



Cite as  
Nano-Micro Lett.  
(2025) 17:41

Received: 1 June 2024  
Accepted: 23 September 2024  
© The Author(s) 2024

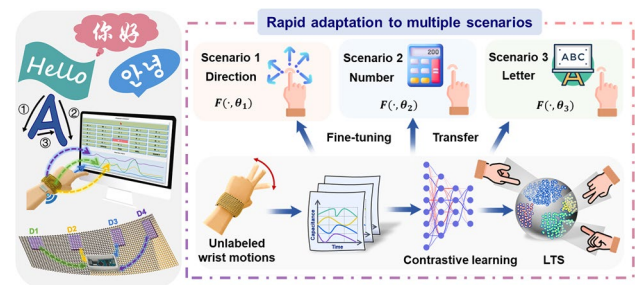
# A Rapid Adaptation Approach for Dynamic Air-Writing Recognition Using Wearable Wristbands with Self-Supervised Contrastive Learning

Yunjian Guo<sup>1</sup>, Kunpeng Li<sup>1</sup>, Wei Yue<sup>2,3</sup>, Nam-Young Kim<sup>2,3</sup>, Yang Li<sup>4,5</sup> ✉, Guozhen Shen<sup>6</sup> ✉, Jong-Chul Lee<sup>1</sup> ✉

## HIGHLIGHTS

- Utilizing self-supervised learning, the proposed wearable wristband with a four-channel sensing array and wireless transmission module is developed for tracking air-writing and dynamic gestures.
- The model can learn prior features from unlabeled signals of random wrist movements, significantly reducing the dependency on extensive labeled data for training.
- The wristband system rapidly adapts to multiple scenarios after fine-tuning using few-shot data, enhancing user interaction through natural and intuitive communication.

**ABSTRACT** Wearable wristband systems leverage deep learning to revolutionize hand gesture recognition in daily activities. Unlike existing approaches that often focus on static gestures and require extensive labeled data, the proposed wearable wristband with self-supervised contrastive learning excels at dynamic motion tracking and adapts rapidly across multiple scenarios. It features a four-channel sensing array composed of an ionic hydrogel with hierarchical microcone structures and ultrathin flexible electrodes, resulting in high-sensitivity capacitance output. Through wireless transmission from a Wi-Fi module, the proposed algorithm learns latent features from the unlabeled signals of random wrist movements. Remarkably, only few-shot labeled data are sufficient for fine-tuning the model, enabling rapid adaptation to various tasks. The system achieves a high accuracy of 94.9% in different scenarios, including the prediction of eight-direction commands, and air-writing of all numbers and letters. The proposed method facilitates smooth transitions between multiple tasks without the need for modifying the structure or undergoing extensive task-specific training. Its utility has been further extended to enhance human-machine interaction over digital platforms, such as game controls, calculators, and three-language login systems, offering users a natural and intuitive way of communication.



**KEYWORDS** Wearable wristband; Self-supervised contrastive learning; Dynamic gesture; Air-writing; Human-machine interaction

Yunjian Guo, Kunpeng Li and Wei Yue contributed equally to this article.

✉ Yang Li, [yang.li@sdu.edu.cn](mailto:yang.li@sdu.edu.cn); Guozhen Shen, [gzshen@bit.edu.cn](mailto:gzshen@bit.edu.cn); Jong-Chul Lee, [jclee@kw.ac.kr](mailto:jclee@kw.ac.kr)

<sup>1</sup> Department of Electronic Convergence Engineering, Kwangwoon University, Seoul 01897, South Korea

<sup>2</sup> Radio Frequency Integrated Circuit (RFIC) Bio Centre, Kwangwoon University, Seoul 01897, South Korea

<sup>3</sup> Department of Electronic Engineering, Kwangwoon University, Seoul 01897, South Korea

<sup>4</sup> School of Microelectronics, Shandong University, Jinan 250101, People's Republic of China

<sup>5</sup> State Key Laboratory of Integrated Chips and Systems, Fudan University, Shanghai 200433, People's Republic of China

<sup>6</sup> School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing 100081, People's Republic of China

Published online: 16 October 2024



SHANGHAI JIAO TONG UNIVERSITY PRESS

Springer

## 1 Introduction

Gesture recognition, acknowledged as an intuitive and natural mode of communication, interprets intentional hand movements to convey significant information and has garnered substantial attention in the field of human–machine interaction [1–3]. Common techniques for capturing hand movements include image recognition, radar systems, and wearable technology [4–6]. Bulky devices such as high-resolution cameras, accelerometers, or radar systems are not suitable for daily wear [7–9]. In contrast, wearable devices offer an attractive alternative for monitoring hand movements and intentional gestures, as they can be seamlessly integrated into various accessories [10–12]. The choice of device placement significantly impacts both wearing comfort and the effectiveness of data acquisition [13, 14]. Given that most tendons and muscle groups responsible for hand movements are located beneath the wrist skin, wristbands offer an optimal placement option compared to positioning devices on the fingers or the back of the hand [15], providing high-sensitivity devices with the opportunity to precisely track subtle movements. The direct mapping of specific gestures has been widely recognized and developed. Wang et al. proposed a gesture recognition wristband by integrating a triboelectric nanogenerator and a piezoelectric nanogenerator, achieving a maximum accuracy of 92.6% in recognizing 26 letters [16]. Similarly, Wu et al. deployed seven triboelectric nanogenerator sensors into a smart wristband, successfully classifying 21 hand motions and enabling wireless control through air gestures [17]. These wristband-integrated systems enable static gesture recognition through specific finger gestures. However, the one-to-one mapping between specific gestures and information restricts the conveyable data, and excessive correspondence places a significant burden on the user. Therefore, there is an urgent need to develop intuitive mapping rules and recognition systems that align with user habits and cognitive processes.

Dynamic gestures based on air-writing utilize direct hand-writing mapping rules for characters, which can optimize user experience and enhance information density [18–20]. Air-writing involves tracing letters or numbers by moving hands or fingers in free space to form a virtual text interface. It has proven valuable in scenarios where conventional writing methods are impractical, especially in translating sign

language, improving experiences in augmented reality or virtual reality, and facilitating various gesture-driven interfaces [21–23]. For example, Liu et al. utilized a stretchable and conductive hydrogel strain sensor to recognize the entire dynamic process of air-writing 26 English letters [24]. This approach of recognizing multi-stroke characters significantly enhances the richness of dynamic information by tracking the spatial writing process while preserving the user's natural habits [25].

