Hindawi
Mobile Information Systems
Volume 2021, Article ID 2955215, 10 pages
https://doi.org/10.1155/2021/2955215

Research Article

Prediction Model of Juvenile Football Players' Sports Injury Based on Text Classification Technology of Machine Learning

Kai He

College of Sports Science and Technology of Wuhan Sports University, Wuhan 430079, Hubei, China

Correspondence should be addressed to Kai He; 20163045@ayit.edu.cn

Received 29 April 2021; Revised 17 May 2021; Accepted 28 May 2021; Published 11 June 2021

Academic Editor: Sang-Bing Tsai

As the level of soccer in our country has improved rapidly, the level of skill has gradually improved, and the requirements for training of athletes have increased. Due to changes in athlete training methods, it has been decided that athletes must bear a great risk of sports injuries. Accurate prediction of injuries is very important for the development of youth soccer. Based on this, this paper proposes a text classification algorithm based on machine learning and builds a sports injury prediction model that can accurately predict athlete injuries, reduce athlete injuries during training, and be effective. We put forward various sports suitable for young athletes, and put forward some measures to prevent and alleviate athletes' injuries. This article selects 48 football players from a college of physical education of a university for testing. The athletes participating in the experiment use professional equipment to collect exercise volume and exercise load data, and real-time records of each athlete's physical fitness data within half a year, through the athlete's exercise volume, exercise load, body metabolism, and physical indicators to predict their sports injury. Experiments show that from the degree of injury, it can be seen that the severe injury is the least, with 5 cases of muscle injury, 2 cases of fascia ligament injury, and 1 case of joint injury. There were 25 cases of mild injuries, accounting for 41.0% of the total. This is because the athlete's sports injury prediction model has better prediction capabilities, allowing athlete coaches and therapists to optimize training courses, ultimately preventing injuries, improving training levels, and reducing rehabilitation costs.

1. Introduction

Soccer is a very attractive sport because it is easily accepted by the general public and has many participants. To reduce the incidence of sports injuries, we should start with prevention. Researchers say that the main methods of prevention are to focus on strength training, raise self-defense awareness, improve self-defense function and physical reserve, and use scientific training methods for practice. In recent years, many scholars and rehabilitation experts have placed sports injury factors on movement restriction and asymmetry, so that when the body does the same movement, the left and right sides of the other parts of the body are also doing the same exercise. This leads to a decrease in muscle strength, leakage of body energy, and an increased risk of injury for athletes.

An acute injury is characterized by rapid onset and short-term illness. Chronic injuries include recurrent injuries and cumulative injuries. Recurrent injuries are injuries that often occur repeatedly in the same place with the same medical condition. A particular symptom is a repetitive attack that is difficult to treat for a long period. Cumulative injury refers to the partial injury of the body that has been accumulated over a long period due to improper training load arrangement, excessive burden, and training beyond the body's ability to bear. Because each sport has its own technical requirements and physiological characteristics, the burden on each part of the athlete's body is not the same.

Therefore, understanding and familiarizing with the injury characteristics of football players is of great significance for the prevention, diagnosis, and treatment of sports injuries.

At present, both domestic and overseas place great importance to research on sports injuries. A series of research results have guaranteed the physical and mental health of athletes and provided an inexhaustible impetus for the sustainable development of various sports. Zhang et al. proposed that several factors contribute to achieving these goals, including athletes' health status, emotional state, exercise load, and physical strength requirements. Injury prediction is an important part of injury prevention, and successful identification of injury prediction is the main part of the effective injury prevention index. The goal of the proposed artificial neural network is to develop and use early work ability and exercise load data to verify a hierarchical machine learning prediction system that can accurately detect athletes' injuries. The framework is used to test the sports information of 21 football players from various sources, including the collected internal burden information, external burden information, and comment information. However, his research did not clearly point out how to detect and avoid the damage that this kind of exercise may cause [1]. Adel et al. proposed that sports medicine data is an indispensable department in the medical field, which is responsible for ensuring sports safety based on the recovery level after injuries caused by sports activities. The method of using a large amount of medical sports data and events to explain reliably and valuable data information has become an important research path for medical data collection and analysis. This article discusses the extraction and research of complex algorithms for key sports medical data, as well as the lack of training and accuracy. It involves an optimized convolutional neural network based on a deep learning model to ensure the successful detection and risk assessment of sports medicine diseases. However, its overall research lacks data support, and more data is needed to support its conclusions [2]. Janani and Vijayarani proposed that sports injuries restrict participants' participation, impose a heavy burden on the economy, and may have a lasting adverse effect on health-related quality of life. When combined with analytical methods, effective use of the internet of things (IoT) can improve player's safety by identifying injury risk factors that can be addressed through targeted risk reduction training activities. The use of IoT devices can promote efficient quantification of related functions before participating in sports, which can greatly promote the popular sports injury management paradigm. This study introduces a framework that uses sensor-derived IoT data to supplement other data to estimate the injury risk level of each rugby player objectively. The experimental results lack more data support so that the athlete's injury risk level remains in doubt [3].

In this paper, the situation of possible sports injuries is classified through the form of text preprocessing, and the text is extracted as a feature item of sports injuries of athletes, so as to meet the predictive effect of the sports injury prediction model of young football players. In this paper, we first collect training load and sports injury data of young soccer players through related equipment and then input the relevant data into a sports injury prediction model for preprocessing. Finally, a text classification algorithm is used to analyze the training load and sports injury characteristics of young soccer players and predict the risk of sports injuries.

