Ministry of Education of the Republic of Belarus
Educational institution
Belarusian State University of
Informatics and Radioelectronics

UDC 004.896

CHEN
ZHEYING

HUMAN ACTIVITY RECOGNITION SYSTEM USING SMARTPHONE ACCELEROMETER DATA AND LONG SHORT TIME MEMORY

Abstract
for a Master’s Degree
in the Specialty 1-45 80 01 Infocommunication Systems and Networks

Supervisor
Doctor of technical sciences,
Professor
Boriskevich Anatoliy Antonovich

Minsk 2023

INTRODUCTION

Traditional sensor devices are bulky and expensive. With the continuous development of smart phones in recent years, the acceleration sensors of mobile phones have also continued to improve. It has the obvious advantages of small size, high penetration rate and lower price, which provides a new idea for the application of intelligent assistance technology and so on. Recognition of human activities from sensor data is at the core of intelligent assistive technologies, such as smart home, rehabilitation, health support, skills assessment or industrial environments [1]. For example, the project of Inooka et al. predicts the energy consumption of users by recognizing their activities [2], and Mathie et al. judges whether users are safe or not by recognizing their actions [3]. This work is motivated by two requirements of activity recognition: improving recognition accuracy and reducing reliance on engineered features to address increasingly complex recognition problems.

Human Activity Recognition (HAR) is based on the assumption that specific body movements translate into characteristic sensor signal patterns, which can be sensed and classified using machine learning techniques. We use data collected from accelerometer sensors. Almost every modern smartphone has a three–axis accelerometer that measures acceleration in all three spatial dimensions.

We selected the data set from the Wireless Sensor Data Mining (WISDM) project, which collected 1,098,207 experimental data generated from 29 volunteers carrying smartphones to perform specified actions every 50 ms, and each piece of data consists of 6 parts: Username, specified action, timestamp and accelerometer values for x, y and z axis. We compared the advantages and disadvantages of different low–pass filters, and finally adopted a third–order Butterworth filter with a cutoff frequency of 4 to filter the noise in the original data. We use a window of size 200 with an overlap of 90 % to divide the x, y and z axis accelerometer and label part in the original data, store them as acceleration data and label data respectively for preprocessing. We get 54901 windows and split both data into a training set (80 %) and a test set (20 %).

We trained a double layer LSTM neural network (implemented in TensorFlow) for HAR from accelerometer data with the purpose of providing an algorithm with higher recognition accuracy. The trained model will be exported/saved and added to the Android app. The network model consists of double layer LSTM network layers and double fully connected layers (FCL), and predicts the corresponding human actions from the x, y and z axis acceleration count values from the data set. The proposed algorithm achieved 98.4 % accuracy and a loss of 0,434 on the test set.

The aim of the work is to increase recognition accuracy of different human activities using the acceleration sensor data from the smartphone and deep learning.

To achieve this aim, the following tasks were solved in the master thesis:

1 Data collection on different human activities by using smartphones.

2 Pre–processing algorithm of the sensor data.

3 Human activity recognition using double FCL and double LSTM algorithm.

4 Evaluation of algorithm performance using confusion matrix and accuracy.

GENERAL DESCRIPTION OF WORK

Relevance of the subject

The work corresponds to paragraph 1 «Digital information and communication and interdisciplinary technologies, production based on them» of the State Program of innovative development of the Republic of Belarus for 2021–2025.

The work was carried out in the educational institution Belarusian State University of Informatics and Radioelectronics within the framework of research work 21–2033 "Processing, coding and transmission of information in network–centric systems".

The aim and tasks of the work

The aim of the work is to increase recognition accuracy of different human activities using the acceleration sensor data from the smartphone and deep learning.

To achieve this aim, the following tasks were solved in the master thesis:

1 Data collection on different human activities by using smartphones.

2 Pre–processing algorithm of the sensor data.

3 Human activity recognition using double FCL and double LSTM algorithm.

4 Evaluation of algorithm performance using confusion matrix and accuracy.

Personal contribution of the author

The content of the dissertation reflects the personal contribution of the author.

1. Preprocessing and temporal feature extraction of acceleration sensor data in public datasets.

2. Construction and implementation of algorithm structure.

3. Evaluate the results of the proposed double FCL and double LSTM algorithm.

Task setting and discussion of the results were carried out together with the supervisor, professor, Anatoliy.Antonovich.Boriskevich.