However, the diversity in character forms and writing styles presents significant challenges for the recognition process. Recent advancements in deep neural networks for processing time-series or array data from wearable devices have opened up new possibilities for feature extraction and classification in complex tasks [26–28]. In combination with convolutional neural network algorithms, Li et al. designed a virtual text-entry interface using gloves integrated with unimodal sensors for finger air-writing applications involving letters and numbers [29]. Although these algorithms have the capability to analyze and learn gestures captured by wearable devices, they are constrained by conventional supervised learning methods, which heavily rely on extensive manually labeled data for single-task execution, making them both time-consuming and labor-intensive [30–33]. Adapting these systems to new users or incorporating new scenarios requires the collection of large amounts of new labeled data or modifications to the model architecture. This raises development and maintenance costs, and extends system deployment time, underscoring their limitations in generalizability and scalability. Consequently, improving data processing efficiency and optimizing learning and training procedures to enhance user experience in air-writing systems is a significant challenge.

This study presents the development of a wearable fiber wristband designed for dynamic gesture recognition in human–computer interactions utilizing the time-series cross-view fusion contrastive (TS-VFC) learning algorithm, as depicted in Fig. 1. This general learning framework can rapidly adapt to multiple scenarios without requiring extensive labeled data collection. The wristband, meticulously designed to accommodate the anatomical and muscular dynamics of the wrist, features an array of four flexible iontronic devices, coupled with a Wi-Fi module for wireless communication, ensuring

comfort during prolonged daily use. Each device consists of ultrathin electrodes (25  $\mu\text{m}$ ) made from silver nanowires (AgNWs)/improved polyvinyl alcohol (PVA) and a hydrogel dielectric layer with microcone structures (Fig. S1). Consequently, its output signal exhibits high sensitivity and precision in capturing subtle gestures. The TS-VFC learning method collects unlabeled random motion data from the wrist, creating a latent time space (LTS) that encapsulates prior features. It enables the model to be fine-tuned for new users or scenarios with few-shot labeled data, thus eliminating the need for extensive training or model redesign. Our experiments demonstrate that transfer learning, with just 5-shot labeled data, can achieve accuracy rates of up to 94.9%. This capability allows the pre-trained model to rapidly adapt to different gesture recognition tasks, including directional movements and air-writing of numbers and letters. Moreover, we showcased practical applications such as game control, calculator operation, and login systems, underscoring the practicality and potential for human–machine interaction. This wristband system effectively translates hand gestures into digital commands, making it a promising candidate to facilitate the interaction of diverse electronic interfaces.

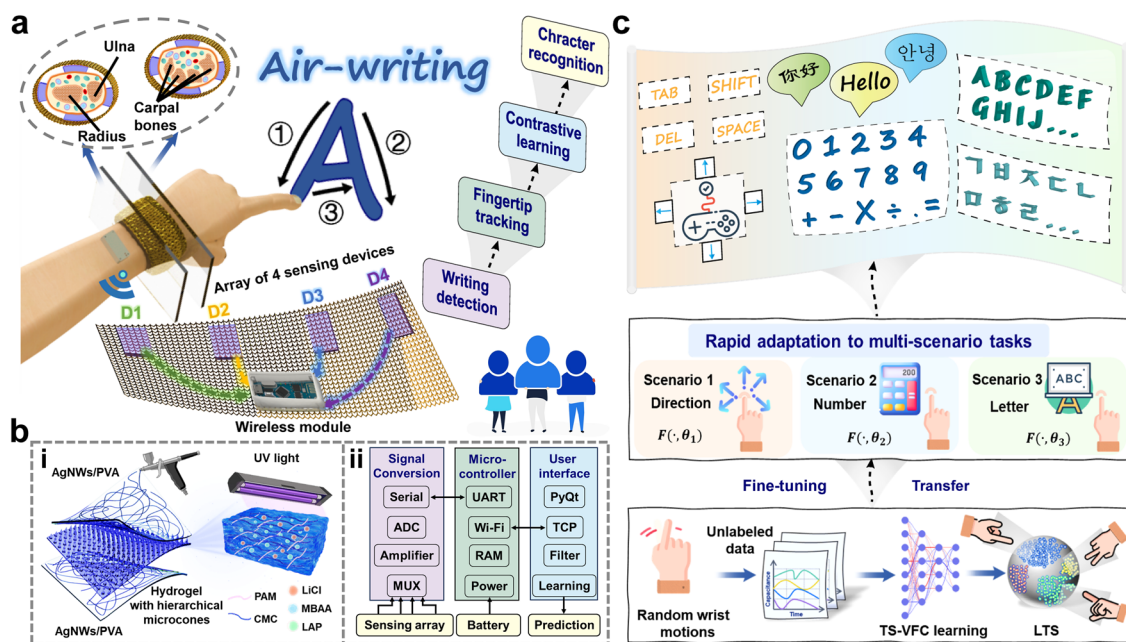
## 2 Experimental Section

### 2.1 Materials

Acrylamide (AM, 99%), sodium carboxymethyl cellulose (CMC), N,N'-methylenebisacrylamide (MBAA, 99%), polyethylene glycol diacrylate (PEGDA), lithium chloride (LiCl, 99%), diphenyl(2,4,6-trimethylbenzoyl)phosphine oxide (TPO) and ethylene glycol (EG) were procured from Sigma-Aldrich. The photoinitiator used was lithium phenyl(2,4,6-trimethylbenzoyl) phosphinate (LAP) obtained from the Tokyo chemical industry, Tokyo, Japan. polydimethylsiloxane (PDMS) precursor and curing agent (Sylgard 184) were purchased from Dow Corning. AgNWs dispersions (DT-AGNW-N30-EOH) were supplied by Ditto Technology Co., Ltd. (South Korea). Improved PVA was purchased from Zhuzhou Shifeng New Blue Sky Materials Co., Ltd. (China). All chemical reagents were analytical grade and used without further purification.

