2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A Wearable Finger-Tapping Motion Recognition System Using Biodegradable Piezoelectric Film Sensors

Shumma Jomyo[1], *Student Member, IEEE*, Akira Furui[1], *Member, IEEE*, Tatsuhiko Matsumoto[2], Tomomi Tsunoda[3] and Toshio Tsuji[1], *Member, IEEE*

***Abstract*— In this paper, we aimed to develop a method for the automatic recognition of individual finger-tapping motion. Biodegradable piezoelectric film sensors were attached to the skin of a forearm near the wrist (16 channels) to measure small movements of the tendons during five-finger tapping. In the proposed method, the segments in which motion occurred were detected by calculating the total activity for all channels. A neural network is trained to classify tapping motion using the extracted data based on the total activity, thereby allowing the accurate classification of flexion/extension of each finger. We collected experimental data from five healthy young adults to verify the motion recognition accuracy of the proposed method. The results revealed that the proposed method can recognize five-finger tapping motions with high accuracy (flexion/extension of each finger: 92.0%; time-series tapping motion: 88.4%).**

I. Introduction

The miniaturization of computers has facilitated the development of wearable devices that can easily be attached to the human body. In recent years, the market for devices such as smartwatches and smartglasses has expanded rapidly; however, as devices become smaller, device operability may be reduced. For example, it is difficult to input text on smartwatches because of their small displays. Voice input is an alternative; however, it tends to be difficult to use voice-based interfaces in public places such as on trains and buses. To solve these problems, human-machine interfaces (HMIs) have been developed to enable direct text input from body movements.

One such system was developed by training a convolutional neural network with hand images captured by a camera attached to the wrist to recognize gestures and tapping motion [1], [2], although the system could not be continuously worn on the wrist because of the device's large size. An alternative approach using electromyogram (EMG) has been used to recognize gestures [3] or to input text [4]; however, because EMG is susceptible to changes in signal characteristics due to muscle fatigue and sweat, EMG is not suitable for long-term use.

Polymeric piezoelectric materials are highly applicable to HMIs because of their thinness, ease of fabrication, high flexibility, and water resistance. In particular, biodegradable piezoelectric film sensors that are made of polylactic acid are well suited to the measurement of biological signals [5]. Therefore, with an HMI that uses biodegradable piezoelectric film sensors, it may be possible to achieve high device operability while retaining device wearability.

This paper proposes a method to recognize finger-tapping motion, such as that used for keyboard input, utilizing biodegradable piezoelectric film sensors attached to the wrist. In this approach, small movements of the underlying tendons are measured by 16 channels of piezoelectric film sensors. The proposed method automatically detects tapping motions based on the total activity of the entire channel and can recognize tapping motions using a neural network.

II. Tapping Motion Recognition Using Piezoelectric Film Sensors

An overview of the finger-tapping motion recognition method is shown in Fig. 1. Piezoelectric film sensors are used to capture the movement of tendons underlying the skin at the wrist, and a data logger was used to store the signals. Signals are filtered and normalized, then combined to calculate total activity. Segments during which finger tapping occurred are detected using a total activity threshold, and the finger motions are classified by machine learning.

A. Biodegradable Piezoelectric Film Sensor

The biodegradable piezoelectric film sensors (Picoleaf; Murata Manufacturing Co., Ltd.), made of polylactic acid, were attached to the skin of the wrist. This piezoelectric sensor is less affected by body temperature because it is nonpyroelectric; therefore, any charge is produced entirely by strain in the material, allowing more accurate measurement [5]. In addition, it is highly sensitive enough to can detect small changes in strain and flexible enough to can be attached to the contours of the human body; hence, it has the advantage of being able to detect bending and twisting [6]. Because of these characteristics, biodegradable piezoelectric is useful for biological applications such as measuring small displacement changes of the skin that are caused by the movement of underlying tendons.

