International Journal of Information Technology and Web Engineering
Volume 18 • Issue 1

Thinking on Construction of Intelligent Auxiliary Physical Exercise Mode Under National Fitness Plan

Cheng Xu, Shao Xing University, China*

ABSTRACT

In recent years, with the increase of work intensity and life pressure, many people are suffering from unhealthy conditions. In this case, it is very necessary for those who care about health to maintain and enjoy an active and healthy life. Therefore, many people tend to improve their health status through nationwide fitness and exercise. With the development of electronic technology, more and more people use smart devices to monitor their movement. In view of the difficulties and challenges people are facing, this paper will propose a multi-user behavior monitoring algorithm based on machine learning, which can improve the frequency and enthusiasm of physical exercise in people's daily life, thus improving the overall health of society. The research on action recognition responds to the call of the World Health Organization and adapts to social development, which has great practical significance.

KEYWORDS

Development Strategy, Intelligent Sports, Machine Learning, National Fitness, Physical Exercise

INTRODUCTION

People are paying more and more attention to their health, and maintaining a good body and mind will enable more people to have positive emotions and attitudes toward life, thus enhancing their happiness and inner confidence (Yang, 2021). More importantly, a strong body will provide the basic security for learning and living. However, diseases, injuries, and unhealthy lifestyles pose challenges to maintaining physical fitness, and people also experience lower quality of life owing to physical limitations caused by injuries and diseases (Feng et al., 2022). In addition, lack of exercise can lead to weakness and obesity. The purpose of a national fitness campaign is to comprehensively improve the physical fitness and health of the nation. It is not only an important symbol of the progress of social civilization but also an important embodiment of comprehensive national power. It also is one of the main contents of spiritual civilization construction. Especially in the era of big data, the construction of intelligent-assisted physical exercise based on national fitness has an important role in promoting the development of a harmonious society (Wang & Zhang, 2022). For the aforementioned situation, physical exercise will be a good medicine for health.

DOI: 10.4018/IJITWE.331080
*Corresponding Author

People are under a lot of pressure at work and need to deal with an abundance of complex information. They have to work for long hours, thus seriously affecting their health (Lake et al., 2017). This situation highlights the importance of fitness and exercise at the right time.

However, fitness and physical exercise need to follow a certain order; otherwise, they not only fail to yield the benefit of a strong body but also may cause secondary damage to the body. For example, some athletes engage in prolonged repetitive load training, and prolonged, excessive load training may increase the risk of injury (Li & Wang, 2021). Some players who undergo ligament reconstruction perform low-intensity recovery procedures in postoperative rehabilitation, but studies have shown that performing high-intensity reps is more beneficial for prolonged recovery. An investigation of the causes of sports injuries among college students revealed that more than 40% of the sports injuries among college students were due to injuries caused by a lack of relevant sports knowledge (Li & Zhang, 2021). People often lack sufficient knowledge about the correct way to play sports and exercise, and they tend to rely on subjective perceptions to carry out the relevant sports activities. Therefore, guidance from professionals is essential in the process of physical activity. Whether in the field of fitness, rehabilitation training, or physical education, the correct exercise under the guidance of professionals is an important guarantee to improve the training effect and prevent sports injuries.

However, obtaining guidance from professionals is more difficult for people when they are performing physical activity. For this reason, the paper proposes a movement recognition model to help people monitor their body movements during exercise and improve the problem of difficulty in obtaining professional guidance to some extent.

Action recognition, also known as activity recognition, is a field in computer vision and artificial intelligence that involves recognizing and classifying the action or activity being performed by a person or object in a video or image sequence (Chaquet et al., 2013). The goal of action recognition is to develop algorithms and models that can automatically analyze visual data and determine what action or activity is occurring. For example, in video clips, action recognition algorithms aim to recognize actions, such as walking, running, jumping, waving, exercising, and cooking. This technology has applications in various fields, including surveillance and security, healthcare, motion analysis, human-computer interaction, autonomous systems, and entertainment (Kulsoom et al., 2022).

In recent years, human motion visual analysis has become a frontier in the field of computer vision. Human motion visual analysis detects, recognizes, and tracks the human body from a sequence of images of the human body and obtains motion parameters to further explain and describe human behavior (Poppe, 2007). It falls under the category of image analysis and understanding. Its research involves computational vision, pattern recognition, image processing, artificial intelligence, and human kinematics, and it is an interdisciplinary research topic (Moeslund et al., 2006). In the broader context of action recognition and human behavior understanding, human activity recognition (HAR), and body movement recognition are closely related fields.

The aim of this study is to develop a physical exercise action recognition system based on LSTM (long short-term memory) technology to improve the effectiveness and safety of physical exercise. By monitoring and analyzing a person's movements during physical exercise, the system can automatically detect whether the movements are performed correctly and provide personalized health and exercise advice. This feedback can help encourage more people to be physically active and thus improve their overall health. The importance of this research lies in the improved efficiency and personalization of exercise supervision. In addition, the use of Artificial Intelligence enables people to enable people to understand and analyze the details and patterns of different physical exercise movements better than with other systems and also provide new data and insights for further research in the field of sports science. LSTM-based physical exercise movement recognition research will bring important improvements and innovations to health, exercise, and scientific research.

MATERIALS AND METHODS

Related Works

Kim et al. (2017) proposed the use of a new class of models known as Temporal Convolutional Neural Networks (TCNs) for 3D HAR. Ma et al. (2021) fused the depth characteristics filtered by the Fourier Pyramid with the bone characteristics and sorted the merged data based on the support engine, thus designing the action recognition unit. Yu et al (2021) proposed a multitask system covering three domains. The traditional action recognition method is easily affected by the action speed, illumination, occlusion, and complex background, leading to the poor robustness of the recognition results (Yu et al., 2021). To solve the aforementioned problems, Yang et al. (2021) used an improved residual dense neural network method to study the automatic recognition of dance action images. Hussain et al. (2022) presented the details of the data acquisition effort using a single chest-mounted tri-axial accelerometer, followed by a novel method for the recognition of a wide range of gym-based free weight exercises. Chen et al. (2022) started from physical education teaching in primary and secondary schools, and from the perspective of modern scientific and technological facilities, discussed the practical sports analysis and action optimization of physical education teaching based on the perception of the Internet of things. Li (2023) proposed a data preprocessing method based on the angle and relative distance feature enhancement and a ball-motion pose recognition model based on LSTM attention.

Related Concepts

HAR focuses on identifying and classifying human activities in daily life according to sensor data that are usually collected from accelerometers, gyroscopes or other devices, and sometimes even cameras (Lara & Labrador, 2012). These sensors can be integrated into wearable devices, such as smartwatches, or embedded in smartphones.

HAR often involves monitoring body movements, posture, and acceleration patterns to infer the type of activity that is being carried out. For example, activities could include walking, running, sitting, standing, climbing stairs, and even more specific actions, such as brushing your teeth or drinking water. HAR finds applications in healthcare, fitness tracking, context-aware computing, and more (Zhang et al., 2017). This technology is especially useful for understanding user behavior and adjusting technology to better meet their needs.