### 2.2 Preparation of Iontronic Device

The device was fabricated by following the procedure outlined in Fig. S1. Initially, a resin mold featuring hierarchical microcones was created using a Stereolithography 3D printer (Formlabs Form 3). Subsequently, the PDMS precursor and curing agent were coated onto the mold at a weight ratio of 10:1 and cured at 80  $^{\circ}\text{C}$  for 1 h. The resulting cured PDMS, with complementary patterns, was then peeled off and utilized as the second mold for the hydrogel. The hydrogel was prepared using AM as the monomer, CMC as the nanofiller, MBAA and PEGDA as crosslinkers, LiCl as conductive salts, and TPO as the photo-initiator. Specifically, the AM (16 wt%), CMC (2 wt%), MBAA (0.3 wt%), PEGDA (0.3 wt%), EG (5 wt%), and TPO (0.1 wt%) were dissolved in a 6 M LiCl aqueous solution at room temperature. Both AM and CMC were used for the hydrogel polymer networks, and the inclusion of EG was beneficial for enhancing the dehydration resistance. EG not only protects the polymer interactions by forming stable ion clusters with  $\text{Li}^+$ , effectively reducing the hydration of the lithium salt, but also enhances the mechanical and electrical stability of the hydrogel system by forming hydrogen bonds with the AM/CMC network. Subsequently, the solution was transferred to the second PDMS mold, which was exposed to UV light (365 nm) for 6 min to polymerize into a cross-linked ionic hydrogel with hierarchical microcones. A commercial AgNWs dispersion was used for the flexible electrodes. The AgNWs were dispersed in ethanol at a concentration of 0.01  $\text{mg mL}^{-1}$ . The typical diameter and length of AgNWs are 20–40 nm and 10–20  $\mu\text{m}$ , respectively. Before use, the solution was stirred vigorously for 3 min until the nanowires were evenly dispersed in the solution. The AgNWs solution was then sprayed onto a 25  $\mu\text{m}$  improved PVA substrate using an air spray gun and left to air dry until the solvent had completely evaporated, resulting in the formation of AgNWs-coated PVA electrodes. As illustrated in the cross section of the SEM image in Fig. S1, the ultrathin electrode effectively converts pressure into capacitance changes with high sensitivity. This improved PVA material remains stable at room temperature and is durable under daily human activities and exposure to sweat, exhibiting solubility only in hot water (above 65  $^{\circ}\text{C}$ ), as shown in Fig. S2. Finally, the iontronic sensing device was assembled by sandwiching the ionic hydrogel with hierarchical microcones between two flexible AgNWs/PVA electrodes. This iontronic device



**Fig. 1** Illustration of the wearable wristband and air-writing prediction process. **a** Wristband containing four sensing devices (D1–D4) equipped with a wireless Wi-Fi module for air-writing recognition. **b** (i) Exploded view of the design layout of iontronic device, including AgNWs/PVA electrodes and photocurable ionic hydrogel with hierarchical microcones. (ii) Brief block diagram of the wristband system and the customized user interface. **c** Real-time prediction and display of air-writing through TS-VFC learning. Rapid adaptation process of directions, numbers, and letters: learning prior features from random wrist movements as LTS via TS-VFC learning, and fine-tuning with few-shot labeled data for rapid adaptation to diverse tasks

demonstrated superior sensitivity compared to traditional sensors using elastomeric dielectrics such as PDMS [34].

### 2.3 Design of Wireless Signal Acquisition Module and Software System

The wireless signal acquisition module contains a microcircuit board that includes a microcontroller, capacitance conversion chip, Wi-Fi wireless component (ESP-01S, see Table S1 for power consumption), and lithium polymer battery. Shielded wiring (AFPF,  $0.035 \text{ mm}^2$ ) was employed in our setup to minimize errors and prevent noise from radio frequency and electromagnetic interference (see Fig. S3 for the quantitative comparison of noise reduction). This compact module captures and measures wrist movements, transmitting them to the host computer at a rate of 12.5 Hz. To facilitate communication with the wireless module, a server based on the TCP/IP protocol was established on a computer for data transmission. Moreover, a graphical interface was developed using the PyQt library in Python to enhance the visualization of digital signals collected

from microcontrollers. Three different applications for human–machine interaction were developed to suit various scenarios. Therefore, the computer continuously collects real-time capacitance signals and transmits them to the learning network for multitask recognition. Informed written consent was obtained from all human subjects before participation in the experiments.

### 2.4 Time-Series Data Augmentation

In self-supervised contrastive learning, the design of diverse data augmentations plays a pivotal role in minimizing the distance between different views of the same sample while maximizing the distance from other samples [35, 36]. Traditional approaches for data augmentation in self-supervised contrastive learning typically create two random variants in the same direction from a sample  $x$ , resulting in two views,  $x_1$  and  $x_2$ , using the same augmentation cluster  $U$ ,  $x_1 \sim U$  and  $x_2 \sim U$ . However, multiple experiments demonstrated that incorporating different data augmentation methods can



enhance model representation robustness (Tables S2 and S3) [37]. Hence, we employed two distinct types of augmentations: strong and weak. Weak augmentations introduced limited variations to the original signal, such as time shifting and scaling. In contrast, strong augmentations introduce more significant perturbations in the shape of the signal while preserving its temporal trends. Upon conducting systematic research on various data augmentation methods (as presented in Table S2), we utilize scaling and jittering to generate weak augmentations, and jittering along with permutation to generate strong augmentations [38]. For each input of unlabeled data  $x$ , we define weak and strong augmentations as  $x_s$  and  $x_w$ , respectively, where  $x_s \sim U_s$  and  $x_w \sim U_w$ .

### 2.5 Details of TS-VFC Learning

Our TS-VFC learning process comprises three main parts: encoder, cross-view fusion module, and projection head. For an input sample  $x \in R^{4 \times 32}$ , where the signal contains 4 channels and the window size is set to 32, high-dimensional features are extracted using an encoder. We define  $Z = f_{enc}(x)$ ,  $Z = [z_1, z_2, \dots, z_T]$ , where  $T$  indicates the total timesteps,  $z_i \in R^d$ , and  $d$  denotes the feature length. In this task, we set the timestep  $T=9$  and feature length  $d=128$ ,  $Z \in R^{128 \times 9}$ . Thus, we acquire two augmented views  $Z^s$  and  $Z^w$ , which are the inputs of the fusion cross-view module.