2. Text Classification Research on the Prediction Model of Football Players' Sports Injury

2.1. Text Classification Algorithm. The purpose of text representation is to use these feature items finally as the representative of the text and then find what method is used to represent the text. At present, it is more commonly used to represent the text through the establishment of a text representation model [4, 5].

2.1.1. Vector Space Model

(1) Feature Item Weight. Among the N feature items contained in the document, the document is denoted as $D(t_1, t_2, \ldots, t_N)$, and the importance of each item t_k in the document D is different, which requires different feature items to have different weights w_k as follows:

$$D = D\{(t_1, w_1), (t_2, w_2), \ldots, (t_N, w_N)\}. \tag{1}$$

where $w_k$is the weight of the feature item t_k; $1 \le k \le N$.

(2) Vector Space Model Construction. In the vector space model, each document $D(w_1, w_2, \ldots, w_N)$ belongs to the N dimensional vector space. In the calculation, $(t_1, t_2, \ldots, t_N)$ is the N dimensional coordinate system, while $(w_1, w_2, \ldots, w_N)$ is the coordinate value in the coordinate system, and $D(w_1, w_2, \ldots, w_N)$ is a vector, usually called a document, the vector space model of D [6, 7].

(3) Document Similarity. Document similarity mainly indicates the degree of relevance between two documents. The calculation methods of the similarity value between two documents are as follows:

The Euclidean distance formula is as follows:

$$\operatorname{sim}(Q, D) = \sqrt{\sum_{i=1}^{m} (w_i - v_i)^2}. \tag{2}$$

The angle cosine formula is as follows:

$$\operatorname{sim}(Q, D) = \cos\theta = \frac{\sum_{i=1}^{N} w_i * v_i}{\sqrt{\sum_{i=1}^{N} w_i^2 * \sum_{i=1}^{N} v_i^2}}. \tag{3}$$

The vector inner product formula is as follows:

$$\mathrm{sim}\,(Q, D) = \sqrt{\sum_{i=1}^{m} w_i v_i}. \tag{4}$$

Here, Q and D represent a document, respectively, and w_t and v_t are the weight values of feature items in documents Q and D, respectively. When the value of the similarity $\mathrm{sim}\,(Q, D)$ is larger, the correlation between the two documents appears to be higher. The smaller the value of the similarity $\mathrm{sim}\,(Q, D)$, the smaller the correlation between the two documents.

2.1.2. Weight Calculation. Currently, there are two main ways to check the weight value of a text feature item. One is to subjectively assign weight values to each functional item through the expert's own knowledge and experience. This method is very time-consuming and labor-intensive for large text sets [8, 9]. The other method is calculated by the weight evaluation function. Currently, the following methods are mainly used:

(1) Boolean weight

$$w_{ij} = \begin{cases} 1, & \mathrm{freq}_{ij} \geq 1, \\ 0, & \mathrm{freq}_{ij} = 0, \end{cases} \tag{5}$$

where w_{ij} is the weight value of the j feature item in the text d_i, and freq_{ij} is the word frequency of the j feature item in the text d_i. The Boolean weight function is the simplest weight function. The weight value of the feature item is 1, and the weight value of the feature item is 0.

(2) Radix function

$$w_{ij} = \sqrt{\mathrm{freq}_{ij}}, \tag{6}$$

where freq_{ij} represents the word frequency of the j feature item in d_i.

(3) IF-IDF function

$$\begin{aligned} w_{ij} &= \frac{\left(tf_{ij} \times \log\left(N/n_i\right)\right)}{\sqrt{\sum\left[tf_{ij} \times \log\left(N/n_i\right)\right]^2}}, \\ w_{ij} &= \frac{\left(\log\left(tf_{ij} + 1\right) \times \log\left(N/n_i\right)\right)}{\sqrt{\sum\left[\log\left(tf_{ij} + 1\right) \times \log\left(N/n_i\right)\right]^2}}, \end{aligned} \tag{7}$$

where j is the text set, i is the feature item, f_{ij} is the frequency of i in the text j, N is the number of texts in j, and n_i is the number of texts where i appears.

(4) TF-IWF function

$$w_{ij} = tf_{ij} \times \left[\log\left[\frac{\sum_{i=1}^{M} nt_i}{nt_i}\right]\right]^2, \tag{8}$$

where nt_i represents the number of occurrences of t_i in the text, and M represents the total number of feature items.

2.1.3. Feature Extraction. After feature selection, the accuracy of document feature selection can be increased more efficiently. After selection, more representative features will be selected from the initial vector space, so as to achieve the goal of new spatial dimensions with fewer dimensions, but not because of these changes, the characteristics of the initial vector space will change [10, 11]. Generally, in document classification, the most common methods of document feature extraction mainly include the following:

(1) Document Frequency Method. Words that appear frequently in documents are generally not representative of the text, and words that appear frequently in very low documents are generally not representative of the text. These words are difficult to distinguish text categories. Therefore, when using the document frequency method to extract features, it is usually necessary to set a certain threshold below and above, so that the entries with too low and too high frequencies can be filtered out [12, 13]. The calculation formula of the document frequency method is as follows:

$$\mathrm{DF}\,(t) = \frac{\text{number of documents with feature } t}{\text{the total number of documents in the training set}}. \tag{9}$$

The document frequency method has the advantages of less calculation and faster speed. Usually, it sets the threshold to extract the words with appropriate frequency as feature items, which can effectively reduce the dimensionality of the feature space, but this method also has certain shortcomings because some documents such as special words may appear very low in frequency, but they have a good effect on distinguishing text categories. If such words are deleted directly, it will affect the accuracy of classification [14, 15].

