Wearable Wristworn Gesture Recognition Using Echo State Network

Weipeng Wang[1], Xiangpeng Liang[1], Maher Assaad[2], Hadi Heidari[1]
[1]Microelectronics Lab (meLAB), School of Engineering, University of Glasgow, G12 8QQ, UK
[2]Department of Electrical Engineering, Ajman University, Ajman, UAE
Hadi.Heidari@glasgow.ac.uk

***Abstract*—This paper presents a novel gesture sensing system for prosthetic limb control based on a pressure sensor array embedded in a wristband. The tendon movement which produces pressure change around the wrist can be detected by pressure sensors. A microcontroller is used to gather the data from the sensors, followed by transmitting the data into a computer. A user interface is developed in LabVIEW, which presents the value of each sensor and display the waveform in real-time. Moreover, the data pattern of each gesture varies from different users due to the non-uniform subtle tendon movement. To overcome this challenge, Echo State Network (ESN), a supervised learning network, is applied to the data for calibrating different users. The results of gesture recognition show that the ESN has a good performance in multiple dimensional classifications. For experimental data collected from six participants, the proposed system classifies five gestures with an accuracy of 87.3%.**

Keywords—Echo State Network; Gesture recognition; Force-Sensing Resistor; Human-machine interaction.

I. INTRODUCTION

Gesture makes up a large part of the body language and has become an efficient way for human-computer interaction. A large number of devices and approaches have been invented to do gesture recognition. Two majority categories for recent advances in gesture recognition is movement-based method and vision-based method [1]. In movement-sensor-based recognition, many sensors such as accelerometer, inertial sensor and electromyography (EMG) have been used to detect the movement of hand and classify the gestures [1-4]. The sensors of this approach have to be attached to the user's body. For vision-based method, the depth image or RGB image will be collected from the camera for extracting gestures [5]. Movement-based method can also be used in the sensing system of the prosthesis. In the sensing system of an active limb prosthesis, the physical movement detected by the sensors can be converted into the electrical signal [6].

A widely used wearable gesture sensing technique is measuring the EMG signal from the skin surface. A voltage difference on the electrode can be detected when the muscle contracted [7]. The sensor is precisely located on the skin and closed to the muscle on the forearm. In [8], Myo armband was used to collected EMG signal, the Support Vector Machine (SVM) was used to classified the feature extracted from the time-domain signal with 86.6% of accuracy. Other movement sensors such as accelerometer, gyroscope and inertial sensor are considered as good candidates in wearable hand gesture sensing. They can also be used in subtle finger gesture sensing. In [9], a smartwatch was used to detect five finger gestures and the accuracy was 87%. Four classification methods were used in classification, the classifier has the highest accuracy for each user is automatically selected. The advantage of this system is that the motion sensors are highly integrated. On the other hand, much more computation resources were consumed.

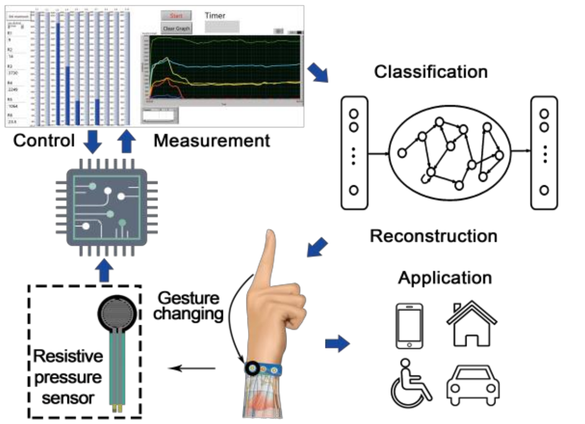

Fig. 1. The conceptual schematic for gesture recognition using resistive force sensors and ESN.

In this paper, we propose a movement-sensor-based gesture recognition method using Force Sensing Resistors (FSR) and Echo State Network (ESN). Fig. 1 illustrates the recognition method together with the data collection and potential applications form a gesture sensing system. The wristband embedded an FSR array which is able to detect the tendons movement around the wrist as a result of varying gestures [10, 11]. The pressure value can be measured and used for reconstructing the gesture. The difficult in gesture recognition using this method is that the tendons movement is relatively small, and there is no specific pattern for different people [12]. One solution to this problem is applying a calibration step by ESN learning algorithm before using.

