

Full length article

A privacy-preserving and unobtrusive sitting posture recognition system via pressure array sensor and infrared array sensor for office workers

Xiangying Zhang^a, Junming Fan^b, Tao Peng^{a,*}, Pai Zheng^{b,*}, C. K. M. Lee^b, Renzhong Tang^{a,c}^a State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou 310027, China^b Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China^c Key Laboratory of Advanced Manufacturing Technology of Zhejiang Province, Zhejiang University, Hangzhou 310027, China

ARTICLE INFO

ABSTRACT

Keywords:

Sitting posture recognition
Pressure array sensor
Infrared array sensor
Ergonomics
Deep learning

Sitting posture recognition is essential in preventing work-related musculoskeletal disorders (WMSDs). WMSDs are of huge concern for office workers whose working process is averagely 81.8% sedentary. Prevailing studies have utilized cameras, wearables, and pressure sensors to recognize sitting postures. The cameras and wearables can achieve accurate recognition results, while personal privacy concerns and inconvenience for long-term use impede their adoption. Meanwhile, the pressure sensors are privacy-preserving and convenient. However, they cannot accurately recognize the sitting posture with different states of the trunk, head, upper extremity, and lower extremity. Considering the pros and cons of those approaches, this study proposes a novel privacy-preserving and unobtrusive sitting posture recognition system, which combines a pressure array sensor with another privacy-preserving sensing technology, i.e., an infrared array (IRA) sensor. Moreover, a deep learning-based sitting posture recognition algorithm is developed, which adopts a feature-level fusion strategy and does not require a complex handcrafted feature extraction process. Based on the ergonomics studies, ten daily sitting postures with the states of different body parts are selected. This system achieved an overall 90.6% accuracy using the leave-subject-out validation approach based on the self-collected dataset from 21 subjects. It has a great potential for privacy-preserving and unobtrusive related applications for sitting posture management.

1. Introduction

The advancement of sensing technologies brings numerous opportunities and challenges to promoting workplace health [1]. Various novel applications such as stress management and remote health monitoring are emerging. These applications significantly contribute to the transformation from curing work-related diseases to daily prevention [2,3]. Office workers are a crucial part of the workforce, and work-related musculoskeletal disorders (WMSDs) management application is of great significance in their health promotion [4]. WMSDs are common work-related health issues, with symptoms such as back pain, neck pain, and improper spine alignment. These health issues lead to a decline in productivity for employees and a rise in employers' costs. It was observed that 80% of bank office workers [5] and up to 53% of the Danish employees [6] suffered from WMSDs. Sitting posture is a major risk factor for WMSDs [7]. Hence, it is necessary to recognize office workers' daily sitting posture to prevent the development of these chronic disorders. In particular, office workers spend massive time

seated, accounting for an average of 81.8% of their regular working hours [8].

Nowadays, a sitting posture recognition system can be developed by utilizing sensing technologies to deliver effective daily health management solutions for office workers. The recognition results help improve office furniture design and even support decision-making by ergonomics experts or healthcare professionals [7,8]. However, office workers' acceptance of sensing technologies should be fully considered to provide daily monitoring.

Wearables devices [11-15], RGB or RGB-D cameras [16-19], and pressure sensors [10,20-24] are the commonly adopted technologies for sitting posture recognition. A wearable device is capable of continuous monitoring, while it needs to be attached to the user's skin or clothes to collect data. Therefore, it may not be practical at the current stage because continuous recognition requires users' acceptance to keep the device operational. An RGB or RGB-D camera is accurate for recognizing the postures of each body part and even the positions of joints. However, it raises users' privacy concerns because their personal information

* Corresponding authors.

E-mail addresses: tao_peng@zju.edu.cn (T. Peng), pai.zheng@polyu.edu.hk (P. Zheng).

could be captured. Pressure sensors are widely adopted, which are attached to a chair to build an unobtrusive and privacy-preserving recognition system. Successful recognition results in existing studies have been made for the pre-defined postures, including the changes in the trunk and lower extremity states, e.g., leaning forward, crossed legs, and leaning to the left. Nevertheless, the changes in the neck and upper extremity states are beyond the sensing capability of this unimodal sensing system. Hence, recently, pressure sensors have been combined with other sensors to obtain complementary modality characteristics [25]. Jeong et al. [26] attached the pressure and distance sensors to the chair's surface and back, respectively. They designed handcrafted features and then adopted the k-Nearest Neighbor (KNN) algorithm as the classifier. However, their system's emphasis mainly lies in recognizing the trunk's different inclination and rotation states.

All four body parts (containing the trunk, head, upper extremity, and lower extremity), based on the ISO 11226–2000 international standard for ergonomic evaluation of working postures, are equally important [27]. For instance, keeping the neck flexed is significantly linked with neck pain [28], and the risk increases as the flexion angle increases [7]. Also, compared with sitting upright, a forward sitting posture may lead to low back pain [29]. Therefore, it is essential to recognize the sitting postures containing different states of the four body parts, which is from the perspective of ergonomics and benefits the investigation of the potential health risks regarding sustained sitting postures. Moreover, a favorable solution is to develop a privacy-preserving and unobtrusive system without cameras or wearables. The solution can be achieved based on multimodal data collected by the pressure sensors and a low-resolution infrared array (IRA) sensor [30]. However, there are no studies to date exploring this issue.

This study aims to develop a privacy-preserving and unobtrusive sitting posture recognition system based on multimodal data. It is the first time to combine the IRA sensor with the pressure array sensor for the sitting posture recognition. The pressure array sensor consists of multiple pressure sensors forming a matrix. Moreover, a deep learning-based approach is proposed to fuse the multimodal data to recognize sitting postures. The postures are selected based on ergonomics studies and involve different states of the four body parts. This paper is organized as follows: Section 2 summarizes the related works on sitting posture recognition and the IRA sensor. Section 3 describes the proposed novel recognition system, including the hardware prototype deployed in the cubicle and the proposed deep learning model. Ten sitting posture categories are predefined in particular based on existing ergonomics studies, demonstrated in Section 3. The experimental data collection, comparative studies, and validation approach are presented in Section 4. More importantly, the experimental results are discussed in Section 5, followed by the conclusion in Section 6.

2. Related works

This section describes related works on the sitting posture recognition and the IRA sensor.

2.1. Sitting posture recognition

Unimodal data have been extensively leveraged for sitting posture recognition based on the wearables, cameras, and pressure sensors [11]. For instance, Feng et al. [12] attached three radio frequency identification (RFID) tags to the user's back and investigated the relationship between radio frequency signals and postures. This system successfully classified seven sitting postures. Kulikajevas et al. [16] proposed a deep recurrent hierarchical network to analyze the data of the RGB-Depth camera. Ho et al. [18] developed a posture classification framework based on a Kinect and the max-margin classifier.

