Zhong *et al.*
Chinese Journal of Mechanical Engineering (2023) 36:126
https://doi.org/10.1186/s10033-023-00951-0

Chinese Journal of Mechanical Engineering

ORIGINAL ARTICLE

Open Access

Construction of Human Digital Twin Model Based on Multimodal Data and Its Application in Locomotion Mode Identification

Ruirui Zhong[1,2], Bingtao Hu[1,2*], Yixiong Feng[1,2*], Hao Zheng[3], Zhaoxi Hong[1,2], Shanhe Lou[4] and Jianrong Tan[1,2]

Abstract

With the increasing attention to the state and role of people in intelligent manufacturing, there is a strong demand for human-cyber-physical systems (HCPS) that focus on human-robot interaction. The existing intelligent manufacturing system cannot satisfy efficient human-robot collaborative work. However, unlike machines equipped with sensors, human characteristic information is difficult to be perceived and digitized instantly. In view of the high complexity and uncertainty of the human body, this paper proposes a framework for building a human digital twin (HDT) model based on multimodal data and expounds on the key technologies. Data acquisition system is built to dynamically acquire and update the body state data and physiological data of the human body and realize the digital expression of multi-source heterogeneous human body information. A bidirectional long short-term memory and convolutional neural network (BiLSTM-CNN) based network is devised to fuse multimodal human data and extract the spatiotemporal features, and the human locomotion mode identification is taken as an application case. A series of optimization experiments are carried out to improve the performance of the proposed BiLSTM-CNN-based network model. The proposed model is compared with traditional locomotion mode identification models. The experimental results proved the superiority of the HDT framework for human locomotion mode identification.

Keywords Human digital twin, Human-cyber-physical system, Bidirectional long short-term memory, Convolutional neural network, Multimodal data

1 Introduction

With the advancement of sensor technology, communication technology, and automation control technology [1–3], the traditional manufacturing industry has been transforming and upgrading to intelligent manufacturing, and the level of automation and intelligence has been continuously improved. Countries have also launched corresponding measures, such as "Made in China 2025" [4], "German Industry 4.0" [5], and so on. Cyber-physical systems (CPS) are multidisciplinary systems with collaborative computing capabilities that are closely related to the surrounding physical systems. It integrates computing, control, and communication technologies for feedback control of distributed computing systems [6–8].

Cyber-physical production system (CPPS) is the key paradigm of CPS in the manufacturing process, which promotes the continuous improvement of the automation level of the manufacturing system. Despite this, the important role of people in manufacturing has been ignored to a certain extent. Human is the most active

*Correspondence:
Bingtao Hu
hubingtao@zju.edu.cn
Yixiong Feng
fyxtv@zju.edu.cn
[1] State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou 310027, China
[2] Engineering Research Center for Design Engineering and Digital Twin of Zhejiang Province, Hangzhou 310027, China
[3] Hangzhou Innovation Institute, Beihang University, Hangzhou 310052, China
[4] School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 637460, Singapore

and dynamic elements in manufacturing. Lack of attention to human beings in the traditional manufacturing process leads to potential safety hazards and reduces the efficiency of human-robot collaboration (HRC). Therefore, both academia and business are exploring how to improve the potential of HRC in smart manufacturing. Digital twin (DT) technology as an emerging paradigm offers new perspectives on the full lifecycle monitoring of machines in smart manufacturing. However, unlike machines, humans themselves are not equipped with sensors, and are difficult to be digitally represented. This poses a challenge to the deeper integration of humans and machines. Zhou et al. [9] proposed the framework of the human-cyber-physical systems (HCPS) and discussed the intelligent manufacturing architecture for the new generation of HCPS. On this basis, Wang et al. [10, 11] expounded the possibility of the relationship and fusion between humans, cyber systems, and physical systems under the HCPS.

To achieve efficient, stable, and safe collaborative work between humans and robots, it is important to enhance the interaction and integration between humans and CPS, and realize the perception and acquisition of feature information of humans. The human body is described in the form of digitalization, to realize the human-centric in the manufacturing process. The human digital twin (HDT) is an important driving way of the HCPS [12]. At present, the construction and application of HDT mainly focus on monitoring the health status of the human body and physiology. Song et al. [13] proposed a digital twin framework for the mechanical properties of the human skeleton in the manufacturing process, and established a digital twin model of the human lumbar spine to monitor the health status of the lumbar spine of workers in the manufacturing process, which can promote harmonious collaboration between human and robot. Makrini et al. [14] established an ergonomic optimization strategy for performing posture optimization based on virtual elements, and the optimization system is monitored by the operator and can analyze the operator's postures and assess the associated musculoskeletal disorder (MSD) risk. In order to monitor the physical state of workers under non-invasive conditions, a fatigue detection system based on perceived exertion (RPE) was constructed [15]. In this system, neural networks and discrete wavelet transform are used to classify fatigue changes and extract features, which can effectively improve the accuracy of the proposed system. There is also decision-making that applies human cognition to CPS. Franceschi et al. [16] developed a production system framework based on multi-agent management and DT monitoring. The operator's cognitive ability is used for supervision and timely intervention in the event of failure, which can effectively improve the fault tolerance performance of the system. In the process of atmospheric plasma spraying, the dynamic changes of the eyes can also reflect the cognitive changes of the operator. Enabling cognitive assessment of complex processes by utilizing eye trackers can integrate operator cognitive abilities and expertise into the atmospheric plasma spray process [17]. Zhang et al. [18] combined a pressure array sensor and infrared array sensor and proposed a deep learning-based human sitting posture recognition algorithm under the premise of protecting privacy and security, which can identify ten human sitting postures. Aiming at the problem of musculoskeletal diseases caused by sedentary and fixed postures of office workers, an ergonomic-based automatic posture assessment method using a depth camera sensor was proposed [19]. The current research mainly focuses on establishing the association and application of certain human behaviors and states with manufacturing systems, and lacks a theoretical system to digitize the human body.

However, the research on the combination of HDT and human locomotion mode identification is less. In the study of human locomotion mode identification, sensors are usually used to obtain the characteristic information of the human body. While sensor data is time series data, the performance of the processing method will affect subsequent recognition tasks. The Hidden Markov Model (HMM) algorithm [20] can effectively deal with time series data, but it has disadvantages in dealing with long-term dependency problems. While the dynamic time warping (DTW) algorithm [21, 22] has a simpler and more effective structure than the HMM algorithm, it is more sensitive to noise. Meanwhile, it is very sensitive to noise. With the rapid development of computer hardware in recent years, deep learning algorithms based on neural networks have provided a new solution to the above problems [23–25]. Lee et al. [26] combined the long short-term memory (LSTM) and the bidirectional long short-term memory (BiLSTM) to build a novel network model, which can enhance the accuracy of the human gait phase estimation. A gait recognition algorithm based on LSTM and convolutional neural network (CNN) for the control of lower-limb exoskeleton robots was proposed [27]. Chen et al. [28] designed different sub-network architectures for different sensors to extract the features of sensor fusion and then used LSTM to obtain the time series data of sensors, thereby improving the accuracy of the algorithm for human activity recognition. These studies mainly leverage one or two human locomotion mode information and improve the overall architecture of the neural network to improve the identification accuracy. However, these methods lack the completeness of the description of human characteristic information due to the simple modal information and

the network structures cannot accurately and efficiently extract human characteristic information.