(2) Information Gain Method. Information gain method is to calculate the difference between the information entropy when a certain feature item is included in the text and when it does not contain a certain feature item. The main formula is as follows:

$$\begin{aligned} \mathrm{IG}\,(t_i) = {} & -\sum_{j=1}^{m} P\left(C_j\right)\log P\left(C_j\right) \\ & + P\,(t_i) \times \sum_{j=1}^{m} P\left(\frac{C_j}{t_i}\right) \times \log P\left(\frac{C_j}{t_i}\right) \\ & + P\,(\bar{t}_i) \sum_{j=1}^{m} P\left(\frac{C_j}{\bar{t}_i}\right) \times \log P\left(\frac{C_j}{\bar{t}_i}\right), \end{aligned} \tag{10}$$

$$\bar{g}_i\,(x', y') = g_i\,(x', y') - \mathrm{mean}\left[g_i\,(x', y')\right],$$

where $P(C_j)$ is the probability of C_j type text in the training text set, $P(t_i)$ is the probability of feature item t_i in the training set, and $P(\bar{t}_i)$ is the text probability without feature item t_i in the training set.

(3) Mutual Information Method

$$\mathrm{MI}\left(t_i, C_j\right) = \frac{\log P\left(t_i, C_j\right)}{P\left(t_i\right)} P\left(C_j\right), \tag{11}$$

where $P(t_i, C_j)$ is the probability value of the text containing t_i and belonging to the C_j category, $P(t_i)$ is the probability of the text containing t_i, and $P(C_j)$ is the probability of the text containing the C_j category in the total training set text [16]. The main meaning of mutual information is to reflect the relationship between functional items and categories. Its main drawback is that it is sensitive to the edge probability density of feature items when calculating mutual information values. From the formula, we can see that if the conditional probability values are the same, words with low word frequency have a large amount of mutual information and cannot accurately reflect the distinctiveness of the feature item.

2.2. Football Player Sports Injury Prediction

2.2.1. Irregular Technical Actions. Mastering the correct basic football skills is vital to the development of young football players' future sports career and the improvement of their competitive level. The factors that cause the irregularities of the athletes' technical movements include the following six aspects. First, the athletes are vague in the concept of the technical movements they have learned at the beginning, and the sequence, essentials, and requirements for completing the movements are unclear. The second is the coaches when they teach technical movements. The explanation is not systematic and clear, and the action demonstration is not standard [17, 18]. The third is that the athletes are not clear about the purpose of learning new technical movements, do not take the movements seriously, and lack self-confidence to complete the movements. Fourth, the coaches have improper organization and teaching methods. Fifth, the athletes do not have enough time to practice technical movements and cannot master them well. Sixth, athletes are taking training under fatigue, which leads to deformation of technical movements.

2.2.2. Inattention. The situation on the field during a football game is complex and changeable. Long-range mobilization requires a quick response and judgment to the opponent's tactical intentions, the direction of the ball, and the speed of the ball in the first time, which requires players to maintain high concentration at all times. Inattentiveness during training or competition will deform the technical movements, and the timing of the shot will not be mastered, which will increase the burden on the body and lead to sports injuries. The factors that affect athletes' attention include internal and external factors: internal factors include athletes' bad mood, low mood or irritability, their own personality characteristics, the influence of physical illness, too much emphasis on the results of the game, or too underestimation of the enemy, and so on, and external factors include interference from outside people or things, the impact of the game weather and environment, the coach's on-the-spot command, and so on [19, 20]. In the usual training process, the coach should arrange some simulation training according to the athlete's own characteristics and the situation that may occur during the competition, which is of great help enhance the athlete's concentration in training and competition.

2.2.3. Excessive Training Load. The bones and muscles of young athletes are not yet fully developed, and their ability to withstand stress is relatively poor compared with adults. If the training load is too large and exceeds the body's ability to withstand, it is easy to cause harm to the body. The training load of football players can only meet or exceed the requirements of the game in order to meet the needs of long-term, high-intensity games [21]. This is a contradictory issue. It can be seen that it is very important to arrange exercise load scientifically and reasonably.

2.2.4. Excessive Physical Fatigue. When it comes to excessive fatigue of athletes, the first reason that comes to mind is caused by excessive training load. This is one of the causes of physical fatigue but not all. Many young athletes lack scientific training experience and are eager to win. They blindly or recklessly carry out training beyond their physical capacity regardless of subjective and objective conditions. Some athletes are forced to conduct training under the requirements of coaches' high-intensity and high-density training programs. Overloaded training is the main cause of excessive fatigue [22, 23]. In addition, most young athletes do not pay attention to relaxation and recovery after training, and the body is prone to be in a state of fatigue for a long time, which increases the probability of sports injuries [24].

2.2.5. Lack of Self-Protection Awareness and Ability. Especially, young football players are young and have little experience in the game. They often have insufficient understanding of the impact of sports injuries on them and lack experience in preventing sports injuries. They have little or no learning or knowledge about sports injury prevention and rehabilitation [25, 26]. In addition, when physical discomfort occurs, young athletes cannot judge the precursors of sports injury, cannot treat them in time, delay their illness, or adopt incorrect treatment and alleviation methods, which turn many acute mild injuries into chronic cumulative injury, which is difficult to heal for a long time and recurs and becomes a stumbling block to the improvement of competitive level and the development of their sports careers.