Body movement identification goes beyond recognizing specific activities and delves into understanding the nuances of human movement patterns (Batchuluun et al., 2018). This technology can involve identifying gestures, postures, or movements that may not necessarily correspond to predefined activities. Body movement identification can include fine-grained analysis of movements, such as identifying gestures translated by sign language, capturing dance movements, or recognizing specific body postures by yoga (Sogon & Masutani, 1989). Because of the diversity and subtlety of human movement, this field usually needs more detailed analysis and potentially more complex models. Advanced machine learning technologies, such as posture estimation and bone tracking, are often used to accurately identify and explain body movements.

An action recognition system typically comprises three main components: action sensors, action recognition algorithms, and data communication modules (Xu & Qiu, 2021). Therefore, in this paper, I focus on the existing HAR research from the aspects of sensor types, action recognition algorithms, and communication methods.

In terms of sensor types, human motion recognition is mainly divided into methods based on surface electromyography (sEMG), methods based on optical tracking, and methods based on micro-electromechanical systems (MEMS) inertial elements (Qiu et al., 2022). The method based on a surface EMG signal can more directly judge the human body's movement intention, but because of the expensive and poor mobility of the sEMG signal acquisition device, few related studies have been published. The application of EMG has been a subject of research since the first observation of EMG in 1922 (Poo & Sundaraj, 2010). Because of the close correlation between an EMG signal and muscle and human movement, EMG has great application potential in sports, sports and rehabilitation medicine, neural rehabilitation engineering, human-computer interaction, and other fields. The methods based on optical tracking first record the information of limb movements or gestures with video acquisition equipment, and then they perform the representation, modeling, and classification of the movements (Tran & Trivedi, 2011). The disadvantage of using the optical tracking method for action recognition is that it must ensure that the moving limbs are within the line of sight of the video capture equipment, which is usually fixed in a certain scene and requires a larger activity space. Therefore, this method is not suitable for physical exercise. In contrast, using MEMS-based inertial sensors for body motion recognition is more flexible, more convenient, and easier to deploy than the optical tracking method. A MEMS sensor is small, consumes little power, and is easy to carry. It can use wireless communication for data transmission with the host computer, thereby reducing space requirements. Therefore, researchers in body motion recognition increasingly use MEMS as a motion acquisition devices.

The subject of human motion detection, recognition, and tracking is mainly used in the field of intelligent video surveillance (Ko, 2008). By observing the moving targets in the monitored scene in real time, such as people or vehicles, as well as analyzing and describing their behavior, this technology saves a lot of human and material resources. More importantly, on some occasions, owing to objective reasons, checking the scene in person may be inconvenient or impossible. At this time, only other methods, such as real-time monitoring with computers, can be used to complete the required work. Backpropagation (BP) neural networks and support vector machines (SVMs) are widely used.

For example, Xing and Marwala (2018) used an SVM and an artificial neural network for a classifier and realized the recognition of six upper limb movements by fusing the data of multiple MEMS pose modules. Neto et al. (2019) studied gesture-based human-robot interaction for human assistance in manufacturing based on an SVM and BP neural networks. Huang and Yu (2019) used the least multiplication-based similarity matching model.

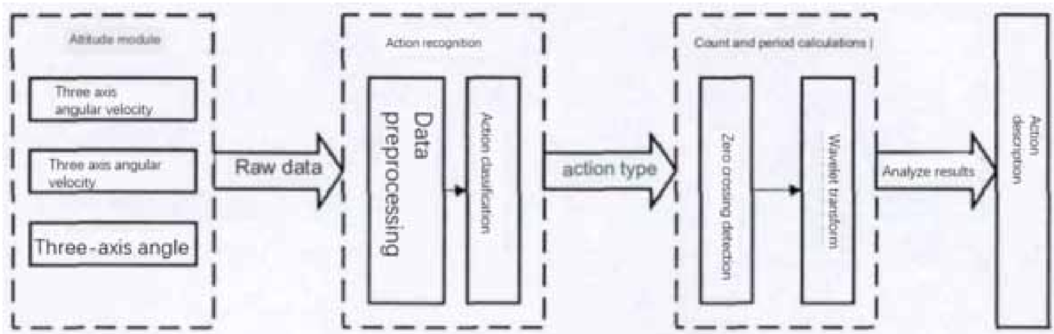
Developed by Yann Lecun, convolutional neural networks (CNNs) are primarily used for tasks involving images and visual data. They are designed to automatically learn hierarchical features from image data through convolutional and pooling layers. However, CNNs also have been successfully applied to other domains beyond images. Researchers have found that their architecture can also be adapted for natural language processing (NLP) tasks. CNNs can be used to extract features from text and sequences, leading to their application in tasks such as sentence classification, sentiment analysis, and text generation. LSTM is a type of recurrent neural network (RNN) architecture designed to capture and process sequential data with long-range dependencies. LSTMs are commonly used in tasks involving time series data, language modeling, speech recognition, and so on.

For this paper, I selected MEMS sensors for collecting motion information, and sensors used by multiple users could upload the data through wireless access points. The data were processed centrally using machine learning algorithms and wavelet analysis methods. The framework for this behavioral recognition model is shown in Figure 1.

Unlike classical machine learning algorithms that need to continue feature extraction or feature selection, deep learning networks can mine high-level features in the data by combining low-level features and have a stronger understanding. MEMS generate raw data containing nine axes, each of which is a time series, and for common deep learning frameworks, only one-dimensional signals are accepted for processing the time series. Therefore, directly feeding the original nine-time series into the deep learning network for processing is not possible; rather, the raw data needs to be preprocessed to convert it from nine axes of data to one.

Deep learning models include several different types, and each network structure is suitable for a different problem domain. For example, CNNs are commonly used for image-processing tasks,

Figure 1. Behavioral recognition model framework



whereas LSTMs are commonly used to process time series data. Typically, deep learning models require large-scale datasets for training and optimization.

However, large-scale datasets have not been widely released in the field of limb movement recognition, especially MEMS-based limb movement recognition. This situation limits the application of deep learning in this field. This limitation is one of the reasons why applying deep learning in this field is difficult.

In this regard, for this study, I established a combined network model based on a one-dimensional CNN (1D CNN) and long- and short-term memory networks. The model combines the features of both and requires less data volume.

RESULTS AND DISCUSSION

Action Recognition Based on Deep Learning

The LSTM model has certain advantages in sequence modeling, has a long-term memory function, is simple to implement, and solves the problems of gradient disappearance and gradient explosion in the process of long sequence training. Long-period memory becomes the other version of the RNN, which was first presented and clarified by Hochreiter and Schmidhuber (1997). Moreover, LSTM improves the fact that the elements are missing in the RNN, and has better performance in long sequences. It has been employed and used in the fields of speech recognition, language modeling, and text translation.