We introduce a cross-view fusion method that facilitates semantic communication between views at different stages [39]. Ablation experiments (Table S3) show that the generalization ability and accuracy of the model can be improved through fusion between views at different stages. The specific method is as follows: in the total range of timestep  $T$ ,  $T_r$  steps are arbitrarily chosen to split the high-dimensional feature vector  $Z$  into two parts, and we define them as  $Z_k = \{k|0 \leq k < T_r\}$ ,  $Z_g = \{g|T - T_r \leq g \leq T\}$ .  $V_k$  is obtained using the Transformer model, where  $V_k = f_{tra}(Z_k)$ , which is used to fuse temporal view information  $Z_g$ . The strong augmentation produces  $V_k^s$  and the weak augmentation produces  $V_k^w$ . For  $V_k^s$  generated by strong augmentation, concatenation fusion is performed with  $Z_g^w$  generated by the weak augmentation, and vice versa. The generated fused feature vectors are defined as  $V_{k,g}^s$  and  $V_{k,g}^w$ . The Transformer is used for further feature extraction from the views to

achieve cross-view fusion. It contains two main parts, multi-head attention and multiplayer perceptron block. We stack four identical layers to generate the regression features. Inspired by the BERT model, we include token  $c$  as a representative vector in the input features of the model [40].

The projection head is responsible for mapping the views to the LTS to learn more discriminative representations. Assuming that  $N$  input samples exist, two different augmentation methods provide two views for each sample,  $2N$  in total. For  $V_{k,g}^i$  from the cross-view fusion module and another augmented view  $V_{k,g}^{i+}$  from the same sample, we consider them as positive pairs, while the remaining  $2N - 2$  samples are treated as negative pairs. Therefore, the objective is to maximize the similarity between positive pairs and minimize the similarity between negative pairs. The loss function  $L_{TS-VFC}$  is defined as:

$$L_{TS-VFC} = - \sum_{i=1}^N \log \frac{\exp(\text{sim}(V_{k,g}^i, V_{k,g}^{i+})/\tau)}{\sum_{m=1}^{2N} \mathbf{1}_{[m \neq i]} \exp(\text{sim}(V_{k,g}^i, V_{k,g}^m)/\tau)}, \quad (1)$$

where  $\text{sim}(u, v) = u^T v / \|u\| \|v\|$  denotes the cosine similarity between  $u$  and  $v$ .  $\tau$  symbolizes a temperature parameter and  $\mathbf{1}_{[m \neq i]} \in \{0, 1\}$  represents an indicator function, only when  $m \neq i$ ,  $\mathbf{1}_{[m \neq i]} = 1$ .

The loss function of fine-tuning stage: upon completing the pre-training of the model, few-shot labeled data for specific tasks were collected to fine-tune the model. The cross-entropy loss function  $L_{fine}$  is expressed as

$$L_{fine} = - \sum_{i=1}^N y_i \log(\hat{y}_i). \quad (2)$$

### 2.6 Code and Data Availability

The source code used for TS-VFC Learning and source data in this study are available at <https://drive.google.com/drive/folders/18fM5DNxor0Ahy7CXhHWjc5MGn12ciYrv?usp=sharing>.

## 3 Results and Discussion

### 3.1 Sensing Mechanism and Properties of Iontronic Device

The detailed illustration of the pressure-sensing mechanism of the iontronic device is presented in Fig. 2a. The

photocurable hydrogel employed as the dielectric layer in the device plays a critical role in sensing. When voltage is applied across the AgNWs electrodes (Fig. 2b), cations and anions within the hydrogel migrate and gather at interfaces between electrodes and hydrogel, creating an electric double layer (EDL) [41, 42]. The crucial characteristic of this layer is its nanometer-scale thickness, resulting in extremely high capacitance. Hierarchical microcones are incorporated within the dielectric hydrogel layer to enhance the contact area variation under pressure [43, 44]. Figures 2b and S1 display hierarchical microcones with two different heights (680 and 500  $\mu\text{m}$ ), diameters (450 and 350  $\mu\text{m}$ ), and a spacing of 1200  $\mu\text{m}$ . As demonstrated in the finite element simulation in Fig. S4, the low pressure primarily influences the contact area variation of the higher microcones. As the pressure increases, smaller microcones sequentially contact the upper electrode after the compression of the taller ones. The resulting high interfacial capacitors ( $C_{\text{EDL1}}$ ,  $C_{\text{EDL2}}$ ...  $C_{\text{EDLn}}$ ) are crucial for achieving high sensitivity because they undergo significant changes with variations in the contact area owing to the pressure-induced deformation of the microcones.