B. Signal Measurement

Eight $20\,\mathrm{mm} \times 3\,\mathrm{mm}$ sensors, placed at $2\,\mathrm{mm}$ intervals and covered with silicone (Ecoflex 00–30; Smooth-On Inc.), were attached to each of the palmar and dorsal sides of the right wrist (Fig. 2). The tendons of the fingers are

This work was not supported by any organization

[1]S. Jomyo, A. Furui and T. Tsuji are with the Graduate School of Advanced Science and Engineering, Hiroshima University, Hiroshima, 739-8527 Japan (e-mail: shummajomyo@hiroshima-u.ac.jp).

[2]T. Matsumoto is with Murata Manufacturing Co., Ltd., Kyoto, 617-8555 Japan.

[3]T. Tsunoda is with Murata Manufacturing Co., Ltd., Kyoto, 617-8555 Japan and also with the Graduate School of Advanced Science and Engineering, Hiroshima University, Hiroshima, 739-8527 Japan.

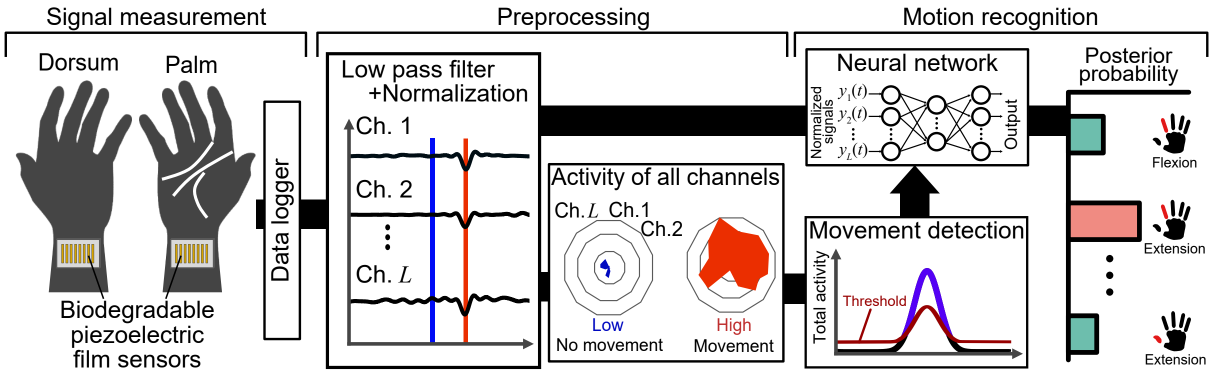
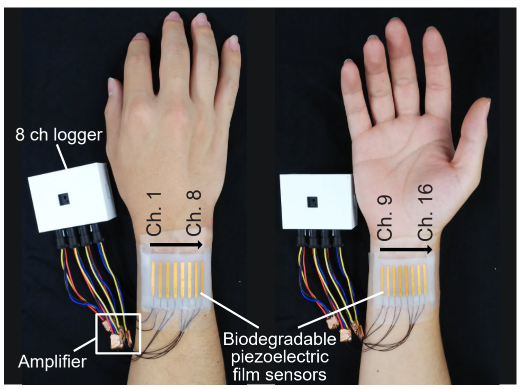
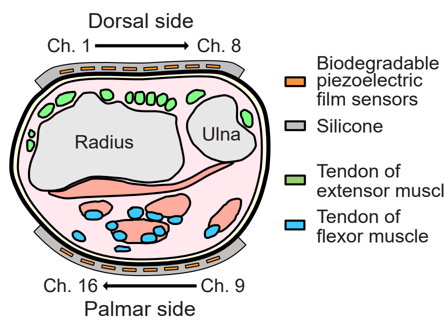


Fig. 1. Overview of the finger-tapping motion recognition method



(a)



(b)

Fig. 2. Signal measurement unit. (a) Placement on the right wrist. (b) Cross-sectional diagram of the right wrist with the piezoelectric sensors attached on the skin.

distributed near the surface, running along the palmar and dorsal forearms to the wrist [7], and each tendon moves to produces a movement in the corresponding finger. The system can capture finger motions by measuring such a movement of tendons at the wrist from the skin surface. The signal obtained from each channel is amplified ($\times$ 54), converted from analog to digital (16-bit resolution) by two data loggers (wireless 8-channel logger; Logical Products, Inc.) and saved to a PC (sampling frequency f_s (Hz)).