Testing and implementation of results

The main provisions and results of the dissertation work were reported and discussed at: 59th scientific conference of postgraduates, undergraduates and students, (Minsk, April 17–21, 2023), International scientific and technical seminar "Technologies of information transmission and processing" (Minsk, March – April 2023) and BIG DATA and Advanced Analytics: collection of scientific articles of the IX International Scientific and Practical Conference, (Minsk, May 17–28, 2023).

Author's publications

According to the results of the research presented in the dissertation, 3 author's works was published, including: 3 articles and abstracts in conference proceedings.

Structure and size of the work

The dissertation work consists of introduction, general description of the work, three chapters with conclusions for each chapter, conclusion, bibliography, eight appendixes.

The total amount of the thesis is 75 pages, of which 54 pages of text, a list of used bibliographic sources (22 titles on 55 pages), a list of the author's publications on the subject of the thesis (3 titles on 56 pages), 1 appendix on 58 pages.

Plagiarism

An examination of the dissertation « Human physical activity recognition algorithm based on smartphone data convolutional neural network and long short time memory» by Chen Zheying was carried out for the correctness of the use of borrowed materials using the network resource «PaperYY» (access address: https://www.paperyy.cn/) in the on–line mode 07.06.2023. As a result of the verification, the correctness of the use of borrowed materials was established (the originality of the thesis is 89.2%)

SUMMARY OF WORK

The **introduction** addresses the problems of physical human activities recognition.

The **general description of work** shows the connection between the work and the priority areas of scientific research, the aim and tasks of the research, the personal contribution of the applicant for a scientific degree, the approbation of the dissertation results.

In the first chapter, we explored the research status of HAR for physical human activities recognition in different domains.

We selected the dataset from the WISDM lab, which data is obtained by volunteers putting the Android phone in the front trouser pocket to complete the 6 specified activities including walking, jogging, upstairs, downstairs, sitting, standing. Each sample from the WISDM dataset includes user number, the type of activities completed by the user, and the time when the sample was collected (in nanoseconds), and acceleration values on the x, y, and z–axis (see figure 1). Accelerometer data is collected every 50ms, and is stored as raw data by specially developed software. It means that when a volunteer completes a given action, we will get 20 samples per second.

Figure 1 – Description of WISDM Raw Data

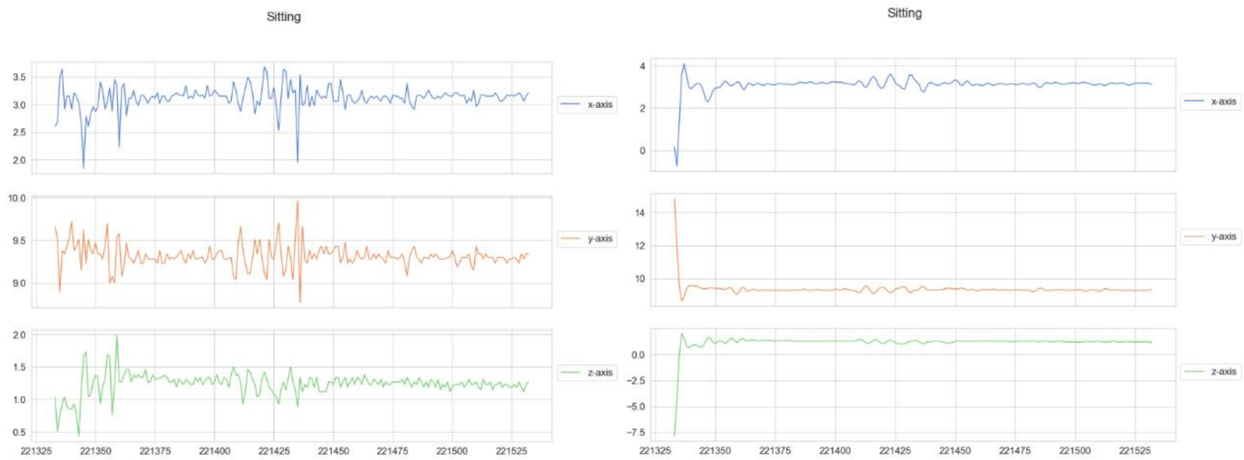
Table 1 shows the sample distribution of each human physical activities in the dataset.

