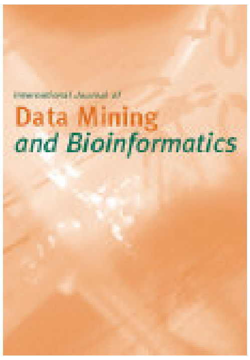


International Journal of Data Mining and Bioinformatics

ISSN online: 1748-5681 - ISSN print: 1748-5673
https://www.inderscience.com/ijdmb

Research on human health status recognition based on association algorithm

Taiping Jiang, Zhibing Wang, Lei Huang

DOI: 10.1504/IJDMB.2023.10058130

Article History:

Received:	13 April 2023
Last revised:	15 May 2023
Accepted:	19 June 2023
Published online:	17 October 2023

Research on human health status recognition based on association algorithm

Taiping Jiang

Xiangzhong Normal College for Preschool Education,
Shaoyang, 422000, China
Email: 512659257@qq.com

Zhibing Wang*

School of Computer Science,
Hunan University of Technology,
Zhuzhou, 412007, China
and
Hunan Key Laboratory of Intelligent Information
Perception and Processing Technology,
Hunan University of Technology,
Zhuzhou, 412007, China
Email: 58913921@qq.com
*Corresponding author

Lei Huang

Xiangzhong Normal College for Preschool Education,
Shaoyang, 422000, China
Email: 278696701@qq.com

Abstract: A human health status recognition method based on association algorithm is proposed to address the problems of low recognition accuracy, low correlation of health data collection, and long recognition time in existing human health status recognition methods. Firstly, a temperature sensor is used to collect human body temperature data. Secondly, the photoelectric capacitance method is used to collect heart rate and blood oxygen data. Once again, by setting the 3D coordinate system of human bone points and using the depth image coordinate system to determine the true distance of bone points, the collection of human bone related data is achieved. Finally, association algorithms are used to analyse the relationship between human health status data. Once a human health status recognition function is constructed, the recognition of health status is then completed. The test results show that the accuracy of the proposed method for identifying human health status remains around 99%.

Keywords: association algorithm; identification of human health status; body temperature data; bone data; identification function.

Reference to this paper should be made as follows: Jiang, T., Wang, Z. and Huang, L. (2023) 'Research on human health status recognition based on association algorithm', *Int. J. Data Mining and Bioinformatics*, Vol. 27, No. 4, pp.267–278.

Biographical notes: Taiping Jiang is an Associate Professor, and graduated from the Department of Mathematics of Hunan Normal University in 1994. Currently, he is engaged in Mathematics education and teaching in the Xiangzhong Normal College for Preschool Education. His main research directions are mathematical modelling and data analysis.

Zhibing Wang received his MS in Computer Science and Technology from the Hunan University of Technology, Zhuzhou, China in 2010. He is currently an Associate Professor with the School of Computer Science, Hunan University of Technology. His research interests include the industrial internet of things, trusted software and knowledge graph.

Lei Huang is an Associate Professor, and graduated from the Hunan City College in 2008. Currently, he is engaged in Computer education and teaching in the Xiangzhong Normal College for Preschool Education. His main research directions are information literacy and data analysis.

1 Introduction

The continuous and rapid development of the social economy has led to the continuous improvement of human living standards. But with the continuous changes in dietary structure and work pressure, there have also been problems with human health (Zhang et al., 2021). With the increasingly fierce social competition, the number of people with sub-health issues is constantly increasing. Human health issues not only include diseases, but also psychological issues. Among them, sub-health states other than diseases refer to the varying degrees of decline in the health vitality state of the human body in the absence of a definite disease state (Lam et al., 2021). It is a state of physiological function reduction between health and disease, and a health imbalance problem that can develop into a healthy state and transform into a disease state. The emergence of this sub-health seriously affects human life and health (Li, 2022). Therefore, effective identification of human health status can effectively detect unhealthy states in the human body. The recognition of human health status can be achieved by analysing human cardiovascular and cerebrovascular data, numerical values, and mental state, in order to preliminarily determine whether the human body is healthy and provide a basis for early detection of human discomfort (Sharma et al., 2022). However, as life becomes busier and social pressure increases, the recognition of human health status is not ideal (Liew and Guo, 2022). Therefore, in order to improve the effectiveness of human health status recognition, relevant researchers have designed many recognition methods and achieved certain results.