Extensive studies focused on the pressure sensors attached to the chair [10]. Tan et al. [24] utilized a pressure array sensor to classify 14 sitting postures. They proposed a Principal Component Analysis (PCA)-

based classification algorithm, and the results indicated that their algorithm performed well on familiar users with an accuracy of 96% and 79% for unfamiliar users. Meyer et al. [31] developed a pressure measurement system using 240 pressure sensing elements. The Preisach model was applied to reduce the measurement error. Their system was tested on a self-collected dataset using leave-one-person-out cross-validation, containing 16 sitting postures' data obtained from 9 subjects. For the sensors attached to the seat surface, 59% accuracy was realized using the Naïve Bayes classifier. Meanwhile, the combination of sensors attached to the seat and back achieved 84% accuracy. Similarly, the sensing chair built by Ma et al. [22] attached pressure sensors to both the seat surface and back. Hu et al. [20] proposed an artificial neural network (ANN) algorithm, which realized a high accuracy of 97.4% on a dataset consisting of 7 sitting postures from 11 subjects. Anwary et al. [32] were centered on asymmetrical sitting and developed a fuzzy rule-based classifier. Their approach was validated on a gathered dataset collected from 10 subjects. Specifically, their posture categories were determined on the postures' asymmetry degrees.

Multimodal data have been leveraged recently to improve the performance of the unimodal systems [33]. Tariq et al. [17] combined a Kinect and a smartwatch to solve a similar joint position issue, which performed better than a single device by 12% accuracy. Jeong et al. [26] utilized pressure and distance sensors to accurately recognize the trunk's fine-grained states. Nevertheless, multimodal data-based studies are still limited regarding sitting posture recognition. The existing studies fail to consider a privacy-preserving and unobtrusive recognition of sitting postures with different states of body parts. For instance, the typical sitting postures usually include sitting straight, leaning to the left/right/forward/back, different types of crossed-legs, and slouched sitting [34]. Moreover, multimodal data fusion relied on handcrafted features and machine learning approaches. The handcrafted features' generation highly relies on human expertise and could encounter challenges for modality representation. Deep learning techniques can solve this issue and achieve end-to-end implementation, e.g., the deep convolutional neural network (DCNN) [14,15].

2.2. Related works on infrared array sensor

The IRA sensor can capture the temperature distribution of the human body directly in a non-contact way. The IRA sensor has a significantly low resolution (24×32) compared to an ordinary high-resolution infrared camera (640×480), preserving users' personal privacy. It has been applied to the domestic occupancy estimation [30] and fall detection [35].

There are limited studies regarding human posture and activity recognition based on the IRA sensor [35–37]. Gochoo et al. [38] used three IRA sensors and developed a privacy-preserving yoga posture recognition system to classify 26 postures using a DCNN, which obtained 99.9% accuracy using 10-fold cross-validation. Furthermore, their experiment validated that a single IRA sensor can also reach similar performance compared with three IRA sensors. Meanwhile, an IRA sensor (MLX 90640) was leveraged by Chen et al. [39] for sleep posture recognition. The histogram of oriented gradient, a well-known object detection method, was applied to generate features of infrared images and realized 86.0% accuracy. Tateno et al. [35] designed a motion detection system for the elderly to detect 8 motions. Afterwards, a three-dimensional convolutional neural network (3D CNN) was deployed and achieved 98.8% accuracy. Mashiyama et al. [37] utilized one IRA sensor to recognize 5 kinds of daily activities. Their proposed algorithm includes three parts: detect the presence of human body, extract four pre-designed features, and classify the postures using the support vector machine (SVM). This algorithm realized an average accuracy of 94.6%.

In summary, the IRA sensor has shown its sensing capability in studies associated with the human body. Nevertheless, to the best of our knowledge, there is a lack of work investigating the IRA sensor's effectiveness for sitting posture recognition as well as the combination of the

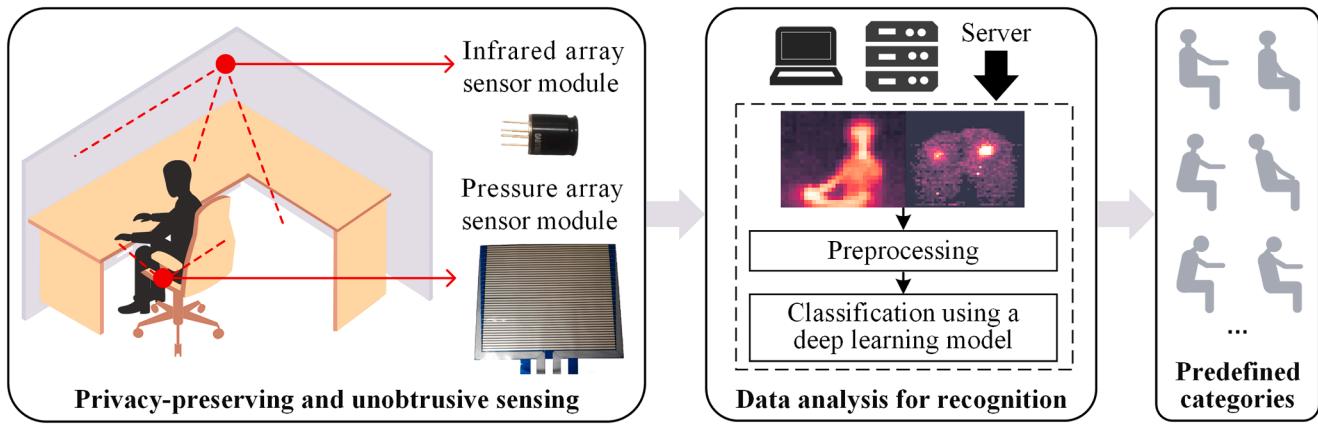


Fig. 1. The structure of the proposed sitting posture recognition system.

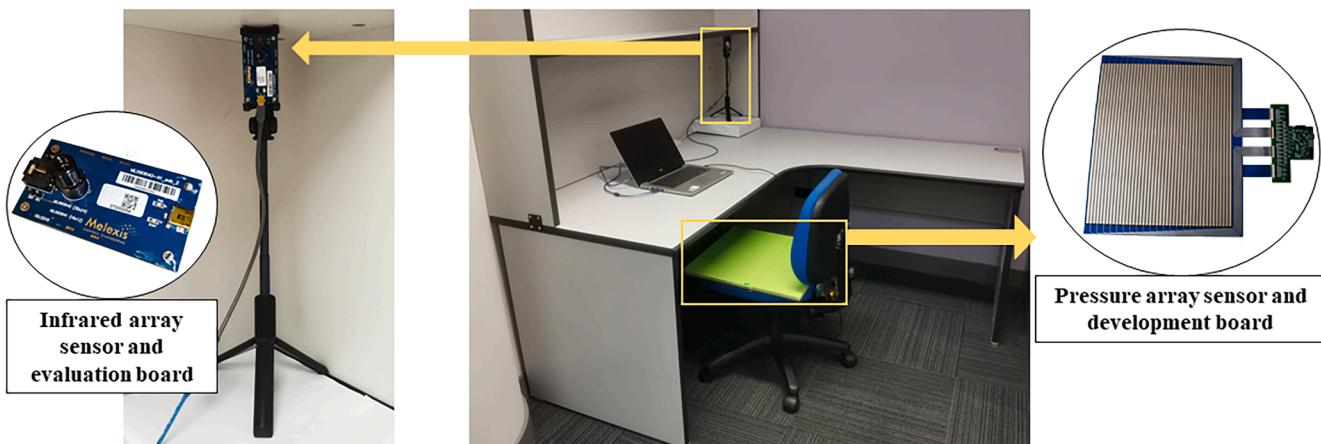


Fig. 2. The hardware prototype and deployment of the pressure sensing module and infrared sensing module.

