Research Article

Design and Analysis of a Smart Sensor-Based Early Warning Intervention Network for School Sports Bullying among Left-Behind Children

Panhong Ren and Mengjian Nie

1Sanquan College of Xinxiang Medical University, Xinxiang 453003, China
2Henan Normal University, Xinxiang 453000, China

Correspondence should be addressed to Mengjian Nie; 2016029@htu.edu.cn

Received 17 May 2022; Revised 27 June 2022; Accepted 1 July 2022; Published 18 July 2022

This paper constructs a sports bullying early warning intervention system using smart sensors to conduct in-depth research and analysis on early warning intervention of school sports bullying behaviors among left-behind children. Unlike daily behavior recognition based on motion sensors, school sports bullying actions are very random and difficult to be described by a specific motion trajectory. For the characteristics of violent actions and daily actions, action features in the time and frequency domains are extracted and action categories are recognized by BP neural networks; for complex actions, it is proposed to decompose complex actions into basic actions to improve the recognition rate. The algorithm of combining action features and speech features to achieve violence recognition is proposed. For the complexity of audio data features, this paper firstly preprocesses the audio data with preweighting, framing, and windowing and secondly extracts the MFCC feature parameters from the audio data and then builds a deep convolutional neural network to design the violence emotion recognition algorithm. The simulation results show that the algorithm effectively improves the accuracy rate of violent action recognition to 91.25% and the recall rate of violent action recognition to 92.13%. Finally, the LDA dimensionality reduction algorithm is introduced to address the problem of the high complexity of the algorithm due to the high number of feature dimensions. The LDA dimensionality reduction algorithm reduces the number of feature dimensions to 7 dimensions, which reduces the system running time by about 52% and improves the recognition rate of specific complex actions by about 12.1% while ensuring the overall system performance.

1. Introduction

In recent years, school bullying has intensified on primary and secondary school campuses around the world, resulting in many tragedies that seriously affect the physical and mental health growth of most of the youth population and challenge the harmony and stability of campuses [1]. The many news reports of school bullying incidents make it easy to see that areas with a high incidence of school bullying are often areas where quality educational resources are lacking and where children are left behind. For many years, primary and secondary schools and education administration have been working to improve school safety and trying to reduce school bullying, but in recent years, many vicious school bullying incidents have still been frequently seen in the media and have caused adverse repercussions in the community. School bullying is a serious threat to the safety of primary and secondary school students, and the lack of family education for parents of children left behind makes their school safety even more of a concern [2]. According to statistics, nearly 50% of left-behind children in school have been bullied in school. This shows that the problem of school bullying among left-behind children has become more serious and needs to be strengthened. However, there is still not enough attention paid to school bullying among left behind children, and there are few measures to address the problem
in different cities. To ensure the school safety of left-behind children, it is necessary to make a targeted study on this group of children subjected to school bullying; explore the current situation, problems, and causes of school bullying among left-behind children; and propose certain preventive and governance measures.

In recent years, inertial sensors have made significant development and played a vital role in the field of pattern recognition, but due to their high cost and limited accuracy requirements, they have only been used in the military field. With the development of the Micro-Electro-Mechanical System (MEMS), the development and application of MEMS microsensors have also been promoted, which combines moderate accuracy, small size, low cost, and easy mass production, allowing inertial sensors to be used in areas other than military. Inertial sensors can be portable, and accurate capture of human motion information and the use of single or multiple inertial sensors fixed in different parts of the human body for human motion data collection and for the motion detection research field opened a new continent [3]. Compared with another major branch of motion recognition-video image-based recognition, inertial sensors based on several advantages: video images are limited by geographic location, and there is a monitoring dead angle problem; motion sensors are easy to carry, low cost, configured on the human body to achieve all-weather dead angle monitoring; in addition, in data processing, compared to processing motion data, image information on equipment performance requirements is high and occupies larger memory; the later hardware implementation is more difficult [4]. Among the inertial sensors, especially, the contribution of accelerometer and gyroscope is the most prominent; its convenient human-computer interaction advantage makes it quickly favored by many researchers; at the same time, this kind of sensor can help various wearable devices capture motion signals and facilitate data access based on the field of motion recognition, so it has developed rapidly.

In recent years, there has been a growing trend of school violence on elementary and middle school campuses, as reported by the news media. Minor forms of school violence include verbal abuse, pushing, and shoving, while serious forms of school violence include beating and abuse. Bullying in schools is a serious problem that makes many students afraid of school life, and verbal bullying and threats can cause great psychological harm to immature minors, resulting in various psychological problems and seriously affecting students’ physical and mental health and academic growth [5]. A common problem in many school bullying incidents in recent years is that students who are bullied are afraid to inform their parents or teachers about the bullying because they are afraid or threatened, thus causing an indelible psychological shadow on the child’s young mind; in addition, because violence is not exposed, the perpetrators become increasingly rampant, which is also an important factor leading to the increase of school bullying. Based on this situation, the study of proactive detection of school bullying was born [6]. This project proposes a system that uses sensors on students such as accelerometers and gyroscopes to collect and process student data, determine whether school bullying has occurred, and proactively and timely transmit the detection results to teachers or parents. This proactive detection avoids students suffering from bullying without being detected and can solve the violence problem in a timely and effective manner. This project is of great significance to conduct fundamental research on bullying detection algorithms from the action perspective using motion sensors.