Liu (2022) proposes a human health state recognition method based on MEMS sensors and Zigbee networks. This method obtains human motion information through multiple sensor nodes, and uses sensor network technology to aggregate and upload these data to the host. A data fusion algorithm based on complementary filtering has been introduced to complete the recognition of human health status. This recognition method can quickly extract human health status data, but there are many recognition errors in the recognition process. Kabir et al. (2021) proposes a human body state recognition method based on CSI signals, which collects human health state data through human-machine interaction sensors and denoises the collected human health state data using a low-pass filter. Finally, the CNN model is used to complete the recognition of human health status. This method can accurately identify human health status, but requires a large amount of data processing, resulting in a longer overall recognition time. Huan et al. (2022) proposes a human health state recognition method based on multi-feature sensor data. This method combines the features of multiple sets obtained through learning networks, manual extraction, and position information to generate mixed features, in order to obtain better recognition results for human health state data. Using CNN and BLSTM networks to extract features from human health data, thus completing the recognition of human health status. However, the recognition performance of this method is significantly better than existing methods, but the recognition process takes longer.

In response to the problems of large errors and long time consumption in the above methods, this paper designs a new human health state recognition method based on correlation algorithms. The detailed research technical route of this method is as follows:

1 The collection of human body temperature, heart rate, blood oxygen, and bone data is carried out using temperature collection sensors, photoelectric converters, and infrared irradiation methods.

2 Based on the collected human health related data, the nearest neighbour data association algorithm is used to update the human health status data and identify the target data of human health status. Construct a human health data state recognition function using correlation algorithms, input the identified state target data into this function, and complete the recognition of human health state data.

3 Experimental verification is conducted by comparing the method proposed in this paper with traditional methods based on the accuracy of human health status recognition, correlation of human health data collection, and time spent on human health status recognition.

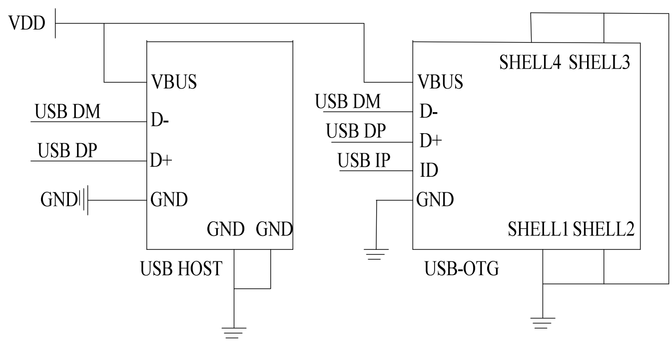
2 Human health status data collection

Collecting body temperature data, heart rate and blood oxygen data, and bone data is to comprehensively evaluate the health status of the human body. These data can reflect the physiological state of the human body, including the body's heat balance, the health status of the cardiovascular system, and the stability of the skeletal system. By monitoring these data, potential health issues can be identified in a timely manner and corresponding measures can be taken for intervention and treatment.

2.1 Human body temperature data collection

Collect human body temperature data through a temperature sensor. The temperature collection sensor is equipped with an infrared temperature sensor and an operational amplifier, which compensates for the environmental temperature through the measurement principle of thermocouples, resulting in better measurement results. The infrared temperature sensor does not pose any danger to human health by absorbing infrared radiation from the human body, and the measured temperature shows a linear change with small errors (Xing et al., 2021; Wang et al., 2021). The circuit of the temperature collection sensor is shown in Figure 1.