The LSTM cell is shown in Figure 2, where xt is the input in the current state, yt is the output in the current state, and Ct (cell state) and ht (hidden state) respectively represent the two types of transmission to the next node. Status output, Ct-1, and ht-1 represent the status output of the previous node.

Ct and ht can be determined using the formulas shown in equations (1) and (2):

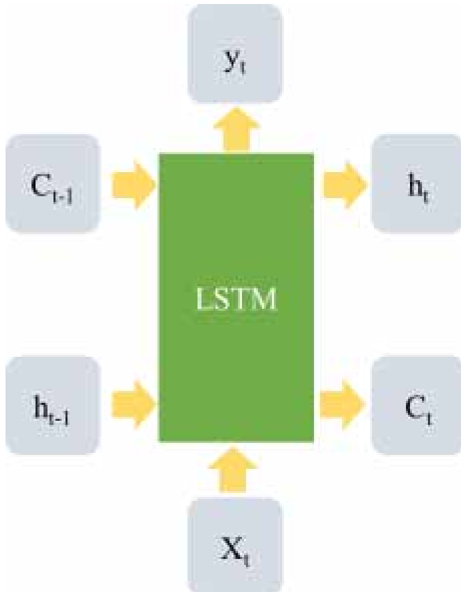
$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C}_t \tag{1}$$

$$h_t = o_t \circ \tanh(C_t) \tag{2}$$

In these equations,formula represents the Hadamard product of matrices, the formula shown in equation (3):

$$\tilde{C}_t = \tanh(W_c \cdot \left[h_{t-1}, x_t\right] + b_c) \tag{3}$$

Figure 2. Input and output of LSTM unit



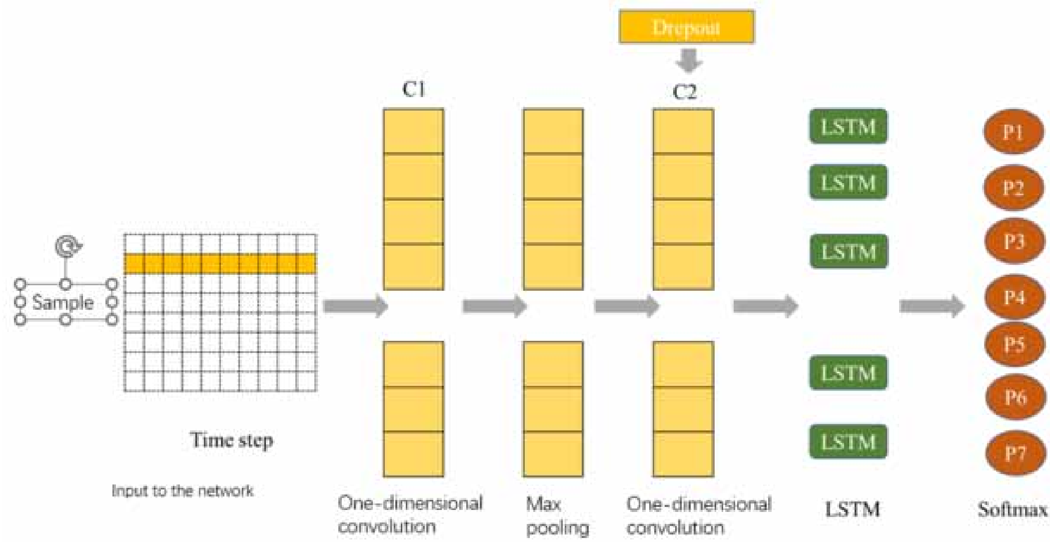
One-dimensional CNNs have the advantage of requiring a low number of training samples. This network uses a one-dimensional convolutional kernel that allows convolutional operations to be performed at each localization of the time series signal, thus mining local features. Combining convolution kernels with different window lengths enables different scale features of the signal to be extracted. This feature makes 1D CNNs perform well on small datasets. In addition, the presence of a pooling layer allows the network to down-sample the signal and shorten the signal length, thereby accelerating the training process and reducing the training time.

However, 1D CNNs are not sensitive to temporal order and cannot fully mine information related to temporal order. In contrast, the LSTM network is a temporal RNN specifically designed to process time-series data. The special gate structure inside the LSTM enables it to selectively retain or forget previous information, and it has a good memory for events with long intervals in the time series.

LSTM requires a large amount of data for training. Therefore, combining a 1D CNN with a long- and short-term memory network can fully use their advantages. Using this combined technology can mine the temporal information in the data, but also does not require large-scale training samples, thus achieving excellent action recognition results.

The structure of the network and its data preprocessing method are shown in Figure 3. The network includes a max pooling layer, an LSTM layer, and an output layer. Among them, the convolutional layer and the LSTM layer both include 32 neurons, represent the probability that the currently input action data belongs to a certain category. The convolution kernel sizes of the one-dimensional convolution layers C1 and C2 are 15 and 17, respectively, and the window length of the maximum pooling layer is 15. The input to this network is a matrix with two dimensions: the time step dimension and the sample dimension. As mentioned earlier, each sample has the nine axes of the raw data end to end. The time step corresponds to the sampled value of the sensor at a specific time. In addition, because the duration of each action recorded by the sensor is different, the total length of each data sample varies. For programming purposes in the application, each data sample may be padded with zero values to have the same length.

Figure 3. Network structure and data preprocessing process



MOVEMENT AND TERM CALCULATION WAYS

The raw data from the posture sensor consists of acceleration, angular velocity, and Euler angles in the X, Y, and Z axes, totaling nine data axes. One of these data axes is known as the sensitivity axis, and it is the axis that best represents the state of limb movement. Each action is characterized by its specific plane of motion, with some actions occurring primarily in the X plane and others primarily in the Z plane. As a result, there will always be one axis with the largest amplitude and most regular waveform in the sensor data, usually identified as the sensitive axis, whereas the other axes usually have very weak or disorganized amplitudes.

The approach I take in this paper is to concatenate the data of nine axes in the first place. The attitude module collects nine values (i.e., acceleration, angular velocity, and Euler angle of the X, Y, and Z axes) at each sampling moment, so the nine axes of the final data have the same length (i.e., the same number of sampling points). After the attitude module joins the data head to tail, the total length of the data becomes nine times that of the original single axis. In this process, no loss of information was caused, but the data was converted from nine axes to one axis, which meets the requirements of the deep learning framework.

Choosing sensitive axes eliminates noise interference and improves the accuracy of action counting and cycle calculation; it also simplifies the calculation process. This step helps you to analyze and understand the state of limb movement more effectively.

Movement counting and period calculation are designed to separate successive limb movements and obtain the period of each movement. A common approach in separating the movements is to use over-zero detection; however, limb jerks create many additional over-zero points, causing interference to the over-zero detection. The method I adopted for this paper was to first carry out the sensitive axis selection to ensure that the signal itself had a regular and obvious waveform. I then carried out the wavelet transform and over-zero detection and excluded the extra over-zero points generated by jitter by comparing the results of wavelet period calculation with the results of over-zero detection. The last step helps to achieve the purpose of anti-jitter and realize the accuracy of action period calculation.