Compared with the existing movement-based methods, this prototype using smaller size sensors with less power consumption. On the other hand, the wristband minimizes the area of on-body attachments, which provides a better experience for users. Compared with our previous work using five sensors and SVM [12-14], this prototype is upgraded to ten sensors for more accurately modelling the tendon movement and enhance the capability of distinguishing gestures. With more sensors used, the sensors are not required to put on a specific location on hand. Furthermore, this prototype applied ESN as the classifier, different from the widely used SVM. The reason is that considering a gesture will be posted and lasted for a certain time, the gesture data can be a time-dependent signal since the gesture would not be changed rapidly. Therefore, we choose ESN that has superior performance in multi-dimensional time-dependent signal to upgrade the wristband.

This paper is organized as follow: Section II introduced the overall system design. The principal of ESN and how it worked in our system is present in Section III. The experimental result, discussion and conclusion are provided in Section IV and Section V respectively.

II. SYSTEM DESIGN AND METHODOLOGY

A. Hardware Design

To detect the tendon movement around the wrist, ten FSRs are embedded in a wristband as a pressure sensor array. The change in pressure can be measured and transmitted into the computer via a microcontroller (Arduino DUE). The signal converted from changing pressure will be classified into different groups according to the network and pre-recorded training data[15].

FSR is a device that decreases in resistance when force increases in its active area. With a simple voltage divider circuit connected, FSR can be used as a sensor to detect pressure change. Fig. 2 shows the circuit where the FSR was connected with a 10kΩ resistor in series. When the resistance decreases, the voltage distributed on R increase so that the signal reading from the input port of the microcontroller increase. With ten sensors were used in this wristband, the layout of the sensors should avoid bones where the tendons movement can hardly be detected.

B. Software Design

The microcontroller can read the analogue signal from the voltage divider of the FSRs. The data acquisition module in Arduino DUE is a 12bit analogue-to-digital converter. Therefore, the data received by the computer yields a maximum value of 4095 and a minimum value of 0. The value from the sensors was read one by one at a sampling rate of 27Hz. The gesture sensing and data collection part are shown in Fig. 3. A LabVIEW program is used to receive the data from the microcontroller as data collection platform and user interface. The waveform graph illustrates the voltage versus time, and these data are stored for further processing.

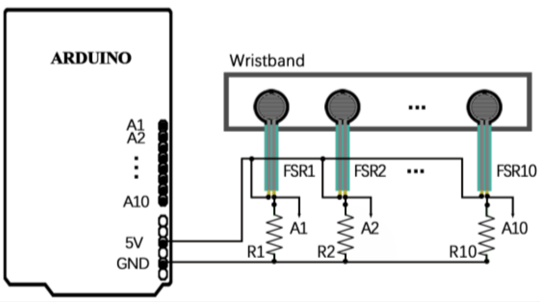

Fig. 2. Circuit for the sensors. *A1* to *A10* represent the analogue-to-digital converter ports on Arduino.

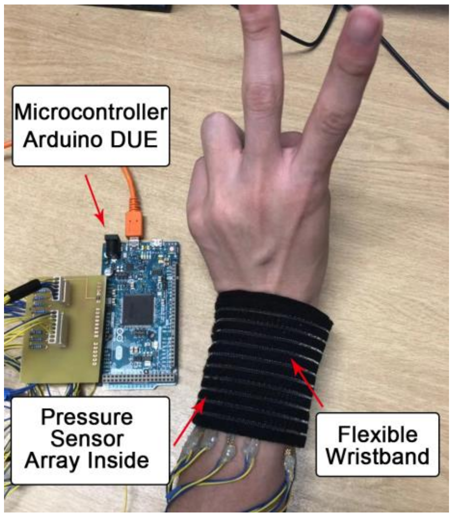

Fig. 3. Data collection using the wristband. The Pressure Sensor Array with ten sensors embedded transmit the data into the microcontroller.

III. ECHO STATE NETWORK DESIGN

ESN is a supervised machine learning model which can be classified into Recurrent neural networks (RNNs). This learning mechanism is suitable for time sequence prediction and classification. As shown in Fig. 4, the structure of ESN is composed of three parts: input layer, reservoir layer and output layer [16]. The characteristics for ESN are: (*I)* the recurrent and input weights are generated randomly and fixed; *(II)* training process only changes the weights connected to the reservoir and output layer using a linear regression algorithm. These make the ESN a fast approach in the training process.