To this end, this paper proposes the construction architecture of the HDT model based on multimodal data for HCPS. By building a data acquisition system to perceive and acquire human body posture signals and physiological signals in the manufacturing process, an HDT model is constructed to describe the human operator in a digital form, so as to enhance the interaction between humans and CPS. Under the proposed HDT model architecture, we develop a novel multimodal data fusion method based on BiLSTM-CNN for human locomotion identification, which could facilitate real-time monitoring and identification of human locomotion mode. The efficiency and accuracy of the human locomotion mode identification are testified.

The remainder of this paper is structured as follows. Section 2 proposes the construction architecture of the HDT model based on multimodal data for HCPS. Section 3 introduces the key technologies and implementation process under the proposed architecture. Section 4 presents the specific application of the proposed HDT model framework in human locomotion mode identification. Meanwhile, a series of optimization experiments are carried out to improve the performance of the proposed BiLSTM-CNN network model. Besides, the proposed BiLSTM-CNN model is compared with the traditional locomotion mode identification model, which verifies the efficiency and accuracy of the proposed architecture in Section 2. The conclusions and the further work of this research are provided in Section 5.

2 Human Digital Twin Model Architecture

The new generation of intelligent manufacturing systems emphasizes the importance of humans in the manufacturing process and puts forward higher requirements for human-machine interaction. In the traditional intelligent manufacturing process, human is a participants independent of CPS, and cannot be integrated. HCPS described humans in the form of data to promote the fusion of humans and CPS, which is of great significance to realize the transformation and upgrading of the intelligent manufacturing system.

Due to the high complexity and uncertainty of the human body, it is necessary to integrate technologies such as human state perception, locomotion mode identification, feature parameter extraction, digital model construction, and deep learning algorithms to construct an efficient and accurate HDT model.

This paper proposes a framework for the construction of an HDT model based on multimodal data, as shown in Figure 1. The HDT model framework includes four basic layers, namely physical entity layer, cyber model layer, data layer, and application layer. Through the effective integration of intelligent sensor perception and processing technology, a human digital model is constructed based on multimodal data, and an intelligent identification model based on a deep neural network is realized. The dynamic update and real-time analysis of HDT are also supported. There is a close relationship between these layers, and the function of each layer is described in detail next.

2.1 Physical Entity Layer

Physical entity layer includes the human body, intelligent sensors, and specific manufacturing environment elements. Intelligent sensors collect human body information as a basis. In an actual manufacturing workshop, it is very important to ensure the normal production of workers based on digitized human body information, so the use of cumbersome external equipment that will increase workers' extra burden should be avoided. Considering that multiple portable, lightweight smart sensors should be used to accurately characterize the human operators, a single sensor cannot comprehensively describe the multi-dimensional feature state of the human body. Using a variety of sensors and obtaining multimodal information can realize the acquisition of human characteristic locomotion information. Intelligent sensors mainly include inertial measurement units (IMU), plantar pressure sensors, a depth camera, and electromyography (EMG) equipment. These sensors can enhance the perception of human body status without interfering with the normal production of workers and have the advantages of being lightweight and portable. Through the use of intelligent sensors, the multi-source heterogeneous data of the human body can be collected and analyzed in real-time, to realize the perception and acquisition of the human body state and physiological signals in the manufacturing process.

Physical entity layer is the hardware foundation of the HDT model framework. Physical entity layer transmits the perceived and acquired body state and physiological signals to cyber model layer and data layer, promoting the construction of the HDT model.

2.2 Cyber Model Layer

Cyber model layer includes a database and a human digital model. The database contains the human body's real-time state data, historical data, and abnormal data. The key features of the human body multimodal data are chosen as the feature parameters of the human digital model, which include the position of human skeleton nodes, EMG signals, body acceleration, and plantar pressure distribution. Therefore, the establishment of a human digital model and the digital expression of human information

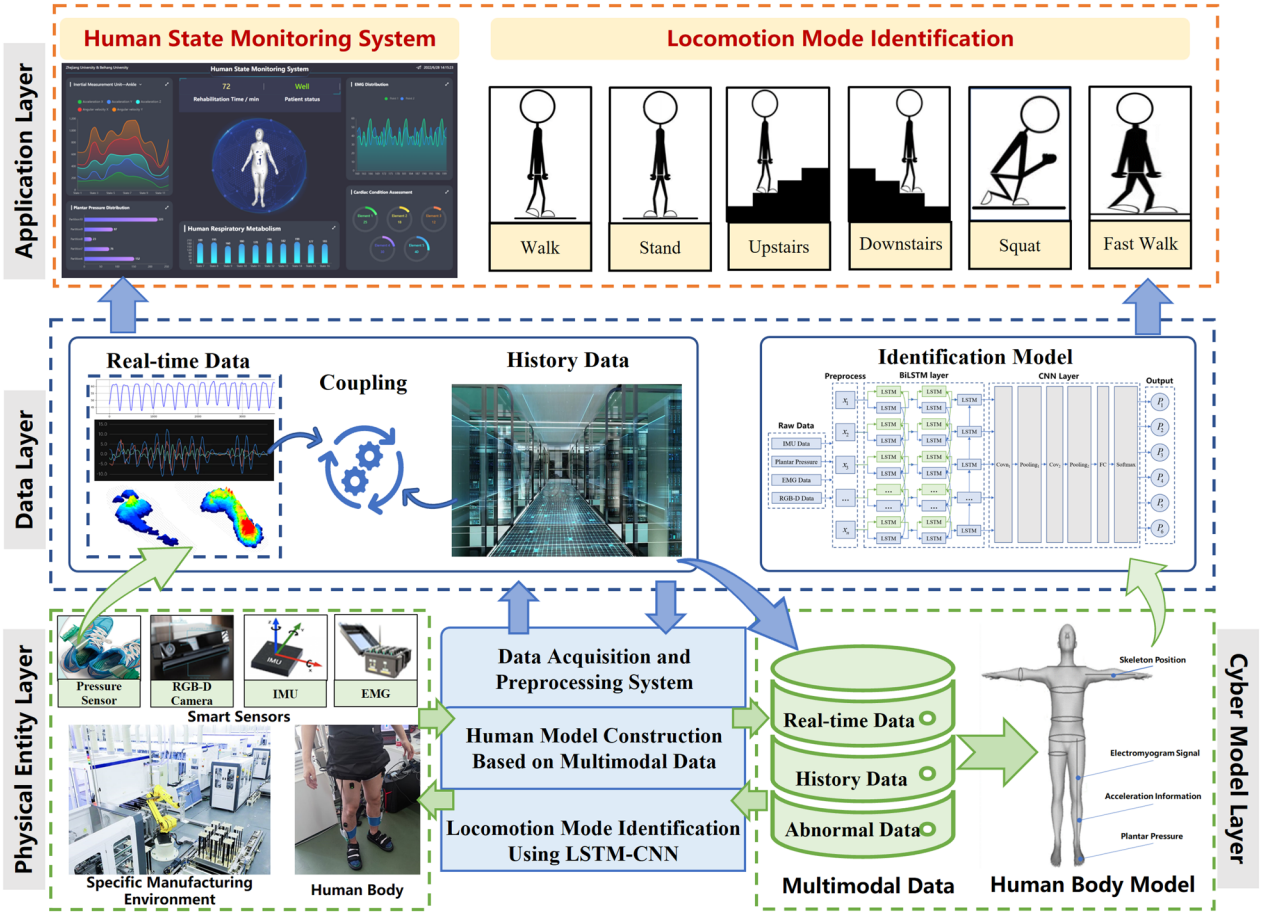

Figure 1 HDT model framework based on multimodal data

in physical space is realized. This multimodal feature information can describe the changes in the human body state over time at different granularities and different spatial scales, which enhances the immersion and authenticity of the human digital model.

