

Fall-portent detection for construction sites based on computer vision and machine learning

Fall-portent
detection for
construction
sites

1499

Xiaoyu Liu, Feng Xu and Zhipeng Zhang

School of Naval Architecture, Ocean and Civil Engineering,
Shanghai Jiao Tong University, Shanghai, China, and

Kaiyu Sun

Shanghai Jianke Engineering Consulting Co., Ltd, Shanghai, China

Received 14 May 2023

Revised 30 July 2023

30 August 2023

Accepted 13 September 2023

Abstract

Purpose – Fall accidents can cause casualties and economic losses in the construction industry. Fall portents, such as loss of balance (LOB) and sudden sways, can result in fatal, nonfatal or attempted fall accidents. All of them are worthy of studying to take measures to prevent future accidents. Detecting fall portents can proactively and comprehensively help managers assess the risk to workers as well as in the construction environment and further prevent fall accidents.

Design/methodology/approach – This study focused on the postures of workers and aimed to directly detect fall portents using a computer vision (CV)-based noncontact approach. Firstly, a joint coordinate matrix generated from a three-dimensional pose estimation model is employed, and then the matrix is preprocessed by principal component analysis, K-means and pre-experiments. Finally, a modified fusion K-nearest neighbor-based machine learning model is built to fuse information from the x , y and z axes and output the worker's pose status into three stages.

Findings – The proposed model can output the worker's pose status into three stages (steady–unsteady–fallen) and provide corresponding confidence probabilities for each category. Experiments conducted to evaluate the approach show that the model accuracy reaches 85.02% with threshold-based postprocessing. The proposed fall-portent detection approach can extract the fall risk of workers in the both pre- and post-event phases based on noncontact approach.

Research limitations/implications – First, three-dimensional (3D) pose estimation needs sufficient information, which means it may not perform well when applied in complicated environments or when the shooting distance is extremely large. Second, solely focusing on fall-related factors may not be comprehensive enough. Future studies can incorporate the results of this research as an indicator into the risk assessment system to achieve a more comprehensive and accurate evaluation of worker and site risk.

Practical implications – The proposed machine learning model determines whether the worker is in a status of steady, unsteady or fallen using a CV-based approach. From the perspective of construction management, when detecting fall-related actions on construction sites, the noncontact approach based on CV has irreplaceable advantages of no interruption to workers and low cost. It can make use of the surveillance cameras on construction sites to recognize both preceding events and happened accidents. The detection of fall portents can help worker risk assessment and safety management.

Originality/value – Existing studies using sensor-based approaches are high-cost and invasive for construction workers, and others using CV-based approaches either oversimplify by binary classification of the non-entire fall process or indirectly achieve fall-portent detection. Instead, this study aims to detect fall portents directly by worker's posture and divide the entire fall process into three stages using a CV-based noncontact approach. It can help managers carry out more comprehensive risk assessment and develop preventive measures.

Keywords Fall-portent detection, Near-miss fall, Unsafe posture, Pre-accidents, Computer vision, Machine learning

Paper type Research paper



Engineering, Construction and
Architectural Management

Vol. 32 No. 3, 2025

pp. 1499-1521

© Emerald Publishing Limited

0969-9988

DOI [10.1108/ECAM-05-2023-0458](https://doi.org/10.1108/ECAM-05-2023-0458)

This research was supported by the Science Research Plan of Shanghai Municipal Science and Technology Committee [Grant No. 20dz1201301], and the 2021 Science Research Plan of Shanghai Housing and Urban-Rural Development Management Committee [Grant No. 2021-002-4049].

1. Introduction

Fall accidents have severe negative effects on the construction industry, including fatal, nonfatal and attempted accidents. They are the leading cause of casualties in the construction industry (Shanti *et al.*, 2022) and contribute to more than 30% of fatal accidents in the construction industries of Europe, America and Singapore (Antwi-Afari *et al.*, 2019; Fang and Dzeng, 2017). Specifically, fall from height (FFH) accidents are often fatal (Ma *et al.*, 2021; Zermane *et al.*, 2023). Meanwhile, nonfatal fall accidents are frequent occurrences (Koc *et al.*, 2022) and cause both economic and personnel losses. Attempted fall accidents are also common at construction sites (Sarkar *et al.*, 2019), often happening alongside nonfatal accidents, although they are not widely reported due to their minor consequences. Therefore, fall accidents affect the construction period and cost, resulting in management problems.

Research on fall portents can track preceding events of fall accidents and provide comprehensive risk assessment on construction sites (Xu and Wang, 2020), especially for nonfatal accidents and near miss falls. Although there may not be enough time to prevent an accident from occurring, the collected data on fall portents can still contribute to site management and accident prevention in the future. Fatal accidents constitute a small proportion of fall accidents, with most being attempted accidents or minor injuries (Koc *et al.*, 2022; Sarkar *et al.*, 2019). Therefore, nonfatal and attempted accidents may not cause serious consequences but also have research significance.

Among all fall portents, loss of balance (LOB) can be considered the root cause of fall accidents, and it is extremely hazardous on construction sites (Wong *et al.*, 2016). The common fall portents leading to FFH in the construction industry involve LOB due to instability of materials (e.g. ladders or scaffolds), or LOB due to slipping and tripping (Dzeng *et al.*, 2014). Fall accidents are caused by fall portents, and the primary fall portent is LOB.

Most studies have applied sensors to detect fall portents (Antwi-Afari *et al.*, 2018; Fang and Dzeng, 2017; Umer *et al.*, 2018). Although the accuracy of sensor-based approaches is high, direct contact between workers and equipment may make workers uncomfortable and even cause danger (Xiong and Tang, 2021; Yu *et al.*, 2019). In addition, the contact approach requires configuration of wearable equipment for each worker, possibly resulting in high costs (Duan *et al.*, 2022). Meanwhile, existing studies using noncontact approaches are either rudimentary for binarily classifying the non-entire fall process (Duan *et al.*, 2023; Zhang *et al.*, 2020) or indirect for risk detection (Fang *et al.*, 2018; Guo *et al.*, 2018; Lee and Han, 2021).

To solve the problems of existing approaches, this study aims to detect fall portents and divide the entire fall process into three stages using a CV-based noncontact approach, providing basis of worker risk assessment on construction sites. First, a three-dimensional (3D) pose estimation technology is applied to obtain the spatial position information of workers. Then, based on the features of the data, we improve traditional K-nearest neighbor (KNN) model into a new model called fusion-KNN model to detect fall portents and divide the entire fall process into three stages (steady-unsteady-fallen). The approach fills the gap in previous researches which lack coverage of the entire fall process of workers, so it can identify high-risk workers with frequent fall portents and hazardous areas prone to unstable poses. Managers can then provide oversight for those workers and reconfigure problematic layouts. Therefore, it is important for construction management and can be extensively applied with the widely installed surveillance cameras on construction sites.

2. Research background

2.1 Detection of fall-related hazard

Many studies have been conducted on fall-related risk detection. According to the method of detection and its impact on workers, it can be divided into contact and noncontact fall risk detection. According to the occurrence time of the risk, it can be divided into fall portent and

fall hazard. By combining these two classifications, the related studies can be divided into four types, as presented in [Table 1](#).

For fall hazard detection, the noncontact approach dominated by computer vision has been applied chiefly by researchers from the computer science field. They used RGB or depth-RGB cameras to obtain data and then applied machine learning models such as convolutional neural networks (CNNs) to process the data, ultimately achieving fall detection ([Espinosa et al., 2019](#); [Núñez-Marcos et al., 2017](#); [Planinc and Kampel, 2013](#)). However, these researchers have paid limited attention to the construction environment and relevant codes. Recently, some researchers have attempted to divide part of the fall process into two stages by computer vision (CV). However, the division was relatively rudimentary and did not cover the entire fall process. For instance, [Duan et al. \(2023\)](#) proposed a personalized method to detect two working states (stable and unstable), using 2D pose estimation, but did not further explore the direct link between instability and falling. [Zhang et al. \(2020\)](#) studied the process of accumulating instability leading to falls based on biomechanics. They built a binary classifier to detect two working states (normal walking and falling), but did not focus on the intermediate process of accumulating instability. In the field of construction, many researchers have used the contact approach to achieve fall detection ([Gelmini et al., 2020](#); [Kim et al., 2021](#); [Santos et al., 2019](#)). In most cases, the contact approach is more accurate than the noncontact one. Nevertheless, all the above studies concentrated on fall hazards after fall accidents.

Some researchers have focused on related factors before fall accidents, i.e. fall portents. Because fall portents are difficult to detect using only a noncontact approach, many researchers have tried the contact approach using multiple sensors, including accelerometers, IMUs and pressure sensors. Some researchers employed accelerometers or IMUs to detect fall portents of workers on construction sites ([Fang and Dzeng, 2017](#); [Jebelli et al., 2016](#); [Lee et al., 2023](#); [Umer et al., 2018](#); [Yang et al., 2017](#)). Others used insole pressure sensors to detect fall portents and fall hazards of construction workers ([Antwi-Afari et al., 2018, 2021](#)). However, it is difficult to detect fall portents using a noncontact approach. [Guo et al. \(2018\)](#) started with pre-actions, which are often followed by fall accidents. [Fang et al. \(2018\)](#) analyzed the correlation between personal protective equipment (PPE), workers and windows to detect fall portents. [Lee and Han \(2021\)](#) proposed an innovative approach for

	Contact approach	Noncontact approach
Fall portent	Sensors (accelerometer, IMU, pressure sensor, etc.)	Cameras (RGB and depth-RGB)
Fall hazard	Use an IMU to detect fall portents (Lee et al., 2023 ; Umer et al., 2018) Use insole pressure sensors to detect fall portents (Antwi-Afari et al., 2018, 2021) Use an accelerometer to detect fall portents (Fang and Dzeng, 2017) Use an IMU-based gait stability metric to assess fall portents (Jebelli et al., 2016 ; Yang et al., 2017)	Detect pose instability of workers working at heights by 2D pose estimation (Duan et al., 2023) Detect fall portents indirectly by measuring the acceleration of scaffolds (Lee and Han, 2021) Detect pre-actions to detect fall portents (Guo et al., 2018) Analyze the correlation between PPE, workers and windows to detect fall portents (Fang et al., 2018)
	Use an IMU to achieve fall detection (Gelmini et al., 2020) Use an accelerometer to achieve fall detection (Fáñez et al., 2022 ; Santos et al., 2019) Use an MEMS motion sensor to build a fall detection system (Hayat and Shan, 2018)	Use multiple cameras and a CNN to achieve fall detection (Espinosa et al., 2019) Use a CNN and optical flow images to achieve fall detection by an RGB camera (Núñez-Marcos et al., 2017) Use Kinect as a 3D depth sensor to achieve fall detection (Planinc and Kampel, 2013)

Source(s): Authors own work

Table 1.
Four types of fall-related hazard detection studies

detecting fall portents indirectly by measuring the acceleration of scaffolds. [Duan et al. \(2023\)](#) detected pose instability of workers working at heights by 2D pose estimation. With the progress of CV technology, it is now more feasible to detect fall portents and other unsafe actions based on workers' body postures ([Yang et al., 2023](#)).