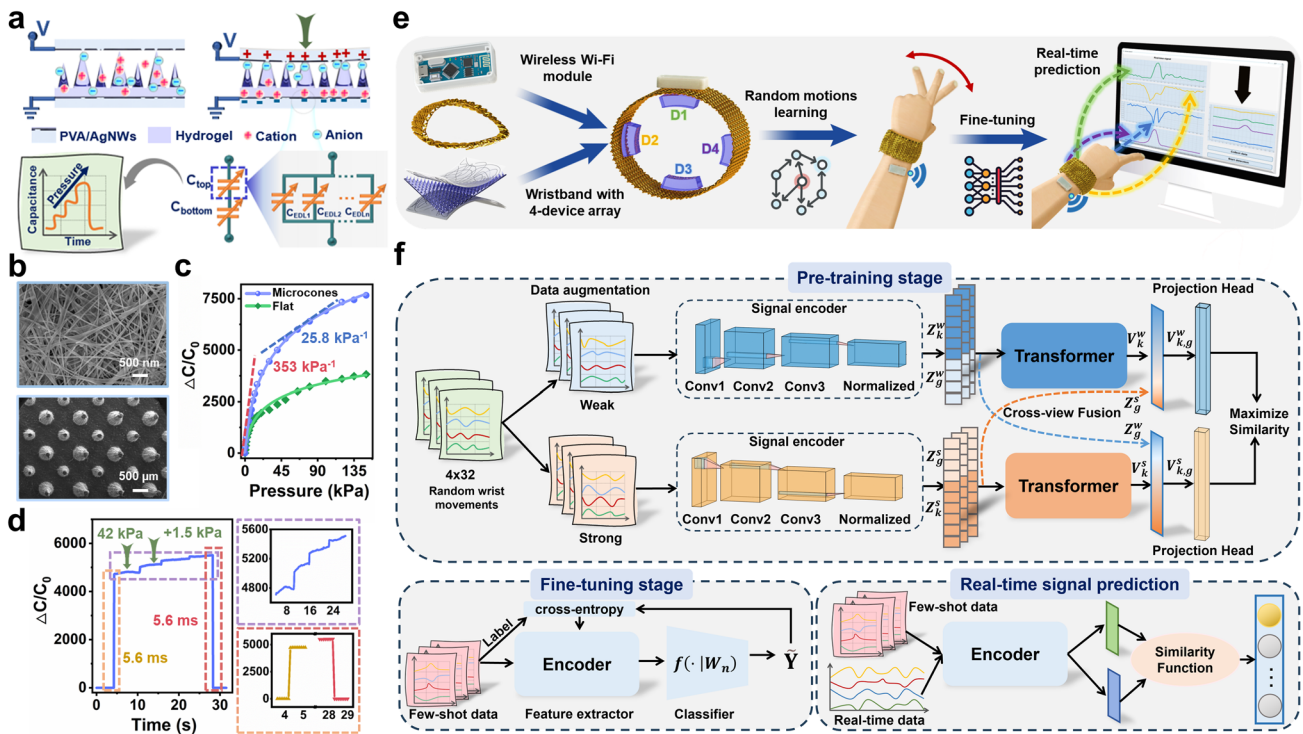
Sensitivity, mathematically expressed as  $S = \partial(\Delta C/C_0)/\partial P$ ,  $C_0$  represents the initial capacitance without any applied pressure,  $\Delta C$  expresses the variation in capacitance after applying pressure, and  $P$  denotes the applied pressure. Figure 2c illustrates the capacitance response curves of the iontronic devices, both with and without hierarchical microcones. In the case of the device featuring the microcones, sensitivity reaches  $353 \text{ kPa}^{-1}$  within the lower-pressure range and  $25.8 \text{ kPa}^{-1}$  in the higher-pressure range, outperforming the device with a flat hydrogel layer across the entire pressure range. Owing to its high sensitivity and broad sensing range (150 kPa), the device maintains a high pressure resolution throughout the pressure range. Thus, even under a preloaded high pressure of 42 kPa, the device can discern slight pressure variations (1.5 kPa), as depicted in Fig. 2d. Furthermore, it shows a fast response and recovery times of 5.6 ms during both loading and unloading pressures. Furthermore, as illustrated in Fig. S5, the device exhibits excellent stability throughout the repeated loading and unloading processes at different pressures (2.5, 8, 13, 22, 43, and 80 kPa). No significant differences in capacitance values were observed at the same pressure levels. Furthermore, the device exhibits stable and synchronous response during repeated bending at different angles (30°, 45°, 60°,

and 80°), as shown in Fig. S6. To assess device durability, a cyclic pressure of 30 kPa was applied over 15,000 cycles, and the capacitive response is shown in Fig. S7. The inset displays the waveforms from the last few cycles of the device, which remained nearly unchanged throughout the test. Additionally, Fig. S8 illustrates that the output signal essentially maintains the same level after 11,000 cycles of bending at 65°, ensuring long-term use in daily activities.

### 3.2 Rapid Adaptation of Multi-Scenario Tasks by Contrast Learning

Previous studies on wrist recognition models were limited by their specificity to individual tasks, lacking the adaptability to handle multiple scenarios [22, 23]. We aim to develop a robust and general model that can rapidly adapt to wrist movements in unfamiliar scenarios for daily tasks. Figure 2e depicts a block diagram that illustrates the entire prediction process. The wristband contains four iontronic devices worn on the wrist, complemented by an attached wireless Wi-Fi module to ensure user comfort and create a self-contained system (Fig. S9 and Movie S1). Instead of using labels to classify the four capacitance signals, the proposed model employs unlabeled random wrist-motion signals for self-supervised contrastive learning, distinguishing it from traditional supervised algorithms. Using this approach, features from the sensor signals are autonomously learned, resulting in a separable feature space that enables classification for different tasks. The fundamental principle involves calculating the similarity between the identical and different samples, creating the LTS with a strong representation of the signals generated by the smart wrist to represent prior index finger movements. Consequently, new users can rapidly engage in transfer learning for various tasks, facilitating the division of distinct samples. When undertaking a new task, only few-shot of wrist-motion data specific to that task need to be collected. These signals are then projected onto the LTS. Through metric calculations, primarily based on cosine similarity, and comparisons of these features with real-time inputs, precise predictions of wrist motion gestures can be displayed on the screen, even for tasks not included in the training set of the model.

To achieve rapid adaption to wrist motions across multiple scenarios, we propose a TS-VFC algorithm trained with a large amount of unlabeled time-series data. Figure 2f



**Fig. 2** Sensing mechanism and properties, and TS-VFC algorithm architecture. **a** Schematic of the sensing mechanism of iontronic device. **b** Scanning electron microscope (SEM) images of ionic hydrogel with microcone structure and AgNWs/PVA electrode. **c** Normalized change of capacitance as a function of pressure over 150 kPa. **d** Response of tiny pressure (1.5 kPa) variations when high pressure (42 kPa) has been preloaded. Response and recovery time of the device (5.6 and 5.6 ms). **e** Illustration of general prediction process, including preparation of wristband with attached sensing devices and wireless module, TS-VFC learning of random motions, few-shot fine-tuning, and real-time signal prediction. **f** Details of TS-VFC learning model architecture. During the pre-training step, the overall architecture of the proposed TS-VFC model performs contrastive training through unlabeled random data. The encoder is then fine-tuned through supervised learning given few-shot labeled data. Finally, the encoder is used to encode real-time signals, metric distance from encoded label data, and complete signal category prediction

illustrates the detailed architecture and workflow of the proposed algorithm, which contains three distinct stages. In the pre-trained stage, unique wrist movements and postures, generated by new users wearing the wristband, result in randomly changing signals from the devices. These unlabeled signals are utilized to create a general LTS through the TS-VFC learning process. First, data augmentation methods are employed to generate two correlated views for each time series. These views are referred to as positive pairs as they originate from the same wrist-motion signal, while sequences with different actions serve as negative pairs. A variety of data augmentation methods, including both ‘weak’ and ‘strong’ approaches, are used to produce these correlated views (Refer to the Experimental Section for details on data augmentation methods). Subsequently, an encoder, comprising three one-dimensional convolutional layers (Table S4), is utilized to extract high-dimensional latent features from the

generated views [45]. A cross-view fusion module is introduced to calculate the time-series contrast loss by merging different views at different stages. The transformer architecture (Fig. S10) serves as the implementation of this module owing to its robust feature extraction capabilities and high efficiency [46]. A neural network projection head composed of two linear layers is employed to map feature vectors to the LTS [47]. Finally, the similarity function minimizes the distance between pairs of positive samples in the LTS, thereby training the potent feature encoder that can be applied to various downstream tasks.