Cyber model layer is the underlying support of the HDT model framework. Cyber model layer obtains the current and historical multimodal data of the human body from the data layer, and then realizes the establishment of a dynamically updated and high-fidelity human digital model. The established human digital model serves as the input basis for the intelligent identification model in data layer.

2.3 Data Layer

Data layer includes real-time human state data, historical data, and the intelligent identification model based on deep neural networks. The real-time state data of the human body is composed of IMU signals, depth camera signals, plantar pressure signals, and EMG signals. By integrating these real-time state data with historical data, the human body information is obtained from both current and historical perspectives, to ensure the comprehensiveness and accuracy of the perception of the human body. The intelligent identification model based on a deep neural network is composed of BiLSTM and CNN.

Data layer is the theoretical basis of the digital twin model of the human body. The real-time multimodal data measured by a variety of Intelligent sensors is obtained from physical entity layer as the input of data layer. All data is transferred through the data transfer module. The data transfer module includes Bluetooth, 5G, etc. These key technologies greatly shorten the data transmission time and provide a guarantee for the efficient transmission of information between the physical entity layer, cyber model layer, and application layer. The intelligent identification model based on the deep neural network can identify the current locomotion mode of the human body according to the human digital model in cyber model layer, and provide theoretical support for application layer.

2.4 Application Layer

Application layer includes a human state monitoring system and locomotion mode identification. Application layer can provide a series of human state monitoring and locomotion mode identification services according to the data and analysis results transmitted from data layer. The human state monitoring system is a real-time condition monitoring platform developed based on the web. According to the different human condition monitoring needs in the manufacturing workshop, it can be customized to generate different solutions. The human state monitoring system mainly includes human acceleration information, plantar pressure distribution, the position of human skeleton nodes, and EMG signals. Since the system is demonstrated through the cloud platform, it can be deployed synchronously on different devices, such as mobile phones and computers. Human locomotion mode identification is based on the intelligent identification model in data layer to analyze the key characteristic parameters of the human digital model, and obtain the real-time locomotion mode of the human body.

Application layer encapsulates the data, algorithms, and results in the HDT model according to different fields and different users, to realize convenient invocation and on-demand access to services. The current and historical multimodal data obtained from data layer and the output results of the intelligent identification model are used as the input of the human state monitoring system and locomotion mode identification, thereby realizing real-time and high-assurance monitoring of the human state and enhancing the perception and acquisition of human locomotion mode by machines, which makes the HDT model reliable, stable and interpretable, and improves the level of collaborative tasks between human and CPS.

3 Human Digital Twin Technology

3.1 Data Acquisition System and Preprocessing

To enable the HDT to perceive and update the human body state and physiological data in real-time, and accurately identify the human locomotion mode, a human body data acquisition, and preprocessing system is set up to obtain the multi-source heterogeneous data of humans.

3.1.1 Data Acquisition System

The system consists of an RGB-D camera, IMU sensors, EMG sensors, and plantar pressure sensors. An RGB-D camera (Kinect V2, USA) is used to obtain data on the overall position of human skeleton nodes. IMU sensors (Xsens Dot, USA) are installed on the human body's arms, legs, and other locations, which can measure 3-axis acceleration data. The EMG sensors (Delsys, USA) are set in the center of muscles, such as arms, legs, and other parts, to measure the changes in muscle signal intensity during human activities. A pair of plantar pressure sensors (Tekscan, USA) is attached inside the shoes to obtain data on the planter pressure distribution of the human body and gait state information during activities. The RGB-D camera communicated with the computer via USB. The IMU sensors, EMG sensors, and plantar pressure sensors are wirelessly connected to the computer terminal through Bluetooth and WIFI, and the data is transmitted to the computer terminal in real-time and then processed.

3.1.2 Data Processing

(1) Normalization

The multi-source heterogeneous data of the human body obtained by the data acquisition system has different units, which may be too large or too small. If the data is directly used to build an HDT model, fuse multimodal data, and train a neural network model, it will lead to large deviations. And then it brings problems such as poor training effect and slow convergence speed. Therefore, it is necessary to normalize the data from different sensors. The normalization calculation formula is as follows.

$$x'_{i,j} = 2 \times \frac{(x_{i,j} - (x_j)_{\min})}{(x_j)_{\max} - (x_j)_{\min}} - 1, \tag{1}$$

where $x_{i,j}$ is the i-th element in the sampling data of the j-th channel; $x_{i,j}$' represents the normalized data of $x_{i,j}$. $(x_j)_{\max}$ and $(x_j)_{\min}$ devote the maximum and minimum values in the sampling data of the j-th channel. After normalization, the maximum and minimum of $x_{i,j}$' are 1 and −1 respectively.

(2) Data segmentation

The sliding window method of segmenting data is one of the main methods used to process time series data and facilitate LSTM to exploit how the data changes over time. The sliding window method is to cut the time series data according to a certain step size and divide the original complete data set into data subsets with window size. Furthermore, it can make different data subsets have a certain correlation and bring the data into the LSTM network structure for better training. Assuming that $\boldsymbol{D}$ represents the initial dataset, it can be expressed as follows:

$$\boldsymbol{D}=\begin{bmatrix} X_1^1 & X_1^2 & X_1^3 & Y_1^1 & Y_1^2 & Y_1^3 & Z_1^1 & Z_1^2 & Z_1^3 & \cdots \\ X_2^1 & X_2^2 & X_2^3 & Y_2^1 & Y_2^2 & Y_2^3 & Z_2^1 & Z_2^2 & Z_2^3 & \cdots \\ X_3^1 & X_3^2 & X_3^3 & Y_3^1 & Y_3^2 & Y_3^3 & Z_3^1 & Z_3^2 & Z_3^3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \\ X_i^1 & X_i^2 & X_i^3 & Y_i^1 & Y_i^2 & Y_i^3 & Z_i^1 & Z_i^2 & Z_i^3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ X_n^1 & X_n^2 & X_n^3 & Y_n^1 & Y_n^2 & Y_n^3 & Z_n^1 & Z_n^2 & Z_n^3 & \cdots \end{bmatrix}, \quad (2)$$

where X_i^j, Y_i^j, and Z_i^j respectively devote the data at the i-th moment under the j-th channel of the X, Y, and Z sensor; n represents the time length of the initial dataset.

In this paper, the size of the sliding window is *len*, and the step size between the windows is *st*. Among them, the calculation formula of the k-th window is as follows:

$$\boldsymbol{D_k}=\begin{bmatrix} X_k^1 & X_k^2 & X_k^3 & \cdots \\ {X_{k+1}}^1 & {X_{k+1}}^2 & {X_{k+1}}^3 & \cdots \\ {X_{k+2}}^1 & {X_{k+2}}^2 & {X_{k+2}}^3 & \cdots \\ \vdots & \vdots & \vdots & \cdots \\ {X_{k+i}}^1 & {X_{k+i}}^2 & {X_{k+i}}^3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ {X_{k+len-1}}^1 & {X_{k+len-1}}^2 & {X_{k+len-1}}^3 & \cdots \end{bmatrix}. \quad (3)$$

3.2 Human Digital Model Construction

By analyzing the human body state and physiological signal data obtained by the data acquisition system and processing, the key feature information is extracted as the human locomotion feature signals, which comprehensively reflect the human body locomotion state and construct a human digital model.