Figure 1 Temperature collection sensor circuit



2.2 Human heart rate and blood oxygen data collection

The heart rate data collection is achieved using the photocapacitance method. The sensors used in this method generally include a light source and a photoelectric converter. Light source can screen oxyhaemoglobin and deoxygenated haemoglobin in arterial blood (Wan, 2021). Extract the data when the light beam passes through the peripheral blood vessels of the human body.

The photoelectric converter senses the reflected optical fibre and converts it into a signal that can simplify applications. Due to the continuous changes in heart rate with human activity, the electrical signal cycle (Karthik and Ramkumar, 2022) can be expressed as heart rate, and the collected heart rate data is represented as:

$$S_i = \frac{c_i}{c_i + c_d} \times 100\% \tag{1}$$

Among them, S_i represents the heart rate data collection results, c_i represents the oxygen saturation, and c_d represents the de-oxygen saturation.

On this basis, the oxygen saturation result collected by the converted photoelectric signal is a quadratic curve fitting (Ahmed et al., 2021), and the result is expressed as:

$$p_i = -45.6 \times R \times R + 30.35 \times R \tag{2}$$

Among them, R represents the change of red light intensity and p_i represents the quadratic fitting curve of oxygen saturation.

The optical capacitance product method is used to determine the absorption effect of oxygen saturation collected at different wavelengths, so as to realise the collection of heart rate and blood oxygen value data.

2.3 Human bone data collection

The data of bones in human health is also crucial (Ding et al., 2021). Human bones can be viewed as a 3D coordinate system, with apogee positions set as:

$$(X = 0, Y = 0, Z = 0) \tag{3}$$

The infrared camera centre of the human bone data collection device follows the left side of the infrared irradiation in the horizontal axis direction, above the infrared irradiation in the vertical axis direction, and in the middle direction, the infrared irradiation direction. In this coordinate system, specific human bone related data is collected by describing the position of depth image pixels (Feng et al., 2021). In the depth image coordinate system of human skeleton points, each pixel point of skeleton points can be obtained to represent the true distance, such as the boundary point position and centre position of human chest, waist and hip (Liu et al., 2021). The contour length calculation formula for the pixel representation of human bone points obtained at this time is set to:

$$v = \sqrt{x^2 + y^2} \tag{4}$$

Among them, v represents the length of the arc at the pixel end and $\sqrt{x^2 + y^2}$ represents the true distance between bone points.

By setting the 3D coordinate system of human bone points and using the depth image coordinate system to determine the true distance of bone points, the collection of human bone related data is achieved.

3 Human health status recognition based on association algorithm

Based on the collected human health data mentioned above, in order to achieve effective judgement of human health status, this article introduces correlation algorithms to identify human health status data, providing certain assistance for human health. Association algorithm is an algorithm that mines the relationships or interrelationships between large amounts of data items (Lu et al., 2021). This algorithm analyses the association relationship of data items in a given dataset based on set rules, mainly describing the degree of closeness between a set of data in the dataset. This algorithm has been widely applied in multiple fields. Therefore, in order to achieve the recognition of human health status, this article uses association algorithms to analyse the relationship between human health status data (Antonello et al., 2021), thereby achieving effective recognition of human health status.

The implementation steps of the human health state recognition algorithm based on association algorithm are as follows:

Step 1 Use the nearest neighbour data association algorithm to update human health status data. This algorithm updates the health status data by setting a tracking gate for the health status data, selecting the measurement value with the smallest weighted norm of information in the tracking gate as the candidate for recognition of the health status data. At this time, the candidate qualification conditions of the determined target meet:

$$\left[q(k+1) - \hat{q}(k+1 \mid k)\right] r^{-1}(k+1) - \hat{q}(k+1 \mid k) \leq \alpha \tag{5}$$

Among them, $q(k + 1)$ represents candidate health status data, $\hat{q}(k+1 \mid k)$ represents predicted values, r^{-1} represents innovation covariance, and α represents gate size.