3. Experimental Design of Sports Injury Prediction Model

3.1. Test Subject. Here, we select 48 football players from a sports college of a university for testing. The average age of the subjects is 20.09; the average height is 179.36 cm; and the average weight is 75.19 kg. The basic conditions of the subjects are shown in Table 1. The 48 football players participating in the experiment used portable positioning systems, 3D gyroscopes, and 3D digital compass integrated equipment to collect exercise volume and exercise load data. Each athlete needs to bring equipment to collect exercise volume and exercise load during training. Record the relevant data of each athlete in real-time within half a year, and predict the sports injury based on the amount of exercise, exercise load, body metabolism, and physical indicators of each athlete.

Table 1: General information about the subject.

Subject	Age	Height	Weight	Years of exercise
Male	20.04 ± 2.56	183.54 ± 6.34	78.90 ± 7.44	6.33 ± 3.67
Female	19.36 ± 3.17	177.45 ± 4.37	70.12 ± 5.39	5.43 ± 2.52

3.2. Main Research Content of Numerical Experiment

3.2.1. Questionnaire Survey Method. The athlete survey for this study was designed to take into account the incidence, nature, type, location, extent of impact, reasons, and other factors of sports injuries in young soccer players. We have supplemented and revised the content of the questionnaire based on the many requests for the opinions of tutors and related experts. Surveys are distributed directly and collected directly. As some young athletes are immature, authors and coaches oversee and guide the entire process of filling out questionnaires, and in time, the author answered the questions posed by athletes to ensure the authenticity and validity of the questionnaires. During the survey, athletes handed out 48 questionnaires with a recovery rate of 100% and an effectiveness rate of 100%.

3.2.2. Test Method. When testing the actions of each item, if there is pain in a part of the body when the action is completed, the score of this test is 0, and the painful part is recorded. If the subject is unable to complete the established action in accordance with the action standard or has deviations in the posture and action of multiple parts when completing the screening action, the score for this test is 1 point. If the subject was able to complete the action according to the screening action standard, the score for this test was 2 points. The injury prediction model test is divided into the sum of the scores of 7 actions, and the score of each action is taken as the lowest score of this item [27].

3.3. Numerical Test Model Establishment. The aspects involved in sports injuries are very complex, and the manifestations are diverse and unstable. In order to comprehensively reflect the information of the injury situation and fully reflect the individual's injury effects, appropriate sports injury characteristics are selected to form a sports injury evaluation index system. At the same time, the questionnaire survey uses the Likert five-level scoring method, with a score ranging from 1 to 5, which, respectively, represent completely disagree with this view, disagree with this view, neither oppose nor agree, agree with this view, and agree completely. For this view, the higher the score, the more agree with the point of view on the topic.

3.4. Statistical Processing. For kinematics data processing, the damage prediction model was used to process the collected data and export the TSV file, and then the Qtools software and Excel software was used to calculate and process the exported raw data to obtain parameters such as time, joint angle, sports injury, and displacement index.

For kinetic data processing, Kistler software was used to screen and export the test results, and Excel software was used for calculation and statistical processing, and parameter indexes such as force value, impulse, and power were obtained.

4. Experimental Sports Injury Prediction Model

4.1. Characteristics of Football Players' Sports Injury

4.1.1. Analysis of the Types and Extent of Sports Injuries. According to the hospital's diagnostic criteria, this article counts the seven types of injuries such as sprains, strains, bone injuries, contusions, abrasions, dislocations, and other injuries: mild injury, moderate injury, and severe injury.

Statistics from the data in Figure 1 show that in terms of the types of athlete injuries, sprains account for the first place in the probability of injury, accounting for 35.3%. Ankle injuries are the most common and most difficult in sports training. It is a type of sports injury that attracts attention. Secondly, strains account for 22.9%. Leg muscle strains are a common sports injury, often caused by violent contraction of the calf muscles or sudden twisting of the posture during exercise. Then, bone injuries account for 18.8%. According to the characteristics of football sports, it can be seen that calf fractures are more common. In order of the remaining injury probability, the proportions of contusion, abrasion, other, and dislocation are 10.4%, 6.3%, 4.2%, and 2.1%, respectively. It shows that the types of football players' injuries are mainly sprains, strains, and bone injuries, and the number of people with moderate injuries is the largest, which mainly occurs in ankle sprains and muscle strains.

4.1.2. Analysis of Sports Injury Tissue. According to the characteristics of football events, this article divides sports injury organizations into several types for detailed analysis: muscle injury, tendon injury, joint injury, fascia ligament injury, fracture, cartilage injury, and epiphyseal injury.

It can be seen from Figure 2 that there are 21 cases of muscle damage in total, accounting for 34.4% of the total

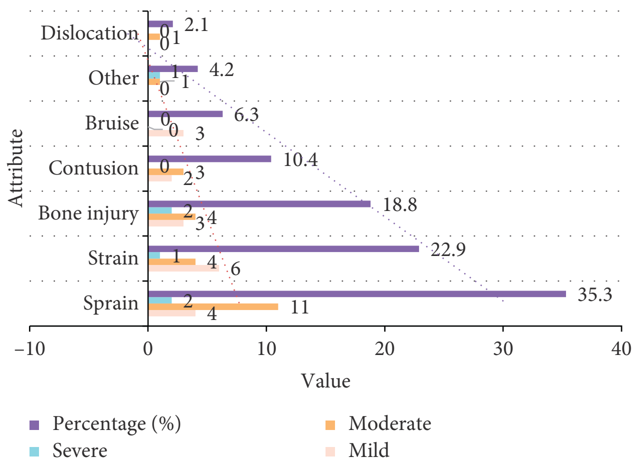

Figure 1: Sports injury type and injury degree distribution map.