“Wavelet” (wavelet) is a mathematical tool and signal processing technique used to analyze and process signals and data. Wavelet analysis is a multiscale analysis method that allows you to simultaneously consider signal features on different time scales or frequency scales. It has a wide range of applications in many fields, including signal processing, image processing, data compression, pattern recognition, seismic analysis, financial analysis, and more.

The basic idea of wavelet analysis is to decompose a signal into wavelet functions at different scales that are obtained from a mother wavelet function by translating and scaling operations. This decomposition allows you to capture both local and global features of the signal, leading to a better understanding of the structure of the signal.

Over-zero detection is a fundamental signal processing technique that plays a key role in a variety of applications. It involves identifying the moment at which a signal crosses a zero-amplitude point and is crucial for rhythm analysis in audio processing, synchronization in power systems, modulation and demodulation in digital communications, and phoneme extraction in speech recognition. This technique serves as the basis for numerous signal processing algorithms capable of extracting important periodic information from different types of signals.

After selecting the sensitive axis, three angles (Euler angles) are first eliminated and then one of the remaining six angles is explored. First, for the original signal, some of the nine axes often have waveform jumps. This phenomenon is concentrated in the three angle axes. For example, suppose the angle of an axis starts to increase from 0, when it increases to +180. When continues to increase, the angle will suddenly change to -180. This phenomenon is caused by the angle calculation method of the attitude module itself, but it will adversely affect the period calculation, so three angle axes should be excluded when selecting the sensitive axis. Next, the specific selection method of the sensitive axis can be calculated as shown in equations (4)–(6):

$$A = \left[A_1, A_2, \cdots, A_6\right]^T \tag{4}$$

$$S_i^2 = \frac{1}{m}\left[(A_{i1} - \bar{A}_i)^2 + (A_{i2} - \bar{A}_i)^2 + (A_{im} - \bar{A}_i)^2\right] \tag{5}$$

$$i' = \underset{i}{\operatorname{argmax}}\, S_i^2 \tag{6}$$

Assuming the signal, its Fourier transform is . When the conditions are met, the result can be calculated as shown in equation (7):

$$C_\psi = \int_R \frac{\left|\hat{\psi}(\omega)\right|^2}{\left|\omega\right|} d\omega < \infty \tag{7}$$

The Fourier Transform (FT) is a mathematical tool that converts a function from its original domain to another. It is primarily used to analyze a signal or function.

It is called a wavelet basis. The wavelet basis I used in this paper is cgau wavelet, which is a complex Gaussian wavelet. When the wavelet base is stretched or shifted, you can get the wavelet sequence, as shown in equation (8):

$$\psi_{(a,b)}(t) = \frac{1}{\sqrt{\left|a\right|}}\psi(\frac{t-a}{a}), a, b \in R; a \neq 0 \tag{8}$$

In this equation, a is the scale factor of the wavelet transform, and b is the translation factor. For the selected key axis signal, its continuous wavelet transform can be calculated using the formula shown in equation (9):

$$W_f(a,b) = < f, \psi_{a,b} > = \left|a\right|^{-1/2} \int_R f(t)\psi(\frac{t-b}{a})dt \tag{9}$$

Inverse transform to the formula shown in equation (10):

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{a^2} W_f(a,b)\psi(\frac{t-b}{a})dadb \tag{10}$$

After the wavelet transform is performed on the key axis signal, the wavelet coefficient matrix can be obtained, where n represents the number of layers of the wavelet transform, and m represents the number of sampling points (that is, the length of the signal), as shown in equation (11):

$$a_{ij} \in A_{m\times n}, 0 < i \le m, 0 < j \le n, i \in N, j \in N \tag{11}$$

The wavelet energy matrix P is shown in equation (12):

$$P = \begin{bmatrix} {a_{11}}^2 & {a_{12}}^2 & \cdots & {a_{1n}}^2 \\ {a_{21}}^2 & {a_{22}}^2 & \cdots & {a_{2n}}^2 \\ \vdots & \vdots & \ddots & \vdots \\ {a_{m1}}^2 & {a_{m2}}^2 & \cdots & {a_{mn}}^2 \end{bmatrix} \tag{12}$$

In this equation, N represents an integer (similarly hereinafter), and the elements in the energy matrix P represent the magnitude of each harmonic component.

Next, use the vector shown in equation (13):

$$C = \left\{ c_1, \cdots, c_m \middle| c_I = \arg\max_k (p_{lk}) \right\} \quad 1 \le l \le m, 1 \le k \le n; l, k \in N \tag{13}$$

In other words, the vector is the position where the component of the maximum energy in each harmonic occurs at each moment; the position of the fundamental frequency of the sensor signal is the wavelet scale. where Pl denotes the lth row of matrix P, and Plk denotes the kth element in the Pl row.

Each wavelet scale in the wavelet transform corresponds to a frequency vector F that is jointly determined by the specific wavelet transform layers and scales. Because the period is the reciprocal of the frequency, T will be like the calculation shown in the formula in equation (14):

$$T = \left[\frac{1}{F_{c1}}, \frac{1}{F_{c2}}, \cdots, \frac{1}{F_{cm}} \right] \tag{14}$$

The action count is calculated by taking the list L into equation (14), resulting in the vector T’, as shown in equation (15):

$$T' = \left[T_{L1}, T_{L2}, \cdots, T_{Ld} \right] \tag{15}$$

In equation (15), the term of the movement is called by the vector. Among it, d will be the extent, and that is to say, the numbers of the movement Ti in T' is the ith factor for the vector T.

RESULT ANALYSIS AND DISCUSSION

Action Recognition Experiment

For this paper, I selected seven types of limb movements for identification. They included four types of dumbbell movements and three types of leg movements; namely, curl, lateral raise, shoulder push, flying, seated leg raise, standing leg raise, and lift weights.

I used a self-constructed dataset that contained a total of 420 samples. These samples came from six subjects, three males and three females, each of whom provided 70 data samples, 10 samples for each action. I used 80% of all samples for the training set and 20% for the test set. For deep learning models, I divided 20% of the training set into a validation set.

In the action recognition experiments, I compared two different data preprocessing methods. For the deep neural network, one preprocessing method is to connect the data of nine axes together, and the other method is to use linear discriminant analysis for processing.

The results of the comparison are shown in Table 1. The data in this table indicate that the accuracy is lower when using linear discriminant analysis for action recognition because although linear discriminant analysis can effectively reduce the dimensionality of the data and convert the original nine-axis data to one-dimensional data, a large amount of information is inevitably lost in the process, resulting in the reduced dimensionality data being ineffective in the classification task. Therefore, I chose the first and last of the nine axes data as the data preprocessing method for the deep neural network. This method retained all the information of the original data and did not destroy the time-series relationship of the data, thus obtaining a high recognition accuracy.