A. Algorithm description

Assume an ESN with K channels as input (*u(n)*)and *x(n)* denotes the states in the reservoir that compose of N neurons (nonlinear node). The dimension of output (*y(n)*) is L. When data fed into the ESN, the state in the reservoir updated for every input using the equation

$$\boldsymbol{x}(n+1) = f(W_{in} \times \boldsymbol{u}(n+1) + W \times \boldsymbol{x}(n)) \quad (1)$$

where W_{in} is an N×K matrix and W is an N×N matrix which represents the weighted connection from the input layer to the reservoir, and every neuron in the reservoir respectively. Both of them are generated randomly and fixed in the whole training process. f is the activation function of the neuron where we take the data from the last state of the reservoir and new input and apply *tanh* function. The reservoir state and input itself should be stored and used in computing output weight for training the network.

In the training process, the desired target value is provided in the output layer. We assume that the reservoir state and input is linearly related to the output in the ideal case. After loading the training data set into the ESN, a linear relationship between the reservoir state and output can be found. W_{out} is an L×(N+K) matrix which represents the output weight. Previous state of neurons in the reservoir will also be used to find W_{out}.

The process of testing is a reverse of the last step in training. The output is calculated using a given input and known output weight:

$$\mathrm{y(n)} = W_{out} \times (x(n); u(n)) \tag{2}$$

y(n) will be an L dimensional data with the same length of testing input. The class of the output can be found by comparing which is the closest value to the target.

B. Training and recognition

The data collected from the sensors will be normalized before fed into ESN. We collected data from 5 gesture using 10 sensors, so there are 10 inputs and 5 outputs. The labels of the data were provided in one-of- five form. For each row of the label, only one element equal to one and the position indicates the gesture.

In the training process, initially, some data is sent to the reservoir. The neurons in the reservoir are set to 0, which may lead to the inappropriate output. We need some input to make the reservoir warm-up. This step makes sure the reservoir works under good condition. The training data set is introduced to the ESN where the signal in the input layer and target value in the output layer are provided. The first 50 input value and reservoir states are discarded in this experiment. The result of training is the output weight, W_{out}.

When new data feed into the trained ESN, the reservoir updates the same way in training. From the reservoir state, input and output weight, the output will be provided in five categories. According to our label method, we need to find out where the output has the value closest to 1. We use:

$$y_l = \min\left(|y(n) - 1|\right) \tag{3}$$

where y_l is calculated by comparing the similarity between prediction output and target value for each category. The ideal value for y_l is 0, so we find the smallest value to indicate the category for predicted result.

C. Performance evaluation

There are three factors that may affect the performance of the ESN: leaking rate, spectrum radius and reservoir size [17, 18]. However, it is difficult to analyses its performance since the ESN generates coefficient randomly, which means the accuracy of recognition is not a constant every time the network is initialized. To optimize the ESN, we stored the weight generated, then change the leaking rate and radius.

The leaking rate represents the speed of updating the reservoir. For every neuron in the reservoir, the leaking rate determined how much the value inherited from the previous state. A greater leaking rate yields better performance in fast-changing input data.

The spectrum radius in the ESN is the largest absolute value of the eigenvalue for the internal weight W. The network has an echo when the spectral radius is smaller than 1 [18]. To make sure the ESN works as it should, we normalized the internal weight:

$$\lambda = \max\left(eig(W)\right) \tag{4}$$

$$W' = \frac{W}{\lambda} \tag{5}$$

where W' denotes the internal weight after normalizing, we use this as internal weight.

Because the size of weight depends on the size of the reservoir, we cannot analyse the reservoir size by loading the stored coefficients. Generally, the accuracy increases when the reservoir has a bigger size, but it will slow down the speed of calculation since the reservoir needs to be updated for every input value. The size of the reservoir is 500 in this paper.

Another factor that should be taken into consideration is the regression equation. Ridge regression was applied here:

$$W_{out} = (X^T X + kI)^{-1} X^T y \tag{6}$$

where k is the ridge parameter, I represents the identity matrix, X denotes the stored reservoir and input states and y is the training labels. Generally, when we keep k within a small value (for example 10^{-6}), the output weight fit the target better. However, it may also lead to overfitting of the classifier, which means a sensitive model may be built because of a small number of unexpected error data [19]. To evaluate how the ridge parameter affects the performance, we test a range of value for k and compare their recognition accuracy. When k=10^{-6}, a smooth match can be found, and it fits most of the input datasets.