2.2 Knowledge gaps and contributions

Present studies on fall-portent detection for pre-accidents have mostly used the contact approach. Although the contact approach has clear advantages in terms of accuracy, it is difficult for all the workers to wear contact sensors because of the inconvenience and high cost. In fact, most of the sensor-based approaches to worker safety management are still at the laboratory level. Therefore, the noncontact approach based on CV has irreplaceable advantages. Moreover, affordable surveillance cameras have become increasingly common at construction sites. Combining CV with common webcams allows a systematic analysis of unsafe behaviors without costly interruptions to ongoing work. However, existing CV-based research have mainly achieved indirect fall-portent detection, such as measuring the scaffold acceleration or detecting preceding events around high-risk areas. Although indirect approaches can work, there are limitations in excluding the worker's posture itself. Other latest researches have applied 2D pose estimation to detect fall process, but they do not cover the entire fall process and the granularity in dividing the fall process is insufficient. Therefore, this study, based on 3D pose estimation using CV, proposes a machine learning-based noncontact approach to detect fall portents directly and divide the entire fall process into three stages (steady-unsteady-fallen). Overall, this study fills the gap in previous researches which lack coverage of the entire fall process of workers, so it can help managers carry out more comprehensive risk assessment and develop preventive measures.

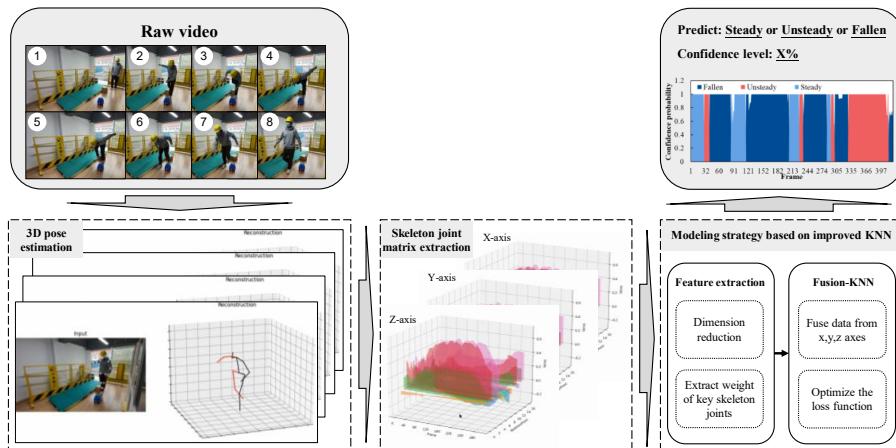
3. Methodology

The proposed approach is built on CV and machine learning to detect fall portents and fall hazards. First, a joint coordinate matrix is generated based on the coordinate values of multiple joints in a 3D coordinate system obtained from a 3D pose estimation model. Using principal component analysis (PCA), the high-dimensional matrix is downsampled, and the process (steady walking - unsteady walking - falling down) is visualized. Pre-experiments are then conducted to select joints that have positive or negative effects on the model. After optimizing the data weights in three axes, a modified fusion-KNN-based machine learning model is built to detect fall portents and output the human pose status (steady, unsteady, or fallen). The architecture of the proposed methodology is shown in [Figure 1](#).

3.1 Data acquisition based on 3D pose estimation

Extracting 3D information from 2D RGB images is a challenging task. The technology of 3D pose estimation based on 2D images mainly benefits from *a priori* knowledge of human skeletal motion ([Zhou et al., 2017](#)). VideoPose3D ([Pavllo et al., 2019](#)) model shows outstanding performance in the Human3.6 M benchmark dataset and has been a baseline model for CV-based 3D human pose estimation ([Chen et al., 2021](#)).

VideoPose3D is adopted in this study for the strong compatibility and efficiency, which is suitable for data acquisition. It is compatible with multiple 2D key-point detections as a two-stage algorithm for single-person videos. In addition, it can achieve real-time operations for approximately 20 FPS on a server with a 1080Ti graphics card ([Peng et al., 2021](#)). Technically, 3D pose estimation's benefits ensure minimal bias in matrix data from varied shooting angles compared to 2D. This "unbiasedness" supports further processing.



Source(s): Authors own work

Figure 1. Architecture of the methodology

However, self-occlusion causes slight differences in results from different angles, but they're not significant.

In the initial stage, algorithms like Hourglass, CPN and Mask R-CNN detect 2D human joint motion trajectories from video. Then, the WaveNet-based network predicts 3D human pose using temporal 2D joint coordinates. True 3D pose data with labels are used for training the trajectory model, and the loss function is the weighted mean position per joint error (Pavlo *et al.*, 2019).

$$E = \frac{1}{y_Z} \|f(X) - y\| \quad (1)$$

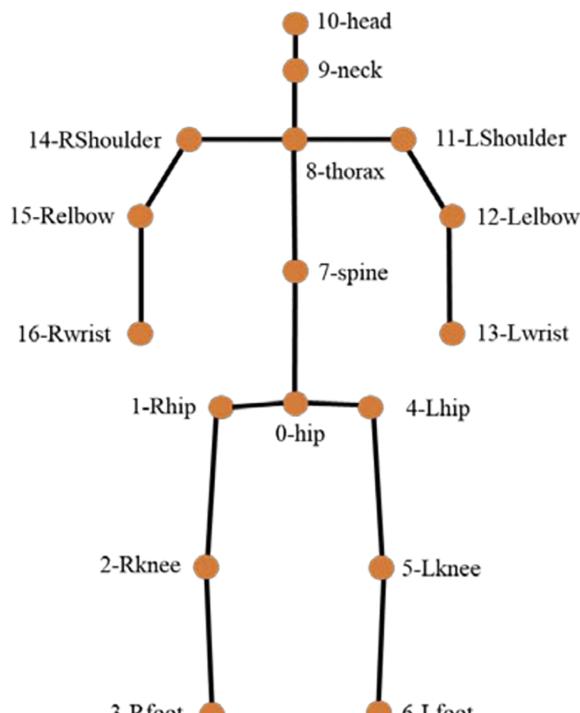
where y_Z denotes the depth of the joint in the camera space. The aim is to reduce the weights of farther joints due to minor screen-projected deviation. With greater distance, trajectory estimation gets tougher, necessitating a decrease in loss function weight.

The pose estimation involved in this study is based on the 2D Human3.6 M dataset. In this dataset, the human skeleton contains 17 joint points, as shown in Figure 2.

After completing the 3D pose estimation using the VideoPose3D model, we obtain the coordinate values of the 17 joints of each frame in the 3D spatial coordinate system. Assume that the total number of frames of the input video is $Frame$. The output format of the raw data is a 3D matrix of $17 \times 3 \times Frame$, where $Frame$ is the time-series dimension and 17×3 is the spatial coordinate dimension. For better presentation, we exclude x -axis and y -axis, and then the matrix from the z -axis, i.e. matrix of $17 \times 1 \times Frame$, is shown in Figure 3.

Each frame is analyzed individually to capture the spatial features. With 17 joints in the Human3.6 M dataset, the 3D pose estimation outputs high-dimensional vectors of 17 dimensions in three axes. However, such high dimensionality can lead to data redundancy and noise for fall-portent detection. Therefore, data preprocessing is necessary. Additionally, as *joint_0-hip* serves as the origin of the 3D spatial coordinate system, its coordinate value remains close to 0, as shown in Figure 3. Consequently, only the remaining joints are considered, resulting in a total of 16 joints for further processing.

Combining visual and physical features, we categorize a worker's postural states during the entire fall process into 3 classes: steady, unsteady and fallen. In the scenario considered in



Source(s): Human3.6M dataset (Ionescu *et al.*, 2013)

Figure 2.
Serial numbers of 17 joints

this study, the worker's process manifests in two patterns - "steady-unsteady-steady" and "steady-unsteady-fallen", as illustrated in Figure 4.

It is worth noting that although the postures of falling behavior can vary randomly among workers, there is a common trend observed: a decrease in the center of gravity of the body, with variations just in the detailed behavioral postures (Zhang *et al.*, 2020). Therefore, we define the state of normal walking, where the center of gravity height is relatively stable, as "steady"; the state where the worker falls down and contacts the ground and afterward, i.e. when the center of gravity of the body contacts the ground, is defined as "fallen"; and all transitional process states in between are defined as "unsteady".

3.2 Modeling strategy based on modified fusion-KNN

3.2.1 Data preprocessing. Principal component analysis (PCA) can be used to perform feasibility studies and feature dimension reduction.

Skeleton joint matrix consists of 16 column vectors, forming a high-dimensional matrix with 16 features. However, high-dimensional data can contain irrelevant and redundant information, negatively impacting modeling (Egecioglu *et al.*, 2004). PCA is an unsupervised linear dimension reduction technique (Pearson, 1901). It transforms the original features into a smaller set of new features that are linear combinations of the originals and linearly independent.

A given training sample $\mathbf{X}_{m \times n}$ consists of m samples $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$, where \mathbf{x}_i represents one sample, $\mathbf{x}_i \in \mathbb{R}^n$.