2. Related Works

The term “bullying” is a common concept of school violence in China, also known as “bullying,” with the meaning of bullying the weak. The earliest conceptual definition of bullying was made by Professor Smith of Smith College, the University of London, who considered bullying to be a special kind of aggressive behavior, and therefore has the basic characteristics of aggression and is a “subset” of aggression [7]. Bullying is a repeated attack by a relatively stronger party on a relatively weaker party without provocation. Bullying is a deliberately aggressive behavior that is perpetrated by a stronger individual against a weaker individual. According to Norwegian education scholar Jiménez-Barbero et al., bullying occurs when a student is exposed to prolonged and regular bullying or harassment led by one or more others. Some scholars also define the concept of school bullying in terms of the forms of school bullying [8]. Wen et al. argue that school bullying refers to persistent and repeated negative behaviors by one or more students against other individuals or groups in a school context [9]. These negative behaviors can be verbal such as teasing and nicknames and physical contact such as hitting, kicking, and pushing, as well as spreading rumors, isolating others in relationships, etc. Regarding the investigation of school bullying behavior, most scholars have used questionnaires to investigate the reality of school bullying behavior. Foreign scholars, represented by Jiménez-Barbero et al., were the first to investigate tens of thousands of primary and secondary school students aged 8-16 in Norway and found that about 15% of students were involved in school bullying “two or three times a month” or more frequently, with the bullied accounting for about 9% and the bully for about 7%. The problem of school bullying of left-behind children has become very serious. To ensure the healthy growth of primary and secondary school students, the problem of school bullying needs to be strengthened urgently. Schools, society, and families must pay great attention to it.

Dake et al. used a cell phone and a three-axis accelerometer to determine and detect different mobile phone-related actions, such as taking the phone out of a clothing pocket, controlling it for dialing, and putting it in a location, and other everyday actions [10]. An online gesture recognition system using accelerometers can recognize up to 95% of 16 different gestures. Fall detection recognition algorithms for visual aspects are investigated for the problem that bullies may be knocked down. Among them, Rigby and Smith propose a class of video-based systems that can automatically recognize human fall actions, which only uses self-set thresholds for two features, tilt angle and human aspect ratio, to identify whether a fall has occurred [11]. This scheme is
simple to implement and has low complexity, so it can be run on modest computing devices, but the system also has many drawbacks, such as the tilt angle and human aspect ratio. Olweus designed a fall detection system that takes a monocular 3D camera to track the human head [12]. After obtaining the estimated head position, these data are transmitted to the fall detection module to calculate the vertical and horizontal velocities of the head, respectively, and finally, the results are compared with a predefined threshold to determine whether the head has fallen. In terms of sensor-based action recognition research, the most typical one is using a smartphone to obtain the signal from the acceleration sensor and designing an improved human action recognition method based on it, which reduced the complexity of the traditional human action recognition algorithm and increased the recognition accuracy. In terms of real-time human action recognition technology based on high accuracy, Lam and Yeoh designed a human action recognition algorithm for acceleration time-domain features [13].

At present, foreign research on on-campus bullying detection is still dominated by passive research. The so-called passive approach requires the aggressor to manually trigger the detection device; however, in the actual operation process, it is almost impossible for the aggressor to have time to trigger the operation, which reduces the practicality of the system [14]. The action recognition for school bullying detection studied in this paper breaks through the previous idea of purely theoretical research in domestic school bullying prevention research and develops recognition algorithms from a practical perspective that can be practically applied to students’ school life [15]. As far as recognition algorithms are concerned, violent actions are more intense than ordinary actions, and the orientation of the sensor may change constantly during the action. Therefore, the feature extraction based on the time domain in most of the previous action recognition is limited in the process of violent recognition. Therefore, this paper will focus on modeling and analyzing the action process under violent action, comparing the difference between violent action and ordinary action, profiling the violent action process, focusing on improving the sensitivity of violent action recognition, and preparing for future hardware implementation by proposing a transition point detection dormancy algorithm to reduce the time and space complexity of system operations. As a school administrator, we must first teach students to establish correct values, so that students can fully understand the harm caused by campus bullying to others and themselves. Secondly, there must be good hardware and software protection measures to prevent the occurrence of school bullying. Finally, teachers must detect abnormal reactions of students to prevent the occurrence and recurrence of school bullying.