TABLE I. ACCURACY FOR GESTURE RECOGNITION

Person	*ESN Accuracy (%)*
#1	90.8
#2	96.0
#3	73.8
#4	89.7
#5	83.4
#6	86.5

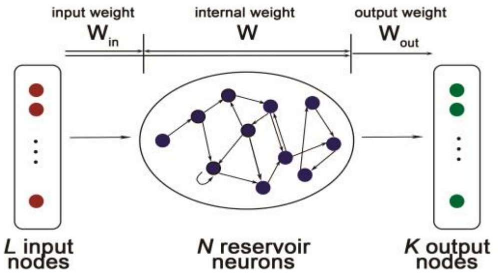

Fig. 4: Schematic diagram of ESN.

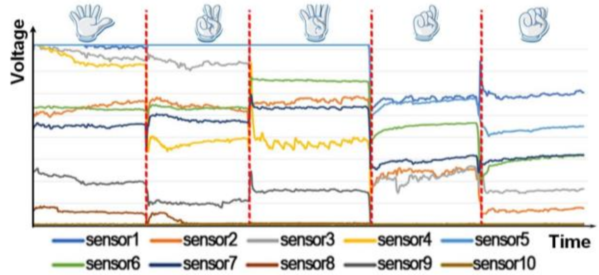

Fig. 5. Input signal change with gestures variations over time. In experimental data collection, participants were asked to show five gestures in 40 seconds.

IV. EXPERIMENTAL RESULT

The data were collected from six participants according to the proposed methodology. To imitate the real calibration and test process, two sets of data were collected from each person: one dataset for training and one for testing. The scenario of five gestures recognition was tested in this experiment, and the gestures were performed in a different order in two datasets for each person, as shown in Fig. 5.

The accuracy of recognition can be shown in a confusion matrix. Fig. 6 shows the result for recognition for all six participants. Since the ESN generates two random matrixes in the training process, the accuracy for the same dataset varies in the tests. To find the accuracy for the recognition, we repeated the training process for ten times and found their accuracy separately, followed by taking their average as the accuracy of recognition. The overall accuracy for all six participants is 86.7%. The accuracy of each person is shown in Table I.

From the confusion matrix, we can see that the misclassified data mainly appeared at Gesture 3,4 and 5, they are shown in the first three gestures in Fig. 5. Because of their similarity, some data on gesture 4 and 5 may be classified into gesture 3, but most of them can be classified into the correct class.

V. CONCLUSION AND FUTURE WORK

This paper demonstrates a novel method for gesture recognition using a wristband based on pressure sensor array which can detect the pressure change caused by tendons and muscles movement. To solve the problem that the value of the sensors varies from person to person, a neural network has been introduced to adapt to different users in the calibration step. The proposed device together with the data acquisition, machine learning and control, forms a wrist-worn gesture sensing system.

This system can provide us with a high accuracy in gesture recognition. However, the repeatability and variance of the recognition can be further improved. One potential improvement for this work is to make this system real-time. All possible gestures are collected and trained in the calibration step before using. Due to the fast computing speed, ESN can classify the data and reconstruct the gesture.

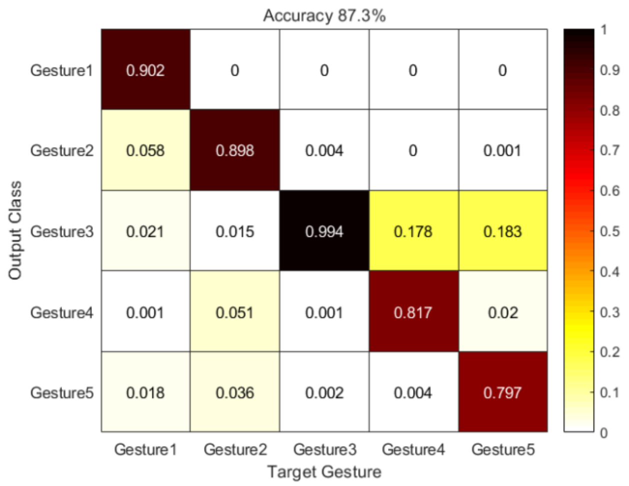

Fig. 6: Gesture recognition result for all participants, the accuracy for all gestures are 87.3%.