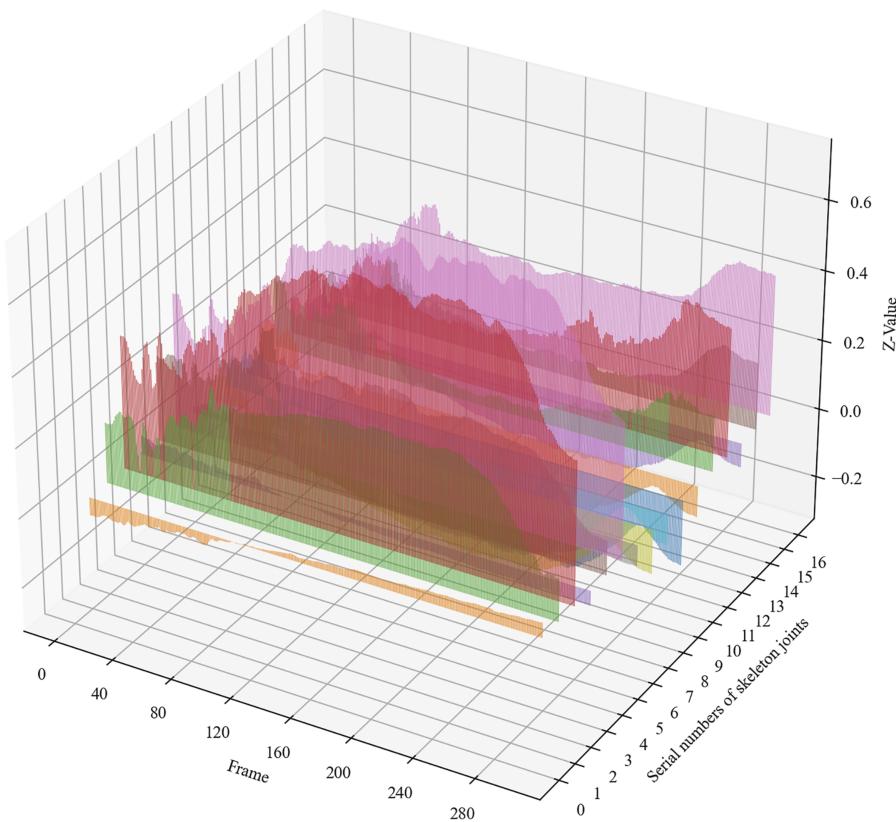


Figure 3.
Shape of the 3D joint
matrix of z -axis

Source(s): Authors own work

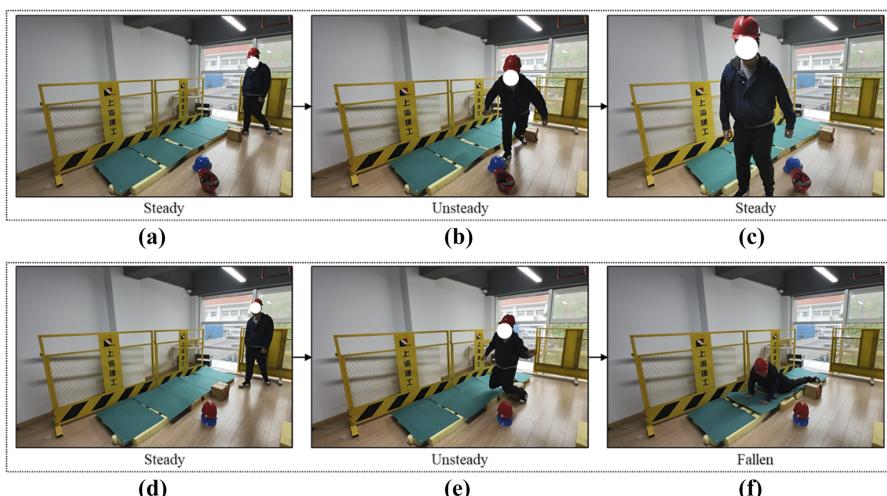


Figure 4.
Samples of workers'
three states during fall
process

Source(s): Authors own work

$$m \text{ samples} = \begin{cases} \mathbf{x}_1 & (x_{11}, \dots, x_{1n}) \\ \vdots & \vdots \\ \mathbf{x}_m & (x_{m1}, \dots, x_{mn}) \end{cases} \Rightarrow \mathbf{X}_{m \times n} = \begin{pmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \dots & x_{mn} \end{pmatrix} \quad (2)$$

Suppose

$$\mathbf{Y} = (\mathbf{X} - \mu)\mathbf{A}. \quad (3)$$

where \mathbf{Y} is the transformed matrix, \mathbf{A} the feature transformation matrix and μ the sample mean. We aim to optimize \mathbf{A} , an orthogonal transformation, to boost feature variance, ensuring uncorrelated features and more information due to higher variance.

Then calculate the covariance matrix \mathbf{C} of the samples, determine the eigenvalues and eigenvectors and build the transformation matrix \mathbf{A} based on the eigenvectors.

$$\lambda_j \mathbf{a}_j = \mathbf{C} \mathbf{a}_j, j = 1, \dots, n, \quad (4)$$

where λ_j is an eigenvalue of \mathbf{C} , \mathbf{a}_j is the corresponding eigenvector and $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_n)$. In the end, matrix \mathbf{A} is used to perform a linear transformation on the original samples, generating new samples with reduced dimensions while preserving more informative features.

Pre-experiments are conducted to achieve weight extraction of key skeleton joints.

Investigating relationships within high-dimensional data is beneficial. Standard clustering groups similar samples. We use transposed K-means clustering to explore feature interactions. By transposing, we investigate feature-feature relationships for different category numbers.

The criterion function of the square sum of the error is used as the clustering criterion function. It is defined as

$$J_e = \sum_{i=1}^k \sum_{X \in C_i} |X - m_i|^2, \quad (5)$$

where C_i is the i th category of the clustering results; m_i is the mean of all data in C_i ; that is, the cluster center; and X is one sample point in C_i .

To obtain the feature-feature relationship, we conducted a pre-experiment based on adversarial feature selection to calculate the accuracy of all combinations of 16 skeleton joints, in order to measure the importance of each joint for the steady-unsteady-fallen process. For instance, *joint_i* is selected, and all combinations with the other 15 skeleton joints are exhausted. For each of the other 15 joints, there are two choices of selection and nonselection. Therefore, there are $(c_2^1)^{15} = 32768$ combinations with *joint_i*. Likewise, there are $(c_2^1)^{15} = 32768$ combinations without *joint_i*. By substituting all the combinations into the model, we can calculate the average accuracy of all combinations containing *joint_i* as $\overline{\text{Acc}}(\text{all combinations with joint}_i)$, and the average accuracy of all combinations not containing *joint_i* as $\overline{\text{Acc}}(\text{all combinations without joint}_i)$.

For each skeleton *joint_i*, if

$$\Delta = \overline{\text{Acc}}(\text{all combinations with joint}_i) - \overline{\text{Acc}}(\text{all combinations without joint}_i) \geq 0, \quad (6)$$

then, skeleton *joint_i* is judged as a positive-effect joint during the steady-unsteady-fallen process. This joint is considered a key joint. On the contrary, if

$$\Delta = \overline{Acc}(\text{all combinations with joint } -i) - \overline{Acc}(\text{all combinations without joint } -i) < 0, \quad (7)$$

then, skeleton *joint_i* is judged as a negative-effect joint during the steady-unsteady-fallen process. This joint is not considered a key joint.

After obtaining positive- and negative-effect joints, when the importance values of the two joints are similar and both Δ are close to 0, referring to the results of K-means, if the two joints are clustered in different categories, both should not be deleted because they carry different information.

3.2.2 Modified fusion-KNN modeling. This study proposes a modified KNN algorithm, called “fusion-KNN”, to handle samples composed of multiple vectors. The traditional KNN algorithm can only handle a sample composed of a single vector, while each sample in this study consists of multiple vectors, so it is necessary to propose an improved algorithm based on the traditional one. The main process for the traditional KNN algorithm is as follows. First, all training samples ($\mathbf{s}^1, \mathbf{s}^2, \dots, \mathbf{s}^m$) are stored in an n-dimensional model space. The model space is then searched for samples to be classified, and the nearest K-training samples are found according to the sample proximity to their K neighbors. Finally, according to the K-nearest neighbor method, a voting strategy is used to predict the sample type to be classified. Different methods can be used to measure the proximity between samples. There are two main measurement methods: Euclidean distance and chord similarity. The calculation formulas are as follows:

$$Distance^{v_1, v_2} = \sqrt{\sum_{i=1}^d (v_{1i} - v_{2i})^2} \quad v_1, v_2 \in i^{d \times 1}, \quad (8)$$

$$Cos(v_1, v_2) = \frac{\sum_{i=1}^d v_{1i} \cdot v_{2i}}{\sqrt{\sum_{i=1}^d v_{1i}^2} \sqrt{\sum_{i=1}^d v_{2i}^2}} \quad v_1, v_2 \in i^{d \times 1}, \quad (9)$$

where $Distance^{v_1, v_2}$ denotes the Euclidean distance between v_1 and v_2 ; $Cos(v_1, v_2)$ is the chord similarity between v_1 and v_2 ; and v_{1i}, v_{2i} are the i th components of each vector.

However, such a method can only handle a sample composed of a single vector, that is, the matrix of $\mathbb{R}^{d \times 1}$. In this study, every sample consists of 3 vectors from three axes (x, y and z). Therefore, the matrix for each sample is $\mathbb{R}^{d \times 3}$. An ordinary vector-based method for calculating the distance between samples is not applicable.

Based on the theory of KNN, we propose a modified model to fuse information from the x , y and z axes, called “fusion-KNN”. In the video data, every frame is a sample. Suppose the sample quantity is m . \mathbf{s}^n is one of the samples, described mathematically as follows:

$$\mathbf{s}^n = (\mathbf{s}_x^n, \mathbf{s}_y^n, \mathbf{s}_z^n) \quad \mathbf{s}_x^n, \mathbf{s}_y^n, \mathbf{s}_z^n \in \mathbb{R}^{16 \times 1} \quad \mathbf{s}^n \in \mathbb{R}^{16 \times 3}, \quad (10)$$

$$\mathbf{s}_j^n = (\mathbf{s}_{j1}^n, \mathbf{s}_{j2}^n, \dots, \mathbf{s}_{jd}^n) \quad d = 16j = x, y, z, \quad (11)$$

where $\mathbf{s}_x^n, \mathbf{s}_y^n, \mathbf{s}_z^n$ are the x, y, z components of each sample in three axes, respectively, and each component is a 16-dimensional vector.