Table 1 – Description of WISDM dataset

Activity distribution	Total number of samples	Percentage
Walking	424400	38.6%
Jogging	342177	31.2%
Up stairs	122869	11.2%
Down stairs	100427	9.1%
Sitting	59939	5.5%
Standing	48395	4.4%

We analyzed several low–pass filters, and finally chose the Butterworth low–pass filter for preprocessing to reduce the impact of noise. After testing different orders and cut–off frequencies, the preprocessing effect of the third–order Butterworth low–pass filter with a cut–off frequency of 4 is the best.

Figure 2 shows the curves of the acceleration data labeled "sitting" from the raw dataset before and after filtering. It can be seen that the filtered graph becomes smoother.



(a) (b)

Figure 2 – Graphs before and after filtering with a Butterworth filter with a 3^{rd}–order cutoff frequency of 4

We also used a 10s sliding window to improve feature extraction from filtered data, and one–hot encoding to digitize features for action labels. We choose an overlap rate of 90%, and get 54901 windows with a data type of 32 bits. Taking the acceleration data with the label "standing" as an example, its sequence segmentation diagram is shown in Figure 3.

In addition, we performed the hyperparameter settings of the neural network model, including learning rate, batch size and Adam optimizer. Finally, we split the data into training (80%) and test (20%) set.

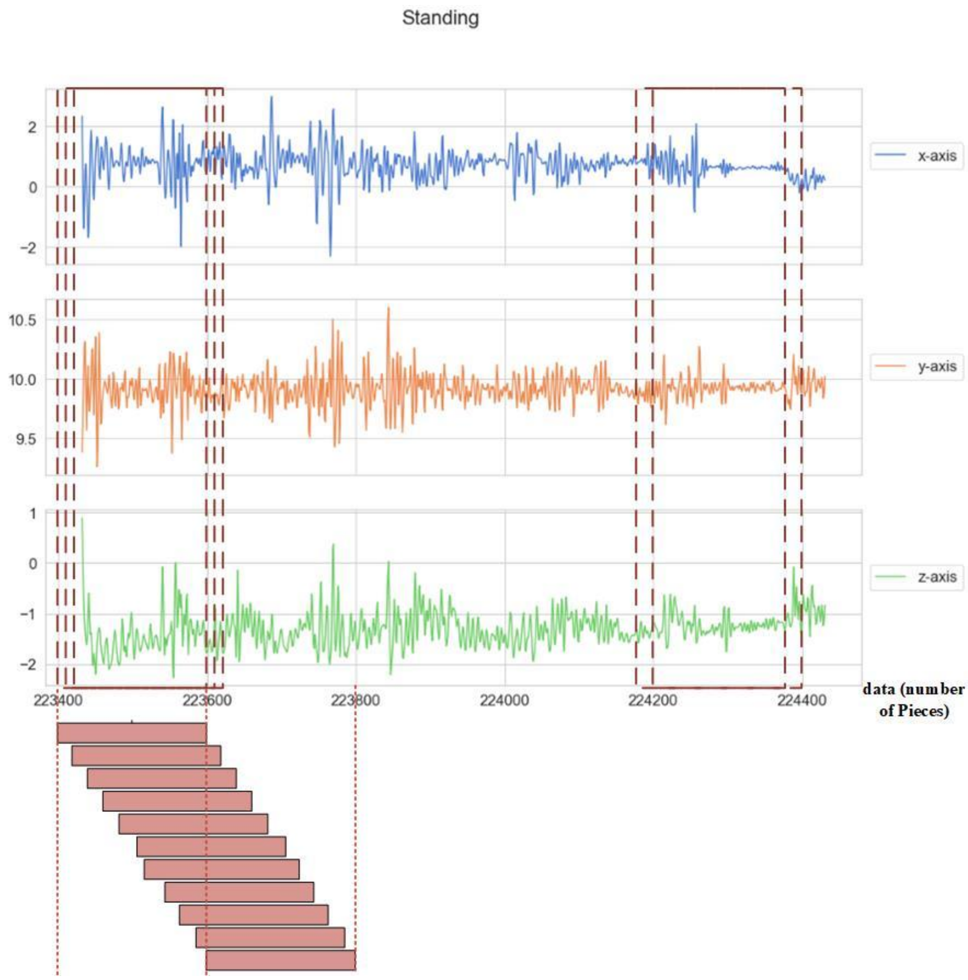


Figure 3 – Data segmentation using a sliding window of size 200

In the second chapter, we analyzed the structures of RNN, LSTM, and GRU respectively, compared their advantages and disadvantages, and introduced the two–layer RNN network structure.