Figure 4 shows that the training and validation losses continue to decrease until they stop at the 200th round of training, and the validation and the training losses remain consistent throughout this process. Figure 5 shows that the training accuracy and validation accuracy are always accurately increasing in the trend, and the difference is not big in the end. This finding means that the model learns the laws in the data better than other methods.

To compare the effects of different types of models and samples, I conducted the test by using different models. I also selected the multilayer perceptron and the method proposed by Xie et al. (2022) and others as two comparison algorithms. The final accuracy of the sample was gained by the test.

Multilayer perceptron is a representative forward neural network that can be used as a comparison algorithm to effectively set the accuracy reference. Because the number of hidden layers of the multilayer perceptron is generally small, the calculation speed is slow, and the effect is not good on the dataset with high dimension. The raw data of the attitude sensor also usually can reach thousands to tens of thousands dimensions, so it is necessary to reduce the dimension data that can be fed into the multilayer perceptron. I adopted the same feature extraction method as SVM, and selected the same 36 statistics as features to compare the performance differences between the two under the same dataset.

The multilayer perceptron I used consists of an input layer, two hidden layers, and an output layer. Too few layers will lead to insufficient learning ability of the multilayer perceptron for complex data, resulting in low classification accuracy, whereas too many layers are more likely to cause

Table 1. Accuracy of different preprocessing methods

Data Preprocessing for Deep Neural Networks	Motion Recognition Accuracy (%)
LDA downscaling	87.79
Data head to tail	97.61

Figure 4. Changes in training loss and validation loss

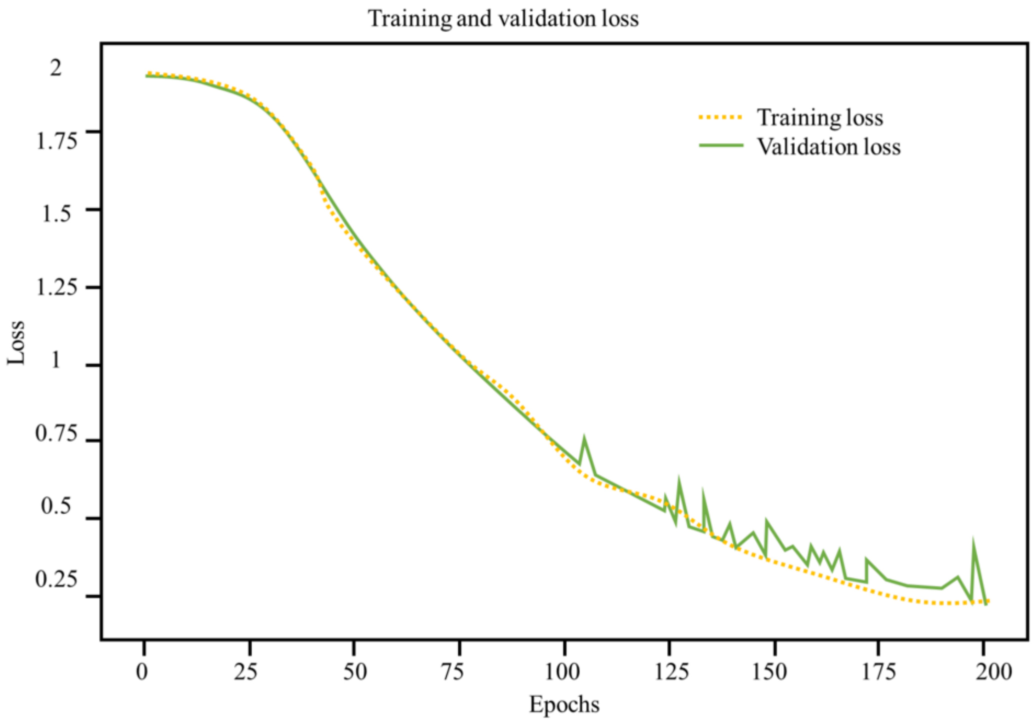
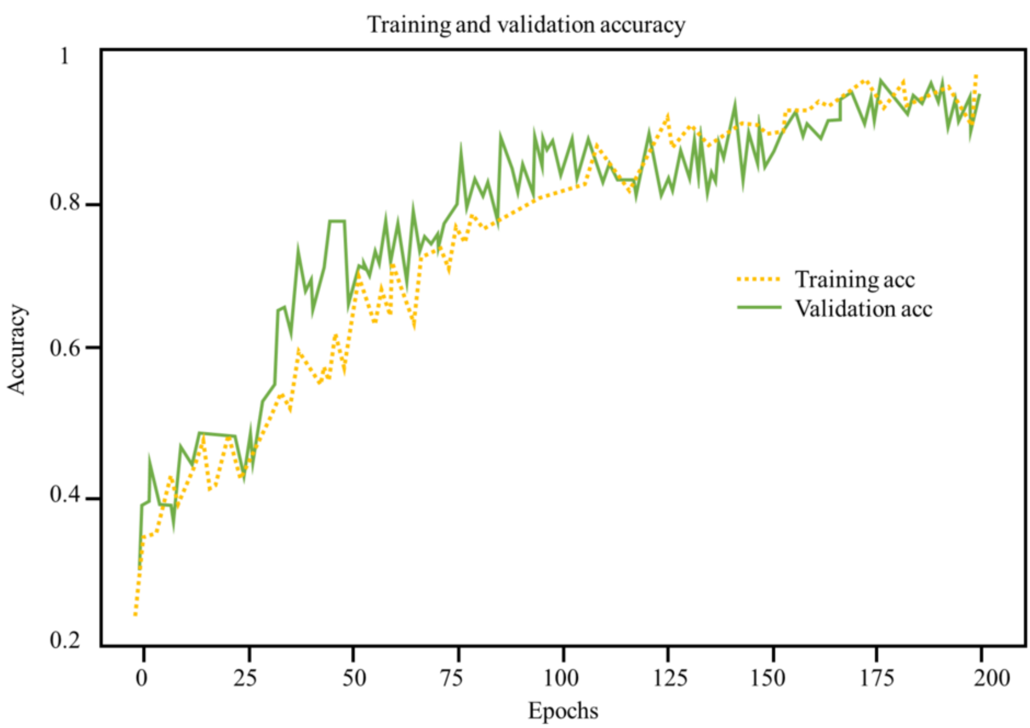


Figure 5. Changes in drilling precision and validation precision



over-fitting; that is, the multilayer perceptron does not learn the dataset, but simply fits each sample, resulting in high training accuracy and low-test accuracy. Therefore, the number of hidden layers should be moderate. In the hidden layer, each layer contains five neurons, and each neuron uses the most commonly used Relu activation function. The output layer has seven neurons, and each neuron uses the commonly used Softmax activation function.

Another comparison algorithm is the similarity matching algorithm based on the least squares method proposed by Yu et al. (2022). This method first performs acceleration decomposition, period normalization, and other operations on the data; it then establishes a feature database. When classifying, the algorithm judges the movement one in line with the similar point for the test one and the standard tested ones from the data.

