

Intelligent Monitoring System for the Elderly based on Posture Recognition

Jinxiang Li, Huateng Liu and Gonghao Nie

North China University of Science and Technology, Tangshan Hebei, 063210, China

Abstract: In recent years, due to the gradual increase in the elderly population, elderly care has become an increasingly urgent issue, and the most common risk faced by the elderly is accidental falls. In this paper, an intelligent monitoring system for the elderly is studied, which includes a communication module, a sensor module and a state detection module. It can monitor the status of the elderly in real time and feed back to the server to activate the buzzer alarm function. In the communication module, WI-FI communication between esp32 microcontroller and PC is established to realize real-time transmission of pressure sensor data. The C/S communication method between Raspberry PI and PC was established to realize the transmission of real-time video data. In the sensor module, the pressure sensor is used for data acquisition, and the RF algorithm is used to realize the binary classification prediction of whether the human body is in bed or out of bed. In the state detection module, the yolov8 algorithm combined with the GSRF algorithm is used to classify the image data into five action categories, namely, falling, waving, sitting, standing and walking, to predict the state of the human body in the real-time video stream with an accuracy of 90%. After the sensor module detects the state of the human body out of bed, the state detection module is started, and the results are transmitted back to the back-end server and uploaded to the Web page to realize the function of real-time monitoring of the state of the elderly.

Keywords: YOLOv8; GSRF Algorithm; Pressure Sensor; State Detection; Web Interface.

1. Introduction

In China's seventh population census in 2020, the elderly population aged 60 years and above accounted for 18.7% of the total population [1], and has entered an aging society. With the increase of the proportion of the elderly, the problem of elderly guardianship is becoming more and more prominent, and in the face of a large number of elderly people, it is difficult to realize the guardianship by relying only on children. With the development of computer vision, sensor technology, remote communication technology, the use of intelligent technology to realize the function of elderly guardianship has become a hot spot. Chen Xiang [2] proposed a study on fall detection based on acceleration sensors, which utilizes a three-axis acceleration sensor to collect velocity data and detects whether an elderly person falls by determining whether the velocity exceeds an advance threshold. Xin Hu [3] et al. studied an OpenPose-based fall detection algorithm for the elderly, and based on the OpenPose human skeleton information recognition network, they proposed to replace some of its convolutional layers with depth-separable convolutional neural network types. The algorithm uses long and short term memory neural network to detect the fall of the elderly. Hao-Hao Huang [4] designed a deep learning based monitoring system for the hazardous conditions of elderly people living alone, proposing the use of DS-MiniVGGNet for expression recognition and fall detection with an accuracy of 99.01% and 98.93%, respectively. The intelligent monitoring system proposed in this paper combines sensor technology, YOLOv8-Pose algorithm, GSRF algorithm, and WebSocket technology, and is able to simultaneously perform the functions of video remote monitoring, real-time status detection, and hazardous action warning. After experiments, it is proved that the system has high accuracy in detecting the fall state and the server runs stably.

2. Principle of Sensor Technology

In this part, the main work is to detect the state of the elderly in bed, and start the state detection module when the elderly leaves the bed. The pressure data obtained by the hardware device of the pressure sensor cannot be directly used to judge the state, so the further processing is adopted. After the data is sent to the PC end, the judgment state is actually a binary classification problem. In this part, the RF algorithm is used to classify the data, which can achieve the advantages of high accuracy and strong generalization ability.

2.1. Sensor Module Hardware Selection

In the process of the experiment, after considering the cost and data accuracy, the 50kg weighing sensor $\times 4$ and the HX711 module $\times 4$ was selected as the data collection devices for the off-bed detection. The 50kg human body weighing module contains a 1000-euro half bridge strain gauge inside, and the deformation resistance bridge circuit [5] can be used to convert the resistance change of the strain gauge into electrical signals of voltage or current. During the measurement, the correct external force is applied on the outside, the bridge arm is bent, and the resistance strain gauge will output a weak voltage signal according to the degree of bending. The HX711 chip has a high-precision 24-bit A/D converter chip, which can amplify the weak voltage signal output by the deformable resistor bridge circuit and convert it into a digital signal. The programming of the chip is relatively simple, and the low-noise programmable amplifier connected to channel A or channel B can be selected by the selection switch. Among them, channel A has a programmable gain of 128 or 64, corresponding to the full-scale difference of $\pm 20\text{mV}$ or $\pm 40\text{mV}$ input signal amplitude, respectively, and channel B has a programmable gain of 32. The HX711 pin diagram is shown in Figure 1.