At this point, the health status data is updated using the minimum measurement value of the weighted norm of innovation (Diamond and Happawana, 2022), and the updated result is:

$$f^2(x) = \vartheta\left[q(k+1) - \hat{q}(k+1 \mid k)\right] r^{-1}(k+1) - \hat{q}(k+1 \mid k) \tag{6}$$

Among them, $f^2(x)$ represents the updated health status data results and ϑ represents the minimum measurement value of information weighted norm.

Step 2 Set the target for identifying human health status data. Assuming that the health status data falling into the set gate mentioned above are all target data to be identified, but the probability of each identified health status data being successfully identified varies. Therefore, in order to reduce the amount of data identified and shorten the recognition time, it is necessary to determine whether the health status data to be identified is the target of recognition. Assuming that the set of candidate health status data to be identified within the target gate of the association algorithm at different times is $z(x)$, the set of health status data to be identified from the initial time to the gate is represented as:

$$z(x) = \left\{z_i(x)_i^{n_k}\right\} \tag{7}$$

Among them, n_k represents the number of health status data to be identified.

On this basis, the following time is defined:

$$\theta_i(x) = \left(z_i(x) \text{ is the measurement from the data to be identified}\right) \tag{8}$$

$$\theta_j(x) = (\text{There is no measurement of data to be identified at this time}) \tag{9}$$

At this time, the conditional probability calculation result of the determined health status data to be identified as the target data is:

$$v_i(x) = P\left\{\theta_i(x) \mid z^x\right\} \tag{10}$$

Among them, $v_i(x)$ represents the conditional probability result of the target data, and P represents the maximum probability that the data volume becomes the target.

Step 3 Calculate the covariance difference of recognition error for human health status target recognition data to reduce the errors in recognition, namely:

$$G(x) = \sum v_i(x) / \beta_0(k) \int \theta_i(x) \tag{11}$$

Among them, $G(x)$ represents the identification error covariance difference of human health status target data, and $\beta_0(k)$ represents the information combination of health status data.

Step 4 Determine the interconnection probability between the target recognition data of human health status, and divide the set of data to be identified into cumulative data and latest data. At this point, the probability of interconnection between human health status target recognition data is:

$$Q\left(G(x) \mid x\right) = \frac{q\left(G(x) \mid x\right) U\left(\beta_0\right)}{\sum q\left(G(x) \mid x\right) U\left(\beta_0\right)} \tag{12}$$

Among them, $Q(G(x) \mid x)$ represents the interconnection probability between human health status target recognition data, q represents the correlation coefficient between data items, and U represents the probability density value.

Step 5 Construct a human health status recognition function based on correlation algorithms, input the identified human health status target recognition data into this function, and achieves the final recognition. The constructed human health status recognition function is:

$$\mu_F\left(z(x)\right) = r^{-\gamma} \sum_{m=1}^{k} \frac{\left(\gamma V_k\right)^k}{z(x)} \sqrt{\frac{\left(\gamma V_k\right)^k}{Q\left(G(x) \mid x\right)}} \tag{13}$$

Among them, r represents the number of iterations for recognition, V_k represents the correlation coefficient between recognition data, and k represents the time of change in human health status data.

In the design of human health status recognition algorithm based on association algorithm, the nearest neighbour data association algorithm is used to update human health status data, set recognition targets for human health status data, calculate the error covariance difference of human health status target recognition data, and determine the interconnection probability between human health status target recognition data. Based on this, a human health status data recognition model based on association algorithm is constructed, Implement final identification research.

4 Experimental analysis

4.1 Experimental data

Conduct an experimental analysis to test the feasibility of the designed recognition method. In the experimental testing, a community in a certain area was used as the research object to identify human health status data in that community. In this community, 100 households were selected, with a fixed number of two individuals per household as the unit, to collect and report daily health status data. Families with a population of less than two individuals were not included in this experimental study. The specific test parameters in the study are shown in Table 1.

