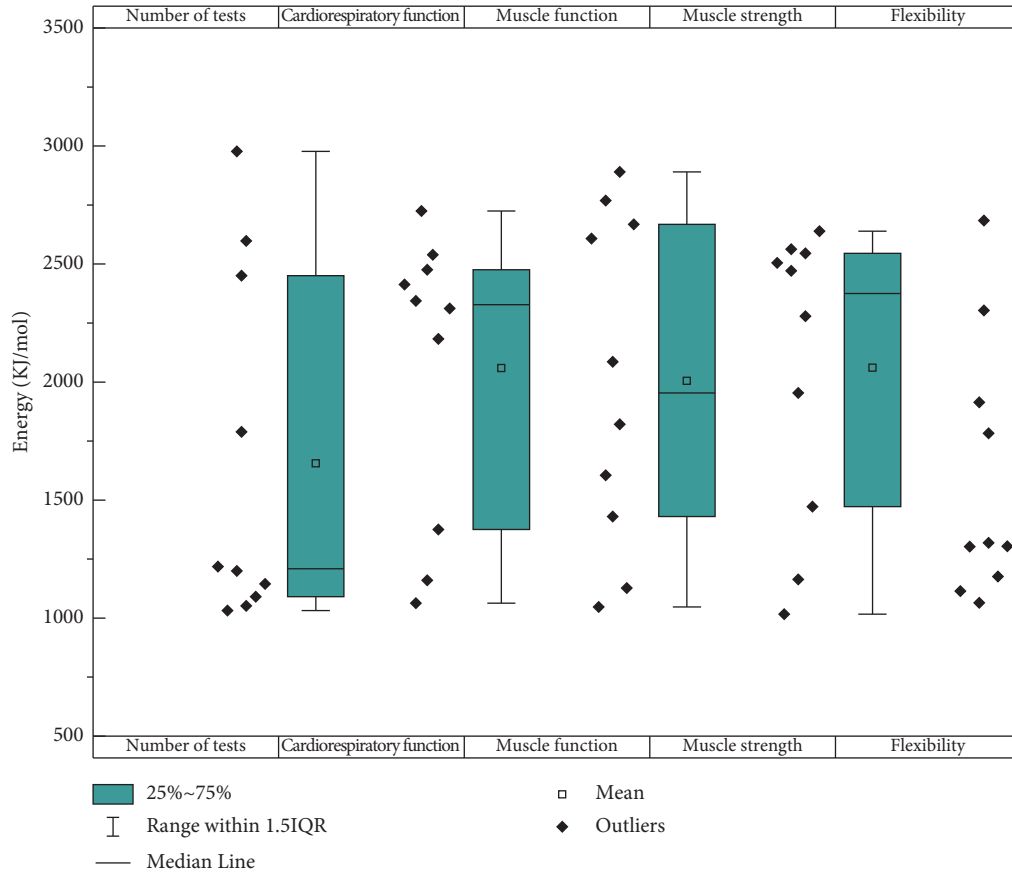


FIGURE 9: Algorithm comparison.

corresponds to its movement record, and the assessment result data are relatively simple and should be placed in the same table with the movement information data. The sample weights are set to.  $W = \{w_1, w_2, \dots, w_N\}$ .

This paper finalizes the implementation of the whole data platform by analyzing the requirements of the overall intelligent motion evaluation system, selecting the technical solution, building the front-end React framework, establishing the server-side Django framework, and designing the MySQL data, as shown in Figure 8.

In this paper, we analyze the characteristics of surface EMG signals to select suitable acquisition devices and conduct experiments designed by ourselves to acquire relevant sensor data and video image data. In this paper, more than 6000 sets of video and sensor data are acquired through several iterations of experiments, which provide reliable support for the subsequent algorithm analysis. We also compare and analyze the advantages and disadvantages of various foreground extraction methods, based on which we propose an optimized hybrid Gaussian background model algorithm to complete the foreground extraction of video images, which makes the foreground extraction results more accurate. Then, after feature extraction and classification training, the recognition of walking and running postures can be realized more accurately. Finally, a probability value of sports fatigue is given according to the number of times the walking state occurs in the whole sports process, which

can better reflect the fatigue state of the sports personnel and thus derive the initial results of sports evaluation, as shown in Figure 9.

Using association rule mining algorithm in the sports training monitoring system module, based on the athletes' daily training data input, the common attribute sets of athletes with common result characteristics can be identified for coaches to analyze the causes of result characteristics to adjust training strategies and competition strategies, providing a reliable decision basis for coaches to develop training plans. This window shows the association between training programs. From the above results, in the training process of players, free pace exercise and electric fencing training, while special exercise and winding exercise can be grouped into one comprehensive training, and this result is outputted through the human-computer interaction system to provide coaches with an auxiliary decision for the development of a training plan.

## 2. Conclusion

The theories related to data mining are introduced, including decision tree and association rule algorithms among data mining algorithms, and the decision tree C4.5 algorithm and association rule, Apriori algorithm is selected to mine the information of physical fitness data, respectively. On this

basis, a specific approach and method of fencing sports training decision support system are proposed. That is, data mining is performed based on establishing a data warehouse. The specific design and development of the system are analyzed in terms of the packaging process development method in conjunction with the development tool SQL Server 2000. The practice shows that the package process development method can effectively achieve module reuse and facilitate system expansion and maintenance, which is a good system design and development method. It provides a data-sharing platform for sports science and technology workers; the second is the popularity of mining tools, and a user interface that can be operated without relevant expertise. Thirdly, they are broadly interdisciplinary through cross-integration with other technologies, such as data mining techniques and various sports system simulations. With the continuous development and growth of data mining technology and in-depth research of sports scientists and technicians, both theoretical research of sports data mining and practical research and development of data mining tools can bring more convenience and considerable benefits to sports management decision-making and scientific research. Therefore, only by successfully solving the above problems, data mining technology can play a greater role in the development of sports science, and thus have a broader development prospect in the field of sports.