Table 1
Specifications of the pressure array sensor and IRA sensor.

Sensor	Item	Specification
Pressure array sensor	Size	40.5 cm × 40.5 cm
	Frame rate	1 Hz – 100 Hz
	Thickness	0.2 mm
	Number of sensing elements	2288 (44 rows × 52 columns)
	Height and width of sensing matrix	30.5 cm × 36.4 cm
IRA sensor	Pressure measure range	Up to 1 Kg for each sensing element
	Frame rate	0.5 Hz – 64 Hz
	Resolution	0.1 °C
	Number of sensors	768 (24 rows × 32 columns)
	Viewing angle	110° × 75°
	Temperature measure range	–40 °C to 300 °C

IRA sensor and pressure sensors.

3. Proposed sitting posture recognition system

3.1. Description of the system structure

The overall structure of the proposed novel sitting posture recognition system is illustrated in Fig. 1. This system consists of three key aspects: hardware prototype, recognition algorithm, and the predefined sitting posture categories. Firstly, a pressure array sensor module and an IRA sensor module are deployed on the cubicle to obtain the sitting posture data unobtrusively with the users' data privacy preserved.

Second, the gathered multimodal data are transmitted to the server, where the data pre-processing and sitting posture classification are executed to output the sitting posture category. Especially, the carefully selected categories with identified different health risks are demonstrated, which have not been well investigated in previous studies of sitting posture recognition. However, as a health-oriented work, this study focuses on recognizing daily sitting postures with different health risks. The system details are as follows.

3.2. Hardware prototype

The first key aspect of the system is the hardware for sensing. Fig. 2 shows the hardware prototype and deployment in a typical cubicle for office work. The prototype has a pressure sensing module and an infrared sensing module. Detailed specifications of the sensors are listed in Table 1.

The pressure sensing module comprises a microcontroller and a pressure array sensor, which has 2288 pressure sensing elements (resistive transducers), forming a 44 × 52 matrix. The pressure array sensor can transform the pressure values into analog voltages and measure up to 1Kg pressure. This is sufficient to gather sitting posture pressure distribution. The microcontroller has multiplexers to control the multiple signals transmission. A 0.2 cm thick wooden board is applied to support the pressure array sensor because it is easily deformed to affect the pressure measurement. Also, a sponge mat is placed on the sensor for comfort.

The infrared sensing module consists of an evaluation board and an IRA sensor (MLX90640) with 24 × 32 infrared detectors, which relies on

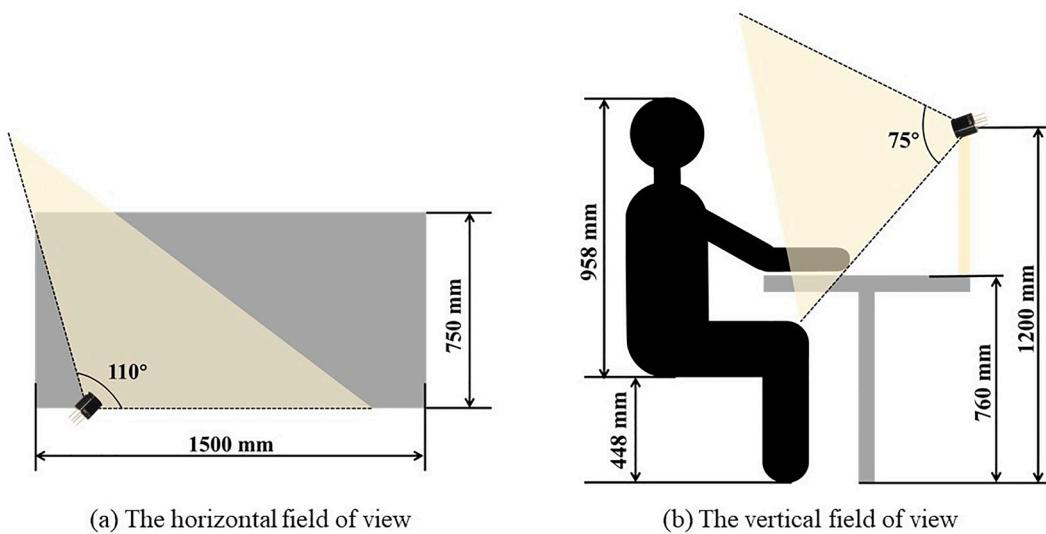


Fig. 3. The deployment illustration of the IRA sensing module.

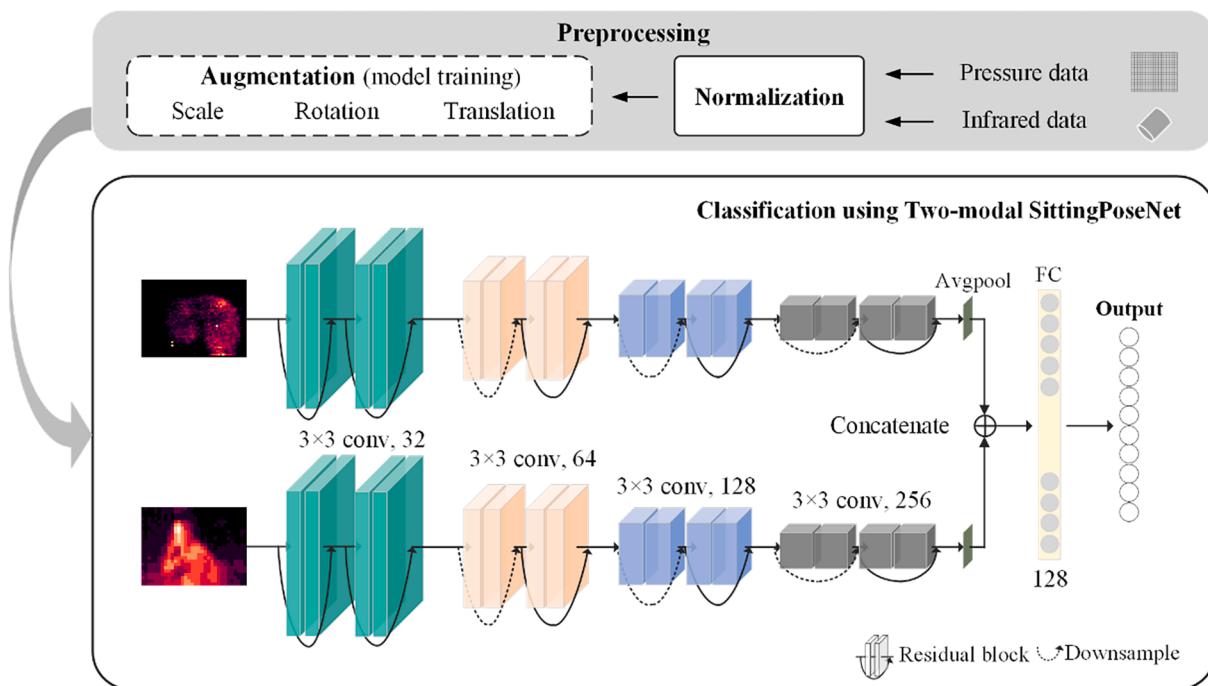


Fig. 4. Sitting posture recognition algorithm.