We proposed a human activity recognition neural network model with double layer LSTM network layers and double FCL (see figure 4), and set the learning parameters and hyperparameters of each layer in the model separately (see Table 2 and Table 3). We partition the training set consist of 3–axis acceleration data (x, y and z) with a batch size of 1024, resulting in 50 iterations as the input. Softmax layer converts the input from the previous layer into probability set of 6 physical human activities as the output.

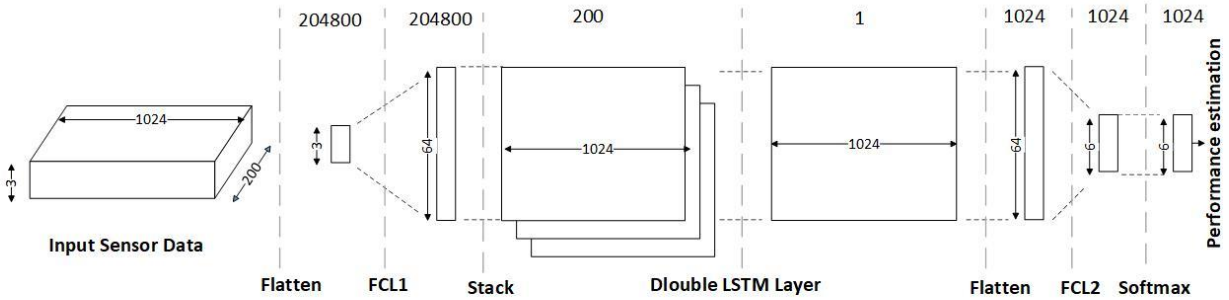


Figure 4 – Block diagram of human activity recognition algorithm based on smartphone data and deep learning

Table 2 – The learning parameters of double FCL and double layer LSTM neural network based on mobile phone accelerometer

Modules	Learning parameters	matrix size
Fully Connected Layer 1	Weight configuration	[3,64]
	bias	[64]
Double LSTM Layer	Number of params	66048
Fully Connected Layer 2	weight	[64,6]
	bias	[6]

The algorithm transforms the input 3 features into 64 through FCL1, so as to better divide different types of data. The double–layer LSTM layer is used to extract the long–term and short–term dependencies between features, FCL2 is used to fuse 64–dimensional features into 6–dimensional, and the softmax layer converts the model prediction results into the probability set of each action category.

Table 3 – The hyperparameter of double FCL and double layer LSTM neural network based on mobile phone accelerometer

Modules	Hyperparameter	Value
Fully Connected Layer 1	hidden unit size	64
	activation function	ReLU
Double LSTM Layer	hidden unit size	64
Fully Connected Layer 2	hidden unit size	6
Training	optimizer	Adam
	batch size	1024
	learning rate	0.0025
	number of epochs	50

In the third chapter, we performed the performance estimation indicators of several recognition algorithms used in this network model using confusion matrix, loss and accuracy.

The confusion matrix our work is shown in Table 4. where A means activity, W means walking, J means jogging, U means upstairs, D means downstairs, SIT means sitting, ST means standing, N is the number of various activities.

Table 4 – Confusion Matrix for Multiple Classification Tasks

Activities		Predicted label						Total
		Walking	Jogging	Upstairs	Downstairs	Sitting	Standing	
True label	Walking	A_{WW}	A_{JW}	A_{UW}	A_{DW}	A_{SITW}	A_{STW}	N_W
	Jogging	A_{WJ}	A_{JJ}	A_{UJ}	A_{DJ}	A_{SITJ}	A_{STJ}	N_J
	Upstairs	A_{WU}	A_{JU}	A_{UU}	A_{DU}	A_{SITU}	A_{STU}	N_U
	Downstairs	A_{WD}	A_{JD}	A_{UD}	A_{DD}	A_{SITD}	A_{STD}	N_D
	Sitting	A_{WSIT}	A_{JSIT}	A_{USIT}	A_{DSIT}	A_{SITSIT}	A_{STSIT}	N_{SIT}
	Standing	A_{WST}	A_{JST}	A_{UST}	A_{DST}	A_{SITST}	A_{STST}	N_{ST}

The confusion matrix of this model is shown in Figure 5. From the confusion matrix, it can be seen that the model predicts well on most actions, and it is worth noting that there are some misclassifications on up and down stairs.