3. Multisensor-Based Action Feature Extraction

To investigate the occurrence of school bullying, this study analyzed the questionnaire "At what time did you experience school bullying A. during school; B. in class; C. between classes; D. after school." To have a comprehensive understanding of the occurrence of school bullying, I conducted a crosstabulation analysis on the occurrence of school bullying among junior high school students by grade level, gender, and geography and used interviews to supplement and refine them. As shown in Table 1, the authors found that school bullying occurred mostly during recess, after school, and during school hours, with recess being the most frequent period for each grade level and classroom being the least frequent period for each grade level. In addition, the number of students experiencing bullying and the number of periods when they experience bullying tend to decrease as students’ cognition increases with grade level. Compared with other actions, motion capture can be performed by setting a smaller sliding window. Even if a capture is missed, its duration is often long, so it will not have much impact on the overall recognition performance.

In school bullying, the number of boys who have a bullied behavior is significantly higher than the number of girls, which has a great correlation with the personality characteristics of men and women. Boys often use gang fights and other attacks, while girls mostly use verbal abuse, cliques, isolation, and in severe cases, hair pulling, slaps, and videotaping.

As shown in Figure 1, the number of students who engage in school bullying is significantly higher than the number of girls, which is inextricably linked to the personality characteristics of male and female students; boys are nurtured by the traditional “martial arts”; they are more aggressive, advocate violence, and mostly use physical attacks, while girls are softer and more delicate and more on verbal bullying and relational bullying. In addition, both boys and girls choose to bully mostly between classes, followed by after school, during school hours, and in the classroom. I found during my visits that most of the girls during recess were in small groups of three or five and that individual conflicts among students were easily transformed into group conflicts, thus laying the groundwork for bullying to occur on campus.

The project team conducted a large amount of data collection through experiments on a variety of daily and violent movements of students, using “jumping,” “playing,” “running,” “walking,” “standing,” “falling,” “pushing,” “falling,” “pushing,” “walking,” “standing,” “falling,” “pushing,” and “pushing down.” The seven categories of action data, “assault”, are used in the final classifier for action recognition. Among these seven categories, “jumping,” “playing,” “running,” “walking,” “standing,” and “falling” are nonviolent data, while “pushing,” “pushing,” and “beating” are nonviolent data. The final accuracy of the test will be measured uniformly by violence or not. The video image is limited by geographical location, and there is a problem of monitoring blind spots; the motion sensor is easy to carry, low cost, and can be configured on the human body to achieve all-weather monitoring without blind spots. At the same time, the data processing of the sensor is more dependent on graphic information.

The initialization stage of data lifting is the initial detection of skeletal points and limbs; due to the different detection difficulties of different skeletal points and limbs, the previous stage identifies the more obvious skeletal points;
some skeletal points are relatively complicated to detect due to clothing, obscuring, etc.; and the subsequent refinement stage is needed to optimize the detection results progressively.

$$S_T = p_t(F, S_{T+1}, L_{T-1}), \quad t \geq 1, \ 0 < T \leq 1.$$  \hspace{1cm} (1)

The refinement stage is to use the feature map $F$ obtained by fusing the output results of the previous stage with the feature extraction stage as the input of the next stage, which is used to improve the detection results step by step. There are two loss functions. Loss is used in each stage, and the loss functions $L_1$ and $L_2$ are used to calculate the difference between the predicted and true values, respectively. $L_2$ is used to calculate the deviation of the minimum absolute value between the two values, and $L_2$ is used to calculate the minimum squared error. $y_i^*$ represents the true labeled value, and $y_i$ represents the predicted value.

$$L_1(y, y^*) = \sum_{i=0}^{n} (y_i^* + y_i),$$

$$L_2(y^*, y) = \sum_{i=0}^{n} |y_i - y_i^*|^2.$$  \hspace{1cm} (2)

The audio data samples in this paper consisted of Finnish speech data, public CASIA data, and a small database of self-recorded speech. The Finnish speech database consists of 82 bullying speech data samples and 81 nonbullying speech data samples; the public CASIA speech database contains audio samples of 6 emotions: angry, scared, happy, neutral, sad, and surprised, with 200 audio samples of each emotion; the self-recorded small speech database contains 50 "angry," 42 "scared," 81 "happy," 115 "neutral," 51 "sad" emotions, and 103 "surprised" emotions. During school bullying, the emotion of the abuser is often "angry," while the emotion of the abused is often "scared" and "sad." Therefore, for the public CASIA voice database and the small voice database recorded by ourselves, this paper classifies "angry," "scared," and "sad" as bullying emotions and "happy," "sad," and "happy" as bullying emotions. "Happy," "neutral," and "surprised" were classified as non-bullying emotions.