To improve the accuracy, future work of the system can be carried out on improving the stability of the pressure sensor array by using high-quality sensors or more reasonable layout accordingly. Another potential improvement is using other methods in classification and proposing a data fusion method to increase accuracy [14, 20].

REFERENCES

[1] J. Kela *et al.*, "Accelerometer-based gesture control for a design environment," *Personal and Ubiquitous Computing,* vol. 10, no. 5, pp. 285-299, 2006.

[2] H. Ghasemzadeh, R. Jafari, and B. Prabhakaran, "A body sensor network with electromyogram and inertial sensors: Multimodal interpretation of muscular activities," *IEEE transactions information technology in biomedicine,* vol. 14, no. 2, pp. 198-206, 2010.

[3] K. Liu, C. Chen, R. Jafari, and N. Kehtarnavaz, "Fusion of inertial and depth sensor data for robust hand gesture recognition," *IEEE Sensors Journal,* vol. 14, no. 6, pp. 1898-1903, 2014.

[4] H. Heidari, S. Zuo, A. Krasoulis, and K. Nazarpour, "CMOS magnetic sensors for wearable magnetomyography," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018: IEEE, pp. 2116-2119.

[5] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, "Robust part-based hand gesture recognition using kinect sensor," *IEEE transactions on multimedia,* vol. 15, no. 5, pp. 1110-1120, 2013.

[6] F. Cordella *et al.*, "Literature Review on Needs of Upper Limb Prosthesis Users," *Frontiers in Neuroscience,* vol. 10, no. 162, 2016.

[7] M. B. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: detection, processing, classification and applications (Correction)," *Biol Proced Online,* vol. 8, no. 1, p. 163, 2006.

[8] J. McIntosh, C. McNeill, M. Fraser, F. Kerber, M. Löchtefeld, and A. Krüger, "EMPress: Practical hand gesture classification with wrist-mounted EMG and pressure sensing," in *Proceed. CHI Conf. Human Factors in Computing Systems*, 2016, pp. 2332-2342.

[9] H. Wen, J. Ramos Rojas, and A. K. Dey, "Serendipity: Finger gesture recognition using an off-the-shelf smartwatch," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2016, pp. 3847-3851.

[10] A. Dementyev and J. A. Paradiso, "WristFlex: low-power gesture input with wrist-worn pressure sensors," in *Proceedings of the 27th annual ACM symposium on User interface software and technology*, 2014, pp. 161-166.

[11] J. Zhao, R. Ghannam, M. Yuan, H. Tan, M. Imran, and H. Heidari, "Design, test and optimization of inductive coupled coils for implantable biomedical devices," *Journal of Low Power Electronics,* vol. 15, no. 1, 2019.

[12] X. Liang, R. Ghannam, and H. Heidari, "Wrist-worn Gesture Sensing with Wearable Intelligence," *IEEE Sensors Journal,* 2018.

[13] X. Liang, H. Heidari, and R. Dahiya, "Wearable capacitive-based wrist-worn gesture sensing system," in IEEE *New Generation of CAS (NGCAS)*, 2017, pp. 181-184.

[14] X. Liang *et al.*, "Fusion of Wearable and Contactless Sensors for Intelligent Gesture Recognition," *Advanced Intelligent Systems,* p. 1900088, 2019.

[15] S. Zuo, K. Nazarpour, and H. Heidari, "Device Modeling of MgO-Barrier Tunneling Magnetoresistors for Hybrid Spintronic-CMOS," *IEEE Electron Device Letters,* vol. 39, no. 11, pp. 1784-1787, 2018.

[16] A. Rodan and P. Tino, "Minimum complexity echo state network," *IEEE trans. neural networks,* vol. 22, no. 1, pp. 131-144, 2011.

[17] D. Verstraeten, B. Schrauwen, M. d'Haene, and D. Stroobandt, "An experimental unification of reservoir computing methods," *Neural networks,* vol. 20, no. 3, pp. 391-403, 2007.

[18] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Computer Science Review,* vol. 3, no. 3, pp. 127-149, 2009.

[19] L. Appeltant, "Reservoir computing based on delay-dynamical systems," *These de Doctorat, Vrije Universiteit Brussel/Universitat de les Illes Balears,* 2012.

[20] H. Li *et al.*, "Multisensor data fusion for human activities classification and fall detection," in *IEEE SENSORS*, 2017, pp. 1-3.