The construction of the human digital model aims to use a variety of Intelligent sensors to obtain rich multimodal information about the human body, which can express the human body in a digital form during the manufacturing process. The construction of the human digital model is shown in Eq. (4):

$$\boldsymbol{S}=\{\boldsymbol{S_I},\boldsymbol{S_F},\boldsymbol{S_E},\boldsymbol{S_R}\}, \quad (4)$$

where S is the data set of the human digital model under the multi-source sensors. In this paper, S mainly includes the IMU acceleration signal dataset $\boldsymbol{S_I}$, the plantar pressure signal dataset $\boldsymbol{S_F}$, the RGB-D camera dataset $\boldsymbol{S_R}$ and the EMG signal dataset $\boldsymbol{S_E}$.

The IMU sensors are used to model the acceleration and angular velocity information of the human body and are composed of 3-axis accelerometers and 3-axis gyroscopes. The accelerometers can detect the acceleration signals of the x, y, and z axes of the installation position in the carrier coordinate system. The accelerometer can detect the acceleration signals of the x, y, and z axes at the installation position in the carrier coordinate system. Using these signals, the posture information of the current installation position of the human body can be obtained. After processing the obtained IMU signals, the acceleration and angular velocity signals of the human body at the corresponding installation position can be obtained, as shown in Eqs. (5) and (6):

$$\boldsymbol{S_I}=\{\boldsymbol{I^1},\boldsymbol{I^2},\cdots,\boldsymbol{I^k},\cdots,\boldsymbol{I^m}\}, \quad (5)$$

$$\boldsymbol{I^k}=\{I_{ax}^k,I_{ay}^k,I_{az}^k,I_{gx}^k,I_{gy}^k,I_{gz}^k\}, \quad (6)$$

where $\boldsymbol{I^k}$ is the data of the k-th IMU sensor; m is the number of IMU sensors; ${I_{ax}}^k$, ${I_{ay}}^k$, ${I_{az}}^k$ represents the acceleration signals of the k-th IMU sensor in x, y, and z axes; ${I_{gx}}^k$, ${I_{gy}}^k$, ${I_{gz}}^k$ represent the angular velocity signals of the k-th IMU sensor in x, y, and z axes.

The plantar pressure sensors model the plantar pressure distribution and gait information of the human body. The plantar pressure distribution map of the human body can be obtained by the pressure measurement system F-scan equipped with plantar pressure sensors. The obtained plantar pressure distribution map is divided into front and rear parts, and the average value of the front and rear sole pressure is taken as the pressure value of the toe and heel, to find the two events of heel-strike and toe-off. The digital model of human plantar pressure is shown in Eqs. (7)–(9):

$$\boldsymbol{S_F}=\{F_{\mathrm{h}},F_{\mathrm{t}}\}, \quad (7)$$

$$F_{\mathrm{h}}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}F_{i,j}^{\mathrm{h}}}{n^2}, \quad (8)$$

$$F_{\mathrm{t}}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}F_{i,j}^{\mathrm{t}}}{n^2}, \quad (9)$$

where F_{h} devotes the average distribution of heel pressure and is used to record the event of heel-strike during human activities; F_{t} is the average distribution of toe pressure, and is used to record the event of toe-off during human activities.

The EMG sensors model the state of the human muscle groups. The EMG signals can effectively reflect the muscle activation level of the human body, and it has a

high correlation with the muscle contraction force [29]. Therefore, EMG signals are applied to identify human locomotion intention in this study. Common muscle groups used for human lower extremity modeling were gluteus medius, right external oblique, semitendinosus, gracilis, biceps femoris, rectus femoris, vastus lateralis, vastus medialis, soleus, tibialis anterior, and gastrocnemius medialis [30]. The digital model formula to obtain the human EMG signals is shown in Eq. (10):

$$\boldsymbol{S_E} = \{E_1, E_2, \cdots, E_i, \cdots, E_{11}\}, \tag{10}$$

where E_i is the activation of the i-th muscle groups.

The final component of the construction of the human digital model is the position information of the human skeleton nodes obtained by the depth camera. In this paper, the depth camera used is the Kinect v2 camera, which is composed of an RGB camera, infrared camera, and infrared projector. The infrared camera and the infrared projector constitute a 3D structured light depth sensor, and then the computer graphics technology can be used to get the skeleton position information of the human. The digital model of the human skeleton position is shown in Eqs. (11) and (12):

$$\boldsymbol{S_R} = \{\boldsymbol{p_1}, \boldsymbol{p_2}, \cdots, \boldsymbol{p_i}, \cdots, \boldsymbol{p_n}\}, \tag{11}$$

$$\boldsymbol{p_i} = \{x_i, y_i, z_i\}, \tag{12}$$

where $\boldsymbol{p_i}$ represents the position information of the i-th human skeleton; n is the total number of collected human skeleton position information; x_i, y_i, and z_i respectively represent the coordinate values of the human skeleton nodes in the Cartesian coordinates.

3.3 Multimodal Data Fusion and Locomotion Mode Identification Using BiLSTM-CNN

Based on the proposed construction method of the human digital model, we design the neural network architecture composed of BiLSTM and CNN for multimodal sensor data fusion and human locomotion mode identification.

Regarding the identification of human locomotion modes, based on the proposed method of constructing a human digital model, the multimodal data mainly consists of acceleration signals, plantar pressure signals, electromyographic signals, and skeletal joint position signals. Among these, acceleration signals are selected from the right thigh and shank. Plantar pressure signals are selected from the right foot. Electromyographic signals are selected from the tibialis anterior muscle and the rectus femoris muscle. Skeletal joint position signals are selected from the anterior superior iliac spines (ASISs). The expression for inputting multimodal data is as follows:

$$\boldsymbol{x_t} = \{I_{ax}^t, I_{ay}^t, I_{az}^t, I_{ax}^s, I_{ay}^s, I_{az}^s, F_h, F_t, E_t, E_r, p_\mathrm{l}, p_\mathrm{r}\}, \tag{13}$$

where ${I_{ax}}^t$, ${I_{ay}}^t$, and ${I_{az}}^t$ represent the acceleration signals at the thigh; ${I_{ax}}^s$, ${I_{ay}}^s$, and ${I_{az}}^s$ represent the acceleration signals at the shank; E_t and E_r devote the electromyographic signals at the tibialis anterior muscle, and the rectus femoris muscle respectively; p_l and p_r are the skeletal joint position signals at the anterior superior iliac spines.

3.3.1 BiLSTM

LSTM is a network structure improved based on recurrent neural network (RNN) structure, so LSTM can effectively deal with temporally changing data. LSTM has the advantage of overcoming the problems of gradient disappearance and explosion in RNN [31]. Thus, it has been widely used in speech recognition, text translation, and state prediction.

LSTM is composed of three gated units: an input gate, a forget gate, an output gate, and a cell unit. σ is the sigmoid function, and is used to place the output range of the forget gate unit at [0,1]. tanh is an activation function that controls the out range between −1 and 1.

The function of the forget gate is to use a forget gate unit to selectively the $\boldsymbol{C_{t-1}}$ sent from the previous node. The calculation formula of the forget gate unit is as follows:

$$\boldsymbol{f_t} = \sigma(\boldsymbol{W_f x_t} + \boldsymbol{U_f h_{t-1}} + \boldsymbol{b_f}), \tag{14}$$

where $\boldsymbol{W_f}$ and $\boldsymbol{U_f}$ devote the weight of the coefficient matrix of the forget gate unit; $\boldsymbol{b_f}$ is the bias parameter of the forget gate unit; $\boldsymbol{C_{t-1}}$ represents the state information of the previous cell.