I chose the confusion matrix as the main metric for evaluating the performance of LSTM-based action-behavior recognition models. The confusion matrix provides detailed performance information, such as precision, recall, and F1 score for each category; all this information helps you to fully understand the performance of the model on different behavioral categories. In addition, the confusion matrix can effectively deal with the category imbalance problem, providing a balanced basis for model performance evaluation. The decision of choosing the confusion matrix as the evaluation metric is intended to ensure that you can accurately and comprehensively evaluate the model's behavior recognition performance.

The comparison results of the four algorithms in this experiment are shown in Figure 6. Figure 7(a) and (b) show the confusion matrix of deep neural network and SVM, respectively.

The confusion matrix is an index to evaluate the results of the model; it is a part of model evaluation. In addition, the confusion matrix is used primarily to judge the advantages and disadvantages of classifiers, which is suitable for classified data models, such as classification tree, logistic regression, linear discriminant analysis, and other methods. Figure 6 shows the model for the deeper studying as well as the SVM reached the identification precision of over 96%, which is almost a little bit higher than the counting ways of comparing, but the precision ratio of the multilayer sensors is the relatively lowest one from them. Among them, the deep neural network studying reached the identification precision as high as 97.61%. The deep learning sample put forward in this paper will

Figure 6. Action recognition accuracy of four algorithms

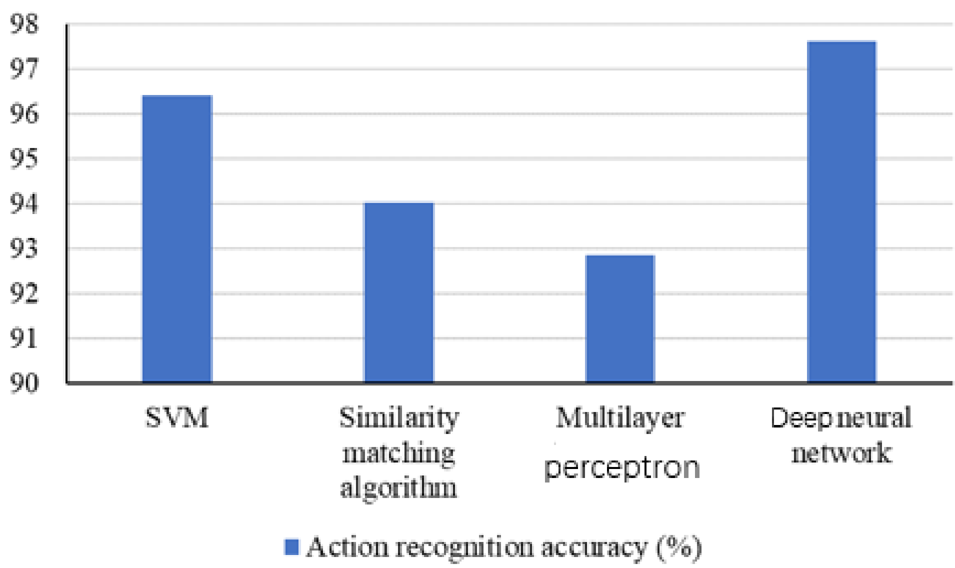
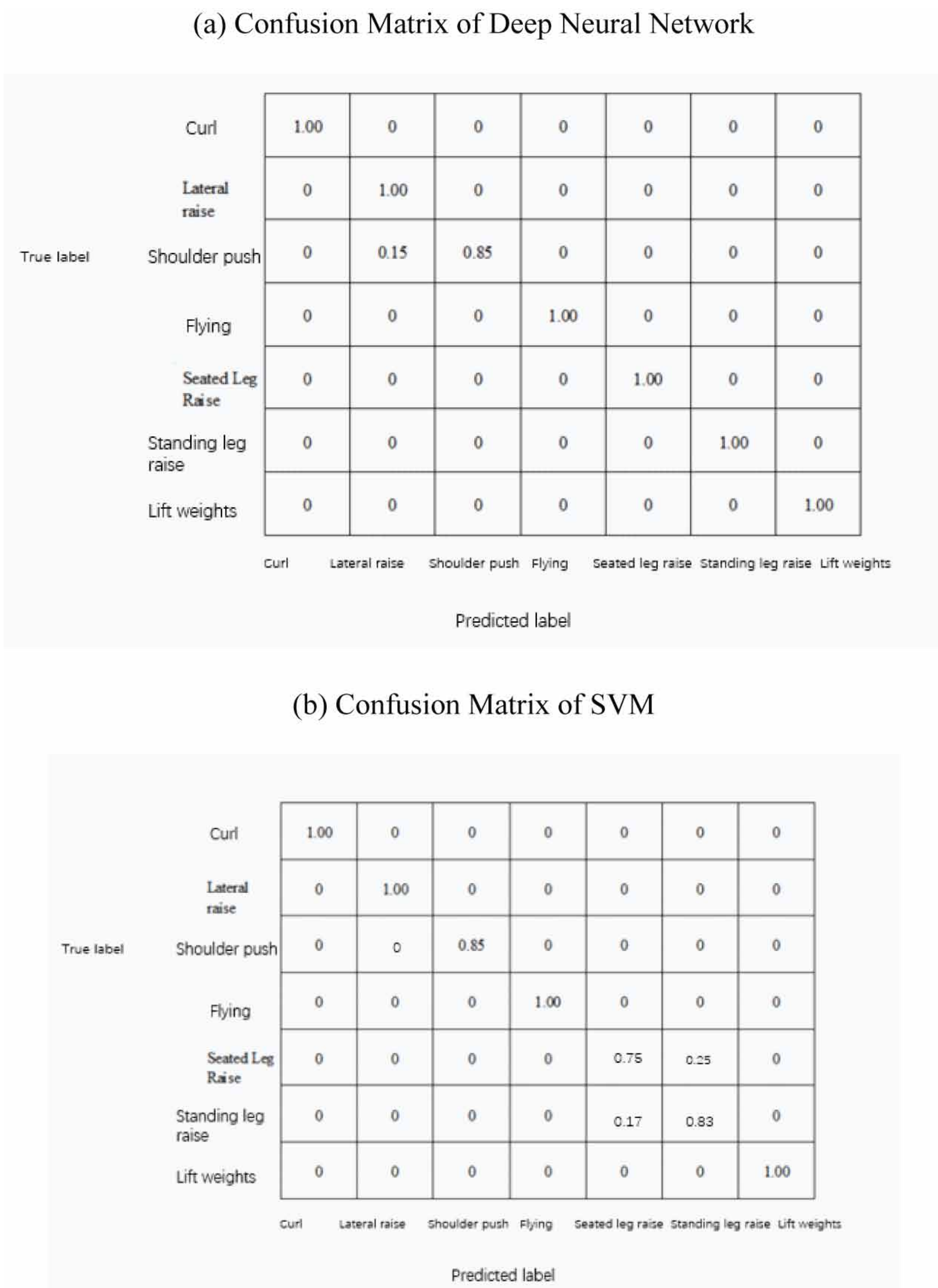