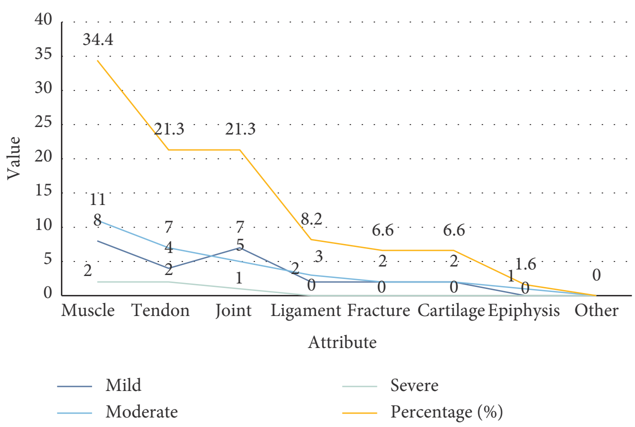

Figure 2: Sports injury tissue distribution map.

number of cases. According to the characteristics of football players, it is very easy to cause muscle damage during long-term high-intensity exercise. Followed by joint injuries and fascial ligament injuries, 13 cases ranked second among the injuries, accounting for 21.3% of the total. There are many special nerve endings in the ligaments, such as the medial collateral ligament and the anterior cruciate ligament of the knee. They play an important role in proprioception and pain perception. Then there were tendon injuries, 5 cases each, accounting for 8.2% of the total; fractures and cartilage injuries also had 4 cases, accounting for 6.6% of the total; and finally, there was 1 case of epiphyseal injury, accounting for 1.6%. From the degree of injury, it can be seen that there are 5 cases of muscle injuries, 2 cases of fascial ligament injuries, and 1 case of joint injuries, which are the least serious injuries. There are 25 minor injuries, accounting for 41.0% of the total. This indicates that tissue damage in soccer players is mainly manifested in muscles, fascial ligaments, and joints. Among them, the knee joints, leg muscles, and ligaments are very prone to injury during exercise.

4.1.3. Analysis of the Characteristics of the Period and Stage of Sports Injury. The occurrence of sports injuries is divided into three categories: prematch training, in-match, and postmatch recovery. Figure 3 shows the statistics of these three situations.

It can be seen from Figure 3 that prematch training has 15 people with mild injuries, 24 people with moderate injuries, 4 people with severe injuries, 43 people with injuries, and 62.3 people with injuries in the entire training period %; In the competition, there were 9 people with mild injuries, 4 people with moderate injuries, and 1 person with severe injuries, and the total number of injuries was 14 people. The injuries accounted for 20.3% of the injuries during the entire training period; the number of people who recovered after the game was light. There were 5 people with moderate injuries, 7 people with moderate injuries, no people with severe injuries, and 12 people with injuries. The number of injuries accounted for 17.4% of the injuries in the entire training period.

The training class is divided into the following four stages: the preparatory activity stage, the special training stage, the quality training stage, and the technical training stage. Figure 4 shows the statistical data of these four situations.

It can be seen from Figure 4 that the number of minor injuries in the preparation phase is 1 person; no people are slanderous; and 2 people are severely injured. The total number of injuries was 3, accounting for 4.0% of the total; during the special training period, there were 7 minor injuries, 33 moderate injuries, and 6 severe injuries. The total number of injured was 46, accounting for 61.3% of the total number of injuries in the entire training course. In the quality training stage, 7 people were slightly injured; 4 people were injured; and no people were seriously injured. The total number of injuries was 11 people, accounting for 14.7% of the total number of injuries; in the technical training phase, the number of minor injuries was 6 people, and the number of people injured was 9 people. The total number of injured was 15, accounting for 20.0% of the total. For football sports, continuous improvement of technical level is more important than physical training. In the youth period of football players, only with a solid technical level, the development of personal competitive ability, and targeted strengthening of physical fitness training, can the overall level of sports be continuously improved.

4.1.4. The Impact of Sports Injuries on Training. It can be seen from Table 2 that 23 cases persist in training, and the reduction in exercise volume affects the training, accounting for 47.9% of the overall proportion; 10 cases who cannot train and can only do some quality exercises, accounting for 20.8% of the overall proportion; There are 7 cases that do not affect normal training, accounting for 14.6% of the total; 4 cases of normal training without sports injuries, accounting for 8.3% of the total; 2 other cases, accounting for 4.2% of the total; due to chronic injury and acute-to-chronic injury, there are more cases of injuries, among which there are more cases of moderate and mild injuries, which leads athletes to develop the habit of not getting under the line of fire for minor injuries. As long as they do not affect normal training, they can continue to be trained, but the amount of exercise

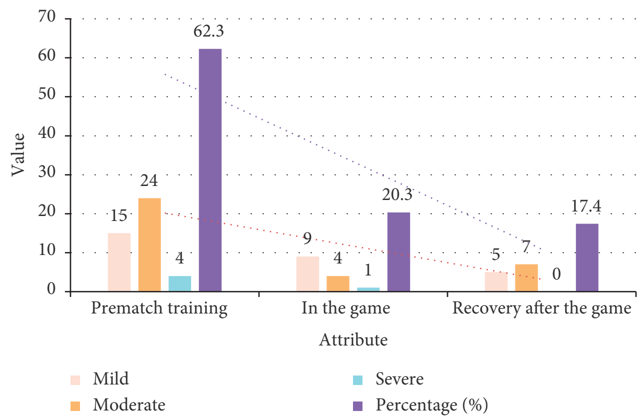

Figure 3: Proportion distribution map during sports injury period.