4.2 Experimental scheme and index

The experimental indexes set in the test were precision of human health status identification, relevance of human health data collection and time spent in identification. Among them, human health status recognition accuracy refers to the effectiveness of human health status data analysis for all samples. The value range of this index is [0, 100]%, and the higher the value, the better the recognition effect. The correlation degree of human health data collection is a key index affecting the final identification effect, which is reflected by the correlation degree of data collection.

Table 1 Test parameter details

Parameter	*Detailed information*
Number of people tested	200
Test object age	16–75
Number of young adults (over 16 to 45 years old)	80
Number of elderly people (over 60 years old)	60
Health status data type	Body temperature data, heart rate, blood oxygen saturation, bone density
Data acquisition sensor model	MAX30205
Data collection cycle/day	0.5
Data reporting cycle/day	1
Number of data identification	100
Health status data volume	20,000

4.3 Analysis of experimental results

4.3.1 Analysis of the accuracy results of human health status recognition

The proposed method, Kabir et al. (2021) method, and Huan et al. (2022) method were used in the test to analyse the recognition accuracy of selected sample human health status data. The results obtained are shown in Figure 2.

According to the test results shown in Figure 2, it can be seen that there are certain differences in the recognition accuracy results of the selected sample's health status data using the proposed method, the method in Kabir et al. (2021), and the method in Huan et al. (2022). As the amount of health status data for recognition varies, all three methods demonstrate certain changes in recognition accuracy. The proposed method maintains a recognition accuracy of around 99% for the sample's health status data, while the other two methods show significant fluctuations in recognition accuracy and consistently perform lower than the proposed method. This verifies the feasibility of the proposed method and highlights the effectiveness of this paper's approach.

4.3.2 Analysis of correlation results for human health data collection

The proposed method, Kabir et al. (2021) method, and Huan et al. (2022) method were used in the test to analyse the correlation of data collection for selected samples of human health status data. The results obtained are shown in Table 2.

According to the test results in Table 2, it can be seen that for the selected sample of human health status data, the proposed method, the method from Kabir et al. (2021), and the method from Huan et al. (2022) were analysed for their correlation in data collection. The correlation coefficients of the three methods in collecting data were different. When the collected sample data was 5,000, the correlation coefficients of the three methods in collecting data were approximately 0.99, 0.86 and 0.84, respectively. When the collected sample data was 8,000, the correlation coefficients of the three methods in collecting data were approximately 0.98, 0.80 and 0.75, respectively. When the collected sample data was 10,000, the correlation coefficients of the three methods in collecting data were approximately 0.97, 0.79 and 0.74, respectively. By comparing the sizes of these correlation coefficients, it can be seen that the correlation coefficient of the proposed method is always higher than the other two methods, indicating that the proposed method has better performance.