Many studies on the hearing mechanism of the human ear share a common finding, which can be briefly described as the human ear has different auditory sensitivity to different frequencies of sound waves; when the frequency of the speech signal is between 200 Hz and 5000 Hz, the frequency has the greatest effect on the clarity of speech. When two sounds of unequal loudness act on the human ear at the same time, the two sounds will affect each other, and the louder sound will cover the lower sound so that the lower sound is not easily detected by the human ear, which is often referred to as the masking effect. The specific reason is that sound propagates in the form of waves in the basilar membrane of the human cochlea, and the sound with a lower frequency travels a greater distance than the sound with a higher frequency, while the sound with a higher loudness

| Grade       | During school | In class | Bullying time | Between classes | After school |
|-------------|---------------|---------|---------------|-----------------|-------------|
| First day   |               |         |               |                 |             |
| Number      | 30            | 12      | 20            | 41              |             |
| Proportion  | 29.13%        | 11.65%  | 19.42%        | 39.81%          |             |
| Second grade|               |         |               |                 |             |
| Number      | 23            | 10      | 12            | 35              |             |
| Proportion  | 28.75%        | 12.50%  | 15.00%        | 43.75%          |             |
| Third grade |               |         |               |                 |             |
| Number      | 45            | 16      | 28            | 53              |             |
| Proportion  | 31.69%        | 11.27%  | 19.72%        | 37.32%          |             |
has a lower frequency, and the sound with a lower loudness has a higher frequency, so the sound with a higher loudness travels a greater distance in the human ear.

Mel-scale Frequency Cepstral Coefficients (MFCC) is specific only in the Mel-scale frequency domain (Mel-scale frequency), which is a frequency domain parameter. The relationship with frequency can be approximated by the following equation:

\[ \text{Mel}(f) = \log_{10} f, \]

where \( f \) denotes frequency in hertz. Figure 2 shows the Mel frequency versus linear frequency. In the relationship between Mel frequency and linear frequency in this paper, a reference value is set, and it can be directly seen that with the increase of frequency, Mel frequency increases linearly.

Preprocessing of audio data starts with preemphasis, which is essentially passing the speech signal through a first-order high-pass filter \( \text{FIR} \) to increase the high-frequency portion of the speech signal, which is represented as follows:

\[ s(n) = s'(n) + as(n + 1)^{\sqrt{2}}. \]

Preemphasis of the speech signal can achieve the following two purposes: (1) the sound gate pulse will cause the high-end spectrum amplitude to decrease, and the preemphasis process not only successfully solves this problem by increasing the zero point but also makes the spectrum of the speech signal very flat, and the amplitude of the resonance peaks will become similar; (2) the first-order high-pass filter used in the preemphasis process makes the high frequency in the speech increase and the low-frequency decrease, so that not only can the dynamic spectrum of the speech signal be narrowed but also can weaken the interference of the larger fundamental frequency amplitude on the resonance peak detection.

\[ x_i^c = \sum_{n \in H_i} x_i \times k_i^{n-1}, \]

\[ W = (1 + \alpha)^{(N-1)} - \cos \alpha. \]

The energy curve and the over-zero rate curve of the speech signal fluctuate in the voiceless region, and this problem can be solved by smoothing, but this method is only effective in the case of a low signal-to-noise ratio. The endpoint detection successfully intercepts the voiced region of the speech signal, and the number of wild points in the waveform of the speech signal after the median smoothing process is significantly reduced, which successfully intercepts the voiced segment of the speech signal to prepare for the next step of feature extraction.

4. Multisensor Motion Recognition and Fusion-Related Technologies

It is well documented that the placement of wearable sensors has a direct impact on the recognition of body movements. The ideal placement of sensors in different application scenarios is still controversial. To ensure accurate early warning of school bullying, this study is based on multisensor fusion recognition, so the location and number of sensors are critical, and research has shown that the location and number of sensors built into wearable devices have a direct impact on recognition results [16]. At present, most human body movements are recognized at the waist, legs, wrists, and chest. In addition, for the multisensor fusion recognition system, there are still a variety of combinations of different positions to choose the results; for the waist, legs, wrist, and other positions of the combination, a lot of research has been made; its recognition action is mainly “lying,” “sitting,” “standing,” “walking,” “running,” “going upstairs,” “going downstairs,” “slow motion,” and “walking downstairs,” “jogging,” “jumping,” and other daily movements. The comparison shows that the performance of action recognition based on a single part sensor is significantly inferior to that of multisensor fusion recognition, and the recognition performance can be improved by fusing different parts. At the same time, the study shows that the number of sensors does not increase the recognition accuracy, but the balance of convenience and recognition rate should be considered. In this project, the data was collected by wearing sensors on the waist, legs, and wrists, and combined with the current research, it was decided to use three parts of the hands, waist, and legs to fuse sensors to detect violent movements in school.

The motion recognition system mainly consists of the following parts: data acquisition and processing, feature extraction and selection, classifier design, etc. Data acquisition and preprocessing refer to the use of sensors such as accelerometers and gyroscopes to capture human motion, thus converting human motion into a signal that can be recognized by the computer and performing preliminary preprocessing. Multisensor fusion can capture the neglected features of things by a single sensor, so the performance of the fused information in terms of accuracy, reliability, and robustness is substantially improved, which is the advantage of multisensor fusion. There are also some issues in a multisensor fusion that need further research to solve, such as reliability, synchronization, integrity, and fusion conflicts of measurement data. In addition, different multiple sensors can generate different noises, which also need to be carefully analyzed and handled. The information fusion process of a multisensor fusion model composed of multiple sensors is shown in Figure 3.