far-infrared radiation and outputs 768 temperature values. The horizontal and vertical fields of view (FOV) are 110° and 75°, respectively. The evaluation board (EVB90649-4, manufactured by Melexis Corp) is responsible for the infrared data collection and transmission. The infrared sensing module is mounted on a bracket and placed on the desk. Its setup is illustrated in Fig. 3. According to the National Standard of the People's Republic of China GB10000-88 (human dimensions of Chinese adults) [40], the 95th percentile anthropometric measurement data of sitting height and calf and foot for the adult male are 958 mm and 448 mm, respectively. The size of the desk based on another Chinese national standard GB/T 3326-2016 (Furniture - Main sizes of tables, chairs and stools) is 1500 mm × 750 mm [41]. The anthropometric measurement data of the user and the size of the desk in Fig. 3 are set as the above values proportionately. It indicates that the FOV of the IRA sensor is slightly larger than the personal desk in most cases, which ensures the

upper body states of users can be fully captured.

3.3. Sitting posture recognition algorithm

The gathered multimodal sitting posture data are preprocessed and then utilized for classification. The whole process for sitting posture recognition is presented in Fig. 4.

3.3.1. Data preprocessing

Two steps are conducted for data preprocessing: normalization and augmentation. The first step uses a min–max normalization to convert the raw data for each frame into 0–1 by $a = \frac{a-min}{a-max}$. The notation of a is the pixel value, and max and min denote each frame's maximum and minimum values. Then, the normalized pressure data and infrared data are converted into 44 × 52 and 24 × 32 images, respectively.

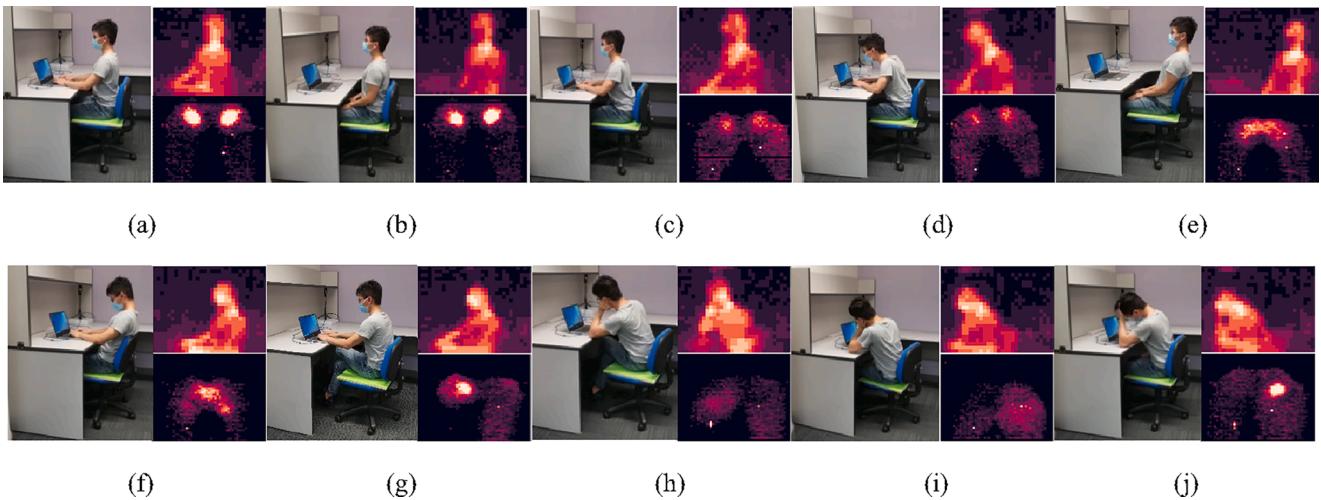


Fig. 5. Examples of the 10 selected sitting postures and the corresponding pressure images and infrared images. Trunk is straight (TS), leaning forward (TF), leaning backward (TB), and leaning to the left (TL). Neck is neutral (NN) or non-neutral (NON). Upper extremity: putting hands on the desk (UD), hands on the legs (UL), holding the chin using one hand (UC), and holding the forehead using one hand (UF). The lower extremity: the legs are crossed (LC) or supported by the ground (LG). (a): TS, NN, UD, and LG. (b) TS, NN, UL, and LG. (c) TF, NN, UD, and LG. (d) TF, NON, UD, and LG. (e) TB, NN, UL, and LG. (f) TB, NON, UD, and LG. (g) TS, NN, UD, and LC. (h) TF, NON, UC, and LC. (i) TL, NON, UC, and LG. (j) TL, NON, UF, and LG.

The following step is augmentation, which is only performed for model training. The deep learning algorithm relies on a large dataset to avoid overfitting. However, there is no access to collect an enormous sitting posture dataset. Hence, geometric data augmentation is applied to provide a more comprehensive dataset and enhance the generalization performance of the proposed model following three steps: rotation, translation, and scale.

First, the images are randomly rotated since the users may sit in different positions. The coordinate of one pixel in the image is (x_a, y_a) and the generated new coordinate after rotation is (x_b, y_b) . The transformation relies on the equations $x_b = x_a \cos\theta - y_a \sin\theta$ and $y_b = y_a \cos\theta + x_a \sin\theta$ where θ is the rotation angle, set between -15° and 15° . Then, the image is translated randomly in the horizontal or vertical direction. The shift value is lower than the 0.1 size of the image's width/height. After the shift, the remaining space of the image is filled with a constant 0. Next, bilinear interpolation is applied to scale the image due to the body shape differences. The image is randomly turned into 0.9 – 1.1 times the size of the original image.

3.3.2. Classification

The classification algorithm construction considers two aspects: feature extraction and multimodal data fusion. The network structure is illustrated in Fig. 3, namely Two-modal SitPoseNet (TSPN).

The sensory data can be viewed as images for processing and the convolutional neural network (CNN) has been extensively adopted for image feature extraction. Especially, a deeper CNN network tends to have a better capability of feature extraction. It is significant because the features of modalities influence the model's performance fundamentally. However, the network degradation prevents the increased number of layers. The residual network (ResNet) was presented by He et al. [42] to solve this issue. Its residual block helps produce a deeper network efficiently by using the "shortcut connection". Compared with other classical CNN models, e.g., VGG and GoogLeNet, ResNet gained better performance on the ImageNet dataset [42]. Hence, eight residual blocks are adopted as a backbone for feature extraction of each modality to avoid complex feature design and selection.

An image for each modality is X_m^i where i denotes the numbers of modalities. The output of n_{th} residual block is S_n^i and $S_n^i = F(X_m^i) + X_m^i$. $F(X)$ is non-linear transformation. Specifically, each residual block comprises two convolutional layers with 3×3 kernels and a stride of 1. Batch normalization is added after each convolutional layer, and recti-

fied linear unit (ReLU) is employed as the activation function. There are four kinds of residual modules containing different numbers of filters (32, 64, 128, and 256). Convolution with 2 strides is performed to downsample feature maps in part residual blocks, as shown in Fig. 4.