### Data Availability

The data used to support the findings of this study are available from the author upon request.

### Conflicts of Interest

The author declares that there are no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

The study was supported by the Department of Sports, Tianjin Vocational Institute.

### References

- [1] X. Schelling and S. Robertson, "A development framework for decision support systems in high-performance sport," *International Journal of Computer Science in Sport*, vol. 19, no. 1, pp. 1–23, 2020.
- [2] B. Salman, N. M. Alhakkak, and M. Musa Jaber, "Football player decision support system baghdad-city as a case study," *International Journal of Engineering & Technology*, vol. 7, no. 3.20, pp. 406–411, 2018.
- [3] C. Yuan, Y. Yang, and Y. Liu, "Sports decision-making model based on data mining and neural network," *Neural Computing & Applications*, vol. 33, no. 9, pp. 3911–3924, 2021.
- [4] S. Latif, F. XianWen, and L. I. Wang, "Intelligent decision support system approach for predicting the performance of students based on three-level machine learning technique," *Journal of Intelligent Systems*, vol. 30, no. 1, pp. 739–749, 2021.
- [5] S. Yang, T. Page, Y. Zhang, and Y. F. Zhao, "Towards an automated decision support system for the identification of additive manufacturing part candidates," *Journal of Intelligent Manufacturing*, vol. 31, no. 8, pp. 1917–1933, 2020.
- [6] M. A. M. Al-Asadi, "Decision support system for a football team management by using machine learning techniques," *Xinyang Teachers College*, vol. 10, no. 2, pp. 1–15, 2018.
- [7] A. Adimabua Ojugo, E. Ekurume, and E. Ekurume, "Predictive intelligent decision support model in forecasting of the diabetes pandemic using a reinforcement deep learning approach," *International Journal of Education and Management Engineering*, vol. 11, no. 2, pp. 40–48, 2021.
- [8] K. Zhang, H. Chen, H.-N. Dai, H. Liu, and Z. Lin, "SpoVis: decision support system for site selection of sports facilities in digital twinning cities," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 1424–1434, 2022.
- [9] A. Awaysheh, J. Wilcke, F. Elvinger, L. Rees, W. Fan, and K. L. Zimmerman, "Review of medical decision support and machine-learning methods," *Veterinary pathology*, vol. 56, no. 4, pp. 512–525, 2019.
- [10] A. Krtalić and M. Bajić, "Development of the TIRAMISU advanced intelligence decision support system," *European Journal of Remote Sensing*, vol. 52, no. 1, pp. 40–55, 2019.
- [11] S. Kasap, D. Abbas, D. Abbas, M. Khajah, and M. Ashknani, "Developing a knowledge-driven decision support system for university/college selection problem in Kuwait," *International Journal of Information and Education Technology*, vol. 10, no. 1, pp. 20–25, 2020.
- [12] M. Omair, S. Noor, M. Tayyab et al., "The selection of the sustainable suppliers by the development of a decision support framework based on analytical hierarchical process and fuzzy inference system," *International Journal of Fuzzy Systems*, vol. 23, no. 7, pp. 1986–2003, 2021.
- [13] K. K. Yung, C. L. Ardern, F. R. Serpiello, and S. Robertson, "Characteristics of complex systems in sports injury rehabilitation: examples and implications for practice," *Sports Medicine-Open*, vol. 8, no. 1, pp. 1–15, 2022.
- [14] M. K. Gupta and P. Chandra, "A comprehensive survey of data mining," *International Journal of Information Technology*, vol. 12, no. 4, pp. 1243–1257, 2020.
- [15] B. Liu, W. Chi, X. Li et al., "Evolving the pulmonary nodules diagnosis from classical approaches to deep learning-aided decision support: three decades' development course and future prospect," *Journal of Cancer Research and Clinical Oncology*, vol. 146, no. 1, pp. 153–185, 2020.
- [16] R. Punj and R. Kumar, "Technological aspects of WBANs for health monitoring: a comprehensive review," *Wireless Networks*, vol. 25, no. 3, pp. 1125–1157, 2019.
- [17] F. Kong and Y. Wang, "Design of computer interactive system for sports training based on artificial intelligence and improved support vector," *Journal of Intelligent and Fuzzy Systems*, vol. 37, no. 5, pp. 6165–6175, 2019.
- [18] Y. Cao and H. Mao, "High-dimensional multi-objective optimization strategy based on directional search in decision space and sports training data simulation," *Alexandria Engineering Journal*, vol. 61, no. 1, pp. 159–173, 2022.
- [19] H. Song, C. E. xiu-ying Han, C. E. Montenegro-Marin, and S. krishnamoorthy, "Secure prediction and assessment of sports injuries using deep learning based convolutional neural

- network,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 3399–3410, 2021.
- [20] O. Enaizan, A. A. Zaidan, N. H. M. Alwi et al., “Electronic medical record systems: decision support examination framework for individual, security and privacy concerns using multi-perspective analysis,” *Health Technology*, vol. 10, no. 3, pp. 795–822, 2020.
- [21] Y. Zeng, “Evaluation of physical education teaching quality in colleges based on the hybrid technology of data mining and hidden markov model,” *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 01, pp. 4–15, 2020.