The current common idea of data segmentation is to set a sliding window, where the data collected by the sensor is stored in the form of a data stream, and by choosing a suitable window length, it can be slid in the time axis of the data stream according to a certain ratio and speed [17]. The choice of the window length is crucial. If the sliding window length is too long, the amount of data processed per interval
is large and the operation speed will be reduced. If the window length is chosen too short, the amount of data processed per interval is not enough and the motion state of a certain action cannot be adequately represented, which leads to a lower recognition rate [18]. The sliding window length is set by observing the data of different movements. In this experiment, a total of “jumping,” “running,” “standing,” “playing,” and “walking” were included. Since running,
standing, and walking are periodic movements, the cycles are short. Therefore, compared to other actions, the action can be captured by a smaller sliding window, and even if capture is missed, the overall recognition performance will not be greatly affected because their duration is often long. While "play," "hit," etc. do not have a fixed cycle and the action is more complex, lacking a certain degree of regularity, so first observe the "push" and "jump" action cycle. If "hitting" is used as a type of action, it is likely to be confused with "standing" and "running." Fortunately, in the course of the experiment, it was found that in the process of "beating," the bullied often had some special action responses, such as rolling over and squatting to protect themselves.

To consider that the sliding window can intercept effective data and not cause too cumbersome data processing and for the subsequent extraction of frequency features, the length of the sliding window in this paper is set to 256, and since the sampling rate of the sensor is 50 Hz, it corresponds to the time length of 5.12 s for each sliding window; according to common sense, 5.12 s can fully reflect the playing and being hitting action. In addition, since the subsequent feature extraction needs to be analyzed in the frequency domain, the design length of the sliding window is chosen as a power of 2 for the convenience of time-frequency conversion. The sliding window was added to the data as shown in Figure 4.

As the data collected by the sensor contains random noise, jitter noise, etc., the data needs to be filtered. Wavelet filtering is based on the different scales of signal and noise to construct the corresponding rules, in the wavelet domain to process the intensity coefficient of noise to reduce or even completely remove the impact of noise while trying to ensure the original useful signal [19]. The primary purpose of wavelet denoising is not smoothing, but trying to remove all the noise and do not consider the frequency range of the noise, according to the different mother functions of the wavelet transform; the transform results are also different; its use should be based on the target selection of the required wavelet basis function; wavelet basis selection principles are based on the support length, similarity, vanishing moments, symmetry, regularity, and other properties. Therefore, a second-order Butterworth filter is used for filtering.

The wavelet denoising method includes three basic steps: perform wavelet transform on the noise-containing signal, perform some processing on the transformed wavelet coefficients to remove the noise contained in them, and perform inverse wavelet transform on the processed wavelet coefficients to obtain the noised signal [20].

Specifically, the success of the wavelet denoising method is mainly due to the following characteristics of the wavelet transform:

1. Low entropy: the sparse distribution of wavelet coefficients reduces the entropy after image transformation

Figure 4: Schematic diagram of sliding window data segmentation.
(2) Multiresolution: due to the multiresolution method, the nonstationary features of the signal, such as edges, spikes, and breakpoints, can be well described.

(3) Decorrelation: because the wavelet transform can decorrelate the signal and the noise tends to whiten after the transformation, so the wavelet domain is more conducive to denoising than the time domain.

(4) Flexibility of base selection: since wavelet transform can flexibly select transform bases, different wavelet mother functions can be selected for different application occasions and different research objects to obtain the best effect.

5. LDA Dimensionality Reduction Algorithm

The motion sensor features plus speech features have 31 dimensions. In this paper, we expect to transfer the algorithm to handheld devices in the future, and if the number of features is too large, it will increase the consumption of hardware resources, so we need to reduce the dimensionality of the feature vector. Since the features extracted from the feature vector have been screened by wrapper method, it is difficult to further screen out the retained features for the neural network classifier, so we can only reduce the dimensionality by projecting the feature vector into the subspace [9, 21]. Feature dimensionality reduction methods are divided into linear and nonlinear dimensionality reduction. Nonlinear dimensionality reduction is generally applicable to flow learning, and this paper is a neural net-based learning method with linear dimensionality reduction.

In this paper, we study multisensor action recognition based on the expectation that it can be implemented in hardware porting afterward. If the number of features used in the system is too high, it will lead to an increase in time complexity, so we introduce a dimensionality reduction algorithm here to reduce the feature dimensionality. Theyandexistence is to design a mapping function: \( f : x \rightarrow y \), where \( x \) represents the original data, which is often represented in vector form and \( y \) represents the new data after mapping to the low-dimensional space. Commonly used dimensionality reduction algorithms include Principal Component Analysis (PCA), Locally Linear Embedding (LLE), and Linear Discriminant Analysis (LDA). In this paper, we focus on LDA dimensionality reduction.