Figure 7. Action recognition confusion matrix of deep neural network and SVM (a): Confusion matrix of deep neural network, (b) confusion matrix of SVM



(a) Confusion Matrix of Deep Neural Network

True label \ Predicted label	Curl	Lateral raise	Shoulder push	Flying	Seated leg raise	Standing leg raise	Lift weights
Curl	1.00	0	0	0	0	0	0
Lateral raise	0	1.00	0	0	0	0	0
Shoulder push	0	0.15	0.85	0	0	0	0
Flying	0	0	0	1.00	0	0	0
Seated Leg Raise	0	0	0	0	1.00	0	0
Standing leg raise	0	0	0	0	0	1.00	0
Lift weights	0	0	0	0	0	0	1.00

(b) Confusion Matrix of SVM

True label \ Predicted label	Curl	Lateral raise	Shoulder push	Flying	Seated leg raise	Standing leg raise	Lift weights
Curl	1.00	0	0	0	0	0	0
Lateral raise	0	1.00	0	0	0	0	0
Shoulder push	0	0	0.85	0	0	0	0
Flying	0	0	0	1.00	0	0	0
Seated Leg Raise	0	0	0	0	0.75	0.25	0
Standing leg raise	0	0	0	0	0.17	0.83	0
Lift weights	0	0	0	0	0	0	1.00

be employed for small datasets and movement identification. For the SVM sample, the data show that the way has the possibility to make mistakes for the leg movements. The reason is that the two types of actions are relatively similar. Based on the features from a single sensor, SVM cannot perfectly separate the two types of actions. The optimal precision is not reached by SVM, but the gap with deep learning models is not obvious, considering that its operation speed and complexity are much lower than those of deep learning models.

ACTION TECHNIQUE AND PERIOD CALCULATION EXPERIMENT

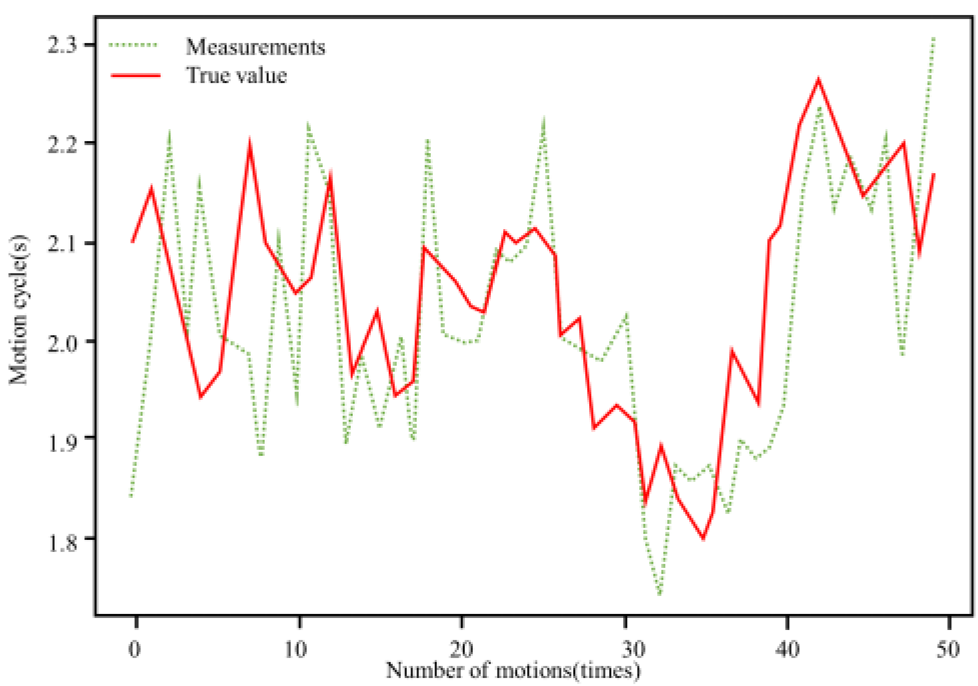
Take dumbbell bending action as an example. Participants completed 50 repetitions, and the number of times statistics and cycle calculation were carried out based on the axis. The analysis was made about the shown data and the real data. Figure 8 shows the 50 actions numbers. The line shaped as a square is the tested value and the line shaped as the angle is the real number. They are measured in seconds.

As the length of the two curves in Figure 8 shows, the algorithm's action counts are very accurate, and the counting results of the 50 counts are exactly the same as the actual values. The overall fit of the two curves on the way is good, with an average error of only 0.08 s and an average error rate of 4.03%. The maximum error of the calculation result is 0.25 s, and the maximum error rate is 13.5%, which occurs only at the endpoint of the curve. This experiment proves that the over-zero detection and wavelet analysis methods are better overall.

CONCLUSION

I introduced some counting ways for the data and discussed a behavior recognition model using deep neural networks. Overall, I made a contribution to the field of behaviour recognition by leveraging the power of deep neural networks, particularly CNNs and LSTMs, to improve accuracy and effectiveness

Figure 8. Action count and cycle calculation results



in classifying and understanding behaviours from data. I then explained the structure and input of the network samples and output characteristics. I also introduced a method on implementing action counting and period calculation using certain tools. Specifically, I first introduced the purpose and specific method of selecting the sensitive axis, and then briefly introduced the development history of wavelet and the commonly used wavelet basis—for example, how to use the zero-crossing detection method to count the number of actions and how to combine wavelet and zero crossing detection to get the cycle of each action.

For the related experiments of the multiuser action monitoring system, I first introduced the preparation work of the experiment and then showed the experimental process and results of the action recognition system. Finally, I presented the experimental process and results of the action counting and introduced the period calculation methods. For deep neural networks, putting the data end to end on each axis was a more effective data pre-processing method than using linear discriminant analysis. In terms of action recognition, the models based on SVM and deep neural network had high action recognition accuracy, whereas the accuracy of multilayer perceptions was not ideal. In terms of action cycle calculation, by perfecting the motion recognition technology, the algorithm can correct people's wrong posture during physical exercise in time and can also greatly promote people's enthusiasm for physical exercise, which is of great significance to promote comprehensive fitness.

For this study, I used a single sensor for limb motion data acquisition. This sensor was very convenient to use, but the limitation was that it was only applicable to the recognition of a single limb motion. In future research, I plan to introduce multiple sensors and extend the application range by data fusion.

In addition, regarding the wavelet-based period computation, I found that energy leakage was easily observed at the beginning and end of the signal, leading to a large error. I plan to address this issue in subsequent research to improve the accuracy of the period calculation without changing the original intention and goal of this research.

DATA AVAILABILITY

The figures used to support the findings of this study are included in the article.

CONFLICTS OF INTEREST

I declare that there are no conflicts of interest.

FUNDING STATEMENT

This work was not supported by any funds.

ACKNOWLEDGMENT

I would like to show sincere thanks to the developers of the techniques that have contributed to this research.