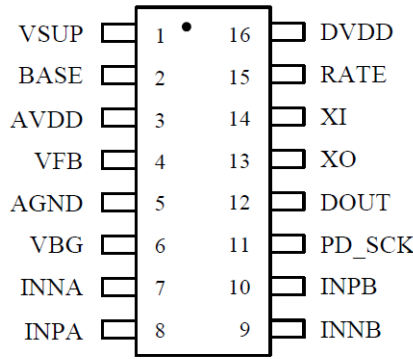


Figure 1. HX711 pin diagram

2.2. MCU Chip Selection

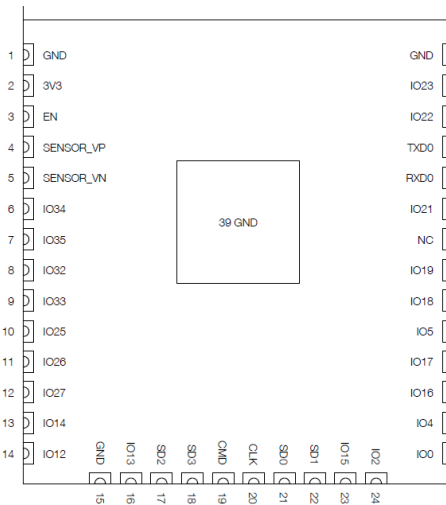


Figure 2. ESP32-WROOM-32 pin layout

The MCU module used in the sensor data reading and transmission part is ESP32-WROOM-32. The chip used in this module is ESP32-D0WDQ6, and this chip has two CPU cores that can be controlled separately. The clock frequency is adjustable from 80MHz to 240MHz, which is suitable for low-power sensor networks. The module also integrates Bluetooth and wide range WI-FI communication functions, and supports data transfer rates up to 150Mbps. This module can complete the function of real-time communication with the PC. The MCU pin diagram is shown in Figure 2.

part of HX711 chip, connect E+ of SIP6 with AVDD, E- with GND, A+ with INA+ of HX711 chip, and A- with INA- of HX711 chip respectively. In the output part of HX711 chip, it is powered by 5V of MCU and connected with GND of MCU and GND of HX711. PD_SCK is connected with IO5 of MCU and DOUT is connected with IO6 of MCU. The voltage signal is converted into digital signal by HX711 chip, and the digital signal is converted into real-time pressure data on MCU, and WIFI communication is established with the PC server, and data transmission is realized through socket protocol. Figure 3 shows the circuit diagram of the sensor module.