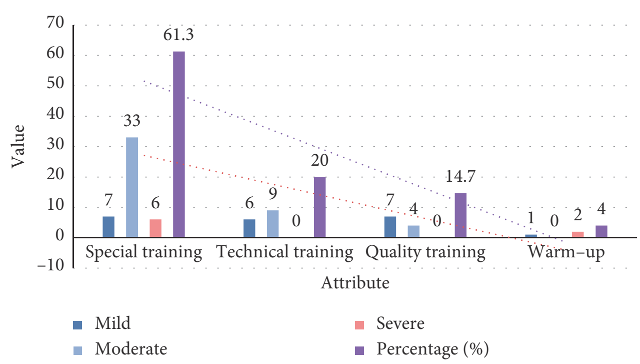

Figure 4: Proportion of the number of injuries in each period of the training session.

can be reduced, or do some strength training, there is no remote mobilization without training at all, so this issue should arouse great attention from coaches, athletes, and scientific researchers.

4.2. Causes of Football Players' Sports Injuries

4.2.1. Physiological Factors. The physiological factors that cause sports injuries of football players were investigated and analyzed, and statistics were made on the three aspects such as insufficient muscle strength, flexibility, poor sensitivity, and insufficient balance ability. The results are shown in Table 3.

It can be seen from Table 3 that there are three factors that affect the physical performance of sports eyes: insufficient muscle strength, poor flexibility, and insufficient balance ability. Among them, 20 coaches and team doctors considered the most important for lack of muscle strength, accounting for 60.6% of the total; 10 considered more important, accounting for 30.3% of the total; 3 considered generally important, accounting for the total 9.1%. There are 12 coaches and team doctors who think that flexibility is the most important, accounting for 36.4% of the total; 13 who think they are more important, accounting for 39.4%; 8 who think they are generally important, accounting for 24.2%. In terms of insufficient balance ability, 11 coaches and team doctors considered the most important, accounting for 33.3%; 11 considered more important, accounting for 33.3%; 11 considered generally important, accounting for 33.3%. Data can show that insufficient muscle strength is the most important factor affecting athletes' injuries in terms of physical fitness.

4.2.2. Training Factors. The training factors that cause sports injuries were investigated, and statistics were made on the following five aspects of insufficient preparation activities: heavy exercise load, partial overload, training or competition with injuries, and insufficient recovery. The results are shown in Figure 5.

From Figure 5, it can be seen that, among these influencing factors, 13 coaches and team doctors think that insufficient preparation is particularly important, accounting for 39.4% of the total; 11 think it is more important, accounting for 33.3%; 9 think it is generally important, accounting for 39.4%. 12 coaches and team doctors thought that exercise load was particularly important, accounting for 36.4% of the total; 15 thought exercise load was more important, accounting for 45.5%; 6 thought it was generally important, accounting for 18.1%. 9 coaches and team doctors think that local overload is particularly important, accounting for 27.3% of the total; 15 think that local overload is more important, accounting for 45.4%; 9 think that local overload is generally important, accounting for 27.3%. In terms of warm-up activities, a large amount of warm-up activities is also an important factor in causing sports injuries. If the amount of warm-up activity is too large, the body will easily enter a state of fatigue sooner or later, and be in a state of exercise fatigue. When participating in training or competition, the possibility of injury will increase; when starting to prepare for practice, the body is in an inert state.

4.2.3. Technical Factors. The technical factors that cause sports injuries were investigated, and statistics were made on the three aspects of lack of skills, incorrect movements, and ignorance of the importance of movements. The results are shown in Figure 6.

It can be seen from Figure 6 that 17 coaches and team doctors considered unskilled skills to be particularly important, accounting for 51.5% of the total; 11 considered more important, accounting for 33.3%; only 5 considered generally important, accounting for 15.2%. Coaches and team doctors considered wrong actions to be particularly important, 23 people, accounting for 69.7% of the total; 6 people thought it was more important, accounting for 18.2%; only 4 people thought it was generally important, accounting for 12.1%. Coaches and team doctors considered the importance of ignoring movements to be particularly important, accounting for 51.5% of the total; 11 considered more important, accounting for 33.3%; only 5 considered generally important, accounting for 15.2%.

Table 2: The impact of sports injuries on training.

Influence level	Number of cases	Percentage (%)
Keep training but reduce the amount of exercise	23	47.9
Cannot take training but can only do quality exercises	10	20.8
Does not affect normal training	7	14.6
No sports injuries during normal training	4	8.3
Others	2	4.2
Stop training completely	2	4.2
Total	48	100

Table 3: The influence degree of sports injury caused by physiological factors.