$d_j^{n,k}$ is based on the Mahalanobis distance, which is based on the Euclidean distance and is more suitable for non-2D sample spaces than the Euclidean distance.

$$d_j^{n,k} = \sqrt{\sum_{i=1}^d \left(\frac{\mathbf{s}_{ji}^n - \mathbf{s}_{ji}^k}{\sigma(\mathbf{s}_{ji}^n, \mathbf{k}_j^n)} \right)^2} \quad j = x, y, z \quad n \neq k \quad \mathbf{s}_{ji}^n, \mathbf{s}_{ji}^k \in \mathbb{R}^{1 \times 1} \quad d = 16, \quad (12)$$

where n and k are two different serial numbers in the sample set, and $\sigma(\mathbf{s}_{ji}^n, \mathbf{k}_j^n)$ is the standard deviation between samples s^n and s^k in a certain axis (x , y , or z).

$\overline{Distance}^{n,k}$ is weighted by $d_x^{n,k}$, $d_y^{n,k}$ and $d_z^{n,k}$. The vertical axis (z) is probably more crucial.

$$\overline{Distance}^{n,k} = w_x d_x^{n,k} + w_y d_y^{n,k} + w_z d_z^{n,k} \quad w_x + w_y + w_z = 1, \quad (13)$$

where $\overline{Distance}^{n,k}$ denotes the weighted distance between s^n and s^k ; w_x , w_y and w_z are the weights of the x , y and z axes, respectively, and the sum is 1.

After obtaining the distance between sample s^n and the other samples to be predicted, the probability of the corresponding category is calculated according to the distance and real label. The predicted probability P_{class_i} is actually the ratio of samples in which the true label \hat{y}^t is $class_i$ in K-nearest neighbor samples.

$$P_{class_i} = \frac{\delta_i(\hat{y}^t = class_i)}{k} \quad t = 1, 2, \dots, k \quad i = 1, 2, 3, \quad (14)$$

where δ_i is the correct number predicted for category i , k is the number of K-nearest neighbor samples, and \hat{y}^t is the true label of sample t .

Finally, each predicted probability P_{class_i} is compared to find the maximum value. The predicted label is obtained from the sample label corresponding to the maximum value.

$$f(\mathbf{s}^n, k) = \underset{class_i}{\operatorname{argmax}} \{P_{class_i}\} \quad i = 1, 2, 3 \quad class_1 = Steady, class_2 = Unsteady, class_3 = Fallen, \quad (15)$$

where k is the hyperparameter of the KNN.

The core problem of both the traditional KNN and modified fusion-KNN models lies in the measurement of the intersample distance. Eq. (13) uses the weighted distance to measure the intersample distance; however, a method to specify the weights has not yet been proposed. Although the weights can be given by subjective judgment through common sense and physical laws, they lack precision. For this reason, optimizing the loss function based on the weighted distance is proposed to obtain the optimal weights in the three axes. It allows fusing the information of the three axes in the most reasonable manner. After the results are obtained, they are substituted into the fusion-KNN model.

Due to the unique characteristics of the proposed fusion-KNN model, an additional optimization function is required to obtain the optimal solution for the weights, as compared to the traditional KNN model. Machine learning models often apply loss functions to optimize the objective function. The loss function proposed in this study is as follows:

$$\begin{aligned} \min Loss(f) &= \sum_n \delta(f(\mathbf{s}^n, k) \neq \hat{y}^n), \\ s.t. \quad &w_x + w_y + w_z = 1 \end{aligned} \quad (16)$$

where δ is the total number of correct predictions for all categories, and \hat{y}^t is the true label of the K-nearest neighbors of s^n . The loss function minimizes the number of prediction errors, taking the current values of w_x , w_y and w_z .

To optimize the code writing and enhance the modeling logic, the loss function is not adopted when modeling the optimization function. Instead, the model is optimized with the score function as the objective, that is, to maximize the score function. The weight coefficients that yield the most accurate predictions are obtained.

$$\begin{aligned} \max Score(f) &= \sum_n \delta(f(s^n), k) = \hat{y}^n \\ \text{s.t. } w_x + w_y + w_z &= 1 \end{aligned} \quad (17)$$

The primary evaluation metric is accuracy, which is the percentage of all correct classifications. The accuracy is calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{\text{Score}}{m}, \quad (18)$$

where TP is the true negative, FP is the false negative, FN is the false positive, TP is the true positive, and m is the total number of samples.

After the optimal values of w_x , w_y and w_z are obtained by optimizing the loss function, the optimal weight coefficients are used to achieve information fusion in the three axes. In the three stages of steady-unsteady-fallen, the weights of the coordinate values of each joint in the vertical axis (z -axis) and the weights of each joint in the horizontal axis (x -axis and y -axis) are theoretically different. In the steady-unsteady-fallen, the z -axis values of each joint of the human body show a significant decreasing trend. Theoretically, the change in coordinate values in the vertical axis may play a dominant role. Therefore, the loss function optimization method is used to accurately quantify the degree of dominance of the three axes. If the experimental results are good, the axes that play a controlling role in the three stages of steady-unsteady-fallen can be obtained, which can help data acquisition and data processing in related studies.

After iterating the loss function to obtain the optimal weights of the x , y and z axes, w_x , w_y and w_z that minimize the loss function are obtained. Based on Eq. (13), the information of the three axes is fused to calculate $\overline{Distance}^{n,k}$. The final prediction categories and their confidence probabilities are then calculated according to Eqs. (14) and (15).

In general, the proposed fusion-KNN model introduces a well-defined optimization function for effectively fusing three-vector format data, which is specifically designed to handle the joint coordinate data obtained from 3D pose estimation. This is in contrast to traditional KNN models that can only handle data in a single-vector format. Therefore, the fusion-KNN approach offers a significant advantage in terms of its high applicability compared to traditional KNN methods. Also, while obtaining the optimal weights respectively for the three axes, we objectively measured the importance of each axis's data. This can provide a basis for research related to workers' fall-related postures. In summary, we introduce a novel and efficient model that utilizes matrix data from workers' 3D pose estimation to extract the three stages (steady-unsteady-fallen) of the entire fall process.

4. Experiments and results

4.1 Experimental design

As stated previously, in a complex construction environment, workers inevitably walk through narrow high-altitude passages or sites with distributed obstacles. In such sites, it is

not only necessary to pay attention to fall accidents but also to near-miss falls (Zhang et al., 2019).

To simulate the hazards at construction sites, which might have led to frequent fall accidents and near-miss falls, an experimental platform composed of a passage is set up in an evaluation experiment, as shown in Figure 5. A construction guardrail is arranged on one side of each passage. This asymmetric layout is different from a real construction site, where a construction guardrail is installed at both sides of the narrow passage. The aim is to leave space for the subject to fall down safely. A cushion is placed on the right of the subject, whereas the construction guardrail is installed on the right of the cushion. Considering that the construction site has many stacked objects, such as steel bars, barrels, toolboxes and miscellaneous objects, random obstacles are set up that are slightly dense to acquire more data.

Three subjects with different body types were invited to take part in the experiment. Each subject was assigned to perform 10 simulated trials of steady-unsteady-fallen and steady-unsteady-not fallen, respectively. Due to the considerable risk involved in the experiment and to ensure the safety of the subjects, the study inevitably collected data in a simulated form.

Because the steady walking time is extremely long, only a part of the steady walking samples is left to balance the size of the steady-unsteady-fallen three-stage sample set. The captured video frames involved with unsteady status are all taken as samples. A total of 1688 frames are eventually used for training. Among them, the total number of frames in the steady, unsteady and fallen statuses are 400, 600 and 688, respectively. The samples are so distributed for sample data balance. Due to the short duration (mostly for several seconds) of the fall action itself, steady-label data might be relatively too large. Therefore, the segments are fused together into 1688 frames in order to make the sample data of three categories balanced.

4.2 Experimental results

4.2.1 Data preprocessing results. Taking advantage of the visualization ability of PCA, a feasibility verification experiment is first conducted. In the 2D visualization based on PCA, PC0 and PC1 are the first and second principal components, respectively. PCA downscales the original 16-dimensional matrix, and the data features are two-dimensionally expressed by the two most prominent principal components. Figures 6 and 7 illustrate the processes of the two types of tripping. The steady-unsteady-fallen process exhibits an apparent linear transition characteristic. The fitting curve is then estimated. It is verified that it is feasible to extract features of fall portents using a mathematical method.