2.3. Sensor Module Design

Connect SIP6 with bridge pressure sensor [6]. In the input

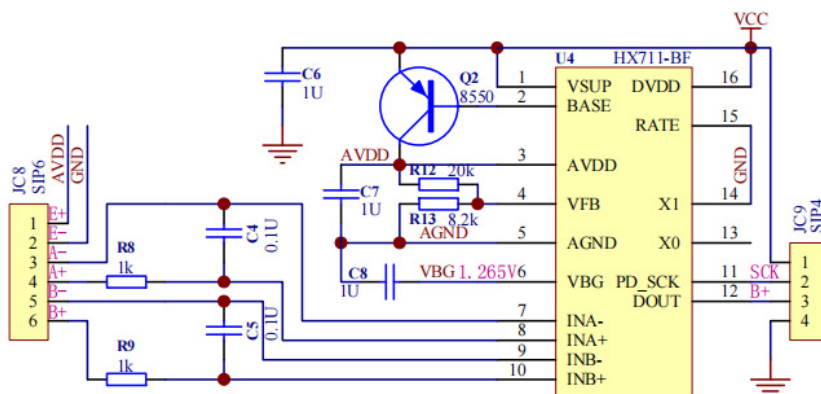


Figure 3. Circuit diagram of the sensor module

3. Main Technologies of the System

The system is mainly divided into four parts: sensor data acquisition, video stream data acquisition, algorithm processing and system interface design. We use WI-FI to transmit sensor data wirelessly, user datagram protocol (UDP) to transmit video streams, YOLOv8-Pose algorithm combined with GSRF algorithm to achieve real-time human body state detection, and then upload the results of the detection to the web front-end, and the whole system interface is based on the development of Websocket technology. The whole system interface is developed based on Websocket technology.

3.1. RTSP Video Streaming

To meet the real-time detection speed in fall detection, the best transmission real-time video streaming protocol Real Time Streaming Protocol (RTSP) [7] is used, using the open source software FFmpeg to push the local camera as an online RTSP video stream, and building an RTSP on a Raspberry Pi server on Raspberry Pi, and a LAN is built through a router to transmit the image data to the PC, which is then invoked by the state detection algorithm.

3.2. Web Server Design

In this paper, the Web interface server with multi-architecture mode is chosen as the feedback component of the system, and the server and the client use WebSocket to establish connection and transmit data, the image transmission achieved by WebSocket has real-time, which is suitable for the scenarios that require real-time monitoring or transmission of image data.

3.3. State Detection Algorithms

In order to better meet the real-time and accuracy of fall

detection for the elderly, this paper uses the YOLOv8 algorithm as well as the addition of RF algorithm to classify the human body's posture.

YOLO target detection algorithm is an end-to-end One-Stage target detection algorithm, the core idea of which is to predict the image in chunks by region. YOLO divides the input image into grids, each of which is responsible for predicting the bounding box containing the target and the category confidence. Unlike traditional two-stage detection algorithms, YOLO completes target detection in a single step with high real-time performance. YOLOv8 makes significant progress in the field of target detection by automatically learning and extracting more representative features through a pre-trained deep neural network. Compared to traditional methods, YOLOv8 performs better in dealing with occlusion and multi-person scenes, thus improving the accuracy of human pose recognition.

3.4. GSRF Algorithm

The GSRF algorithm consists of a random forest algorithm and a grid search algorithm. Random forest is an integrated learning algorithm based on the concept of labelling and decision trees, originally proposed by Bell Labs, USA. The basic building block of Random Forest is the classification decision tree. Random forest constructs multiple decision trees randomly by modelling and classifying the data. After training, the final result is determined by combining the output results through weighted voting. Unlike traditional random forest algorithms, the Grid Search Random Forest (GSRF) algorithm uses the optimal combination of hyperparameters for training and prediction of the random forest model, which significantly improves its performance and alleviates problems such as overfitting or underfitting, and enables human posture classification. The algorithmic workflow of the GSRF is shown in Figure 4.