In the fine-tuning stage, building upon the pre-trained model weights, we collected few-shot label data and introduced a traditional supervised learning model to fine-tune the LTS. The fine-tuning process is essential for achieving transfer learning, ensuring that the LTS can effectively handle complex wrist movements across different scenarios. To

accomplish this, a classifier in the form of a linear layer was integrated into the last layer of the encoder. This supervised training step facilitates the adaptation of the model to the specific task at hand. Recognizing the diversity in user writing habits within the same scenario, which can manifest as variations in writing time, intensity, and pause time for each stroke, it is essential to normalize these writing habits. The normalization process helps minimize signal deviations within the same action, as illustrated in Figs. S11 and S12. For real-time signal prediction, the encoder is used to map the signals generated by the real-time wrist movements of the user into high-dimensional features. These features are then compared with the prior movements already embedded in the LTS through metric calculations, enabling the prediction of wrist-motion categories [48, 49].

### 3.3 Prediction of Four/Eight Directions and Game Control

When the user points their index finger in different directions with the wristband, the corresponding capacitance signals from the four devices are captured from the wireless module. Real-time direction predictions can be achieved by fine-tuning the labeled motion data specific to directions. Figure 3a illustrates the corresponding signal waveforms when the index finger moves in eight different directions. Moving the index finger horizontally to the right represents a right command ( $0^\circ$ ), and moving it diagonally upward to the right by  $45^\circ$  indicates an upper right command ( $45^\circ$ ). Notably, discernible differences exist in the time-varying outputs of Signal 2, even for similar gestures (such as left and lower left). The proposed model is designed to effectively adapt to new actions introduced into a few-shot labeled dataset while preserving the information from the original actions. The training process does not need to be strengthened, and the model does not need to be redesigned for newly added actions. While initially trained to recognize four basic directions (left, right, up, and down), the proposed model can effortlessly expand its capabilities to recognize four additional directions ( $45^\circ$ ,  $135^\circ$ ,  $225^\circ$ , and  $315^\circ$ ) by simply supplementing the corresponding few-shot labels. Images of real-time predictions are illustrated in Fig. 3b, demonstrating continuous changes in the four-channel signals and corresponding predicted results on the interface. The entire recognition process for both four and eight directions,

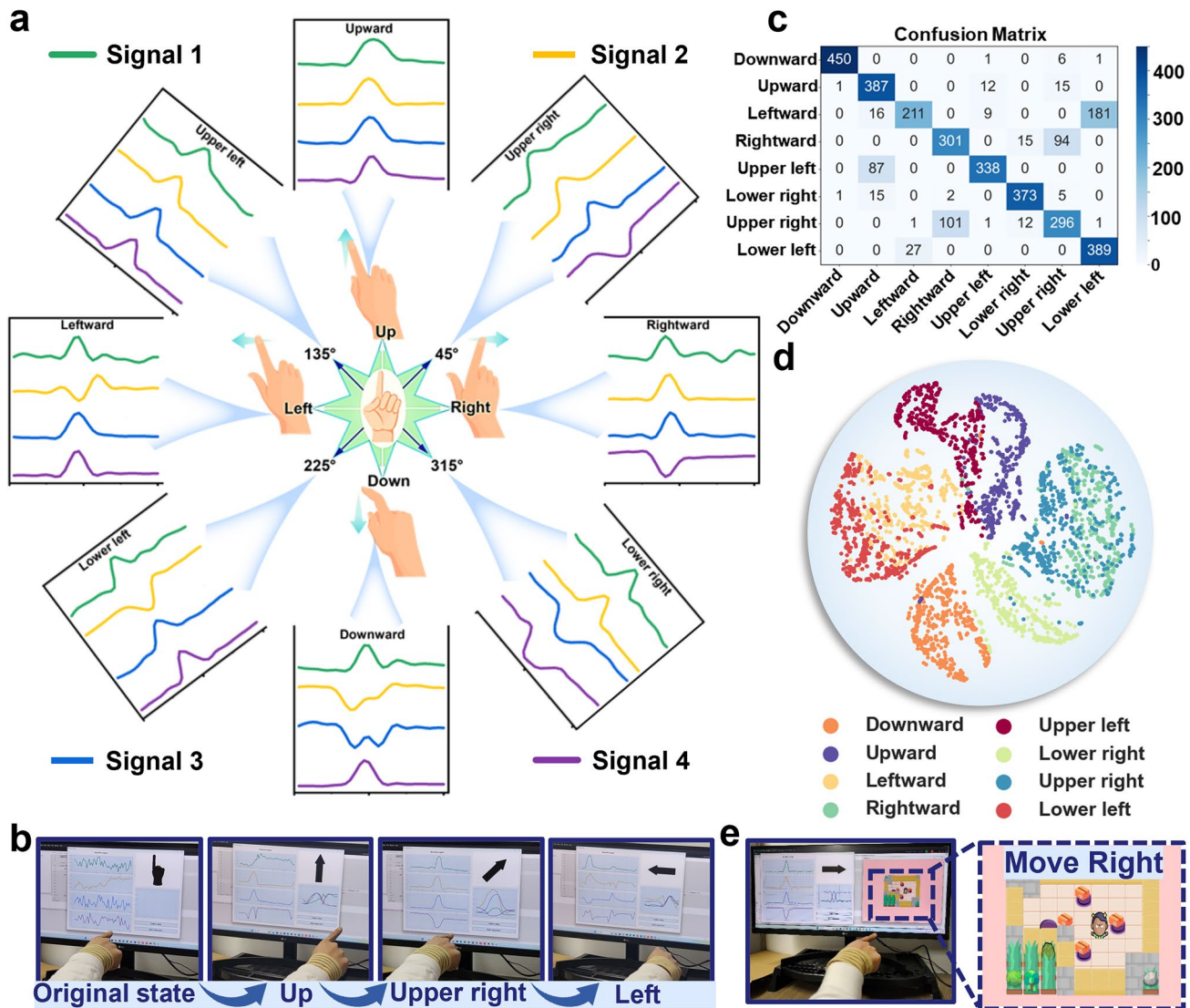
encompassing prior motion learning, is demonstrated in Movies S2 and S3, respectively. The distinction is that for the four directions, only 3-shot data are required in each direction for a total of 12 actions for transfer learning. Conversely, for the eight directions, only 5-shot data are required in each direction to complete the fine-tuning.