REFERENCES

Batchuluun, G., Naqvi, R. A., Kim, W., & Park, K. R. (2018). Body-movement-based human identification using convolutional neural network. *Expert Systems with Applications*, *101*, 56–77. doi:10.1016/j.eswa.2018.02.016

Chaquet, J. M., Carmona, E. J., & Fernández-Caballero, A. (2013). A survey of video datasets for human action and activity recognition. *Computer Vision and Image Understanding*, *117*(6), 633–659. doi:10.1016/j.cviu.2013.01.013

Chen, M., & Jiang, J. (2022). Sports analysis and action optimization in physical education teaching practice based on Internet of Things sensor perception. *Computational Intelligence and Neuroscience*, *2022*, 1–8. Advance online publication. doi:10.1155/2022/7152953 PMID:35814534

Feng, X., Liu, Y., & Wang, X. (2022). Analysis of interdisciplinary characteristics and knowledge structure from multidimensional perspective: A case study of national fitness topic in China. *Library Hi Tech*. Advance online publication. doi:10.1108/LHT-06-2021-0189

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735 PMID:9377276

Huang, H., & Yu, H. (2019). *Compact and fast machine learning accelerator for IoT devices* (Vol. 149). Springer Singapore., doi:10.1007/978-981-13-3323-1

Hussain, A., Zafar, K., Baig, A. R., Almakki, R., AlSuwaidan, L., & Khan, S. (2022). Sensor-based gym physical exercise recognition: Data acquisition and experiments. *Sensors (Basel)*, *22*(7), 2489. doi:10.3390/s22072489 PMID:35408104

Kim, T. S., & Reiter, A. (2017). Interpretable 3D human action analysis with temporal convolutional networks. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 1623–1631). IEEE. doi:10.1109/CVPRW.2017.207

Ko, T. (2008, October). A survey on behavior analysis in video surveillance for homeland security applications. In *Proceedings of the 2008 37th IEEE Applied Imagery Pattern Recognition Workshop* (pp. 1–8). IEEE. doi:10.1109/AIPR.2008.4906450

Kulsoom, F., Narejo, S., Mehmood, Z., Chaudhry, H. N., Butt, A., & Bashir, A. K. (2022). A review of machine learning-based human activity recognition for diverse applications. *Neural Computing & Applications*, *34*(21), 18289–18324. doi:10.1007/s00521-022-07665-9

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, e253. doi:10.1017/S0140525X16001837 PMID:27881212

Lara, O. D., & Labrador, M. A. (2013). A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials*, *15*(3), 1192–1209. doi:10.1109/SURV.2012.110112.00192

Li, D., & Zhang, J. (2021). Computer aided teaching system based on artificial intelligence in football teaching and training. *Mobile Information Systems*, *2021*(9562782), 1–10. doi:10.1155/2021/9562782

Li, X. (2023). Study on volleyball-movement pose recognition based on joint point sequence. *Computational Intelligence and Neuroscience*, *2198495*, 1–8. Advance online publication. doi:10.1155/2023/2198495 PMID:36844697

Li, Z., & Wang, H. (2021). The effectiveness of physical education teaching in college based on artificial intelligence methods. *Journal of Intelligent & Fuzzy Systems*, *40*(2), 3301–3311. doi:10.3233/JIFS-189370

Ma, C., & Shou, M. (2021). Sports competition assistant system based on fuzzy big data and health exercise recognition algorithm. *Mobile Information Systems*, *2021*(7687178), 1–10. doi:10.1155/2021/7687178

Moeslund, T. B., Hilton, A., & Krüger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, *104*(2–3), 90–126. doi:10.1016/j.cviu.2006.08.002

Neto, P., Simão, M., Mendes, N., & Safeea, M. (2019). Gesture-based human-robot interaction for human assistance in manufacturing. *International Journal of Advanced Manufacturing Technology*, *101*(1-4), 119–135. doi:10.1007/s00170-018-2788-x

Poo, T. S., & Sundaraj, K. (2010, December). Design and development of a low cost EMG signal acquisition system using surface EMG electrode. In *Proceedings of the 2010 IEEE Asia Pacific Conference on Circuits and Systems* (pp. 24–27). IEEE. doi:10.1109/APCCAS.2010.5774896

Poppe, R. (2007). Vision-based human motion analysis: An overview. *Computer Vision and Image Understanding*, *108*(1–2), 4–18. doi:10.1016/j.cviu.2006.10.016

Qiu, S., Zhao, H., Jiang, N., Wang, Z., Liu, L., An, Y., Zhao, H., Miao, X., Liu, R., & Fortino, G. (2022). Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. *Information Fusion*, *80*, 241–265. doi:10.1016/j.inffus.2021.11.006

Sogon, S., & Masutani, M. (1989). Identification of emotion from body movements: A cross-cultural study of Americans and Japanese. *Psychological Reports*, *65*(1), 35–46E. doi:10.2466/pr0.1989.65.1.35 PMID:2928452

Tran, C., & Trivedi, M. M. (2011). 3-D posture and gesture recognition for interactivity in smart spaces. *IEEE Transactions on Industrial Informatics*, *8*(1), 178–187. doi:10.1109/TII.2011.2172450

Wang, X., & Zhang, Z. (2022). Model construction of urban public sports service system under the background of information technology. *Mobile Information Systems*, *5207857*, 1–10. Advance online publication. doi:10.1155/2022/5207857

Xie, S., Li, Y., Tan, H., Liu, R., & Zhang, F. (2022). Multi-scale and multi-layer perceptron hybrid method for bearings fault diagnosis. *International Journal of Mechanical Sciences*, *235*, 107708. doi:10.1016/j.ijmecsci.2022.107708

Xing, B., & Marwala, T. (2018). *Smart maintenance for human–robot interaction*. Springer Cham., doi:10.1007/978-3-319-67480-3

Xu, Y., & Qiu, T. T. (2021). Human activity recognition and embedded application based on convolutional neural network. *Journal of Artificial Intelligence and Technology*, *1*(1), 51–60. doi:10.37965/jait.2020.0051

Yang, Y. (2021). Piano performance and music automatic notation algorithm teaching system based on artificial intelligence. *Mobile Information Systems*, *2021*(3552822), 1–13. doi:10.1155/2021/3552822

Yu, G., Liu, C., Tang, B., Chen, R., Lu, L., Cui, C., Hu, Y., Shen, L., & Muyeen, S. M. (2022). Short term wind power prediction for regional wind farms based on spatial-temporal characteristic distribution. *Renewable Energy*, *199*, 599–612. doi:10.1016/j.renene.2022.08.142

Yu, Q., Wang, H., Laamarti, F., & El Saddik, A. (2021). Deep learning-enabled multitask system for exercise recognition and counting. *Multimodal Technologies and Interaction*, *5*(9), 55. doi:10.3390/mti5090055

Zhang, S., Wei, Z., Nie, J., Huang, L., Wang, S., & Li, Z. (2017). A review on human activity recognition using vision-based method. *Journal of Healthcare Engineering*, *2017*(3090343), 1–31. doi:10.1155/2017/3090343 PMID:29065585