The feature-level fusion strategy is applied to fuse features of each modality, which obtains features from different sensors or channels and fuses them. Therefore, each modality owns a tailored backbone in the proposed model, which helps preserve the characteristics of each modality and gain their complementarity [43]. Specifically, the S_n^i is converted into V_n^i through average pooling, where $V_n^i \in \mathbb{R}^{C \times H \times W}$. C is the number of channels. H and W are the height and width, respectively. V_n^i is reshaped as a feature vector with B dimensions and $B = C \times H \times W$. The representative of two modalities (X_{multi}) can be obtained by $X_{multi} = concat(V_n^1, V_n^2)$. Accordingly, the final feature vector is $2 \times B$ in length, which serves as the input of a fully-connected (FC) layer with 128 neurons. Finally, the softmax function is adopted to obtain the probabilities of sitting posture categories.

3.4. Predefined sitting posture categories

A total of ten common sitting postures are selected by considering the potential health risks and daily working states of office workers based on the existing ergonomic studies, as shown in Fig. 5. The states of four body parts (the trunk, head, upper extremity, and lower extremity) are all involved among the categories [25]. The trunk states are straight, leaning forward, leaning backward, and leaning to the left while the neck can keep neutral or non-neutral (e.g., axial rotation and lateral flexion). The states of the upper extremity include hands on the desk, hands on the legs, holding the chin using one hand, and holding the forehead using one hand. As for the lower extremity, the legs are crossed or supported by the ground uniformly.

Each sitting posture contains the states of the four body parts described above, which involves different health risks. The commonly suggested sitting postures are (a), (b), and (e), where the trunk keeps straight or supported by a backrest without slumping back. These can be considered as low-risk categories [26]. The others are considered into high-risk categories. The leaning forward or leaning to the left (sitting postures (c), (d), (h), (i), and (j)) mean that the trunk is flexed or in a lateral bending state. When the trunk flexes, the gravity center is shifted anteriorly with a simultaneous forward pelvic tilt and flexion of the spine [44]. It increases the activities of back muscles and is associated

with muscle fatigue. There is also a potential risk of increasing load for intervertebral discs [29]. The lateral flexion of the trunk also contributes to the development of low back disorders and may lead to scoliosis [45]. Keep the neck in a neutral state (sitting postures (a), (b), (c), (e), and (g)) refers to that the ears are aligned with the shoulders. The non-neutral neck (sitting postures (d), (f), (h), (i), and (j)) may lead to increased stress on the cervical spine and weaken deep neck flexors, cervical erector spinae, etc. [46]. Especially, the forward head posture is significantly linked with neck pain [28], and the risk of neck pain increases as the flexion angle increases [7]. Compared with sitting with legs supported by the ground, the crossed legs in sitting postures (g) and (h) result in higher gluteal pressure and loads on the lumbar joints [47,48].

4. Experimental setup

This section describes the experimental data collection, comparative studies, and validation approach.

4.1. Experimental data collection

Experimental data collection was conducted to verify the performance of the proposed system, which was approved by the Human Subjects Ethics Sub-committee (HSESC) of The Hong Kong Polytechnic University (Reference number: HSEARS20201110002). A total of 21 subjects (12 males and 9 females) participated in the experiment, of which the average values of age, weight, and height were 26 (range 19–33) years, 1.67 (range 1.53–1.78) m, and 61.46 (range 45–83) Kg, respectively. The standard variances were 3.82 years, 0.07 m, and 12.54 Kg. Their body mass indexes (BMI) ranged from 18.52 to 27.04. The environmental temperature of the cubicle ranged from 22 °C to 26 °C during the data collection process. Hence, most subjects wore t-shirts and pants except 3 wore long skirts, and one wore a coat.

The sampling rates of the two sensors were set to 1 fps. Ten seconds of data were collected for each sitting posture; this process was repeated 3 times for each subject. The subjects were asked to walk around after performing one sitting posture. Finally, a total of 6300 frames with manual labels were collected. Additionally, there were circumstances where other people or devices may appear in the background and were detected by the IRA sensor in the shared office environment. They were shown as small hot spots in the infrared images. These images with background noises were preserved in the dataset, of which the number is 1910.

4.2. Comparative studies

Two comparative experiments were conducted to verify the performance of the proposed system, which examined the effectiveness of the proposed algorithm and the adopted sensor modalities.

The first experiment focused on the effectiveness of the proposed recognition algorithm. The performance of the recognition approaches, backbones, and data augmentation approaches were examined. The DCNN algorithm proposed by Gochoo et al. [38] was selected because it had shown good performance for yoga posture recognition based on the IRA sensor. This DCNN model has 2 convolutional layers and 3 FC layers. One-branch TSPN was adopted for comparison, containing only one backbone followed by FC layers. The one-channel pressure and infrared images were concatenated into a two-channel image to serve as the DCNN model and one-branch model's input, called data-level fusion strategy. Additionally, we tested another recognition approach, which adopted the pressure images for the lower extremity recognition and the infrared images for the head, trunk, and upper extremity recognition, considering the capability of modalities. Hence, four unimodal models were built, and their recognition results were concatenated for the final performance evaluation. Each model's input is the one-channel images.

The second experiment intended to compare the performance of each

Table 2

The deep learning algorithms' performance based on multimodal data.

Model	Accuracy
DCNN using two sensors	73.4%
One-branch TSPN using two sensors	86.6%
Four unimodal models	76.7%
TSPN using two sensors	90.6%
TSPN using two sensors (no data augmentation)	89.1%

modality and validate the effectiveness of the sensors' combination. Hence, two benchmark sensor systems were adopted: the pressure array sensor and the IRA sensor systems. Two datasets that contain only pressure data and infrared data were used for model training. The one-branch TSPN was selected to develop the unimodal models, which output one of the ten sitting postures.

These models were eventually implemented using PyTorch and all trained for 200 epochs, of which the optimal learning rate was set as 0.0001 with a batch size of 8. Loss calculation applied cross-entropy, and model training utilized Adam optimization based on a server with an Nvidia RTX3090 GPU.

4.3. Validation approach

The validation adopted a 7-fold cross-validation approach. Since there could be new users, the leave subject out approach was applied particularly to validate the robustness of the sitting posture recognition system. Therefore, twenty-one subjects were randomly divided into seven groups, and the data of 6 groups were utilized for model training while the other for testing. Thus, the dataset for model testing was not applied to model training. The validation process was accomplished until seven groups were used as the testing dataset, which means that seven trained models were generated. Each model has four measures, including accuracy, precision, recall, and F1-score, where the accuracy served as the primary indicator. The average values of the seven models' measures were calculated for the final performance evaluation. This validation approach has lower accuracy than the commonly adopted k-fold cross-validation [39].

5. Experimental results

5.1. Effectiveness of the proposed recognition algorithm

The detailed results are presented in Table 2. The results of DCNN and the one-branch TSPN based on the multimodal data were acquired. The former reached 73.4% accuracy, while the latter obtained 86.6%.