Figure 2 Precision results of human health status recognition

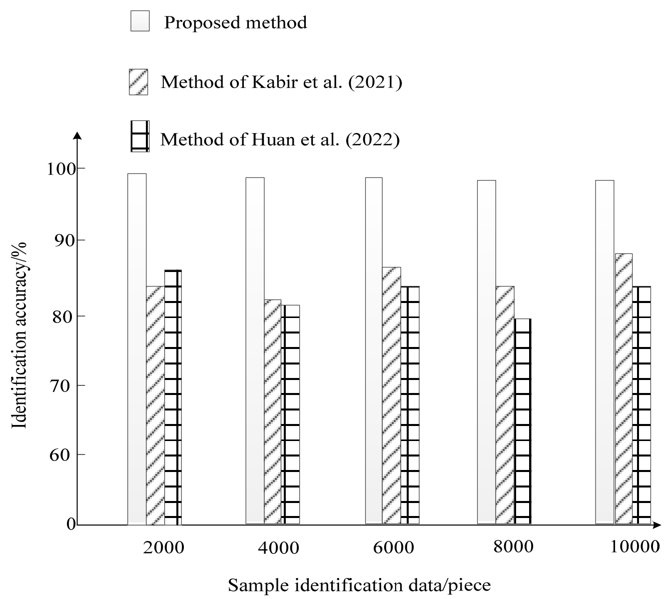


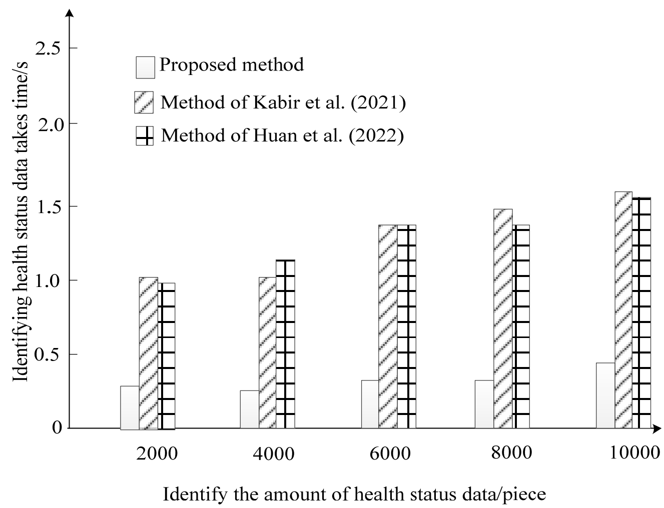
Table 2 Analysis of correlation results for human health data collection

Identify health status data / piece	*Proposed method*	*Kabir et al. (2021) method*	*Huan et al. (2022) method*
1,000	0.99	0.90	0.89
2,000	0.99	0.90	0.89
3,000	0.99	0.87	0.87
4,000	0.98	0.87	0.86
5,000	0.99	0.86	0.84
6,000	0.98	0.84	0.87
7,000	0.98	0.80	0.80
8,000	0.98	0.80	0.75
9,000	0.98	0.79	0.76
10,000	0.97	0.79	0.74

4.3.3 Analysis of time consuming results for identifying human health status

In the test, the proposed method, Kabir et al. (2021) method, and Huan et al. (2022) method were used to recognise and analyse the human health status data of the selected samples. The results obtained are shown in Figure 3.

Figure 3 Time consuming results of human health status recognition



The test results in Figure 3 show that the time required for identifying the selected sample's health status data using the proposed method, the method in Kabir et al. (2021), and the method in Huan et al. (2022) varies with the number of samples. Among them, the proposed method consistently requires less than 0.5 seconds for identification, while the method in Kabir et al. (2021) requires over 1.5 seconds. The identification time for the method in Huan et al. (2022) fluctuates greatly but is always longer than that of the proposed method. Therefore, it can be concluded that the proposed method has the shortest identification time.

5 Conclusions

Human health status recognition helps to increase people's attention to health. Therefore, this paper proposes a human health status recognition method based on association algorithms, and verifies the performance of the method from both theoretical and experimental perspectives. This method has high accuracy in human health state recognition, short time consumption for state recognition, and high correlation in data collection. The test results show that the highest accuracy of the proposed method in identifying data is about 99%, the highest correlation coefficient of the collected data is 0.99, and the recognition time is less than 0.5 seconds, highlighting the recognition performance of the proposed method.

References

Ahmed, M., Antar, A.D. and Ahad, M. (2021) 'Static postural transition-based technique and efficient feature extraction for sensor-based activity recognition', *Pattern Recognition Letters*, Vol. 147, No. 6, pp.25–33.

Antonello, F., Baraldi, P., Shokry, A., Zio, E. and Serio, L. (2021) 'A novel association rule mining method for the identification of rare functional dependencies in complex technical infrastructures from alarm data', *Expert Systems with Applications*, Vol. 170, No. 15, pp.1145–1152.

Diamond, B.J. and Happawana, K.A. (2022) 'Association rule learning in neuropsychological data analysis for Alzheimer's disease', *Journal of Neuropsychology*, Vol. 16, No. 1, pp.116–130.