To obtain the interrelationships between the motion features of the 16 joints, the original 16-dimensional matrix of 1688 frames is transposed, and then K-means clustering is performed. Finally, it is transformed into 16 1688-dimensional matrices. The purpose is to obtain the relationship between the 16-dimensional matrices and use them to aid decision

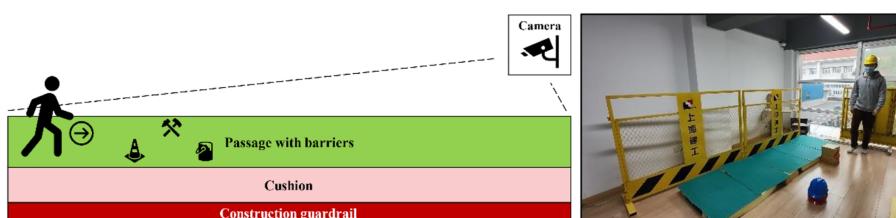
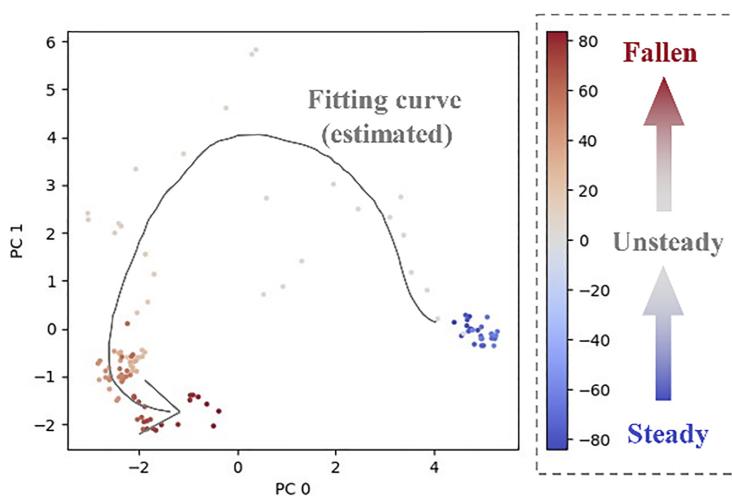


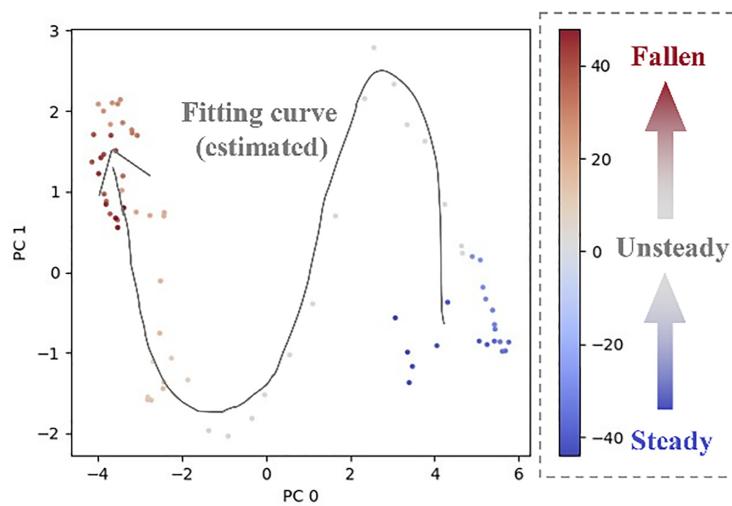
Figure 5.
Layout of the
experimental platform

Source(s): Authors own work



Source(s): Authors own work

Figure 6.
Visualization of
banana trip by PCA



Source(s): Authors own work

Figure 7.
Visualization of
walking trip by PCA

making when selecting key joints. The clustering results for $K = 2$ to 7 are shown in [Figure 8](#). The figures are drawn from the front view for better demonstration, but the video data are acquired from the side view, as mentioned previously.

From the clustering results, it can be observed that the motion characteristics of both legs are particularly remarkable during the steady-unsteady-fallen process. From $K \geq 2$, the joints of both legs are not always in the same classification as the other joints.

When K is larger ($K \geq 5$), the clustering results are not entirely symmetrical, especially when there is a difference between the left and right arms. This asymmetry can be attributed

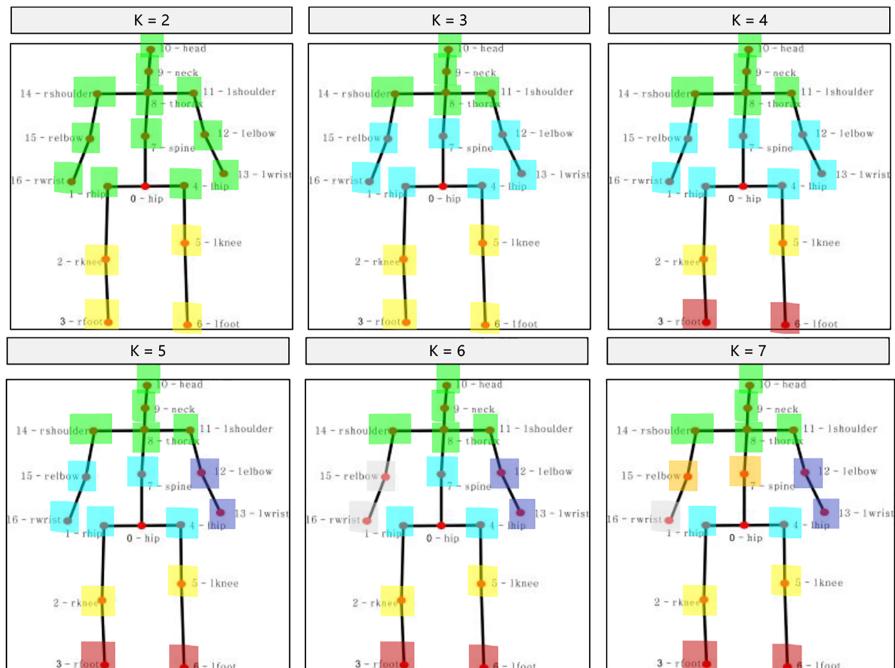


Figure 8.
Experimental results
of K-means clustering

Source(s): Authors own work

to the shooting angle issue. Since the shots are taken from one side, the coordinate information obtained from 3D pose estimation on the camera side may have larger deviations. As a result, uneven clustering between the left and right sides of the human body becomes apparent.

Because of the large number of joints, part of the joints may have a negative impact on the steady-unsteady-fallen three-stage detection because of the shooting angle and occlusion. According to the previously described method, to filter the joints with negative effects among the 16 joints, the experimental results are depicted in Figure 9. From the results of Eqs. (6) and (7), joint_9, joint_16, joint_7, joint_12, joint_11 and joint_5 have a negative effect on the model accuracy. As the camera angle is on one side, there is a problem that the other side is obscured. Therefore, the joints on the other side are not predicted accurately, which negatively affect the model performance. This is consistent with the experimental results of K-means.

4.2.2 Optimization results and prediction results of fusion-KNN. Fusion-KNN performance hinges on optimally blending motion data from x , y and z axes. An optimal function model, derived from Eq. (17), determined these axis weights by assessing accuracy variations, securing optimal weights at maximum accuracy.

Adversarial validation is applied to the model to evaluate the effectiveness and reduce overfitting (Bengio and Grandvalet, 2003). The validation of four-fold cross-validation is adopted. The experimental results are shown in Figure 10. Notably, from the horizontal axis, the experimental results consist of 10,000 discrete sets and each column represents the experimental result under the current values of the x , y and z weights. After taking the optimized weights, PCA dimensionality reduction and filtering of the positive/negative

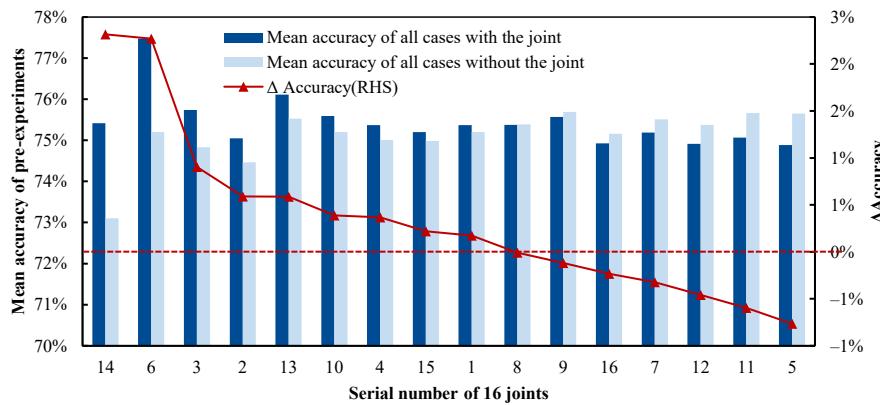


Figure 9.
Experimental results of pre-experiment

Source(s): Authors own work

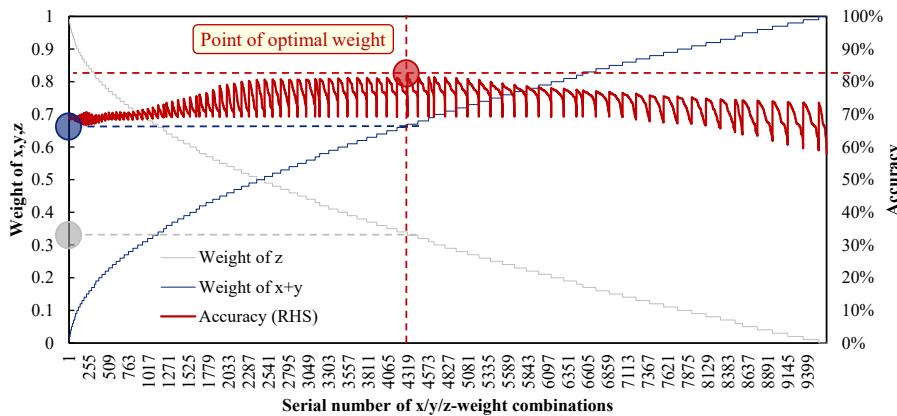


Figure 10.
Process of solving the point of optimal weight

Source(s): Authors own work

skeleton joints, the accuracy reaches 81.40%. Before the PCA dimensionality reduction process, the highest accuracy is 78.26%, which is not shown in Figure 10. Therefore, it can be concluded that PCA dimensionality reduction can accelerate the computing efficiency of high-dimensional data and play a role in data cleaning. In Figure 10, the values of x and y are not shown for a better illustration of the experiment. Actually, the exact values of x and y shown in Figure 10 are known.

The accuracy is higher when the z-axis weight is 100%, that is, when only the z-axis information is used, compared to that when only the x- or y-axis information is used. After fusing the x- and y-axis information, the accuracy is improved and reached the highest value. This highlights the importance of fusing data in multiple axes.

However, the model accuracy can be higher than 81.40% because some obvious errors can be avoided by the threshold-based method. One typical error is that there are several short-duration of fall portents, i.e. status of unsteady walking, lasting only for approximately 5

frames (5/30 s) in the long-duration steady status. This is not in consistent with physical laws and should be rectified to the correct status. Moreover, detecting an extremely short unsteady status has limited significance for both application and research. To address this, a threshold is set to filter the noticeable incorrect results. Referring to previous research (Cartt *et al.*, 2011), the swing phase duration of LOB is approximately 0.18 s, corresponding to approximately 5.4 frames. Therefore, a threshold of 6 frames is adopted, and the process of threshold application is visually shown in Figure 12(b) and (c). After filtering using the threshold, the accuracy reached 85.02%. The confusion matrix of the predicted output before and after filtering using the threshold is depicted in Figure 11.