The input gate can selectively memorize the input, memorize useful information and reduce the memory of useless information, to determine how much information to put in the current state. The calculation of the input gate is divided into two steps. The first step is to use the sigmoid function to obtain the corresponding $\boldsymbol{i_t}$. The second step needs to generate alternative data for updating, which is obtained by using tanh function. The calculation formula for the input gate unit is as follows:

$$\boldsymbol{i_t} = \sigma(\boldsymbol{W_i x_t} + \boldsymbol{U_i h_{t-1}} + \boldsymbol{b_i}), \tag{15}$$

$$\widetilde{\boldsymbol{C_t}} = \tanh(\boldsymbol{W_c x_t} + \boldsymbol{U_c h_{t-1}} + \boldsymbol{b_c}), \tag{16}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C_t}, \tag{17}$$

where W_i, W_c, U_i, and U_c respectively represent the weight coefficient matrix of the input gate unit; b_i and b_c represent the bias parameters of the input gate unit; C_t is used to update the state information of the next cell.

The output gate decides which part of the information needs to be output according to the current cell state. Firstly, the output gate gets an o_t to control the content of the output par. Then use the tanh function to process the cell state, so that the calculation formula of the output gate unit can be obtained as follows:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \tag{18}$$

$$h_t = o_t \cdot \tanh(C_t), \tag{19}$$

where W_o, U_o respectively devotes the weight coefficient matrix of the output gate unit; b_o represents the bias parameters of the output gate unit.

Considering the process of training LSTM, the time series data is arranged in chronological order and then input into the network. Therefore, LSTM only considers the forward propagation of time series data, and ignores the backpropagation. BiLSTM adds backward learning based on the original LSTM model and obtains knowledge information from the backpropagation, which forms a two-way learning framework. In practical application scenarios, the process of identification and prediction often involves the information of the entire input sequence, so BiLSTM can well overcome the problem of one-way learning of LSTM. BiLSTM is proposed to process multimodal data fusion and model in HDT model.

3.3.2 CNN

CNN is a kind of feedforward neural network, which is composed of convolutional layers, pooling layers, and a fully connected layer.

Convolutional layers and pooling layers are the most important differences between CNN and traditional deep neural networks. The convolution layer is composed of a convolution kernel and an activation function. The local feature extraction of the input data can be achieved by using the convolution kernel. Through the parameter-sharing mechanism, the training parameters of the network are greatly reduced and the training efficiency is improved. The pooling layer is located after the convolutional layer. By extracting the main features of a certain area, the feature map and the number of parameters are reduced to prevent the model from overfitting. Currently, the commonly used pooling methods are maxpooling and average pooling. It is mainly necessary to extract the feature information of the data and ignore the interference of useless information on the accuracy in the process of locomotion mode identification. This paper chooses the maxpooling method. The fully connected layer is used to map the features extracted from the data by the convolutional and pooling layers with the data classification labels in the sample space. By connecting all neurons in the fully connected layer with the neurons in the previous layer, all local features are combined into global features.

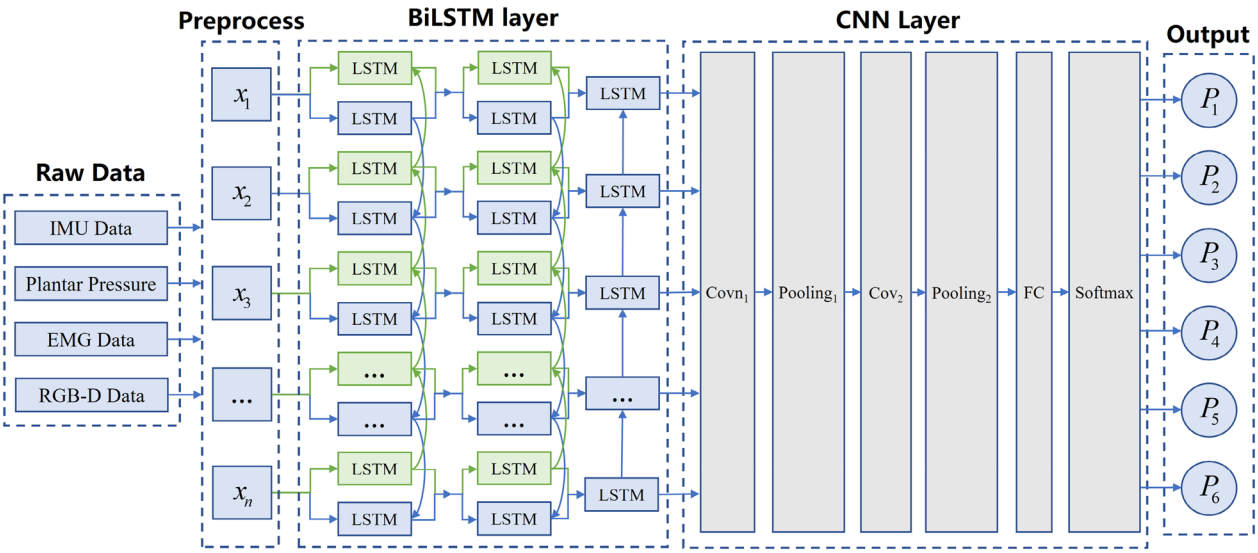

Figure 2 The structure of BiLSTM-CNN network

CNN can more efficiently extract the features in the data, thereby improving the accuracy of locomotion mode identification. The HDT model selects CNN to extract feature information in human locomotion mode identification.

3.3.3 Proposed Neural Network

This paper comprehensively considers the characteristics of BiLSTM and CNN. BiLSTM can process and model the input time series data from a bidirectional perspective. CNN can quickly and efficiently extract the characteristics of data features. Therefore, a BiLSTM-CNN network structure is proposed. The network structure is shown in Figure 2.

BiLSTM-CNN model consists of the raw data layer, preprocess layer, BiLSTM layer, CNN layer, and output layer. The BiLSTM layer is formed by stacking two BiLSTM units and one LSTM unit, in which the number of BiLSTM units and LSTM units is 128. CNN layer includes two convolutional layers, two pooling layers, and one fully connected layer. The two convolutional layers are $conv_1$ and $conv_2$ respectively. The number of convolution kernels of $conv_1$ is 128, and the parameters and step size of the convolutional kernel are [1, 3] and [1, 1] respectively. The number of convolution kernels of $conv_2$ is 64, and the parameters and step size of the convolutional kernel are [1, 3] and [1, 1] respectively. The pooling method used by the two pooling layers is maxpooling. The size of the pooling units in the two pooling layers is [1, 3] and the stride is set to 2. The second pooling layer is connected with a fully connected layer. Finally, the six locomotion mode results are output by the Softmax connected with the fully connected layer. The sensitivity analysis of the hyperparameters from the proposed network is described in Section 4.

The optimizer used in the training process of the BiLSTM-CNN network model is the Adam optimizer, where the learning rate is set to 0.001. The loss function of the network model is cross-entropy, and its calculation formula is as follows:

$$Loss = -\sum_{i} y_i' \log(y_i), \tag{20}$$

where y_i is the vector of the locomotion mode classification result after Softmax output; y_i' represents the vector of the actual locomotion mode label. Both y_i and y_i' are vector results obtained by encoding according to one-hot.

The training epoch of this network model is 100. The batch size is 50, and the step length is 128. To ensure the training effect of the network model, the data set is divided into two parts, of which 75% is used as the training set and 25% is used as the testing set. To prevent overfitting during model training, the L2 normalization is added to the loss function.