The TSPN realized the highest accuracy of 90.6% on the predefined ten sitting postures. This result is higher than the performance of the one-branch TSPN. Moreover, the four unimodal models-based recognition approach obtained 76.7% accuracy, lower than the TSPN. This result indicates the effectiveness of the adopted feature-level fusion strategy, which can help obtain the complementarity of multimodal data based on one end-to-end model.

Additionally, the accuracies of models with backbone using 2, 4, 6, and 8 residual blocks were 76.4%, 86.9%, 89.6%, and 90.6%, respectively. Therefore, it is concluded that a deeper network can improve the classification accuracy, which brings stronger fitting capabilities of learning modality characteristics with more parameters.

The importance of the adopted data augmentation approach was also examined. The TSPN trained without data augmentation obtained an accuracy of 89.1%. This result shows that the adopted data augmentation approach enhances the generalizability ability of our model by generating more comprehensive data points, which shortens the distance between the training dataset and the testing dataset.

Table 3
Comparison of different sensor modalities.

Model	Accuracy
DCNN using pressure array sensor	43.4%
DCNN using IRA sensor	67.5%
One-branch TSPN using pressure array sensor	50.0%
One-branch TSPN using IRA sensor	78.3%
TSPN using two sensors	90.6%

5.2. Comparison of different sensor modalities

Table 3 lists the results of applying one modality for sitting posture recognition. The unimodal one-branch TSPN realized a relatively higher accuracy, i.e., 50.0% and 78.3% for the pressure array sensor and the IRA sensor. The multimodal system achieved the highest accuracy (90.6%) compared to the two unimodal systems. Fig. 6 presents the detailed confusion matrixes.

The accuracies that the pressure array sensor obtained of all selected sitting postures were lower than the multimodal system. The recognition of "crossed legs" accomplished the best performance among all sitting postures, i.e., posture (g) and (h). A lot of sitting posture (a) data were massively misclassified as posture (b), and a similar situation also occurred between sitting postures (c) and (d), (e) and (f), (i) and (j). It can be concluded that the same category of trunk state led to the occurrence of misclassification. Specifically, the trunk states of sitting posture (a) and (b) were straight, while their upper extremity states differed. Sitting posture (c) and (d) were leaning forward. However, the neck in (c) was neutral and in (d) was non-neutral. This result demonstrates that the pressure array sensor attached to the chair can recognize different trunk states, while the upper extremity and head changes are beyond its sensing capability.

Additionally, the sitting posture (a) and (b) were often misclassified as (c), (d), (e), and (j). There are two potential reasons: First, different upper body configurations and the contact area between the subject and the pressure array sensor may result in a similar pressure distribution [26]. Second, the physical characteristics of the human body also impact the pressure distribution. The differences between the pressure of the ischium tuberosities and the other body part of the low BMI subjects are more significant than those with high BMI. The pressure differences remain high when the subjects slightly lean forward or left, and the corresponding pressure image was observed to be similar to the image of subjects who have a straight trunk and higher BMI. A potential solution to avoid this issue is to develop a personalized recognition model suiting human physical characteristics.

Therefore, relying only on the pressure array sensor is insufficient to support ergonomics-oriented sitting posture recognition. The IRA

sensor's performance was superior to the pressure array sensor. It achieved the highest accuracies on the sitting postures (e) and (j). Also, the performance of the IRA sensor on sitting postures (a) and (b) has shown its noticeable potential to recognize the states of the upper extremity. The reason is that the IRA sensor captures the temperature of the human body, which directly reflects the posture. However, the pressure array sensor captures the pressure distribution of the human body on the seat surface, which makes it inexact to recognize the states of the head and the upper extremity.

Nevertheless, the IRA sensor showed unsatisfactory performance on the sitting postures (h) and (g) because the desk covered the lower extremities. The pressure array sensor can serve as a supplement to solve this problem based on the experiment analysis. Moreover, the recognition accuracy of sitting posture (b) increased with reducing misclassified instances of (e), indicating the pressure array sensor's capability of enhancing the trunk's recognition. However, it is found that the multimodal system slightly increases the number of misclassification cases between sitting postures (a) and (b) compared with the IRA sensor system. Due to the pressure images for the two sitting postures being similar, adjusting the weights of each modality in the model may help to solve this problem. Overall, the combination of sensors improves the accuracy significantly compared with the unimodal systems holistically.

Additionally, it is significant to consider the non-human infrared sources in the personal workspace for future study. Although there were some noises in part of the collected infrared maps, they were mainly shown in the background. The output of the IRA sensor contains the temperature distribution of both the human and the non-human objects. The temperature of some non-human objects, e.g., a glass of hot water, can be close to or much higher than that of human body. If the glass is in front of the human, it could directly influence the infrared maps. Future study would pay closer attention to provide a comprehensive dataset and examine the robustness of the DL-based model, which can adapt to the dynamic user, sudden infrared sources, and static background.

6. Conclusion

The ultimate goal of this study is to prevent the potential health risks due to the prolonged and awkward sitting postures. The contributions are summarized: 1) This study proposed a novel privacy-preserving and unobtrusive sitting posture recognition system. To the best of our knowledge, it is the first time that an IRA sensor and a pressure array sensor are combined to recognize sitting postures containing different states of four main body parts, which is proven to be superior to the unimodal systems based on experiments. 2) A deep learning-based sitting posture recognition algorithm is developed, which is free from complex handcrafted feature extraction and enables end-to-end

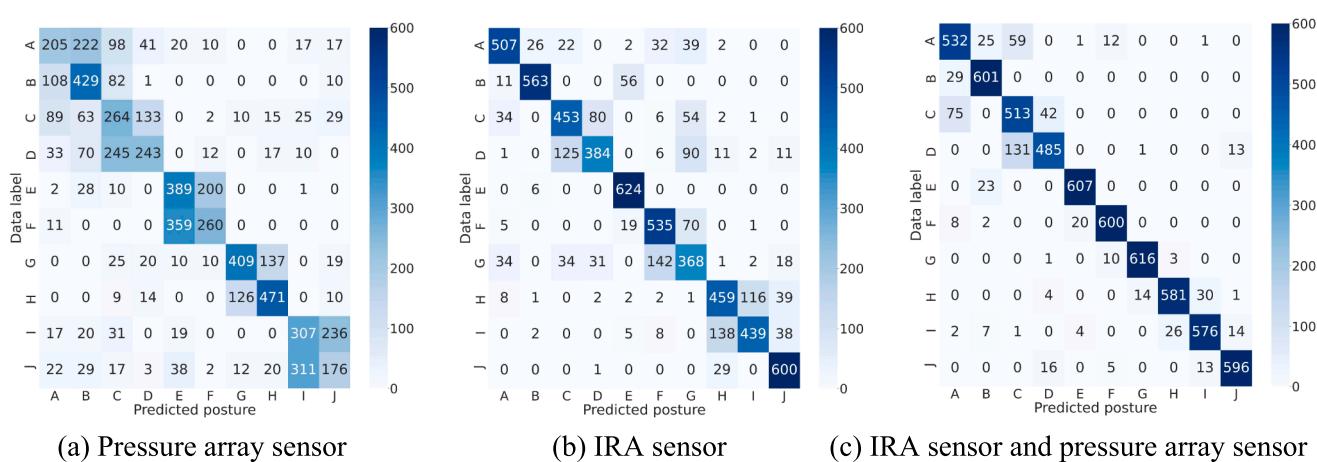


Fig. 6. Confusion matrixes of different sensor modalities. The unimodal sensor system adopted the one-branch TSPN as classification approach.

implementation. The recognition accuracy of the proposed algorithm was 90.6% on the pre-defined 10 sitting postures selected from an ergonomics point of view. Additionally, this paper found that using a pressure array sensor is insufficient to capture dissimilarities of the sitting postures with the same trunk state, which can support the future improvement of the sensing chair.