After obtaining the optimal weight values of x , y and z in the fusion-KNN, the new x , y and z weights are substituted into the model. From the prediction results of the sample set collected in the experiment, it can be observed that the model can predict the steady-unsteady-fallen three-stage correctly in a high proportion. Because the status boundary of the unsteady status is relatively vague, 60/600 is misjudged as steady, and 58/600 is misjudged as fallen in the prediction, as illustrated in Figure 11(b). Furthermore, 64 of 400 steady samples are misjudged as unsteady.

To visualize the prediction results of the steady-unsteady-fallen three-stage process, a 422-frame video is used for demonstration, as displayed in Figure 12. Figure 12(a) shows the true label of the video clip, whereas Figure 12(b) indicates the unmodified prediction results that are immediately output by the model. Figure 12(c) depicts the final results, after postprocessing using the threshold-based method. In the diagram, the vertical axis represents the confidence probability of the model output, of which the maximum value is 1. If the confidence probability is low, the output is more likely to be incorrect, as shown on the rightmost side of Figure 12(b).

5. Discussion

The experimental results of this study are difficult to compare with the existing works because of the different research objectives and evaluation indicators. Existing works mainly include two aspects, sensor-based approach and CV-based approach. As for sensor-based approach, for instance, the research purpose of Umer *et al.* (2018) was to pre-test the worker's balance ability by an IMU, rather than to identify fall-portent actions in the field. Fang and

		Predicted class					Predicted class		
		Steady	Unsteady	Fallen		Steady	Unsteady	Fallen	
True class	Steady	302	96	2		336	64	0	
	Unsteady	74	456	70		60	478	58	
	Fallen	31	45	612		31	39	618	

Before filtering by threshold
(accuracy = 81.40%)

(a)

After filtering by threshold
(accuracy = 85.02%)

(b)

Source(s): Authors own work

Figure 11.
Confusion matrix

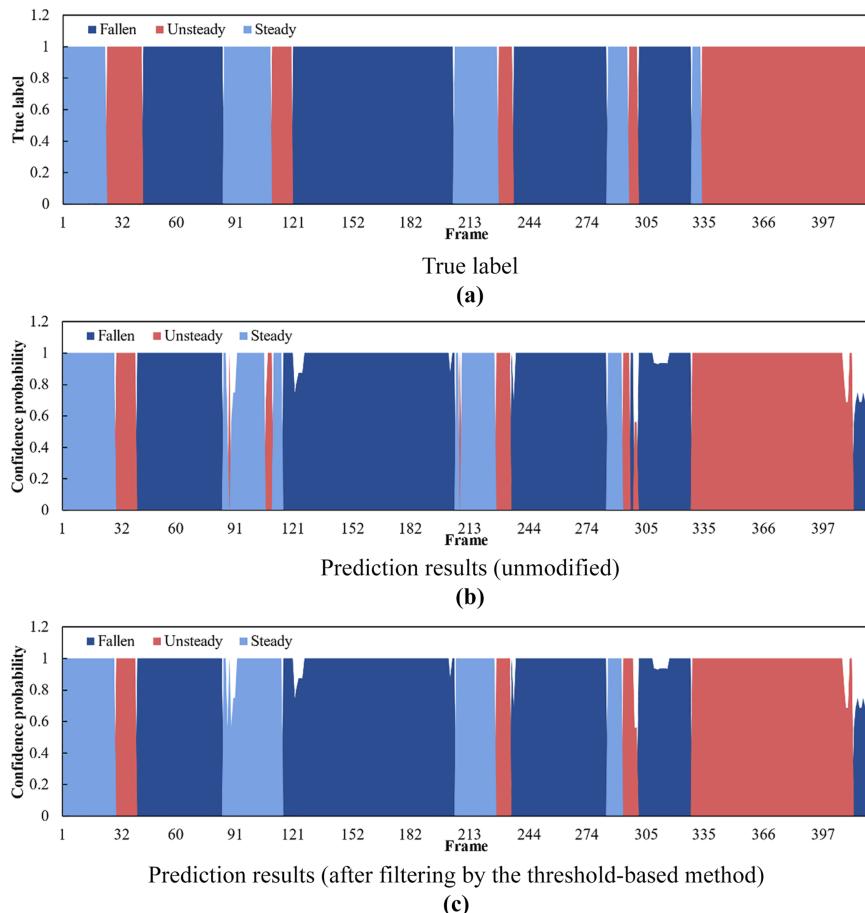


Figure 12.
True label and prediction results

Source(s): Authors own work

Dzeng (2017) developed an accelerometer-based fall portent detection system, but they focused on how to detect fall portents when workers performed tasks under different conditions and achieved an accuracy of 79.13%. Instead, this study aims to detect fall-portent posture by CV when walking and we propose a model to output three stages of steady-unsteady-fallen, achieving an accuracy of 85.02%. Antwi-Afari *et al.* (2018) adopted wearable insole pressure sensors to detect fall portents of workers, achieving an accuracy of 97.1%. This approach showed high-performance thanks to high-precision contact sensors but sensor-based approach has its limitations and hard to be applied broadly on construction sites. As for CV-based approach, Duan *et al.* (2023) used OpenPose to assess the postural stability of workers performing elevated tasks, and their proposed model achieved an accuracy of 84.38%. However, their emphasis was on personalized safety monitoring for workers, and the pose estimation model was limited to 2D. The authors noted that in the future, 3D pose estimation methods will be employed. Other CV-based existing works mostly relied on indirect approaches to achieve the function of fall-potent detection (Fang *et al.*, 2018;

(Guo *et al.*, 2018; Lee and Han, 2021), which means it is hard to present a specific figure of accuracy.

Regarding the approach's generalization capability, we have the following considerations. Firstly, we have adopted k-fold cross validation during model verification. Considering that the sample size in this study is not very large and can be divided by 4, we select $k = 4$ from among commonly used k values (Arlot and Celisse, 2010). 4-fold cross validation used in this study reduces overfitting and provides an approximately unbiased estimate of the model's generalization error (Cawley and Talbot, 2010). Therefore, it can exhibit consistent performance across different dataset splits, demonstrating strong generalization capabilities (Bengio and Grandvalet, 2003). Secondly, as referenced in related research, the fall process demonstrates a common trend across different populations (Zhang *et al.*, 2020). Thus, Duan *et al.* (2023) only used 1 subject and 5 subjects to verify, while other researchers (Guo *et al.*, 2018; Lee and Han, 2021) also used fewer than 10 subjects in their experiments. Therefore, given the prerequisite of achieving 3D pose estimation of workers on construction sites, the proposed approach and experiments in this study have generalizability.

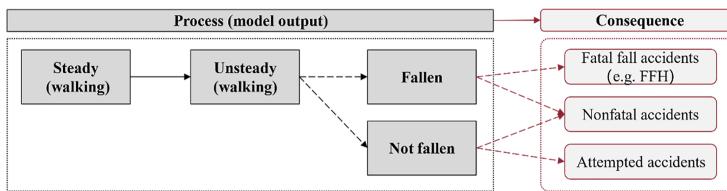
From the standpoint of application, the positive or negative effects of skeleton joints extracted from data preprocessing are related to the shooting angle. The joints closer to the camera and less obscured have more weight in the steady-unsteady-fallen process owing to the higher accuracy of the acquisition. This issue is also evidenced in the results of clustering as Figure 8. Therefore, for engineering applications, a surveillance camera needs to be selected as a capture device for fall-portent actions based on an appropriate shooting angle. It is important to emphasize that this conclusion does not contradict the "unbiasedness" possessed by 3D pose estimation mentioned earlier. "Unbiasedness" only serves as a fundamental condition for achieving the goal through fusion-KNN and does not imply that this foundation is flawless. In addition, the experiment is only a simulation; it would be better if we could conduct a case study. The environmental conditions can be much more complex on a real construction site. For instance, workers' steady walking gait may undergo alterations when walking on surfaces that are either slippery or non-slippery (Mohandes *et al.*, 2022). Meanwhile, the probability of fall portents or fall hazard may also change.

Some problems need to be addressed in the 3D pose estimation technology. In addition to cross-validation, there are some problems when using the validation set of a new environment: false positives (FPs) occasionally occur when the model performs predictions. The main reason for this is that biases occur in the 3D reconstruction of the human skeleton owing to the estimation accuracy of the 3D estimation model (extremely thick clothing, poor lighting, blurred color features, etc.). CV-based approaches are limited by occlusion problems (Antwi-Afari and Li, 2018; Wong *et al.*, 2021). Considering the laws of physics, the human body cannot have an ultrashort time of fall portent in a steady walking interval, so FPs can be filtered during prediction. Setting the filter rule with 5 frames as the threshold, can avoid most of the FP misdetection cases and the expected accuracy can be approximately 85%. The advancements in 3D pose estimation technology may address previous challenges related to ambiguity and the handling of multiple-person frames.

The proposed model's frame-by-frame modular analysis theoretically enables it to handle both scenarios of "steady-unsteady-fallen" and "steady-unsteady-not fallen". As shown in Figure 13, the scenario of "steady-unsteady-not fallen" may result in nonfatal accidents or attempted accidents. Due to individual differences, some workers with strong balance ability have a higher possibility of returning to steady status through the adjustment of their body when losing balance. It presents a process of "steady-unsteady-fallen". For workers with weak balance ability or without protective equipment nearby, it is difficult to adjust themselves once they are affected by certain factors to lose balance.