C. Preprocessing

Preprocessing is performed on each channel prior to calculating the total activity, which represents motion magnitude. The DC component is removed from each measured signal by subtracting the average value of the full time series. The signal is then filtered by a digital Nth-order low-pass Butterworth filter (cutoff frequency: f_c (Hz)) to extract the low-frequency components related to the displacement of the skin surface involved in tendon movements. Each signal $x_l(t)$ ($l = 1, 2, \cdots, L$; L is the number of channels) at time t is normalized as

$$y_l(t) = \frac{x_l(t) - \mu_l}{\sigma_l}, \tag{1}$$

where μ_l and σ_l are the mean and standard deviation, respectively, of each channel.

To classify movement as that of a specific finger motion from the set of signals, it is necessary to detect when the tapping motion occurs. We create a radar chart by radially arranging the absolute values of $y_l(t)$, and an area of this radar chart is calculated and defined as the total activity $s(t)$:

$$s(t) = \frac{1}{2}\sin\frac{2\pi}{L}\left|\sum_{l=1}^{L} y_l(t)y_{l+1}(t)\right|, \tag{2}$$

where $y_{L+1}(t) = y_1(t)$. During tapping, the larger value of the total activity $s(t)$ because the internal area of the radar chart expands. This total activity is used to determine when the tapping motion occurs.

D. Motion Recognition

In the proposed method, each finger's flexion and extension are defined as two different classes, and a neural network is used to classify tapping motions with the normalized signals $y_l(t)$ as input.

1) Training data creation: To classify each motion accurately, it is necessary to correctly extract a series of $y_l(t)$ during the tapping motion from the measured data, and use it as training data for the neural network. Therefore, training data for the neural network are created; segments in which the tapping motion occurred are extracted by fitting a Gaussian function to the total activity $s(t)$.

In the interval $[t_1, t_2]$ containing a single tapping motion, the time-series waveform $s(t)$ has two peaks corresponding

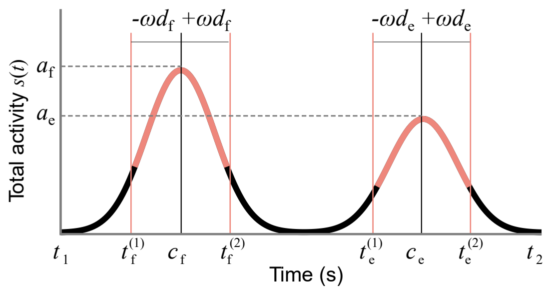


Fig. 3. Signal segmentation based on ω where the total activity $s(t)$ during a single tapping motion. The first and second peaks correspond to the finger flexion and extensions, respectively. The orange lines indicate the motion segments extracted using the Gaussian function fitting.

to finger flexion and extension, respectively. A function consisting of the sum of the two Gaussian functions,

$$R(s(t)) = a_{\mathrm{f}}\exp\left[-\frac{(s(t)-c_{\mathrm{f}})^2}{2d_{\mathrm{f}}^2}\right] + a_{\mathrm{e}}\exp\left[-\frac{(s(t)-c_{\mathrm{e}})^2}{2d_{\mathrm{e}}^2}\right] \quad (t_1 \leq t \leq t_2), \tag{3}$$