4 Case Study

4.1 Experimental Protocol

IMU sensors, plantar pressure sensors, EMG sensors, and depth cameras are used to collect the corresponding acceleration information, plantar pressure signals, EMG signals, and the position of human skeleton nodes to build an HDT model under six different locomotion modes and contribute to realizing the digital description and visualization of the human body and physiological data in the cyber space. Operators and managers can monitor the current physical state and movement of the human body in real-time through the digital monitoring system, and use the BiLSTM-CNN-based locomotion mode identification model to realize real-time monitoring and identification of human locomotion modes.

The BiLSTM-CNN model proposed is trained using the TensorFlow deep learning framework. The computer hardware parameters for this model training are CPU I5-12400F (2.5 GHz) and GPU NVIDIA RTX 3060 (12GB). The data acquisition system is used to record the data of the IMU sensors, the plantar pressure sensors, the EMG sensors, and the depth camera in six different human locomotion modes, as shown in Fig. 3. The arrangement of the sensors is as follows. The IMU sensors are respectively installed on the thigh and calf on the right side of the human body. The plantar pressure sensor is installed in the human right foot shoe. The EMG sensors are chosen to be installed in the center of the tibialis anterior and rectus femoris muscles. The depth camera collects the position of the human waist node. The sliding window algorithm is used to segment the data of different sensors, where the window length is 128 and the step length is 5 ms. The data

Figure 3 Data acquisition system

acquisition system collected a total of 9000 sample data, including six locomotion modes, namely walking, standing, squatting, going upstairs, going downstairs, and fast walking, with 1500 sample data for each locomotion mode. 75% of the data is used to train the model and 25% of the data is used to test the model.

To verify the importance of the proposed HDT model framework in human locomotion mode identification, the accuracy and F1 score are used as the evaluation indicators of the performance of the BiLSTM-CNN model. The *Accuracy* is used to evaluate the proportion of correctly identified data to the total data in the locomotion mode identification data set, and its formula is as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}, \tag{21}$$

where *TP*, *TN*, *FP*, and *FN* respectively devote the number of true positive, true negative, false positive, and false negative in the process of classifying sample data.

While the formula for calculating the F1 score is based on precision and recall. The formulas for calculating *Precision* and *Recall* are as follows:

$$Precision = \frac{TP}{TP + FP}, \tag{22}$$

$$Recall = \frac{TP}{TP + FN}. \tag{23}$$

Targeting only precision or recall has limitations and can lead to extremes in model training, and the accuracy can also be affected by imbalanced samples. Therefore, to take into account the precision and recall, while avoiding the impact of sample imbalance, the F1 score needs to be used [32–34]. The F1 score is a harmonic mean. Its calculation formula is as follows:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \tag{24}$$

4.2 Experimental Result

In this section, we first discuss the influence of hyperparameters in the BiLSTM-CNN model on its locomotion mode identification performance. The number of BiLSTM layers and the number of hidden units, convolution kernel size, and learning rate in the BiLSTM-CNN model are optimized. Then, the BiLSTM-CNN model proposed in this paper is compared with the conventional locomotion mode identification models, including LSTM and CNN. Finally, the superiority of the BiLSTM-CNN model in the accuracy of locomotion mode identification is proved. The training process of the BiLSTM-CNN model is shown in Figure 4.

To study the influence of the number of BiLSTM layers and the number of hidden units in the BiLSTM-CNN model on the proposed human locomotion mode model, the number of BiLSTM layers is set to 1, 2, and 3. The number of hidden units of BiLSTM is set to 16, 24, 32, and 64. The remaining parameters of the BiLSTM-CNN model remain the same. Comparative experiments are carried out by adjusting the number of BiLSTM layers and the number of hidden units in BiLSTM units.

The experimental results are shown in Figure 5. According to the change in the accuracy rate in Figure 5, it can be found that when the number of hidden units is the same and the number of BiLSTM layers is 2, the accuracy of the corresponding model recognition is higher. At the same time, when the number of BiLSTM layers is set to 2 and the number of hidden units is 32, the accuracy of model recognition is the highest. Therefore, the number

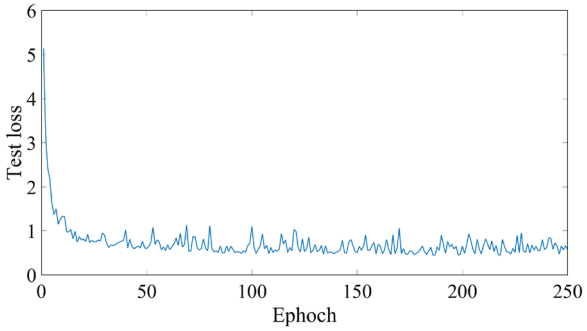

Figure 4 Convergence change of test loss during model training

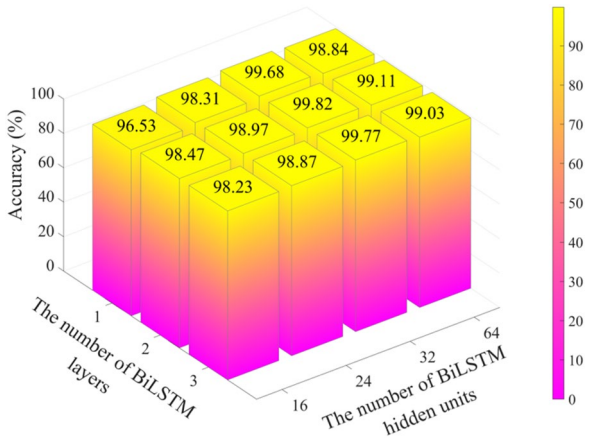

Figure 5 The influence of BiLSTM layers and hidden units on recognition accuracy

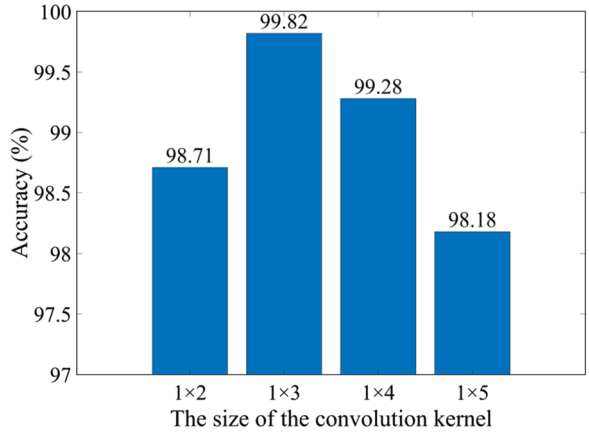

Figure 6 The influence of convolution kernel size on recognition accuracy

Table 1 The influence of learning rate on recognition accuracy

Learning rate	0.1	0.01	0.005	0.001
Accuracy (%)	95.12	97.74	98.21	99.82

of BiLSTM layers in the BiLSTM-CNN model is set to 2, and the number of hidden units is set to 32.

To verify the influence of the size of the convolution kernel of the convolutional layer in the BiLSTM-CNN model on the performance of the model, the size of the convolution kernel is set to [1, 2], [1, 3], [1, 4], [1, 5]..

When the remaining parameters of the BiLSTM-CNN model remain the same, the accuracy of the BiLSTM-CNN model with different convolution kernel sizes is compared. The experimental results are shown in Figure 6. When the convolution kernel size is [1, 3], the accuracy of model recognition is the highest. Therefore, the BiLSTM-CNN model selects convolution kernels of size [1, 3].