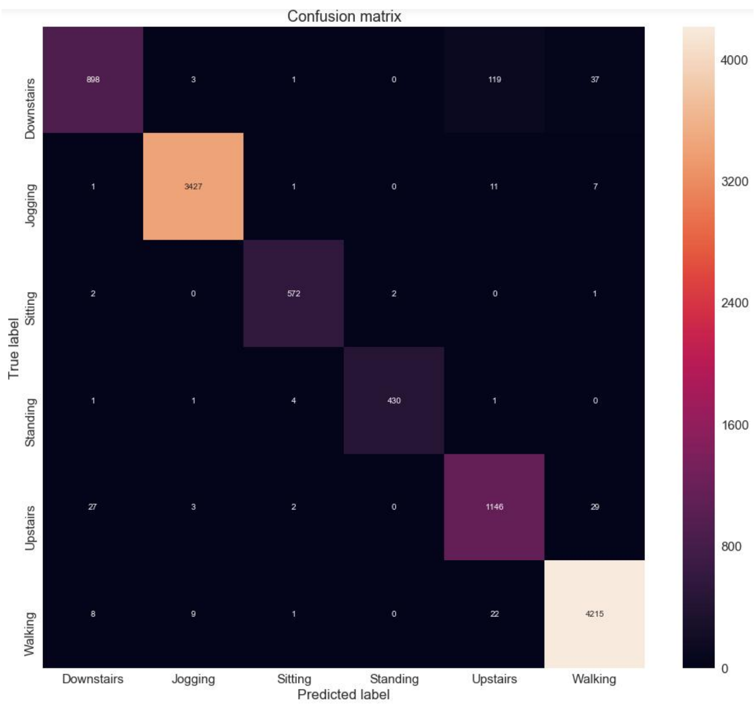


Figure 5 – confuse matrix of the double layer LSTM neural network model

In addition, we separately evaluated and tested models consisting of double FCL with the single–layer RNN, single–layer LSTM, single–layer GRU, double–layer RNN, double–layer LSTM, and double–layer GRU.

It was set that the network model composed of double FCL and double layer LSTM, which achieved an accuracy rate of 98.4% and a loss of 0.434 had the best HAR effect. The training process and the accuracy and loss curves obtained by the model on the test set are shown in Figure 6.