In conclusion, it can be convinced that the proposed multimodal system has a great potential for privacy-preserving applications related to sitting postures, e.g., sitting posture evaluation and daily stress detection. Moreover, the sensing technology can also be implemented for other privacy-preserving applications in general, such as smart home applications. Despite those achievements, future research can be conducted on: 1) sensor layout optimization to reduce the sensor array resolution, and 2) exploring the multimodal system's capability of recognizing more sitting postures related to health risks and testing it on the human body of various physical characteristics. The wearable devices can assist in developing such a dataset because they can accurately capture users' joint states to provide annotation.

Declaration of Competing Interest

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

Acknowledgements

This research was funded by the National Natural Science Foundation of China (Grant No. 72071179 and No. 52005424), Joint Supervision Scheme with the Chinese Mainland, Taiwan and Macao Universities-Zhejiang University, The Hong Kong Polytechnic University (Project code: G-SB2E), Laboratory for Artificial Intelligence in Design, Hong Kong Special Administrative Region (Project Code: RP2-1), and ZJU-Sunon Joint Research Center of Smart Furniture, Zhejiang University. The authors would also like to appreciate Mr. Liqiao Xia from the Department of Industrial and Systems Engineering at The Hong Kong Polytechnic University, Mr. Wangchujun Tang from the University of Cambridge, and Ms. Qiqi He and Ms. Hongling Ye from the Institute of Industrial Engineering at Zhejiang University, for their dedicative contributions.