To validate the performance of the model, confusion matrices were generated for recognizing either four or eight directions in the test set after 80 transfer training epochs. These matrices indicate that the proposed model achieved average accuracies of 94.7% for the four directions (Fig. S13) and 82.0% (Fig. 3c) for the eight directions, respectively. To improve the accuracy of predicting eight directions, adding more labeled data (e.g., from 5-shot to 10-shot) during the fine-tuning phase is effective. As shown in Fig. S14, the accuracy of predicting the eight directions reached 87.5% with 10-shot. Our proposed algorithm allows users to customize the number of training samples, enabling adjustments based on personal needs. For example, users can choose a smaller number of labeled data for rapid adaptation or a larger number of labeled data to improve accuracy. We employed the t-distributed stochastic neighbor embedding (t-SNE) algorithm to perform dimensionality reduction on the high-dimensional features extracted from the encoder to provide further insight and analysis of the recognition of these actions (Figs. 3d and S15) [50]. In this feature space, the feature points of each category are accurately mapped to their corresponding positions for these directions, underscoring the robustness of the proposed model in this scenario. To demonstrate the practical interaction capabilities of the wristband, we utilize direction recognition for controlling a Sokoban game. As indicated in Fig. 3e, the capacitive outputs generated by the index finger motions are captured and transmitted by the wireless module, subsequently read by a computer program to predict the direction. This prediction is then used to control the movement direction of the game character in a customized game (Movie S4).

### 3.4 Prediction of Air-Writing Numbers and Letters

Based on the excellent direction recognition ability of the wristband, we extend its functionality to handle more complex tasks, such as recognition of numbers and letters, by analyzing the features of the signals generated when the index finger moves in different directions. Figure 4a

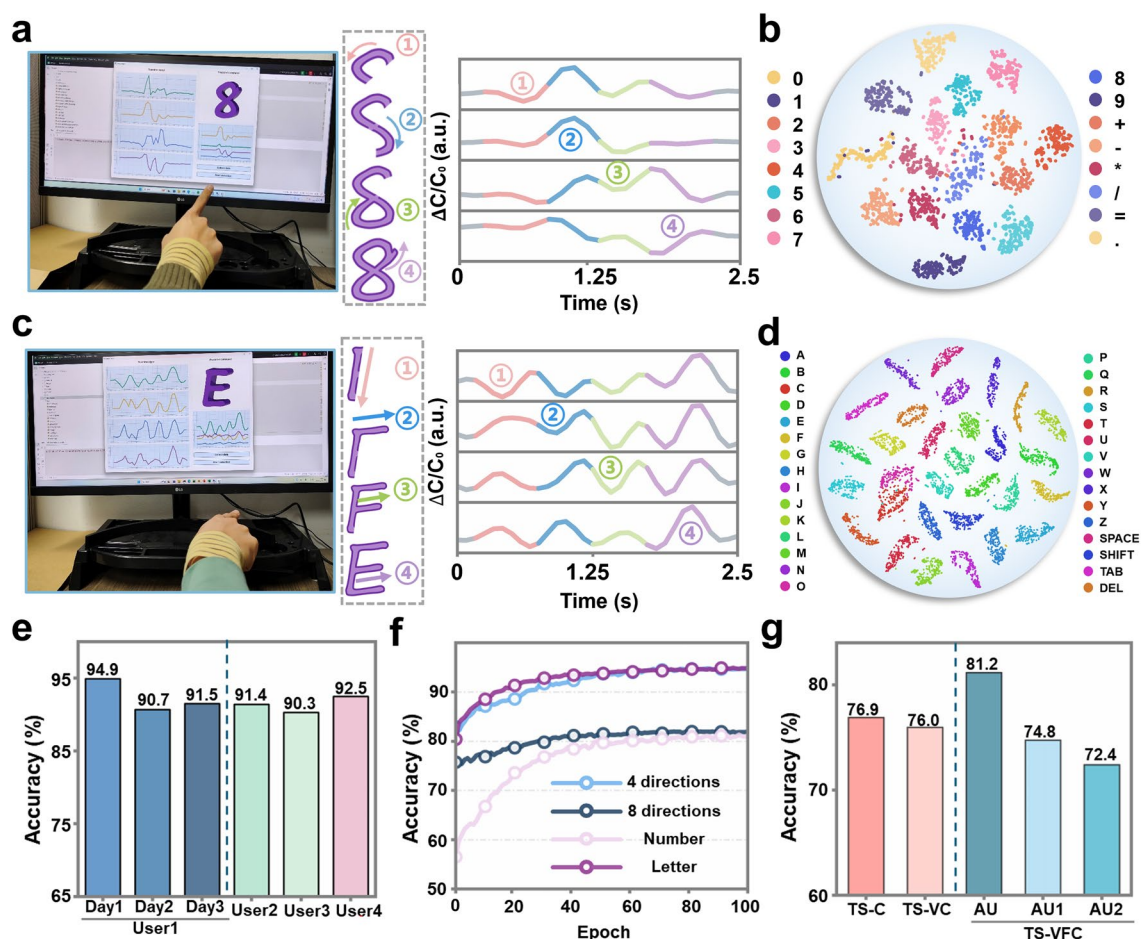




**Fig. 3** Demonstration of direction prediction and game control. **a** Waveforms of real-time four-channel signals of eight directions. **b** Photos of real-time prediction for direction recognition with the wristband. **c** Confusion matrix for prediction of eight different directions, with an average accuracy of 82.0%. **d** t-SNE projection of high-dimensional latent features of labeled data for eight-direction recognition using transfer learning. **e** Demonstration of game control and screenshots of the game interface

illustrates the scenario of air-writing numbers with the wristband, along with the interface displaying the four-channel signal and prediction. The number “8” is identified by extracting the capacitance signals during the writing process. The consecutive writing process is divided into four steps (①–④), with each sensor producing corresponding signals that exhibit continuous but different waveforms. In addition, mathematical operators (such as +, −, =) are included in the learning scope to develop the calculator system, and their waveforms differ from those of the numbers, following the writing format, as illustrated in Fig. S11. To standardize

the writing methods, the finger returns to its initial position after completing the last stroke of each character. The four-channel signal diagram for all numbers “0–9” and six operators are presented in Fig. S16. The proposed model achieves an average accuracy of 81.2% in predicting 16 categories of numbers and symbols with only 5-shot learning for each category (Fig. S17), eliminating the necessity for a large volume of labeled data for training. As illustrated in Fig. 4b, employing the t-SNE algorithm to visualize these features and project them onto the feature space reveals that each category forms distinct clusters.