In the training process of the BiLSTM-CNN model for human locomotion mode identification, the learning rate is closely related to the training duration and convergence speed. On the one hand, when the learning rate is low, the overall update and convergence speed of the model training process is slow. On the other hand, when the learning rate is high, there will be oscillations during the model training process, failing to converge to the optimal solution. Therefore, an appropriate learning rate is crucial to the training of the model. According to the data in Table 1, it can be found that when the learning rate is 0.001, the accuracy of human locomotion mode identification is the highest. Therefore, the learning rate during BiLSTM-CNN model training is set to 0.001.

To prove the superiority of the BiLSTM-CNN model proposed in this paper in human locomotion mode identification, the proposed model is compared with several existing human locomotion mode identification models, including LSTM and CNN. In the process of comparing different models, LSTM and CNN have the same structure and hyperparameters. The confusion matrix of three different models in human locomotion mode identification is shown in Figure 7. Among them, the human locomotion mode identification of the BiLSTM-CNN model proposed in this paper has the highest accuracy rate of about 99.82%. The identification accuracy of LSTM, and CNN models are all below 96%. At the same time, we compared the F1 scores of different models and found that the F1 score of the BiLSTM-CNN model exceeded 99%, which was significantly better than the other two types of models. Therefore, under the framework of the HDT model proposed in this paper, the proposed BiLSTM-CNN model for human locomotion mode identification has a good recognition accuracy.

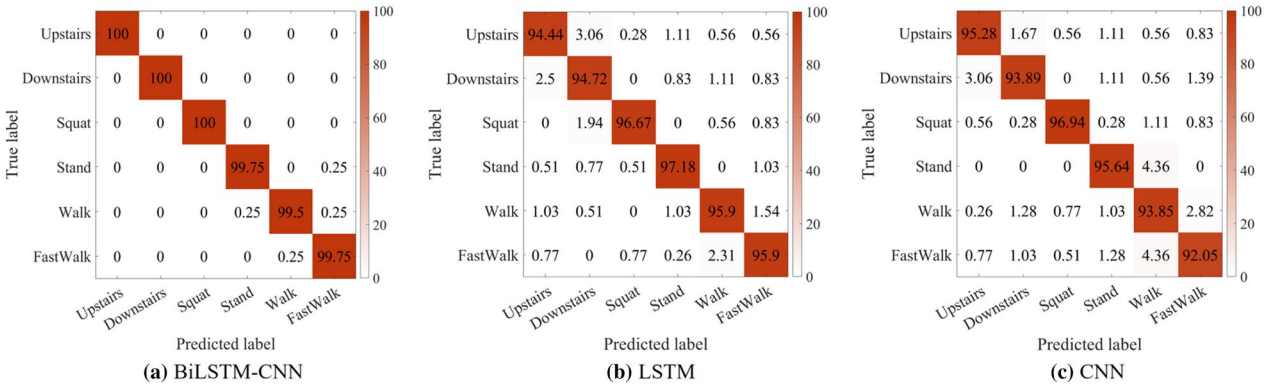

Figure 7 The confusion matrix of three different models in human locomotion mode identification

5 Conclusions

With the widespread implementation of human-centered intelligent manufacturing, it has become an inevitable requirement to improve the efficiency of collaborative work between humans and manufacturing systems. To address the above challenges, this paper proposes a framework for building a human digital twin model to enhance the in-depth integration of humans and CPS. In summary, the main contributions of this paper are as follows:

(1) A framework for modeling an HDT based on multimodal data is proposed in this paper. To realize the dynamic update and real-time analysis of the HDT, the framework integrates key technologies for the construction of the HDT. The data acquisition system and preprocessing technology realize the dynamic perception of the human body. The human digital model construction technology facilitates the digital representation of complex human features.
(2) A multimodal data fusion method based on BiLSTM-CNN is devised to realize real-time monitoring of human and locomotion mode identification.
(3) Furthermore, we also conduct experiments on hyperparameter optimization in the BiLSTM-CNN network and compare the proposed model with traditional locomotion mode identification models. The results indicate that the proposed model has an accuracy rate of about 99.83% for six human locomotion modes.

The HDT model building framework based on multimodal data can achieve high-quality modeling of the complex human body and help realize dynamic mapping of human physical and virtual entities. The framework obtains good results in human motion pattern recognition experiments and outperforms traditional locomotion mode identification algorithms.

The construction of the HDT model is to better serve the HCPS and improve the automation level of manufacturing. However, the construction of the HDT model framework only focuses on the identification of human locomotion modes in manufacturing scenarios, and does not establish the connection between the results of the HDT model and the intelligent control of the machine. In the next step, we will focus on combining the constructed HDT model with the CPS so that the machine can perceive the changes in the human body and physiological data, and then promote the adaptive control of HCPS.

Acknowledgements
No applicable.

Author Contributions
RZ and BH conceived the model constructing framework. RZ write the manuscript. BH was responsible for guiding and reviewing the writing of the manuscript. YF, HZ, ZH, and SL reviewed and supervised the study. JT supervised the study. All authors read and approved the final manuscript.

Authors' Information
Ruirui Zhong, born in 1998, received his bachelor's degree from *Jiangnan University, China*, in 2017. He is currently a PhD candidate at the *State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, China*. His research interests include human digital twin and human-cyber-physical system.
Bingtao Hu, born in 1992, received his B.S. and Ph.D. degrees in mechanical engineering from *Zhejiang University, China*, in 2014 and 2021, respectively. He is currently a Research Assistant with the *Department of Mechanical Engineering, Zhejiang University, China*, and a Member of the *State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, China*. His research interests include big data and design theory.
Yixiong Feng, born in 1975, received his B.S. and M.S. degrees in mechanical engineering from *Yanshan University, China*, in 1997 and 2000, respectively, and a Ph.D. degree in mechanical engineering from *Zhejiang University, China*, in 2004. He is currently a Professor at the *Department of Mechanical Engineering, Zhejiang University, China*. He is also a member of the *State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, China*. His research interests include mechanical product design theory, intelligent automation, and advanced manufacturing technology.
Hao Zheng, born in 1988, received his B.S. degree in mechanical engineering from *Southwest Jiaotong University, China*, in 2010, and a Ph.D. degree in mechanical engineering from *Zhejiang University, China*, in 2017. He is currently an Associate Research Fellow with the *Hangzhou Innovation Institute of Beihang University, China*. His research interests include data-driven design, decision-making and optimization, multi-objective evolutionary algorithms, and human-machine integration.
Zhaoxi Hong, born in 1990, received her Ph.D. degree in mechatronic engineering from *Zhejiang University, China*, in 2020. She is a member of the *State Key Lab of Fluid Power and Mechatronic Systems, Zhejiang University, China*. Her research focuses on low carbon mechanical product design and uncertain optimization in intelligent manufacturing.
Shanhe Lou, born in 1993, received his Ph.D. degree from the *School of Mechanical Engineering, Zhejiang University, China*, in 2020. He is currently a Research Fellow with the *Department of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore*. His research interests include intelligent design and manufacturing, and human-cyber-physical systems for autonomous vehicles.
Jianrong Tan, born in 1954, received the B.S. and first M.S. degrees in mechanical engineering and electronic engineering from *The Open University of China, China*, in 1982, the second M.S. degree in engineering from *Huazhong University of Science and Technology, China*, in 1985, and the Ph.D. degree in mathematics from *Zhejiang University, China*, in 1987. He is an Academician of the *Chinese Academy of Engineering* and a member of the *State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, China*. His research focuses on mechanical design theory, and digital design and manufacturing.