The LDA dimensionality reduction technique belongs to the category of supervised learning; in other words, it uses the labels of the dataset to produce a categorical output. The method tries to add more confidence to the classification by reconstructing the columns of the data. Specifically, it selects the feature subspace basis vector matrix \( W \) to maximize the ratio of the hash of the data points between categories to the hash within the data.

\[
E_j = \sum_{x \in X_j} (x + \mu_i)^T (x + \mu_i) \quad (i = 1, 2, \ldots, n). \tag{6}
\]

To derive the multiclass LDA dimensionality reduction principle, we first explain the two-class LDA dimensionality reduction principle, because it is two-class data, so after the dimensionality reduction, we only need to project the data to a straight line.

The high dimensionality of the features inevitably leads to an increase in the time complexity and space responsibility of the network. By reducing the features to 7 dimensions through the LDA dimensionality reduction algorithm, the number of neurons needed to reach convergence is greatly reduced and the required error is also reduced, which greatly reduces the time and space complexity of the classification calculation and improves the efficiency of the system, allowing the algorithm to run on resource-constrained terminal devices such as smartphones. The convergence performance is shown in Figure 5.

The analysis shows that LDA dimensionality reduction can reduce the feature dimension to 7 dimensions while ensuring the recognition rate, which greatly reduces the running time and spatial complexity of the system, and the action recognition conversion matrix shows that the recognition rate is higher for violent and complex actions and 12.1% for specific complex actions such as “fall” and “push.” The recognition rate for specific complex actions such as “fall” and “push” is improved by 12.1%. In addition, the analysis of the network structure model and convergence performance graph shows that the reduction of the feature dimension accelerates the convergence of the network and improves the system operation speed.

6. Analysis of Results

6.1. Experimental Simulation of Violence. By observing the experimental video, we found that the similarity of the performers’ actions during “fall” and “push” is very high, and it is very easy to be confused from the classification point of view, so the following will separately identify the actions without “fall/push” and with “fall/push.” Therefore, the following section identifies the action data without “fall/push down” and with “fall/push down” separately. The features are obtained by extracting and filtering the action features, and the number of neurons in the implicit layer of the neural network is set to 6, the transfer function of the implicit layer is log sig, and the transfer function of the output layer is purelin through multiple debugging [22, 23]. To improve the accuracy of the model, “push down” and “fall” actions are added in this section. However, for the fall action, the action process is short, as shown in Figure 6, which shows the change of acceleration sensor during the “push to” violent action.

Using “violent action” as the positive action, we obtained the following four performance indicators: “precision,” “accuracy,” “recall,” and “F1-measure”, “Recall” and “F1-measure”: precision 85.29%, accuracy 83.58%, recall 75.77%, and F1-measure 80.06%. This is partly because increasing the number of categories does lead to performance degradation. Thanks to the data analysis, the four categories of “hitting,” “pushing,” “pushing,” and “playing” have a higher conversion rate to recognize each other [24].
The conversion rate of mutual recognition is large and easily confused. So, there is a need to take certain measures to improve the system recognition performance. The “hitting” action is easily confused with some nonviolent actions. By observing the experimental data, we can see that the “hit” action is not simple but is composed of several kinds of actions. Therefore, if “striking” is considered a type of action, it is likely to be confused with “standing” and “running.” Fortunately, during the experiment, it was found that during the “hitting” process, the

\[
\begin{align*}
0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40
\end{align*}
\]

\[
\begin{align*}
-4, -2, 0, 2, 4
\end{align*}
\]

\[
\begin{align*}
0, 100, 200, 300, 400, 500, 600, 700
\end{align*}
\]

**Figure 5:** Relationship between the convergence error and the number of samples in the LDA dimensionality reduction network.

**Figure 6:** “Push down”-type action acceleration sensor change schematic.
bullied often reacted with some special movements, such as rolling on the ground and crouching down to protect themselves.

6.2. Joint Action-Verbal Bullying Identification. In the motion sensor-based classification algorithm, the “play” action is easily confused with violent action, and the “fall” action is very similar to “push,” to improve the differentiation of violent action. To improve the differentiation of violent actions, speech signals are introduced in this section [25]. The speech signal collected during the experiment is combined with the features extracted from the motion sensor, where the combined feature vector includes the 15-dimensional motion data features extracted in Section 3 and the 16-dimensional speech features extracted in Section 4, and the combined 31-dimensional features are used as the input of the neural network, and the implied layer is still a single layer with the transfer function of log sig, and the number of neurons in the implied layer is set to 9. The four types of actions, “roll,” “squat,” “push,” and “push down,” are used as violent actions, and the four types of actions, “jump” and “jump,” are used as violent actions. “Jumping,” “playing,” “running,” “standing,” “walking,” “jumping,” “playing,” “running,” “standing,” “walking,” and “falling” were classified as nonviolent actions, and a confusion matrix of violence recognition was obtained. In bullying, “hitting” is not a single action; it is accompanied by the perpetrator’s “cry” and “forward attack” and the victim’s “wailing” and “avoidance.”