is fitted to the total activity segments using the least-squares method, and a_i, c_i, and d_i are respectively the peak amplitude, location, and spread of each Gaussian function, and $i \in \{\mathrm{f}, \mathrm{e}\}$ is the indicator variable that specifies the flexion and extension. We define the range of $\pm\omega d_i$ around each peak c_i as the segments of finger flexion $[t_{\mathrm{f}}^{(1)},\ t_{\mathrm{f}}^{(2)}]$ and extension $[t_{\mathrm{e}}^{(1)}, t_{\mathrm{e}}^{(2)}]$ intervals (Fig. 3). The normalized signals $y_l(t)$ corresponding to these motion segments are extracted and used as training data for the neural network pattern classifier to learn the characteristics of the flexion and extension of the tapping motions.

The above procedure is performed for all tapping movements, and we created a total of J training data sets. Each set of training data contained at least one flexion and extension movement for each finger.

2) Time-series tapping recognition: To classify specific tapping motions in time-series data, the normalized signal $y_l(t)$ is used as input to a multilayer perceptron neural network, and the class with the maximum prediction score is determined as the classified motion. Here, to prevent misclassifications due to small noises generated at rest and unpredictable noises such as body movements, we define the threshold $s_{\mathrm{th}}(t)$ based on the total activity $s(t)$ as

$$s_{\mathrm{th}}(t) = k(s(t) + \bar{s}_{\mathrm{th}}), \tag{4}$$

$$\bar{s}_{\mathrm{th}} = \frac{1}{J}\sum_{j=1}^{J} s_{\mathrm{min}}^{(j)}, \tag{5}$$

where k is an arbitrary gain and $s_{\mathrm{min}}^{(j)}$ is the minimum value of $s(t_i^{(1)})$ in j-th dataset $(j = 1, 2, \cdots, J)$. The classification is only performed when $s(t)$ is satisfied $s(t) \geq s_{\mathrm{th}}(t)$. Here, this conditional expression can be reduced to the following condition:

$$s(t) > \kappa\bar{s}_{\mathrm{th}}, \tag{6}$$

where

$$\kappa = \frac{k}{1-k}, \tag{7}$$

and accordingly, $0 \leq k < 1$.

Network outputs may involve instantaneous misclassification as a result of artifacts that will eventually affect the stability of classification. Therefore, to mitigate these effects, at a given time t, the classification result is adopted if the network outputs during the segment $[t-p, t]$ are the same label.

Although for each finger the neural network classifies movement into two separate classes—flexion and extension—, the flexion and extension of each finger should be performed in a set when inputting text on a keyboard. Therefore, the tapping motion is only recognized when flexion and extension movements of the same finger are classified consecutively. This condition reduces the influence of instantaneous misclassification and enables stable outputs for the recognition of tapping motion.

III. Experiments

To verify the effectiveness of the proposed method, experiments were conducted on five healthy young adults (males, right-handed, age range: 23–25 years). Study procedures were carried out in accordance with the Declaration of Helsinki. Informed consent was obtained from all subjects before the experiments were performed, and the study was approved by the Hiroshima University Ethics Committee (Registration number: E-840). A total of 16 channels were captured from biodegradable piezoelectric film sensors attached on the dorsal side (channels 1–8) and the palmar side (channels 9–16) of the right wrist of each subject. The sampling frequency was set to $f_s = 200\,\mathrm{Hz}$, and the order and cutoff frequency of the low-pass filter were set to $N = 12$ and $f_c = 10\,\mathrm{Hz}$, respectively. The number of hidden layers in the neural network was set to one, and the number of units in the hidden layer was set to 16. The parameter ω, which determines the range of waveform extraction during neural network training, was set to 0.5 based on the results of a pilot experiment. The threshold multiplier k was set to 0.5.