Funding
Supported by National Natural Science Foundation of China (Grant Nos. 52205288, 52130501, 52075479) and Zhejiang Provincial Key Research & Development Program (Grant No. 2021C01110).

Declarations

Competing Interests
The authors declare no competing financial interests.

Received: 3 October 2022 Revised: 13 September 2023 Accepted: 15 September 2023
Published online: 20 October 2023

References

[1] K Xu, Y G Li, C Q Liu, et al. Advanced data collection and analysis in data-driven manufacturing process. *Chinese Journal of Mechanical Engineering*, 2020, 33: 43.
[2] Y Fu, G Zhu, M L Zhu, et al. Digital twin for integration of design-manufacturing-maintenance: An overview. *Chinese Journal of Mechanical Engineering*, 2022, 35: 80.
[3] S Sahoo, C Y Lo. Smart manufacturing powered by recent technological advancements: A review. *Journal of Manufacturing Systems*, 2022, 64: 236-250.
[4] J Zhou. Intelligent manufacturing——Main direction of "Made in China 2025". *China Mechanical Engineering*, 2015, 26(17): 2273. (in Chinese)
[5] N Bhagwan, M Evans. A review of industry 4.0 technologies used in the production of energy in China, Germany, and South Africa. *Renewable and Sustainable Energy Reviews*, 2023, 173: 113075.
[6] B A Salau, A Rawal, D B Rawat. Recent advances in artificial intelligence for wireless internet of things and cyber–physical systems: A comprehensive survey. *IEEE Internet of Things Journal*, 2022, 9(15): 12916-12930.
[7] D G Pivoto, L F D Almeida, R D R Righi, et al. Cyber-physical systems architectures for industrial internet of things applications in Industry 4.0: A literature review. *Journal of Manufacturing Systems*, 2021, 58: 176-192.
[8] S M Liu, J S Bao, P Zheng. A review of digital twin-driven machining: From digitization to intellectualization. *Journal of Manufacturing Systems*, 2023, 67: 361-378.
[9] J Zhou, Y H Zhou, B C Wang, et al. Human–cyber–physical systems (HCPSs) in the context of new-generation intelligent manufacturing. *Engineering*, 2019, 5(4): 624–636.
[10] B C Wang, S H Huang, B Yi, et al. State-of-art of human factors/ergonomics in intelligent manufacturing. *Journal of Mechanical Engineering*, 2020, 56(16): 240-253. (in Chinese)
[11] D Mourtzis, N Panopoulos, J Angelopoulos, et al. Human centric platforms for personalized value creation in metaverse. *Journal of Manufacturing Systems*, 2022, 65: 653-659.
[12] B C Wang, H Y Zhou, G Yang, et al. Human digital twin (HDT) driven human-cyber-physical systems: Key technologies and applications. *Chinese Journal of Mechanical Engineering*, 2022, 35: 11.
[13] X G Song, X W He, K P Li, et al. Construction method and application of human skeleton digital twin. *Journal of Mechanical Engineering*, 2022, 58(18): 218-228. (in Chinese)
[14] I E Makrini, G Mathijssen, S Verhaegen, et al. A virtual element-based postural optimization method for improved ergonomics during human-robot collaboration. *IEEE Transactions on Automation Science and Engineering*, 2022, 19(3): 1772-1783.
[15] E Escobar-Linero, M Domínguez-Morales, J L Sevillano. Worker's physical fatigue classification using neural networks. *Expert Systems with Applications*, 2022, 198: 116784.
[16] P Franceschi, S Mutti, K Ottogalli, et al. A framework for cyber-physical production system management and digital twin feedback monitoring for fast failure recovery. *International Journal of Computer Integrated Manufacturing*, 2022, 35(6): 619-632.
[17] F Bocklisch, G Paczkowski, S Zimmermann, et al. Integrating human cognition in cyber-physical systems: A multidimensional fuzzy pattern model with application to thermal spraying. *Journal of Manufacturing Systems*, 2022, 63: 162-176.
[18] X Y Zhang, J M Fan, T Peng, et al. A privacy-preserving and unobtrusive sitting posture recognition system via pressure array sensor and infrared array sensor for office workers. *Advanced Engineering Informatics*, 2022, 53: 101690.
[19] P B Rodrigues, Y J Xiao, Y E Fukumura, et al. Ergonomic assessment of office worker postures using 3D automated joint angle assessment. *Advanced Engineering Informatics*, 2022, 52: 101596.
[20] L L Wang, Y Zhou, R Li, et al. A fusion of a deep neural network and a hidden Markov model to recognize the multiclass abnormal behavior of elderly people. *Knowledge-Based Systems*, 2022, 252: 109351.
[21] L P Huang, J B Zheng, H C Hu. A gait phase detection method in complex environment based on DTW-mean templates. *IEEE sensors journal*, 2021, 21(13): 15114-15123.
[22] L P Huang, J B Zheng, H C Hu. Online gait phase detection in complex environment based on distance and multi-sensors information fusion using inertial measurement units. *International Journal of Social Robotics*, 2022, 14(2): 413-428.
[23] Ç B Erdaş, S Güney. Human activity recognition by using different deep learning approaches for wearable sensors. *Neural Processing Letters*, 2021, 53: 1795-1809.
[24] I Kang, D D Molinaro, S Duggal, et al. Real-time gait phase estimation for robotic hip exoskeleton control during multimodal locomotion. *IEEE Robotics and Automation Letters*, 2021, 6(2): 3491-3497.
[25] X Y Wu, Y Yuan, X K Zhang, et al. Gait phase classification for a lower limb exoskeleton system based on a graph convolutional network model. *IEEE Transactions on Industrial Electronics*, 2021, 69(5): 4999-5008.
[26] J Lee, W Hong, P Hur, et al. Continuous gait phase estimation using LSTM for robotic transfemoral prosthesis across walking speeds. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2021, 29: 1470-1477.
[27] C F Chen, Z J Du, L He, et al. A novel gait pattern recognition method based on LSTM-CNN for lower limb exoskeleton. *Journal of Bionic Engineering*, 2021, 18: 1059-1072.
[28] L Chen, X Z Liu, L Y Peng, et al. Deep learning based multimodal complex human activity recognition using wearable devices. *Applied Intelligence*, 2021, 51: 4029-4042.
[29] D Z Xiong, D H Zhang, X G Zhao, et al. Deep learning for EMG-based human-machine interaction: A review. *IEEE/CAA Journal of Automatica Sinica*, 2021, 8(3): 512-533.
[30] J Camargo, A Ramanathan, W Flanagan, et al. A comprehensive, open-source dataset of lower limb biomechanics in multiple conditions of stairs, ramps, and level-ground ambulation and transitions. *Journal of Biomechanics*, 2021, 119: 110320.
[31] Z H Cheng, B Chen, R Y Lu, et al. Recurrent neural networks for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(2): 2264-2281.
[32] K Kawintiranon, L Singh, C Budak. Traditional and context-specific spam detection in low resource settings. *Machine Learning*, 2022, 111(7): 2515-2536.
[33] P Yang, C Yang, V Lanfranchi, et al. Activity graph based convolutional neural network for physical activity recognition using acceleration and gyroscope data. *IEEE Transactions on Industrial Informatics*, 2022, 18(10): 6619-6630.
[34] Y Dong, X Li, J Dezert, et al. Dezert-Smarandache theory-based fusion for human activity recognition in body sensor networks. *IEEE Transactions on Industrial Informatics*, 2020, 16(11): 7138-7149.