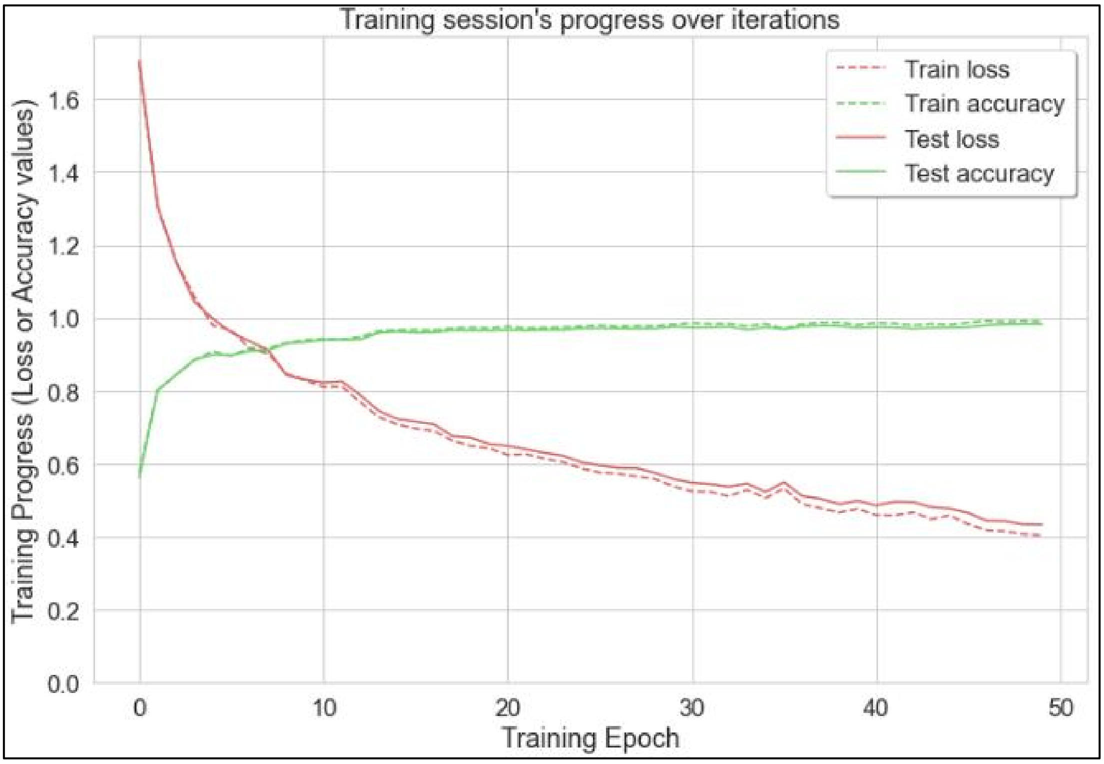


Figure 6 – Loss and accuracy of the double layer LSTM neural network model

CONCLUSION

The modeling result of this paper is based on the WISDM dataset, which data is obtained by volunteers putting the Android phone in the front trouser pocket to complete the 6 specified actions (Walking, Jogging, Upstairs, Downstairs, Sitting, Standing). In this paper, we mainly completed the following work:

1 We analyzed the WISDM dataset, as well as data preprocessing methods. In terms of data preprocessing, we used a 3rd–order 4 Hz Butterworth low–pass filter for filtering to remove high–frequency noise. In terms of feature generation, the sliding window technique is used to improve extraction of features from the filtered data. Furthermore, we digitize the features of action labels using one–hot encoding;

2 We performed the hyperparameter settings of the neural network model, including learning rate 0.0025, batch size 1024, and Adam optimizer;

3 We propose a human activity recognition neural network model with double layer LSTM network layers and double FCL. The algorithm transforms the input 3 features into 64 through FCL1, so as to better divide different types of data. The double–layer LSTM layer is used to extract the long–term and short–term dependencies between features, FCL2 is used to fuse 64–dimensional features into 6–dimensional, and the softmax layer converts the model prediction results into the probability set of each action category;

4 We set the learning parameters and hyperparameters of FCL1, double layer LSTM, and FCL2 respectively, as shown in Tables 2.3 and 2.4, including the following parameters:

– Learning parameters: FCL1 weight matrix [3,64], offset matrix [64]; double LSTM layer parameter quantity 66048; FCL2 weight matrix [64,6], offset matrix [6].

– Hyperparameters: The number of hidden units in FCL1 is 64, the activation function ReLU; the number of hidden units in the double LSTM layer is 64; the number of hidden units in FCL2 is 64;

5 We separately evaluated and tested the single–layer RNN, single–layer LSTM, single–layer GRU, double–layer RNN, double–layer LSTM, and double–layer GRU,

6 It was set that the network model composed of double FCL and double layer LSTM, which achieved an accuracy rate of 98.4% and a loss of 0.434 had the best HAR effect.

LIST OF AUTHOR’S PUBLICATIONS

1–A. Chen, Z. Y. Human physical activity recognition algorithm based on smartphone data and long short time memory neural network / Z. Y.Chen, Z. X. Yang, H. Li // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сборник научных статей IX Международной научно–практической конференции, Минск, 17–18 мая 2023 г. : в 2 ч. Ч. 1 / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: В. А. Богуш [и др.]. – Минск, 2023. – С. 21–28.

2–A. Yang, Z. C. Human physical activity recognition algorithm based on smartphone data convolutional nerual netwoek and long short time memory / Z. X. Yang, Z. Y.Chen // Международный научно–технический семинар «Технологии передачи и обработки информации»г. Минск, март – апрель 2023 г. / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: В.В. Чепикова [и др.]. – Минск, 2023. – Р. 102–107.

3–А. Chen, Z. Y. Human physical activity recognition algorithm based on smartphone data and long short time memory neural network / Z. Y.Chen, Z. X. Yang // 59–я научная конференция аспирантов, магистрантов и студентов БГУИР, Минск, 17–21 апреля 2023 г. / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: В.В. Рабцевич . – Минск, 2023. – Р. 160–162.