Subjects were asked to sit and place their right hand comfortably on a desk. We showed the subjects a video that demonstrated the timing of the tapping and had them perform a task in which they tapped each of their five fingers once in random order; each subject performed ten trials of the task. The interval between each tap was set to $1\,\mathrm{s}$, and the measurement time per trial was set to $30\,\mathrm{s}$. Subjects were instructed to keep their wrists on the desk during the trials. The position where all fingertips were raised was set as the neutral position, and the tapping was defined as the movement of finger flexion to touch the desk followed by finger extension to return to the neutral position.

First, we examined classification accuracy for the flexion and extension of each finger. The 10 target classes were

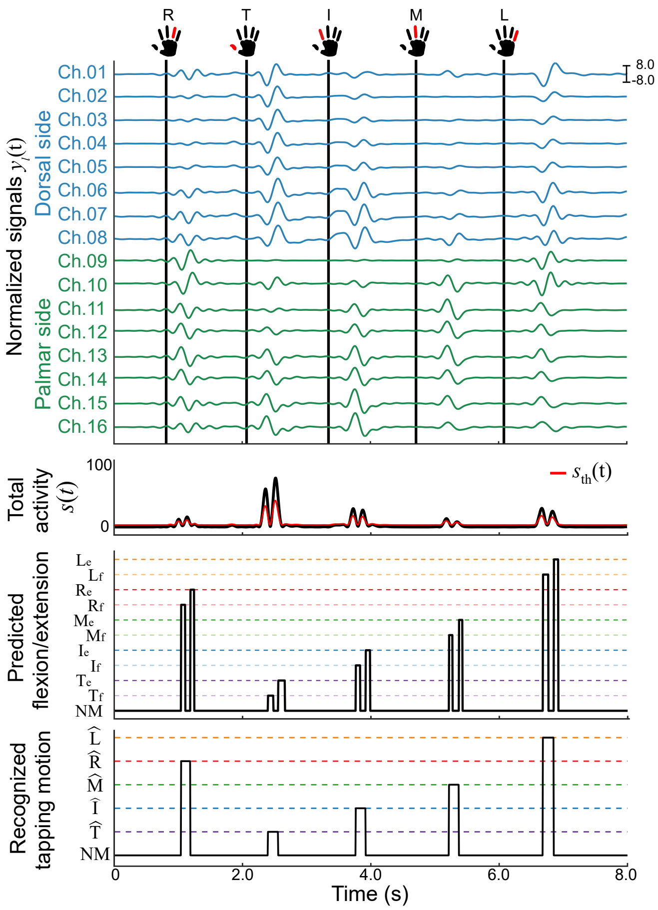


Fig. 4. Results of time-series recognition of tapping motions. From the top, normalized signals (Ch. 1–8: dorsal, Ch. 9–16: palmar side), the total activity of all channels $s(t)$, time-series prediction of flexion/extension, and recognized tapping motions are shown. Note that NM represents the no motion.

flexion and extension each of the thumb (T_i), index finger (I_i), middle finger (M_i), ring finger (R_i), and little finger (L_i), where $i \in \{\mathrm{f}, \mathrm{e}\}$. Five out of ten trials for each subject were used as training data (trials 1–5), and the remaining five trials (trials 6–10) were used for accuracy verification. For all combinations of trials, the average classification accuracies were calculated. Note that we extracted the interval data from the motion segments in advance and used them for classification.

Next, to confirm the applicability of our method for text input tasks, we verified the recognition accuracy of each specific tapping movements on time-series data: $\widehat{\mathrm{T}}$, thumb; $\widehat{\mathrm{I}}$, index; $\widehat{\mathrm{M}}$, middle; $\widehat{\mathrm{R}}$, ring; and $\widehat{\mathrm{L}}$, little. In this verification, the training data (trial 1–5th) were extracted from the motion segments in advance, and the test data (trial 6–10th) were time-series data including the resting periods.