Comparing the average conversion rate of action recognition without joint speech, we can see that the recognition rate of “play” action increased from 70.83% to 78.33%, and the recognition rate of “push down” as violent action increased from 36.00% to 78.00%. The probability of recognizing “pushing” as “falling” decreased from 44.00% to 10.00%. This indicates that due to the different speech environments of violent and nonviolent actions, the discrimination between violent and nonviolent actions with similar performance can be significantly improved by combining action features and speech features [26]. In addition, among the four classification performance metrics, although accuracy decreased from 91.09% to 89.42%, the decrease was only 1.67%. The remaining three metrics improved, especially for recall, which increased from 84.63% to 93.90%, a performance improvement of 9.27%. The increase in recall improves the sensitivity of the system to recognize violent actions and enhances the reliability of the system.

After building the structure of the speech emotion 6-classification neural network, the model was trained, and the training results are shown in Figure 7.

The model starts to overfit after 50 rounds of training when the training accuracy reaches 70% and the validation accuracy reaches 80%, and finally, after 200 rounds of training, the training accuracy reaches 97% and the validation accuracy reaches 80%. The model structure and parameters with the best performance are preserved in 200 rounds of training by monitoring the changes in the loss function values in the test set in this classification algorithm. When the frequency of the speech signal is between 200 Hz and 5000 Hz, the frequency has the greatest impact on the clarity of the speech.

The output result of the above speech emotion 6-classification algorithm represents the probability of showing 6 emotions in speech samples, and then, the recognition
result is converted into the probability of showing bullying emotion and nonbullying emotion in speech samples, and the conversion process increases the computation of the algorithm and reduces the recognition efficiency of the classification algorithm [27]. Therefore, in this paper, we design a speech emotion 2-classification neural network model based on the speech emotion 6-classification neural network model, and the output of this model represents the probability of showing bullying emotion and nonbullying emotion in speech samples without the conversion process of recognition results.

7. Conclusion

To achieve the detection of violent actions in school, we propose a multisensor data fusion-based algorithm for school violence detection, which involves three violent actions including “hitting,” “pushing,” and “shoving” and six daily behavioral actions. The proposed algorithm is based on multisensor data fusion, which includes three kinds of violent actions: “hitting,” “pushing,” and “shoving,” and six kinds of daily actions, and achieves a more comprehensive capture of human actions by fusing the sensor data of multiple body parts and improves the distinction between violent actions and daily actions by fusing them at the decision level to achieve redundancy and complementarity. Simulation results show that the multisensor data fusion algorithm can achieve an average recognition accuracy of 88.35% for display actions and a recall rate of 84.63% for violent actions. MFCC is selected as the feature parameter of speech data, and the speech emotion classification algorithm is designed based on successful feature parameter extraction, which is based on three different speech databases. The experimental results show that the classification algorithm has a correct recognition rate of 87.92% and an F1-score value of 80.95% on the homemade speech database. Although the classification of actions and speech recognition at the algorithm level is carried out in this paper, however, if we want to recognize actions effectively and accurately in practice, we still need to build a more complete hardware and software platform from data acquisition, action and cloud-implicit change judgment, feature extraction, and classification. It is believed that the relevant hardware and software facilities based on the school violence recognition algorithm will appear in the hands of millions of students soon.

Data Availability

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

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the project of training plan for key teachers of Sanquan College of Xinxiang Medical University (No. SQ2022GGJS10), Henan Education Science Planning Project (Effect of a school-based sport intervention on bullying behavior of rural left-behind children, No. 2021YB0068), and Henan Provincial Philosophy and Social Science Planning Project (No. 2021CTY031).