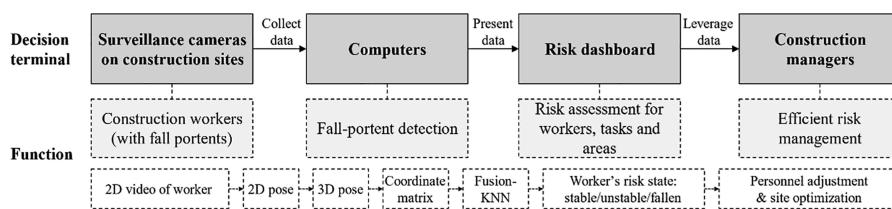
Instead, they are likely to fall subsequently, possibly causing fatal or nonfatal accidents (Antwi-Afari and Li, 2018), as shown in Figure 13. Compared to existed works, this study is based on CV and has a larger scope, including fatal, nonfatal and attempted accidents. Since the risk to workers is also determined by the specific types of jobs and varying complex situations, results of this study can provide a significant basis for comprehensive and thorough risk assessments. Moreover, it should be noted that although the proposed approach has the potential for real-time detection in theory, it is not the core significance of this study. The processing speed for the whole approach is 3.125 frames per second, which suggests the potential for real-time monitoring. However, this study aims to conduct more granular research into workers' entire fall process, in order to accomplish more accurate risk assessment and provide a basis for future construction site management during post-accident stage.

Overall, this study proposes an approach for detecting the entire fall process of workers on construction sites using 3D pose estimation. Specifically, this method can achieve application value at construction sites through the following steps as shown in Figure 14. First, operation videos from workers are captured using surveillance cameras on construction sites, with the data transferred to a computer for processing. Then 2D pose estimation followed by 3D pose estimation is conducted to extract the worker's joint coordinate matrix. Subsequently, the proposed fusion-KNN model detects fall portents and classify the three stages of the entire fall process. The results will be presented on a risk dashboard along with other factors to provide objective evidence for subsequent management. In addition to actual fall accidents detected by the system, construction managers may need to provide specialized oversight for high-risk workers and high-risk areas. In particular, construction workers can be individually analyzed and managed according to their physical condition and the work intensity. For instance, workers frequently exposed to fall portents may be alerted and educated afterward. Meanwhile, construction sites can also be assessed using fall-portent information from workers, so layouts in areas where fall portents frequently occur can be reconfigured in time.



Source(s): Authors own work

Figure 13.
Consequences of different processes



Source(s): Authors own work

Figure 14.
Application process on construction sites

6. Conclusion

Owing to the high frequency of fall accidents in the construction field, it is vital to proactively manage fall risk at construction sites. As sensor-based approaches have application limitations because of high cost and inconvenience, CV-based approaches can be widely applied with the numerous surveillance cameras on construction sites. However, existing CV-based approaches either oversimplify the fall process by binary classification or indirectly detect fall portents by pre-actions. Therefore, this study proposes an improved fusion-KNN machine learning model to detect fall portents directly and divide the entire fall process into three stages (steady-unsteady-fallen) based on CV.

The proposed approach is based on a joint coordinate matrix generated from a 3D pose estimation model. A machine learning model based on the matrix is proposed. Using PCA, K-means and pre-experiments, the data are preprocessed and cleaned. After optimizing the weights of the three axes, a modified fusion-KNN-based machine learning model is built to output the worker pose status (steady, unsteady, or fallen) and divide the entire fall process into three stages. Experiments are conducted to evaluate the proposed approach. A threshold-based postprocessing method is adopted, and the final accuracy of the model reached 85.02%.

The proposed machine learning model determines whether the worker is in a status of steady, unsteady, or fallen using a CV-based approach. From the perspective of construction management, when detecting fall-related actions on construction sites, the noncontact approach based on CV has irreplaceable advantages of no interruption to workers and low cost. It can make use of the surveillance cameras on construction sites to recognize both preceding events and happened accidents. The detection of fall portents can help worker risk assessment and safety management. It is possible for managers to timely remove sources of danger on construction sites and identify workers who are at high risk or need training. In addition, the modified KNN model after 3D pose estimation is a nondeep learning approach with high interpretability, saving computing resources and time. The model achieves an accuracy of 85.02% and can help promote low-cost proactive risk assessment and management on construction sites.

However, this study has several limitations that can provide directions for future research. First, 3D pose estimation needs sufficient information, which means it may not perform well when applied in complicated environments or when the shooting distance is extremely large. Second, solely focusing on posture-related factors of fall accidents may not be comprehensive enough. Future studies can incorporate the results of this research as an indicator into the risk assessment system to achieve a more comprehensive and accurate evaluation of worker and site risk.

References

- Antwi-Afari, M.F. and Li, H. (2018), "Fall risk assessment of construction workers based on biomechanical gait stability parameters using wearable insole pressure system", *Advanced Engineering Informatics*, Vol. 38, pp. 683-694, doi: [10.1016/j.aei.2018.10.002](https://doi.org/10.1016/j.aei.2018.10.002).
- Antwi-Afari, M.F., Li, H., Seo, J. and Wong, A.Y.L. (2018), "Automated detection and classification of construction workers' loss of balance events using wearable insole pressure sensors", *Automation in Construction*, Vol. 96, pp. 189-199, doi: [10.1016/j.autcon.2018.09.010](https://doi.org/10.1016/j.autcon.2018.09.010).
- Antwi-Afari, M.F., Li, H., Wong, J. K.-W., Oladinrin, O.T., Ge, J.X., Seo, J. and Wong, A.Y.L. (2019), "Sensing and warning-based technology applications to improve occupational health and safety in the construction industry: a literature review", *Engineering, Construction and Architectural Management*, Vol. 26 No. 8, pp. 1534-1552, doi: [10.1108/ECAM-05-2018-0188](https://doi.org/10.1108/ECAM-05-2018-0188).
- Antwi-Afari, M.F., Li, H., Seo, J., Anwer, S., Yevu, S.K. and Wu, Z. (2021), "Validity and reliability of a wearable insole pressure system for measuring gait parameters to identify safety hazards in

- construction”, *Engineering, Construction and Architectural Management*, Vol. 28 No. 6, pp. 1761-1779, doi: [10.1108/ECAM-05-2020-0330](https://doi.org/10.1108/ECAM-05-2020-0330).
- Arlot, S. and Celisse, A. (2010), “A survey of cross-validation procedures for model selection”, *Statistics Surveys*, Vol. 4, pp. 40-79, doi: [10.1214/09-SS054](https://doi.org/10.1214/09-SS054).
- Bengio, Y. and Grandvalet, Y. (2003), “No unbiased estimator of the variance of k-fold cross-validation”, *Advances in Neural Information Processing Systems*, Vol. 16, p. 8.
- Carty, C.P., Mills, P. and Barrett, R. (2011), “Recovery from forward loss of balance in young and older adults using the stepping strategy”, *Gait Posture*, Vol. 33 No. 2, pp. 261-267, doi: [10.1016/j.gaitpost.2010.11.017](https://doi.org/10.1016/j.gaitpost.2010.11.017).
- Cawley, G.C. and Talbot, N.L. (2010), “On over-fitting in model selection and subsequent selection bias in performance evaluation”, *The Journal of Machine Learning Research*, Vol. 11, pp. 2079-2107, doi: [10.5555/1756006.1859921](https://doi.org/10.5555/1756006.1859921).
- Chen, T., Fang, C., Shen, X., Zhu, Y., Chen, Z. and Luo, J. (2021), “Anatomy-aware 3d human pose estimation with bone-based pose decomposition”, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 32 No. 1, pp. 198-209, doi: [10.1109/TCSVT.2021.3057267](https://doi.org/10.1109/TCSVT.2021.3057267).
- Duan, P., Zhou, J. and Tao, S. (2022), “Risk events recognition using smartphone and machine learning in construction workers’ material handling tasks”, *Engineering, Construction and Architectural Management*, Vol. 30 No. 8, pp. 3562-3582, doi: [10.1108/ECAM-10-2021-0937](https://doi.org/10.1108/ECAM-10-2021-0937).
- Duan, P., Goh, Y.M. and Zhou, J. (2023), “Personalized stability monitoring based on body postures of construction workers working at heights”, *Safety Science*, Vol. 162, 106104, doi: [10.1016/j.ssci.2023.106104](https://doi.org/10.1016/j.ssci.2023.106104).
- Dzeng, R.-J., Fang, Y.-C. and Chen, I.C. (2014), “A feasibility study of using smartphone built-in accelerometers to detect fall portents”, *Automation in Construction*, Vol. 38, pp. 74-86, doi: [10.1016/j.autcon.2013.11.004](https://doi.org/10.1016/j.autcon.2013.11.004).
- Egecioglu, O., Ferhatosmanoglu, H. and Ogras, U. (2004), “Dimensionality reduction and similarity computation by inner-product approximations”, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16 No. 6, pp. 714-726, doi: [10.1109/TKDE.2004.9](https://doi.org/10.1109/TKDE.2004.9).
- Espinosa, R., Ponce, H., Gutierrez, S., Martinez-Villasenor, L., Brieva, J. and Moya-Albor, E. (2019), “A vision-based approach for fall detection using multiple cameras and convolutional neural networks: a case study using the UP-Fall detection dataset”, *Computers in Biology and Medicine*, Vol. 115, 103520, doi: [10.1016/j.combiom.2019.103520](https://doi.org/10.1016/j.combiom.2019.103520).
- Fáñez, M., Villar, J.R., De la Cal, E., González, V.M. and Sedano, J. (2022), “Improving wearable-based fall detection with unsupervised learning”, *Logic Journal of the IGPL*, Vol. 30 No. 2, pp. 314-325, doi: [10.1093/jigpal/jzaa064](https://doi.org/10.1093/jigpal/jzaa064).
- Fang, Y.-C. and Dzeng, R.-J. (2017), “Accelerometer-based fall-portent detection algorithm for construction tiling operation”, *Automation in Construction*, Vol. 84, pp. 214-230, doi: [10.1016/j.autcon.2017.09.015](https://doi.org/10.1016/j.autcon.2017.09.015).
- Fang, Q., Li, H., Luo, X., Ding, L., Luo, H. and Li, C. (2018), “Computer vision aided inspection on falling prevention measures for steeplejacks in an aerial environment”, *Automation in Construction*, Vol. 93, pp. 148-164, doi: [10.1016/j.autcon.2018.05.022](https://doi.org/10.1016/j.autcon.2018.05.022).
- Gelmini, S., Strada, S., Tanelli, M., Savaresi, S. and Guzzon, A. (2020), “Automatic detection of human’s falls from heights for airbag deployment via inertial measurements”, *Automation in Construction*, Vol. 120, 103358, doi: [10.1016/j.autcon.2020.103358](https://doi.org/10.1016/j.autcon.2020.103358).
- Guo, H., Yu, Y., Ding, Q. and Skitmore, M. (2018), “Image-and-Skeleton-Based parameterized approach to real-time identification of construction workers’ unsafe behaviors”, *Journal of Construction Engineering and Management*, Vol. 144 No. 6, doi: [10.1061/\(asce\)co.1943-7862.0001497](https://doi.org/10.1061/(asce)co.1943-7862.0001497).
- Hayat, A. and Shan, M. (2018), “Fall detection system for labour safety”, *2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*, pp. 1-4, doi: [10.1109/ICEAST.2018.8434476](https://doi.org/10.1109/ICEAST.2018.8434476).