As an evaluation measure for time-series recognition, we calculated the tapping recognition accuracy based on character recognition accuracy [8], $1-(N_{\mathrm{err}}/N_{\mathrm{total}})$, where N_{err} is the total number of misrecognized tapping (the sum of the number of replacements, insertions, and deletions) and

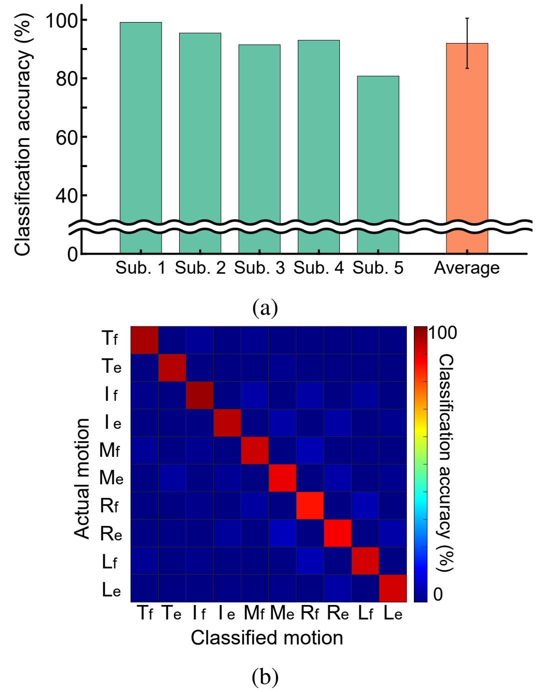


Fig. 5. Results of classification accuracy for the five-finger motions of flexion and extension. (a) Average classification accuracy. The green bars are average accuracy of Sub. 1–5 and the orange bar is the average over all subjects. The error bar in the average represents the 95% confidence interval for all subjects. (b) Average confusion matrix.

N_{total} is the total number of taps. With this scale, we can quantitatively evaluate whether the system can recognize the tapping movements in the correct order.

IV. Results

Fig. 4 shows an example of the measurement and recognition results of subject 1. From the top, the normalized signals $y_l(t)$, the total activity $s(t)$ with the threshold $s_{\mathrm{th}}(t)$, the predicted results of flexion/extension of each finger, and the recognized movement are presented. Note that the vertical black line in the figure represents the start time of each tap presented to the subject, and the actual time when the subject performed the tapping motion was approximately 0.1–$0.5\,\mathrm{s}$ later than this line. In this example, the tapping order was ring finger, thumb, index finger, middle finger, and little finger.

Fig. 5(a) shows the average classification accuracy over verification trials for each subject. The average classification accuracy over all subjects is also shown. All subjects had classification accuracies greater than 80%, and the average accuracy over all subjects was 92.0% (95% confidence interval [CI]: 83.4–100.0). Fig. 5(b) shows the confusion matrix for the average classification accuracy of all subjects.

Fig. 6 shows the tapping recognition accuracy for the time series data. The average recognition accuracy for all subjects was 88.4% (95% CI: 76.3–100.0).

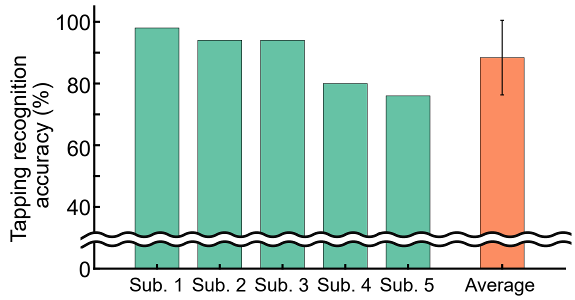


Fig. 6. Tapping recognition accuracy for the time-series data of tapping motions. The error bar in the average represents the 95% confidence interval for all subjects.

V. Discussion

Piezoelectric sensors attached to the wrist captured waveforms with different characteristics for each tapping movement (Fig. 4). In particular, dorsal channels represented thumb and little-finger tapping, palmar channels represented middle-finger and ring-finger tapping, and the entire channel represented index-finger tapping. Two peaks corresponding to flexion and extension were observed in the total activity $s(t)$ after the start time of the tapping motion of each finger; therefore, we can conclude that the method can detect the tendon movements corresponding to each finger's tapping.