Ding, W., Ding, C., Li, G. and Liu, K. (2021) 'Skeleton-based square grid for human action recognition with 3D convolutional neural network', *IEEE Access*, Vol. 20, No. 13, pp.54078–54089.

Feng, D., Wu, Z.C., Zhang, J. and Ren, T.T. (2021) 'Multi-scale spatial temporal graph neural network for skeleton-based action recognition', *IEEE Access*, Vol. 10, No. 14, pp.58256–58265.

Huan, R., Jiang, C., Ge, L., Shu, J., Zhan, Z., Chen, P., Chi, K. and Liang, R. (2022) 'Human complex activity recognition with sensor data using multiple features', *IEEE Sensors Journal*, Vol. 22, No. 1, pp.757–775.

Kabir, M.H., Rahman, M.H. and Shin, W. (2021) 'CSI-IANet: an inception attention network for human-human interaction recognition based on CSI signal', *IEEE Access*, Vol. 16, No. 21, pp.166624–166638.

Karthik, B. and Ramkumar, G. (2022) 'Comparison of feature extraction technique for segmentation in human iris recognition under uncontrolled environment using CNN algorithm with SVM classifier', *ECS Transactions*, Vol. 107, No. 1, pp.16785–16795.

Lam, K.F., Lee, C.Y., Wong, K.Y. and Bandyopadhyay, D. (2021) 'Marginal analysis of current status data with informative cluster size using a class of semiparametric transformation cure models', *Statistics in Medicine*, Vol. 40, No. 10, pp.2400–2412.

Li, Z. (2022) 'Quantifying exposure source allocation factors of pesticides in support of regulatory human health risk assessment', *Journal of Environmental Management*, Vol. 309, No. 1, pp.114–119.

Liew, Z. and Guo, P. (2022) 'Human health effects of chemical mixtures', *Science*, Vol. 375, No. 12, pp.1–10.

Liu, M., Ye, Y., Jiang, J. and Yang, K. (2021) 'MANIEA: a microbial association network inference method based on improved Eclat association rule mining algorithm', *Bioinformatics*, Vol. 37, No. 20, pp.3569–3578.

Liu, Q. (2022) 'Human motion state recognition based on MEMS sensors and Zigbee network', *Computer Communications*, Vol. 181, No. 11, pp.164–172.

Lu, P.H., Keng, J.L., Tsai, F.M., Lu, P.H. and Kuo, C.Y. (2021) 'An apriori algorithm-based association rule analysis to identify acupoint combinations for treating diabetic gastroparesis', *Evidence-based Complementary and Alternative Medicine*, Vol. 36, No. 17, pp.1–9.

Sharma, S., Shree, B., Aditika, Sharma, A., Irfan, M. and Kumar, P. (2022) 'Nanoparticle-based toxicity in perishable vegetable crops: molecular insights, impact on human health and mitigation strategies for sustainable cultivation', *Environmental Research*, Vol. 212, No. 3, pp.168–179.

Wan, Y. (2021) 'Abnormal patterns recognition in bivariate autocorrelated process using optimized random forest and multi-feature extraction', *ISA Transactions*, Vol. 109, No. 1, pp.102–111.

Wang, Q., Tao, B. and Han, F. (2021) 'Extraction and recognition method of basketball players' dynamic human actions based on deep learning', *Mobile Information Systems*, Vol. 55, No. 2, pp.1–6.

Xing, S., Jiao, Y., Salehzadeh, M., Soma, K.K. and Huan, T. (2021) 'SteroidXtract: deep learning-based pattern recognition enables comprehensive and rapid extraction of steroid-like metabolic features for automated biology-driven metabolomics', *Analytical Chemistry*, Vol. 93, No. 14, pp.5735–5743.

Zhang, J., Hu, J., You, H., Jia, R. and Zhang, X. (2021) 'Characterization method of IGBT comprehensive health index based on online status data', *Microelectronics Reliability*, Vol. 116, No. 9, pp.1–8.