- Jebelli, H., Ahn, C.R. and Stentz, T.L. (2016), "Comprehensive fall-risk assessment of construction workers using inertial measurement units: validation of the gait-stability metric to assess the fall risk of iron workers", *Journal of Computing in Civil Engineering*, Vol. 30 No. 3, doi: [10.1061/\(ASCE\)CP.1943-5487.0000511](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000511).
- Kim, J., Chi, S. and Ahn, C.R. (2021), "Hybrid kinematic-visual sensing approach for activity recognition of construction equipment", *Journal of Building Engineering*, Vol. 44, 102709, doi: [10.1016/j.jobe.2021.102709](https://doi.org/10.1016/j.jobe.2021.102709).
- Koc, K., Ekmekcioglu, Ö. and Gurgun, A.P. (2022), "Prediction of construction accident outcomes based on an imbalanced dataset through integrated resampling techniques and machine learning methods", *Engineering, Construction and Architectural Management*, Vol. ahead-of-print No. ahead-of-print. doi: [10.1108/ECAM-04-2022-0305](https://doi.org/10.1108/ECAM-04-2022-0305).
- Lee, K. and Han, S. (2021), "Convolutional neural network modeling strategy for fall-related motion recognition using acceleration features of a scaffolding structure", *Automation in Construction*, Vol. 130, 103857, doi: [10.1016/j.autcon.2021.103857](https://doi.org/10.1016/j.autcon.2021.103857).
- Lee, H., Lee, G., Park, S., Lee, S., Jacobs, J.V. and Ahn, C.R. (2023), "Collective sensing of workers' loss of body balance for slip, trip, and fall hazard identification: field validation study", *Journal of Computing in Civil Engineering*, Vol. 37 No. 1, 04022052, doi: [10.1061/JCCEE5.CPENG-4938](https://doi.org/10.1061/JCCEE5.CPENG-4938).
- Ma, Z., Chong, H.-Y. and Liao, P.-C. (2021), "Development of a time-variant causal model of human error in construction with dynamic Bayesian network", *Engineering, Construction and Architectural Management*, Vol. 28 No. 1, pp. 291-307, doi: [10.1108/ECAM-03-2019-0130](https://doi.org/10.1108/ECAM-03-2019-0130).
- Mohandes, S.R., Durdyev, S., Sadeghi, H., Mahdiyar, A., Hosseini, M.R., Banihashemi, S. and Martek, I. (2022), "Towards enhancement in reliability and safety of construction projects: developing a hybrid multi-dimensional fuzzy-based approach", *Engineering, Construction and Architectural Management*, Vol. 30 No. 6, pp. 2255-2279, doi: [10.1108/ECAM-09-2021-0817](https://doi.org/10.1108/ECAM-09-2021-0817).
- Núñez-Marcos, A., Azkune, G. and Arganda-Carreras, I. (2017), "Vision-based fall detection with convolutional neural networks", *Wireless Communications and Mobile Computing*, Vol. 2017, pp. 1-16, doi: [10.1155/2017/9474806](https://doi.org/10.1155/2017/9474806).
- Pavllo, D., Feichtenhofer, C., Grangier, D. and Auli, M. (2019), "3d human pose estimation in video with temporal convolutions and semi-supervised training", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7753-7762, doi: [10.1109/CVPR.2019.00794](https://doi.org/10.1109/CVPR.2019.00794).
- Pearson, K. (1901), "LIII. On lines and planes of closest fit to systems of points in space", *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Vol. 2 No. 11, pp. 559-572, doi: [10.1080/14786440109462720](https://doi.org/10.1080/14786440109462720).
- Peng, Y., Lan, L., Yun, C., Yu, H. and Songhai, Z. (2021), "Home fitness monitoring system based on monocular camera", *Journal of Zhejiang University (Science Edition)*, Vol. 48 No. 5, pp. 521-530, doi: [10.3785/j.issn.1008-9497.2021.05.001](https://doi.org/10.3785/j.issn.1008-9497.2021.05.001).
- Planinc, R. and Kampel, M. (2013), "Introducing the use of depth data for fall detection", *Personal and Ubiquitous Computing*, Vol. 17 No. 6, pp. 1063-1072, doi: [10.1007/s00779-012-0552-z](https://doi.org/10.1007/s00779-012-0552-z).
- Santos, G.L., Endo, P.T., Monteiro, K.H.C., Rocha, E.D.S., Silva, I. and Lynn, T. (2019), "Accelerometer-based human fall detection using convolutional neural networks", *Sensors (Basel)*, Vol. 19 No. 7, p. 1644, doi: [10.3390/s19071644](https://doi.org/10.3390/s19071644).
- Sarkar, S., Vinay, S., Raj, R., Maiti, J. and Mitra, P. (2019), "Application of optimized machine learning techniques for prediction of occupational accidents", *Computers and Operations Research*, Vol. 106, pp. 210-224, doi: [10.1016/j.cor.2018.02.021](https://doi.org/10.1016/j.cor.2018.02.021).
- Shanti, M.Z., Cho, C.S., de Soto, B.G., Byon, Y.J., Yeum, C.Y. and Kim, T.Y. (2022), "Real-time monitoring of work-at-height safety hazards in construction sites using drones and deep learning", *Journal of Safety Research*, Vol. 83, pp. 364-370, doi: [10.1016/j.jsr.2022.09.011](https://doi.org/10.1016/j.jsr.2022.09.011).
- Umer, W., Li, H., Lu, W., Szeto, G.P.Y. and Wong, A.Y.L. (2018), "Development of a tool to monitor static balance of construction workers for proactive fall safety management", *Automation in Construction*, Vol. 94, pp. 438-448, doi: [10.1016/j.autcon.2018.07.024](https://doi.org/10.1016/j.autcon.2018.07.024).

- Wong, L., Wang, Y., Law, T. and Lo, C.T. (2016), "Association of root causes in fatal fall-from-height construction accidents in Hong Kong", *Journal of Construction Engineering and Management*, Vol. 142 No. 7, doi: [10.1061/\(asce\)co.1943-7862.0001098](https://doi.org/10.1061/(asce)co.1943-7862.0001098).
- Wong, P. K.-Y., Luo, H., Wang, M., Leung, P.H. and Cheng, J.C.P. (2021), "Recognition of pedestrian trajectories and attributes with computer vision and deep learning techniques", *Advanced Engineering Informatics*, Vol. 49, 101356, doi: [10.1016/j.aei.2021.101356](https://doi.org/10.1016/j.aei.2021.101356).
- Xiong, R. and Tang, P. (2021), "Pose guided anchoring for detecting proper use of personal protective equipment", *Automation in Construction*, Vol. 130, 103828, doi: [10.1016/j.autcon.2021.103828](https://doi.org/10.1016/j.autcon.2021.103828).
- Xu, W. and Wang, T.-K. (2020), "Dynamic safety prewarning mechanism of human-machine-environment using computer vision", *Engineering, Construction and Architectural Management*, Vol. 27 No. 8, pp. 1813-1833, doi: [10.1108/ECAM-12-2019-0732](https://doi.org/10.1108/ECAM-12-2019-0732).
- Yang, K., Ahn, C.R., Vuran, M.C. and Kim, H. (2017), "Collective sensing of workers' gait patterns to identify fall hazards in construction", *Automation in Construction*, Vol. 82, pp. 166-178, doi: [10.1016/j.autcon.2017.04.010](https://doi.org/10.1016/j.autcon.2017.04.010).
- Yang, M., Wu, C., Guo, Y., Jiang, R., Zhou, F., Zhang, J. and Yang, Z. (2023), "Transformer-based deep learning model and video dataset for unsafe action identification in construction projects", *Automation in Construction*, Vol. 146, 104703, doi: [10.1016/j.autcon.2022.104703](https://doi.org/10.1016/j.autcon.2022.104703).
- Yu, Y., Li, H., Yang, X., Kong, L., Luo, X. and Wong, A.Y.L. (2019), "An automatic and non-invasive physical fatigue assessment method for construction workers", *Automation in Construction*, Vol. 103, pp. 1-12, doi: [10.1016/j.autcon.2019.02.020](https://doi.org/10.1016/j.autcon.2019.02.020).
- Zermane, A., Tohir, M.Z.M., Zermane, H., Baharudin, M.R. and Yusoff, H.M. (2023), "Predicting fatal fall from heights accidents using random forest classification machine learning model", *Safety Science*, Vol. 159, 106023, doi: [10.1016/j.ssci.2022.106023](https://doi.org/10.1016/j.ssci.2022.106023).
- Zhang, M., Cao, T. and Zhao, X. (2019), "Using smartphones to detect and identify construction workers' near-miss falls based on ANN", *Journal of Construction Engineering and Management*, Vol. 145 No. 1, doi: [10.1061/\(asce\)co.1943-7862.0001582](https://doi.org/10.1061/(asce)co.1943-7862.0001582).
- Zhang, J., Wu, C. and Wang, Y. (2020), "Human fall detection based on body posture spatio-temporal evolution", *Sensors (Basel)*, Vol. 20 No. 3, p. 946, doi: [10.3390/s20030946](https://doi.org/10.3390/s20030946).
- Zhou, X., Huang, Q., Sun, X., Xue, X. and Wei, Y. (2017), "Towards 3d human pose estimation in the wild: a weakly-supervised approach", *Proceedings of the IEEE International Conference on Computer Vision*, pp. 398-407, doi: [10.48550/arXiv.1704.02447](https://doi.org/10.48550/arXiv.1704.02447).

Corresponding author

Feng Xu can be contacted at: F.Xu@sjtu.edu.cn