	Statistics	Very important	Important	General
Insufficient muscle strength	N	20	10	3
	%	60.6	30.3	9.1
Poor flexibility	N	12	13	8
	%	36.4	39.4	24.2
Insufficient balance	N	11	11	11
	%	33.3	33.3	33.6

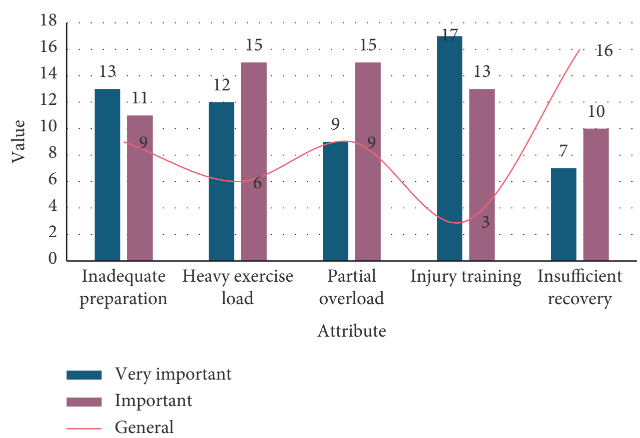

Figure 5: Diagram of the impact of training factors on sports injuries.

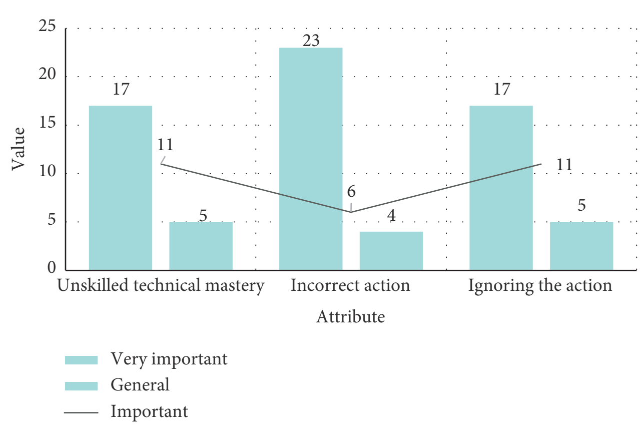

Figure 6: Diagram of the impact of technical factors on sports injuries.

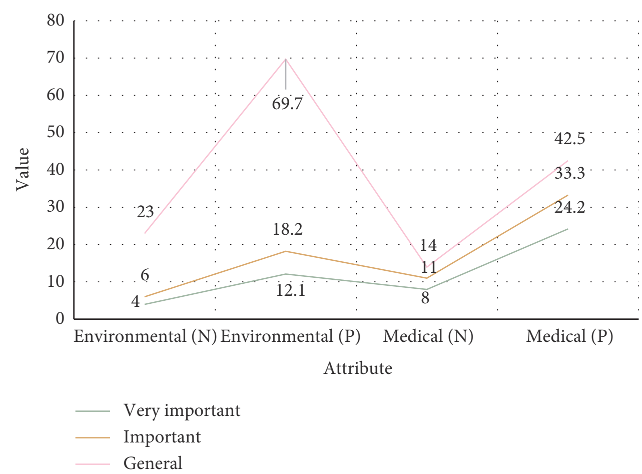

Figure 7: Diagram of the impact of other factors on sports injuries.

4.2.4. Other Factors. Other factors that cause injury are investigated, and the site factors and the incomplete medical supervision are analyzed. The results are shown in Figure 7.

It can be seen from Figure 7 that only 4 coaches and team doctors believe that sports injuries caused by venue and environmental factors are particularly important, accounting for 12.1%; while 23 people believe that they are generally important, accounting for 69.7%. After analysis, it can be seen that there are fewer sports injuries due to site and environmental factors, but the influence of factors such as the surrounding environment and weather cannot be neglected. In addition, some training equipment is relatively slow to update; some have been damaged; or the mood caused by the weather is not happy, which can easily cause sports injuries. 8 coaches and team doctors thought that sports injuries caused by incomplete medical insurance were the most important, accounting for 24.2%; 11 thought it was more important, accounting for 33.3%; 14 thought it was generally important, accounting for 42.5%. Perfect medical security is an important factor to protect athletes from sports injuries. Players can easily ignore minor injuries that occur under the condition of incomplete medical supervision, and if they will not receive timely treatment, the probability of injury will increase accordingly.

5. Conclusions

This article analyzes the characteristics and causes of sports injuries of football players and uses a text classification algorithm based on machine learning to establish a simple prediction model of sports injuries. The selection of data fitting methods takes full account of the complexity of sports injuries and influencing factors. The relationship was finally determined. The experimental results in this article show that the athlete's sports injury prediction model has better predictive capabilities and provides better methods for sports trainers, coaches, and physical therapists to optimize training courses, so as to prevent injuries and improve training levels and reduce rehabilitation costs. Innovation in this article is the bold extraction of features and classification of situations where sports injuries may occur through the form of text preprocessing to achieve predictive effects. The disadvantages of this article are that the number of models is small, and more research data is needed in a real sports injury scenario to complete and validate a predictive model. In addition, the number of samples has been increased to improve accuracy.

Data Availability

The data underlying the results presented in the study are included within the manuscript.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] X. Zhang, B. Wu, L. Dong et al., "Application of spark parallelization technology in architectural text classification," *Journal of Computational Methods in Sciences and Engineering*, vol. 18, no. 5, pp. 1–14, 2018.

[2] A. Adel, N. Omar, M. Albared et al., "Feature selection method based on statistics of compound words for Arabic text classification," *The International Arab Journal of Information Technology*, vol. 16, no. 2, pp. 178–185, 2019.

[3] B. Janani and M. S. Vijayarani, "Artificial bee colony algorithm for feature selection and improved support vector machine for text classification," *Interlending & Document Supply*, vol. 47, no. 3, pp. 154–170, 2019.