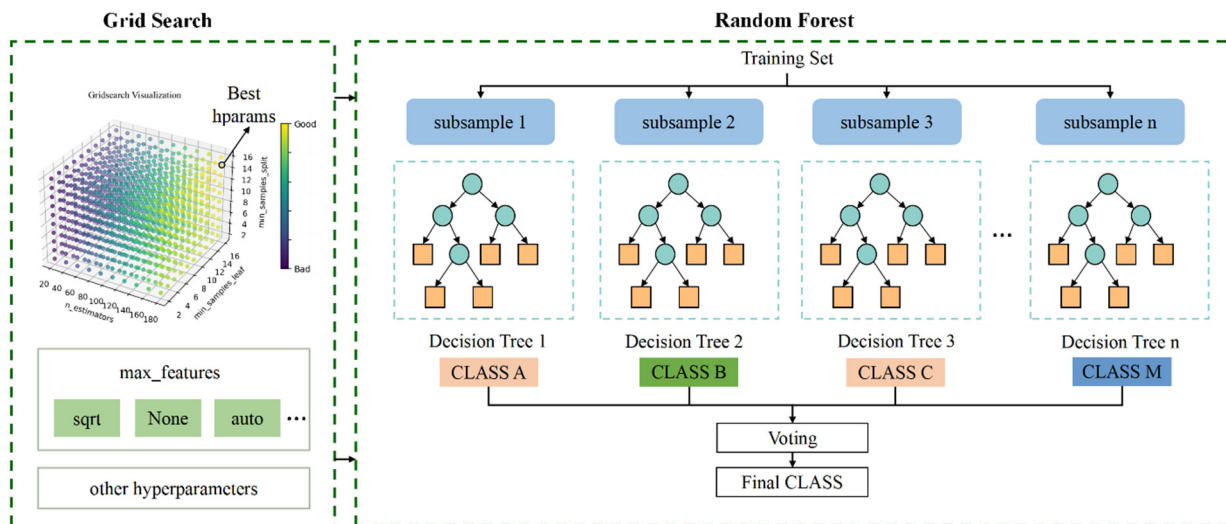


Figure 4. The working flow diagram of the GSRF algorithm

3.5. Size Adaptive Input Algorithm

In order to make the image have a uniform scale, it is necessary to carry out the size adaptive operation, at the same time, in order to cope with the problem of decreasing the intensity of the image features after the operation, and to stabilise the detection accuracy of the model, it is necessary to carry out the enhancement processing of the image first.

Image enhancement is first standardised, and common standardisation methods include mean-variance normalization [8], histogram equalisation [9], etc. In this paper, the mean-variance normalisation method is used.

Mean-variance normalisation (Z-score Normalization) is the process of assigning all data to a distribution with mean 0 and variance 1, calculated as shown in Eq. (1):

$$x_{scale} = \frac{x - \mu}{S} \quad (1)$$

Where x_{scale} is the normalised value, x is the value to be normalised, μ is the mean value of the image pixels, S is the standard deviation of the image, and the standard deviation is calculated as shown in Eq. (2) as follows:

$$S = \max(\sigma, \frac{1.0}{\sqrt{N}}) \quad (2)$$

where σ is the standard variance and N is the number of pixels in the image.

The normalisation process is followed by normalisation to map the original data onto the [0, 1] interval to obtain the image Z , calculated as shown in Eq. (3):

$$Z = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (3)$$

Where x_i is the value of the image pixel point, $\max(x)$ and $\min(x)$ are the maximum and minimum values of the image pixel point respectively.

The normalised image is scaled and fixed to 640×640 pixels, the calculation process is as follows According to the target size and the size of the original image, calculate the proportion of the required scaling, which can be based on the proportion of the long side or the short side. As shown in Eq. (4), Eq. (5):

$$scale_1 = \max(\frac{w}{W}, \frac{h}{H}) \quad (4)$$

$$scale_2 = \max(\frac{w}{W}, \frac{h}{H}) \quad (5)$$

Where $scale_1$ is the long side scaling ratio, $scale_2$ is the short side scaling ratio, w is the target image, W is the length of image Z , h is the height of the target image and H is the height of image Z .

4. Experiments and Analysis of Results

4.1. Dataset and Experimental Environment

The training dataset of YOLOv8-Pose model is used

Table 1. Key Point Correspondence Rule

0	1	2	3	4	5	6	7	8
Nose	Left eye	Right eye	Left ear	Right ear	Left shoulder	Right shoulder	Left elbow	Right elbow
9	10	11	12	13	14	15	16	
Left wrist	Right wrist	Left hip	Right hip	Left knee	Right knee	Left ankle	Right ankle	