The overall classification accuracy for tapping motion was high, greater than 90% on average (Fig. 5(a)). This indicates that the neural network could correctly learn the characteristics of the flexion and extension of each finger and that the proposed method could accurately extract intervals with motion for the training of the classifier. The overall accuracies for the index finger and the thumb were relatively high, while those for the middle, ring, and little fingers tended to be low. Anatomically, the thumb and index finger are independent, whereas the middle, ring, and little fingers are difficult to move completely independently [9]. This difference in the independence of each finger may have affected the classification accuracy. If a model considering finger dependency could be introduced into the classifier, the classification accuracy of the middle, ring, and little fingers may be improved.

For time-series recognition of each finger tapping motion, the average accuracy reached higher than 88% (Fig. 6), which suggests that the proposed method is effective even in situations where virtual keyboard interface is assumed. Although recognition accuracies for some subjects were relatively low, they can be expected to improve with sufficient training. The piezoelectric film sensor that was used is thin, flexible, lightweight, and less susceptible to the effects of muscle fatigue and sweat than EMG sensors; therefore, an HMI based on our method has the potential for application in practical wearable devices. However, there were some cases in which the tapping recognition accuracy was reduced by incorrect classification, most likely as a result of transient signals. This may be improved by optimizing the threshold for motion determination and using time-series classification models.

VI. Conclusions

This paper proposed a novel method for finger-tapping motion recognition that uses biodegradable piezoelectric film sensors to detect finger motions from the skin at the wrist, which is used as input to a neural network to recognize tapping motion. The experiments with five subjects demonstrated that the proposed method can accurately recognize tapping motion, suggesting its applicability for virtual keyboard interfaces.

In the future, we would like to adapt our method to interfaces in actual online environments. In addition, we will verify classification accuracy when tapping position and speed vary and when unknown subjects are used for recognition.

References

[1] H.-S. Yeo *et al.*, "Opisthenar: Hand poses and finger tapping recognition by observing back of hand using embedded wrist camera," in *Proc. 32nd Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 2019, pp. 963–971.

[2] F. Hu *et al.*, "Fingertrak: Continuous 3D hand pose tracking by deep learning hand silhouettes captured by miniature thermal cameras on wrist," in *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 4, no. 2, jun. 2020.

[3] X. Zhang *et al.*, "A framework for hand gesture recognition based on accelerometer and EMG sensors," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 41, no. 6, pp. 1064–1076, Nov. 2011.

[4] Z. Fu *et al.*, "Typing everywhere with an EMG keyboard: A novel myo armband-based HCI tool," in *Proc. Int. Conf. Algorithms Architectures Parallel Process.*, 2020, pp. 247–261.

[5] M. Ando *et al.*, "Pressure-sensitive touch panel based on piezoelectric poly(l-lactic acid) film," *Jpn. J. Appl. Phys.*, vol. 52, no. 9S1, p. 09KD17, Sept. 2013.

[6] M. Ando *et al.*, "Film sensor device fabricated by a piezoelectric poly(L-lactic acid) film," *Jpn. J. Appl. Phys.*, vol. 51, p. 09LD14, Sept. 2012.

[7] M. D. Miller *et al.*, *Orthopaedic Surgical Approaches*. Saunders, Aug. 2014.

[8] P. Lech and K. Okarma, "Prediction of the optical character recognition accuracy based on the combined assessment of image binarization results," *Elektron. ir Elektrotechnika*, vol. 21, pp. 62–65, Dec. 2015.

[9] C. Häger-Ross and M. H. Schieber, "Quantifying the independence of human finger movements: Comparisons of digits, hands, and movement frequencies," *J. Neurosci.*, vol. 20, no. 22, pp. 8542–8550, 2000.