References

- [1] M. Shain, D.M. Kramer, Health promotion in the workplace: framing the concept; reviewing the evidence, *Occup. Environ. Med.* 61 (2004) 643–648, <https://doi.org/10.1136/oem.2004.013193>.
- [2] S. Papagiannidis, D. Marikyan, Smart offices: a productivity and well-being perspective, *Int. J. Inf. Manage.* 51 (2020), <https://doi.org/10.1016/j.ijinfomgt.2019.10.012>.
- [3] X. Yan, H. Li, H. Zhang, T.M. Rose, Personalized method for self-management of trunk postural ergonomic hazards in construction rebar ironwork, *Adv. Eng. Informatics.* 37 (2018) 31–41, <https://doi.org/10.1016/j.aei.2018.04.013>.
- [4] X. Zhang, P. Zheng, T. Peng, Q. He, C.K.M. Lee, R. Tang, Promoting employee health in smart office: a survey, *Adv. Eng. Informatics.* 51 (2022) 101518, <https://doi.org/10.1016/j.aei.2021.101518>.
- [5] Q.A.S. Akrouf, J.O. Crawford, A.S. Al Shatti, M.I. Kamel, Musculoskeletal disorders among bank office workers in Kuwait, *East Mediterr Health J* 16 (01) (2010) 94–100.
- [6] C. Jensen, L. Finsen, K. Søgaard, H. Christensen, Musculoskeletal symptoms and duration of computer and mouse use, *Int. J. Ind. Ergon.* 30 (4–5) (2002) 265–275, [https://doi.org/10.1016/S0169-8141\(02\)00130-0](https://doi.org/10.1016/S0169-8141(02)00130-0).
- [7] S. Wu, L. He, J. Li, J. Wang, S. Wang, Visual display terminal use increases the prevalence and risk of work-related musculoskeletal disorders among Chinese office workers: A cross-sectional study, *J. Occup. Health.* 54 (1) (2012) 34–43, <https://doi.org/10.1539/joh.11-0119-OA>.
- [8] S. Parry, L. Straker, The contribution of office work to sedentary behaviour associated risk, *BMC Public Health.* 13 (2013) 1, <https://doi.org/10.1186/1471-2458-13-296>.
- [10] S. Matuska, M. Paralic, R. Hudec, A smart system for sitting posture detection based on force sensors and mobile application, *Mob. Inf. Syst.* 2020 (2020), <https://doi.org/10.1155/2020/6625797>.
- [11] Y. Zhang, Y. Huang, B. Lu, Y. Ma, J. Qiu, Y. Zhao, X. Guo, C. Liu, P. Liu, Y. Zhang, Real-time sitting behavior tracking and analysis for rectification of sitting habits by strain sensor-based flexible data bands, *Meas. Sci. Technol.* 31 (2020), <https://doi.org/10.1088/1361-6501/ab63ea>.
- [12] L. Feng, Z. Li, C. Liu, X. Chen, X. Yin, D. Fang, SitR: sitting posture recognition using RF signals, *IEEE Internet Things J.* 7 (12) (2020) 11492–11504, <https://doi.org/10.1109/JIOT.2020.3019280>.
- [13] N.D. Nath, T. Chaspuri, A.H. Behzadan, Automated ergonomic risk monitoring using body-mounted sensors and machine learning, *Adv. Eng. Informatics.* 38 (2018) 514–526, <https://doi.org/10.1016/j.aei.2018.08.020>.
- [14] J. Zhao, E. Obonyo, Convolutional long short-term memory model for recognizing construction workers' postures from wearable inertial measurement units, *Adv. Eng. Informatics.* 46 (2020), 101177, <https://doi.org/10.1016/j.aei.2020.101177>.
- [15] J. Zhao, E. Obonyo, Applying incremental Deep Neural Networks-based posture recognition model for ergonomics risk assessment in construction, *Adv. Eng. Informatics.* 50 (2021), 101374, <https://doi.org/10.1016/j.aei.2021.101374>.
- [16] A. Kulikajevas, R. Maskeliunas, R. Damasevicius, Detection of sitting posture using hierarchical image composition and deep learning, *PeerJ. Comput. Sci.* 7 (2021), <https://doi.org/10.7717/peerj.cs.442>.
- [17] M. Tariq, H. Majed, M.O. Beg, F.A. Khan, A. Derhab, Accurate detection of sitting posture activities in a secure IoT based assisted living environment, *Futur. Gener. Comput. Syst.* 92 (2019) 745–757, <https://doi.org/10.1016/j.future.2018.02.013>.
- [18] E.S.L. Ho, J.C.P. Chan, D.C.K. Chan, H.P.H. Shum, Y.-M. Cheung, P.C. Yuen, Improving posture classification accuracy for depth sensor-based human activity monitoring in smart environments, *Comput. Vis. Image Underst.* 148 (2016) 97–110, <https://doi.org/10.1016/j.cviu.2015.12.011>.
- [19] T. Kong, W. Fang, P.E.D. Love, H. Luo, S. Xu, H. Li, Computer vision and long short-term memory: learning to predict unsafe behaviour in construction, *Adv. Eng. Informatics.* 50 (2021), 101400, <https://doi.org/10.1016/j.aei.2021.101400>.
- [20] Q. Hu, X. Tang, W. Tang, A smart chair sitting posture recognition system using flex sensors and FPGA implemented artificial neural network, *IEEE Sens. J.* 20 (14) (2020) 8007–8016, <https://doi.org/10.1109/JSEN.2020.2980207>.
- [21] A.R. Anwary, D. Cetinkaya, M. Vassallo, H. Bouchachia, Smart-cover: a real time sitting posture monitoring system, sensors actuators, *A Phys.* 317 (2021), <https://doi.org/10.1016/j.sna.2020.112451>.
- [22] C. Ma, W. Li, R. Gravina, G. Fortino, Posture detection based on smart cushion for wheelchair users, *Sensors.* 17 (2017) 6–18, <https://doi.org/10.3390/s17040719>.
- [23] M. Huang, I. Gibson, R. Yang, Smart Chair for Monitoring of Sitting Behavior, *KN Eng. Eng.* 2 (2017) 274–280. 10.18502/keg.v2i2.626.
- [24] H.Z. Tan, L.A. Slivovsky, A. Pentland, A sensing chair using pressure distribution sensors, *IEEE/ASME Trans. Mechatronics.* 6 (2001) 261–268, <https://doi.org/10.1109/3516.951364>.
- [25] C. Ma, W. Li, J. Cao, J. Du, Q. Li, R. Gravina, Adaptive sliding window based activity recognition for assisted livings, *Inf. Fusion.* 53 (2020) 55–65, <https://doi.org/10.1016/j.inffus.2019.06.013>.
- [26] H. Jeong, W. Park, Developing and evaluating a mixed sensor smart chair system for real-time posture classification: combining pressure and distance sensors, *IEEE J. Biomed. Heal. Informatics.* 25 (5) (2021) 1805–1813, <https://doi.org/10.1109/JBHI.2020.3030096>.
- [27] International Organization for Standardization, ISO 11225:2000(E): Ergonomics - Evaluation of Static Working Postures, (2000).
- [28] R. Mousavi-Khatir, S. Talebian, N. Maroufi, G.R. Olyaei, Effect of static neck flexion in cervical flexion-relaxation phenomenon in healthy males and females, *J. Bodyw. Mov. Ther.* 20 (2) (2016) 235–242, <https://doi.org/10.1016/j.jbmt.2015.07.039>.
- [29] P. Waengengnarm, B.S. Rajaratnam, P. Janwantanakul, Perceived body discomfort and trunk muscle activity in three prolonged sitting postures, *J. Phys. Ther. Sci.* 27 (7) (2015) 2183–2187, <https://doi.org/10.1589/jpts.27.2183>.
- [30] A. Naser, A. Lotfi, J. Zhong, Adaptive thermal sensor array placement for human segmentation and occupancy estimation, *IEEE Sens. J.* 21 (2) (2021) 1993–2002, <https://doi.org/10.1109/JSEN.2020.3020401>.
- [31] J. Meyer, B. Arnrich, J. Schumm, G. Troster, Design and modeling of a textile pressure sensor for sitting posture classification, *IEEE Sens. J.* 10 (8) (2010) 1391–1398, <https://doi.org/10.1109/JSEN.2009.2037330>.
- [32] A.R. Anwary, D. Cetinkaya, M. Vassallo, H. Bouchachia, Smart-Cover: a real time sitting posture monitoring system, *Sensors Actuators, A Phys.* 317 (2021) 112451, <https://doi.org/10.1016/j.sna.2020.112451>.
- [33] A. Rezaei, M. Khoshnam, C. Menon, Towards user-friendly wearable platforms for monitoring unconstrained indoor and outdoor activities, *IEEE J. Biomed. Heal. Informatics.* 25 (3) (2021) 674–684, <https://doi.org/10.1109/JBHI.2020.3004319>.
- [34] J. Wang, B. Hafidh, H. Dong, A. El Saddik, Sitting posture recognition using a spiking neural network, *IEEE Sens. J.* 21 (2) (2021) 1779–1786.
- [35] S. Tateno, F. Meng, R. Qian, Y. Hachiya, Privacy-preserved fall detection method with three-dimensional convolutional neural network using low-resolution infrared array sensor, *Sensors.* 20 (2020) 5957.
- [36] T. Kawashima, Y. Kawanishi, I. Ide, H. Murase, D. Deguchi, T. Aizawa, M. Kawade, Action recognition from extremely low-resolution thermal image sequence, in: AVSS, Lecce, Italy, 2017: pp. 1–6.
- [37] S. Mashiyama, J. Hong, T. Ohtsuki, in: *Activity recognition using low resolution infrared array sensor in: ICC*, IEEE, London, UK, 2015, pp. 495–500.
- [38] M. Gochoo, T.-H. Tan, S.-C. Huang, T. Batjargal, J.-W. Hsieh, F.S. Alnajar, Y.-F. Chen, Novel IoT-based privacy-preserving yoga posture recognition system using low- resolution infrared sensors and deep learning, *IEEE Internet Things J.* 6 (4) (2019) 7192–7200.
- [39] Z. Chen, Y.A. Wang, Remote recognition of in-bed postures using a thermopile array sensor with machine learning, *IEEE Sens. J.* 21 (9) (2021) 10428–10436, <https://doi.org/10.1109/JSEN.2021.3059681>.
- [40] S.A. of China, GB/T 10000-1988: Human dimensions of Chinese adults, (1988).

- [41] C.N.L. Industry, GB/T 3326-2016: Furniture—Main sizes of tables, chairs and stools, (2016).
- [42] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in, CVPR (2016) 770–778, <https://doi.org/10.1109/CVPR.2016.90>.
- [43] S. Münzner, P. Schmidt, A. Reiss, M. Hanselmann, R. Stiefelhagen, R. Dürichen, CNN-based sensor fusion techniques for multimodal human activity recognition, Proc. - Int. Symp. Wearable Comput. ISWC. Part F1305 (2017) 158–165. 10.1145/3123021.3123046.
- [44] L. Oddsson, A. Thorstensson, Fast voluntary trunk flexion movements in standing: primary movements and associated postural adjustments, Acta Physiol. Scand. 128 (1986) 341–349, <https://doi.org/10.1111/j.1748-1716.1986.tb07987.x>.
- [45] W.S. Marras, K.P. Granata, Spine loading during trunk lateral bending motions, J. Biomech. 30 (7) (1997) 697–703, [https://doi.org/10.1016/S0021-9290\(97\)00010-9](https://doi.org/10.1016/S0021-9290(97)00010-9).
- [46] K.K. Hansraj, Assessment of stresses in the cervical spine caused by posture and position of the head, Surg. Technol. Int. 25 (2014) 277–279.
- [47] M. Huang, K. Hajizadeh, I. Gibson, T. Lee, Analysis of compressive load on intervertebral joint in standing and sitting postures, Technol. Heal. Care. 24 (2) (2016) 215–223, <https://doi.org/10.3233/THC-151100>.
- [48] J.-S. Yu, D.-H. An, Differences in lumbar and pelvic angles and gluteal pressure in different sitting postures, J. Phys. Ther. Sci. 27 (5) (2015) 1333–1335, <https://doi.org/10.1589/jpts.27.1333>.