[4] P. L. Prasanna and D. Rao, "Text classification using artificial neural networks," *International Journal of Engineering & Technology*, vol. 7, no. 1, pp. 603–606, 2018.

[5] R. P. Rajeswari, K. Juliet, and A. Hana, "Text classification for student data set using naive bayes classifier and KNN classifier," *International Journal of Emerging Trends & Technology in Computer Science*, vol. 43, no. 1, pp. 8–12, 2017.

[6] S. Bahassine, A. Madani, and M. Kissi, "Arabic text classification using new stemmer for feature selection and decision trees," *Journal of Engineering Science and Technology*, vol. 12, no. 126, pp. 1475–1487, 2017.

[7] R. Pisot, U. Marusic, and B. Simunic, "Sports injury model for effective prevention, diagnostic and rehabilitation," *Medicine & Science in Sports & Exercise*, vol. 52, no. 7S, p. 1057, 2020.

[8] Z. Jiang, B. Gao, Y. He, Y. Han, P. Doyle, and Q. Zhu, "Text classification using novel term weighting scheme-based improved TF-IDF for Internet media reports," *Mathematical Problems in Engineering*, vol. 2021, no. 6, 30 pages, Article ID 6619088, 2021.

[9] H.-J. Nam and E. Chang, "A study on the improvement guide for sports injury management system: an analysis of survey on the management, satisfaction of facilities and job satisfaction of athletic trainers," *Exercise Science*, vol. 29, no. 3, pp. 291–299, 2020.

[10] M. Hildingsson, U. T. Fitzgerald, and M. Alricsson, "Perceived motivational factors for female football players during rehabilitation after sports injury - a qualitative interview study," *Journal of Exercise Rehabilitation*, vol. 14, no. 2, pp. 199–206, 2018.

[11] M. Grygorowicz, M. Michalowska, and T. Piontek, "Appraisal of the functional movement screen in football injury prediction," *British Journal of Sports Medicine*, vol. 51, no. 4, pp. 325-326, 2017.

[12] D. Janueviius, A. Sniekus, M. Mickeviius et al., "Integration of high velocity elastic band for hamstring training in pre-season routine of football players," *Baltic Journal of Sport and Health Sciences*, vol. 4, no. 119, pp. 31–39, 2021.

[13] B. Wojciechowska-Maszkowska, D. Borzucka, and A. M. Rogowska, "Comparison of balance skills, personality, and temperament of elite sports athletes and football players," *Journal of Physical Education and Sport*, vol. 20, no. 6, pp. 3671–3683, 2020.

[14] H. Silvers-Granelli, M. Bizzini, A. Arundale et al., "Does higher compliance to the fifa 11+ injury prevention program improve overall injury rate in male soccer (football) players?" *British Journal of Sports Medicine*, vol. 51, no. 4, pp. 382–390, 2017.

[15] R. Izzo, M. Giovannelli, and T. D'Isanto, "Original Article the injury prevention program WTA functional primitive movement in professional football players: a case study," *Journal of Physical Education and Sport*, vol. 19, pp. 1885–1889, 2019.

[16] A. Carapinheira, P. Mendes, P. G. Carvalho et al., "Sports career termination in football players: systematic review," *Revista Iberoamericana de Psicologia del Ejercicio y el Deporte*, vol. 14, no. 1, pp. 61–65, 2018.

[17] F. Guo, R. Fei, Y. M. Wang et al., "Screening of different metabolites in teenage football players after exercise fatigue," *Zhongguo Ying Yong Sheng Li Xue Za Zhi Zhongguo Yingyong Shenglixue Zazhi Chinese Journal of Applied Physiology*, vol. 36, no. 5, pp. 465–470, 2020.

[18] P. Kundu, "Impact of motivation on anxiety and tactical knowledge of young football players," *International Journal of Physical Education & Sports Sciences*, vol. 13, no. 2, pp. 57–61, 2018.

[19] D. J. Richards, "Comparison of SCAT-3 baseline testing and P300 ERP between seasons in university football players," *Western Undergraduate Research Journal Health and Natural Sciences*, vol. 8, no. 1, pp. 1–11, 2017.

[20] S. Mullainathan and J. Spiess, "Machine learning: an applied econometric approach," *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 87–106, 2017.
[21] A. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2017.
[22] C. Helma, T. Cramer, S. Kramer et al., "Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds," *The Journal for Chemical Information and Computer Scientists*, vol. 35, no. 4, pp. 1402–1411, 2018.
[23] J. X. Wang, J. L. Wu, and H. Xiao, "Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data," *Physical Review Fluids*, vol. 2, no. 3, pp. 1–22, 2017.
[24] H. Song and M. Brandt-Pearce, "A 2-D discrete-time model of physical impairments in wavelength-division multiplexing systems," *Journal of Lightwave Technology*, vol. 30, no. 5, pp. 713–726, 2012.
[25] C. Voyant, G. Notton, S. Kalogirou et al., "Machine learning methods for solar radiation forecasting: a review," *Renewable Energy*, vol. 105, no. MAY, pp. 569–582, 2017.
[26] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, "Machine learning on big data: opportunities and challenges," *Neurocomputing*, vol. 237, pp. 350–361, 2017.
[27] K. Shankar, M. Elhoseny, S. K. Lakshmanaprabu et al., "Optimal feature level fusion based ANFIS classifier for brain MRI image classification," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 1, Article ID 24887, 2020.