Table 2. Parameter combinations

Number	Models	Epochs	Imgsz	Batch	Learning rate	momentum	Optimizers
1	YOLOv8n-pose	20	640	16	Lr=0.002	0.9	AdamW
2	YOLOv8s-pose	20	640	16	Lr=0.002	0.9	AdamW
3	YOLOv8m-pose	20	640	16	Lr=0.002	0.9	SGD
4	YOLOv8n-pose	10	640	32	Lr=0.002	0.9	AdamW
5	YOLOv8n-pose	16	640	10	Lr0=0.1, Lrf=0.1	0.937	'auto'
6	YOLOv8n-pose	32	640	10	Lr0=0.1, Lrf=0.1	0.937	'auto'

COCO2017 dataset and the dataset has been labelled with key points. The experimental environment is tested using windows 11 system, CPU model AMD EPYC 7453, graphics card model NVIDIA RTX 4090 with 24G of video memory hardware

4.2. Evaluation Indicators

In order to test the detection and classification ability of the model, the mAP (meanAverage Precision) method is used to evaluate the correctness and accuracy of the model for human posture detection. mAP mainly consists of P (precision), R (recall), AP , and mAP as shown in Eq. (6), Eq. (7), Eq. (8), and Eq. (9):

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$R = \frac{TP}{TP + FN} \quad (7)$$

$$AP = \int_0^1 P(r) d(r) \quad (8)$$

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} \quad (9)$$

Where, TP (True Positive) indicates that the predicted sample is positive and the actual sample is positive; FP (False Positive) indicates that the predicted sample is positive and the actual sample is negative; FN (False Negative) indicates that the predicted sample is negative and the actual sample is positive.

4.3. Key Point Test Results

In YOLOv8, six sizes of pre-trained model structures are provided for human key point detection as shown in Table 2. All YOLOv8-Pose models are pre-trained based on the COCO dataset, which is mainly used for the pose estimation task. Nsmplx represent the different sizes of the models, respectively, and the accuracy, parameter count, and inference speed of the models gradually increase with the increase of the model size.

Table 3. Model results

Number	Box				Pose			
	P	R	Map50	Map0-95	P	R	Map50	Map0-95
1	0.857	0.755	0.863	0.594	0.788	0.628	0.673	0.323
2	0.812	0.773	0.855	0.593	0.771	0.636	0.680	0.337
3	0.815	0.775	0.852	0.599	0.726	0.645	0.660	0.312
4	0.834	0.746	0.844	0.566	0.768	0.579	0.624	0.287
5	0.807	0.741	0.832	0.554	0.704	0.595	0.602	0.260
6	0.832	0.706	0.817	0.546	0.710	0.589	0.591	0.26

YOLOv8-Pose can realize the detection of 17 key points of the human body and return [17*3] position information, which represents the x-coordinate, y-coordinate and whether the 17 key points are visible or not. The corresponding rules of the 17 key points are shown in Table 1.

Since YOLOv8 has already provided a pre-training parameter model, we can further save time and obtain higher model accuracy by adjusting the model parameters and pre-training on this basis. By selecting different pre-training parameter models and adjusting the corresponding model parameters, six sets of experiments were conducted, and their corresponding results are shown in Table 3. Through the experiments, it is found that the selection of more complex pre-training models can effectively improve the accuracy, but its inference speed also increases. After comprehensive consideration, we select parameter combination 4 as the key extraction model for the subsequent part of the pose recognition, and export its training weights as onnx model for

subsequent calls.

4.4. Posture Recognition Results

Five main actions are recognized, namely falling, waving, sitting, standing and walking. The raw training data are all obtained by self-taking and processing. We further flatten the feature map information of the five [17,3] keypoints to form 255 feature columns, and perform category labeling. We experimented with models such as Random Forest, Support Vector Machine, and so on, and the classification accuracies of the models were stable at about 90% on the test set, and for this reason, we chose the Random Forest model as the gesture recognition model after the extraction of the keypoints, and the main parameters of the model are n_estimators The main parameters of the model are n_estimators=100,max_depth=5, min_samples_split=3,min_samples_leaf=5. The performance of the model on the test set is shown in Figure 5:

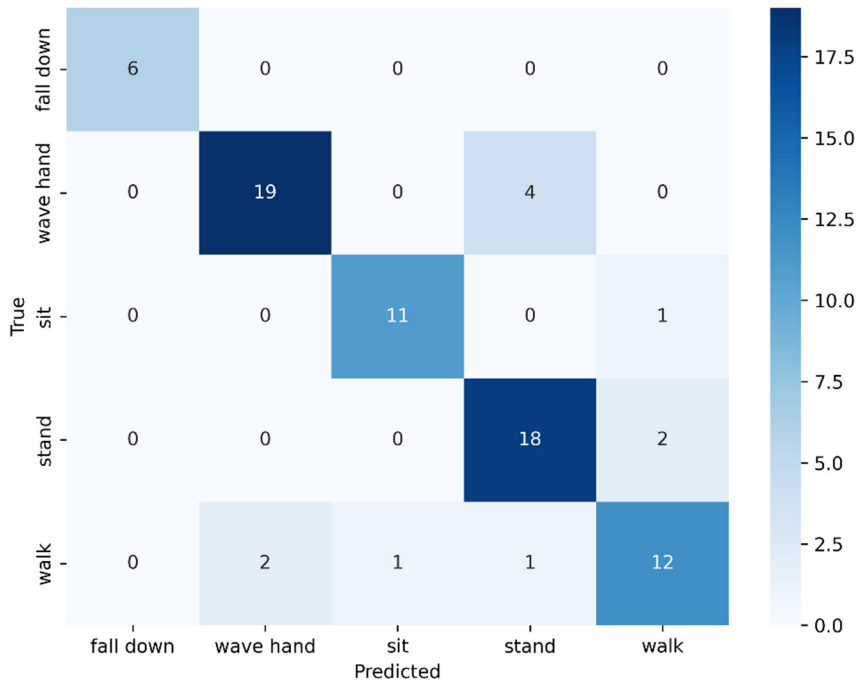


Figure 5. GSRF model test set confusion matrix

The final test on the RTSP video stream, all five actions can be effectively detected, as shown in Figure 6:



Figure 6. Human posture prediction effect diagram

Acknowledgments

This paper is for the 2023 Hebei College Students' Innovation and Entrepreneurship Training Program project "Intelligent Monitoring System Based on Posture Recognition" [S202310081137].

References

- [1] Zhang Xianling, Ming Yan. Analysis of the accuracy of age data in the Seventh National Population Census[J]. Population Research, 2022, 46(04): 27-39.
- [2] Chen Xiang. Research on fall detection based on acceleration sensor [D]. Fuzhou University, 2018.
- [3] Hu Xin, Liu Ruian, Huang Yulan. Improved Fall detection Algorithm for the Elderly Based on OpenPose [J]. Modern Information Technology, 2023, 7 (23): 73-78+82.
- [4] Huang, Haohao. Research and Design of Deep Learning Based Monitoring System for Dangerous Conditions of Elderly Living Alone [D]. Supervisor: Fuchun Zhang. Yan'an University, 2023.
- [5] Li Chaoyang. Drinking water metering device based on HX711 with memory function[J]. Computer Knowledge and Technology, 2023, 19 (26): 95-98.
- [6] Liang Fangfang. Design of Intelligent Electronic Scale Based on HX711 Multi-function [J]. Engineering Technology Research, 2021, 6 (15): 136-140.
- [7] Zhang Min, Zhang Huan, Shi Xiaojuan, Liang Zhuowen, Zhang Na. Design and implementation of fall detection system for elderly patients. China Medical Equipment, 2024, 21 (02): 157-161.
- [8] SHI Xiaoyuan, JING Xinxing, Zeng Min, et al. Robustness of speaker Recognition Based on Improved PNCC and i-vector [J]. Computer Engineering and Design, 2017, 38 (4).
- [9] Ma Ronggui, WEI Zizhong, He Yixu, et al. Backlight traffic scene image enhancement based on foreground and background characteristics [J/OL]. Radio Engineering: 1-9 [2023-10-04].