References

[1] M. Woźniak, M. Wieczorek, J. Ślka, and D. Polap, "Body pose prediction based on motion sensor data and recurrent neural network," IEEE Transactions on Industrial Informatics, vol. 17, no. 3, pp. 2101–2111, 2021.
[2] Z. Wu, W. Ding, Y. Dai et al., “Self-powered multifunctional motion sensor enabled by magnetic-regulated triboelectric nanogenerator,” ACS Nano, vol. 12, no. 6, pp. 5726–5733, 2018.
[3] F. Deng, J. Dong, X. Wang et al., “Design and implementation of a noncontact sleep monitoring system using infrared cameras and motion sensor,” IEEE Transactions on Instrumentation and Measurement, vol. 67, no. 7, pp. 1555–1563, 2018.
[4] A. Das, N. Borisov, and E. Chou, “Every move you make: exploring practical issues in smartphone motion sensor fingerprinting and countermeasures,” Proceedings on Privacy Enhancing Technologies, vol. 2018, no. 1, pp. 88–108, 2018.
[5] G. Zhao, S. Gong, H. Wang et al., “Ultrathin biocompatible electrospun fiber films for self-powered human motion sensor,” International Journal of Precision Engineering and Manufacturing-Green Technology, vol. 8, no. 3, pp. 855–868, 2021.
[6] A. Yurtman and B. Barshan, "Novel noniterative orientation estimation for wearable motion sensor units acquiring accelerometer, gyroscope, and magnetometer measurements," IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 6, pp. 3206–3215, 2020.
[7] G. Fellmeth, K. Rose-Clarke, C. Zhao et al., "Health impacts of parental migration on left-behind children and adolescents: a systematic review and meta-analysis," The Lancet, vol. 392, no. 10164, pp. 2567–2582, 2018.
[8] J. A. Jiménez-Barbero, A. Jiménez-Loaisa, D. González-Cutre, V. J. Beltrán-Carrillo, L. Llor-Zaragoza, and J. A. Ruiz-Hernández, “Physical education and school bullying: a systematic review,” Physical Education and Sport Pedagogy, vol. 25, no. 1, pp. 79–100, 2020.
[9] Y. J. Wen, W. P. Hou, W. Zheng et al., “The neglect of left-behind children in China: a meta-analysis,” Trauma, Violence, & Abuse, vol. 22, no. 5, pp. 1326–1338, 2021.
[10] J. A. Dake, J. H. Price, and S. K. Telljohann, “The nature and extent of bullying at school,” Journal of School Health, vol. 73, no. 5, pp. 173–180, 2003.
[11] K. Rigby and P. K. Smith, “Is school bullying really on the rise?,” Social Psychology of Education, vol. 14, no. 4, pp. 441–455, 2011.
[12] D. Olweus, “A profile of bullying at school,” Educational Leadership, vol. 60, no. 6, pp. 12–17, 2003.
[13] T. Lam and B. S. A. Yeoh, “Parental migration and disruptions in everyday life: reactions of left-behind children in Southeast
[14] L. Ding, L. W. Yuen, E. S. Buhs, and I. M. Newman, "Depression among Chinese left-behind children: a systematic review and meta-analysis," *Child: Care, Health and Development*, vol. 45, no. 2, pp. 189–197, 2019.

[15] K. Cui and X. Jing, "Research on prediction model of geotechnical parameters based on BP neural network," *Neural Computing and Applications*, vol. 31, no. 12, pp. 8205–8215, 2019.

[16] Y. Zhang, J. Tang, R. Liao et al., "Application of an enhanced BP neural network model with water cycle algorithm on landslide prediction," *Stochastic Environmental Research and Risk Assessment*, vol. 35, no. 6, pp. 1273–1291, 2021.

[17] P. K. Smith, K. Ananiadou, and H. Cowie, "Interventions to reduce school bullying," *The Canadian Journal of Psychiatry*, vol. 48, no. 9, pp. 591–599, 2003.

[18] L. Xu, J. Wang, H. Wang, T. Aaron Gulliver, and K. N. le, "BP neural network-based ABEP performance prediction for mobile Internet of Things communication systems," *Neural Computing and Applications*, vol. 32, no. 20, pp. 16025–16041, 2020.

[19] Y. Chen, "Prediction algorithm of PM2.5 mass concentration based on adaptive BP neural network," *Computing*, vol. 100, no. 8, pp. 825–838, 2018.

[20] B. Li, X. Tian, and M. Zhang, "Thermal error modeling of machine tool spindle based on the improved algorithm optimized BP neural network," *The International Journal of Advanced Manufacturing Technology*, vol. 105, no. 1-4, pp. 1497–1505, 2019.

[21] X. Geng, S. Lu, M. Jiang et al., "Research on FBG-based CFRP structural damage identification using BP neural network," *Photonic Sensors*, vol. 8, no. 2, pp. 168–175, 2018.

[22] J. Juan, L. Weijun, H. Guifeng, G. Xiaojing, C. Zhaoxia, and S. Li, "Prevalence of unintentional injury among left-behind children in mainland China: evidence from epidemiological surveys," *Child: Care, Health and Development*, vol. 47, no. 3, pp. 387–399, 2021.

[23] S. Zhang and H. Li, "The construction of an action-speech feature-based school violence recognition algorithm and occupational therapy education model for adolescents," *Occupational Therapy International*, vol. 2022, 11 pages, 2022.

[24] Z. Ballard, C. Brown, A. M. Madni, and A. Ozcan, "Machine learning and computation-enabled intelligent sensor design," *Nature Machine Intelligence*, vol. 3, no. 7, pp. 556–565, 2021.

[25] V. Bijalwan, V. B. Semwal, and T. K. Mandal, "Fusion of multi-sensor-based biomechanical gait analysis using vision and wearable sensor," *IEEE Sensors Journal*, vol. 21, no. 13, pp. 14213–14220, 2021.

[26] T. Vaillancourt, H. Brittain, A. Krygsman, A. H. Farrell, S. Landon, and D. Pepler, "School bullying before and during COVID-19: results from a population-based randomized design," *Aggressive Behavior*, vol. 47, no. 5, pp. 557–569, 2021.

[27] D. Fraguas, C. M. Díaz-Caneja, M. Ayora et al., "Assessment of school anti-bullying interventions," *JAMA Pediatrics*, vol. 175, no. 1, pp. 44–55, 2021.