**Fig. 4** Demonstration of air-writing numbers and letters. **a** Photo of real-time number prediction, as well as detailed process and waveform of air-writing “8”. **b** t-SNE projection of high-dimensional latent features for the test set using transfer learning. **c** Photo of real-time letter prediction, as well as detailed process and waveform of air-writing “E”. **d** t-SNE projection of high-dimensional features for test set of 26 letters and four function keys. **e** Comparison of accuracy in air-writing 26 letters and 4 function keys by the same user at different times and by 4 different users. **f** Accuracy trends for different tasks during fine-tuning training epochs: under the weights of a pre-trained model, supervised training is conducted with few-shot labeled data by adding a linear layer behind the encoder model. **g** Ablation experiment study of different variants of the TS-VFC model and using different data augmentation methods for number prediction. TS-C model without view fusion and Transformer architecture, and TS-VC model with the Transformer architecture. AU (TS-VFC model) uses “scale+jitter” for weak augmentation and “jitter+permutation” for strong augmentation. AU1 uses “scale+jitter” for both weak and strong augmentation, and AU2 uses “jitter+permutation” for both weak and strong augmentation

Similar to the number prediction approach described earlier, the wristband can recognize all 26 English letters through air-writing pattern tracking. Considering the letter “E” as an example, the waveform characteristics of the corresponding signal are analyzed. Figure 4c demonstrates that it takes four strokes to write the letter “E,” specifically down-right-right-right (①–④), with the finger returning to the initial position after the last stroke. Since the strokes from ② to ④ are identical, it is evident that the separated signal features are also similar. Furthermore, four function keys were added to develop the keyboard

input system encompassing 26 letters and four function keys. The specific writing methods for these function keys are defined in Fig. S12. The four-channel signal diagram for all letters “A–Z” and four functional keys is presented in Fig. S18. Owing to the increased complexity of these actions and the consequently greater semantic information collected from the four-channel signals, the proposed model achieves an average accuracy of 94.9% for letter and function key recognition with only 5-shot learning (Fig. S19). Figure 4d displays the test set containing these 30 categories, extracted using the proposed model, and

projected into the latent space using the t-SNE algorithm for dimensionality reduction. A significant distance exists between each category within this space, enabling character recognition within this scenario by calculating the distances between each category.

Movie S5 showcases the steady and continuous recognition of the letters “K B G I F M V P.” When the same user writes these letters at different times, as shown in Fig. S20, the four-channel signal changes are only slightly different. While prolonged wear may lead to sliding between the wristband and skin, which could affect device output and recognition precision, the simultaneous operation of the four sensing devices and TS-VFC learning algorithms ensures consistent performance even when slight positional shifts of the wristband occur over time (Figs. 4e, S21 and S22). Despite potential variations among individuals, the proposed model can consistently learn the writing characteristics of new users. When four users with different wrist sizes write these eight letters, the changes in the four-channel signals are illustrated in Fig. S23. As shown in Fig. S24, the wrist of the overweight male (user 4) is relatively thick, with a circumference greater than 20 cm, while the slender female (user 3) has a thinner wrist with a circumference less than 15 cm. The wrist sizes of users 1 and 2 are average, approximately 17.5 cm. Although these signals may be different, new users only need to provide few-shot labeled data to achieve the desired prediction effect (Figs. 4e and S25–S27). Considering that some symbols have the same strokes but are arranged in different relative positions, to demonstrate that the wristband is capable of restoring, reproducing, and reflecting the spatial trajectory of the strokes rather than just recognizing the strokes, we designed additional air-writing symbol experiments. The writing sequence of the symbols with the same stroke features and their corresponding four-channel signals are shown in Fig. S28. Through our TS-VFC learning with 5-shot, the prediction accuracy can reach 88.6% (Fig. S29). Given the widespread use of commercial accelerometers for motion tracking, a comparison was conducted to evaluate the performance of our proposed device. We collected signals from air-writing letters A to H using the accelerometer, as shown in Fig. S30. The similarity matrices for the signals corresponding to these 8 letters from

both sensors were calculated (Fig. S31), and the accuracy of predictions was trained using TS-VFC learning under the same conditions (Fig. S32). The results demonstrate that the flexible sensing device provides superior performance for continuous and accurate monitoring of subtle wrist movements during air-writing applications.

Figure 4f illustrates the capability of the pre-trained model learned by the TS-VFC algorithm to adapt to different scenarios. Without any fine-tuning training (epoch = 0), the pre-trained model already initially adapts, and it can achieve over 80% accuracy within the transfer training period of 40 epochs. Following 80 epochs of fine-tuning training, the accuracy and MF1-score results that correspond to each scenario are presented in Table S5. Our pre-trained TS-VFC model demonstrates robust prediction capabilities across multiple scenarios. In addition, taking the number prediction scenario as an example, we perform ablation experiments by constructing different model variants to compare with the TS-VFC model, aiming to demonstrate the effectiveness of each component. The TS-C module lacks both view fusion and Transformer architecture, while the TS-VC module incorporates the Transformer architecture. As illustrated in Fig. 4g, the experiment results reveal that the proposed TS-VFC model improves accuracy by over 4% for these two variants. This is attributed to the use of the Transformer to achieve view fusion at different stages, which helps positive samples from the same class generate more discriminative features. The comparison of ablation experiments for other tasks, including direction and letter prediction, is provided in Table S3. Subsequently, we explore the effects of weak and strong data augmentation on the TS-VFC model. As indicated in AU1 and AU2 (Fig. 4g), when only two weak or two strong augmentations are used, the accuracy falls by approximately 9% compared to the proposed TS-VFC model. Evidently, merely using weak augmentation (AU1) fails to learn more discriminative features owing to minor data variations. Conversely, relying solely on strong augmentation (AU2) prevents the model from recognizing the original signal during prediction. To validate the rationality of the proposed data augmentation method selection, we conducted a detailed study on the impact of various augmentation techniques on the TS-VFC model in number and letter prediction, as shown in Table S2.

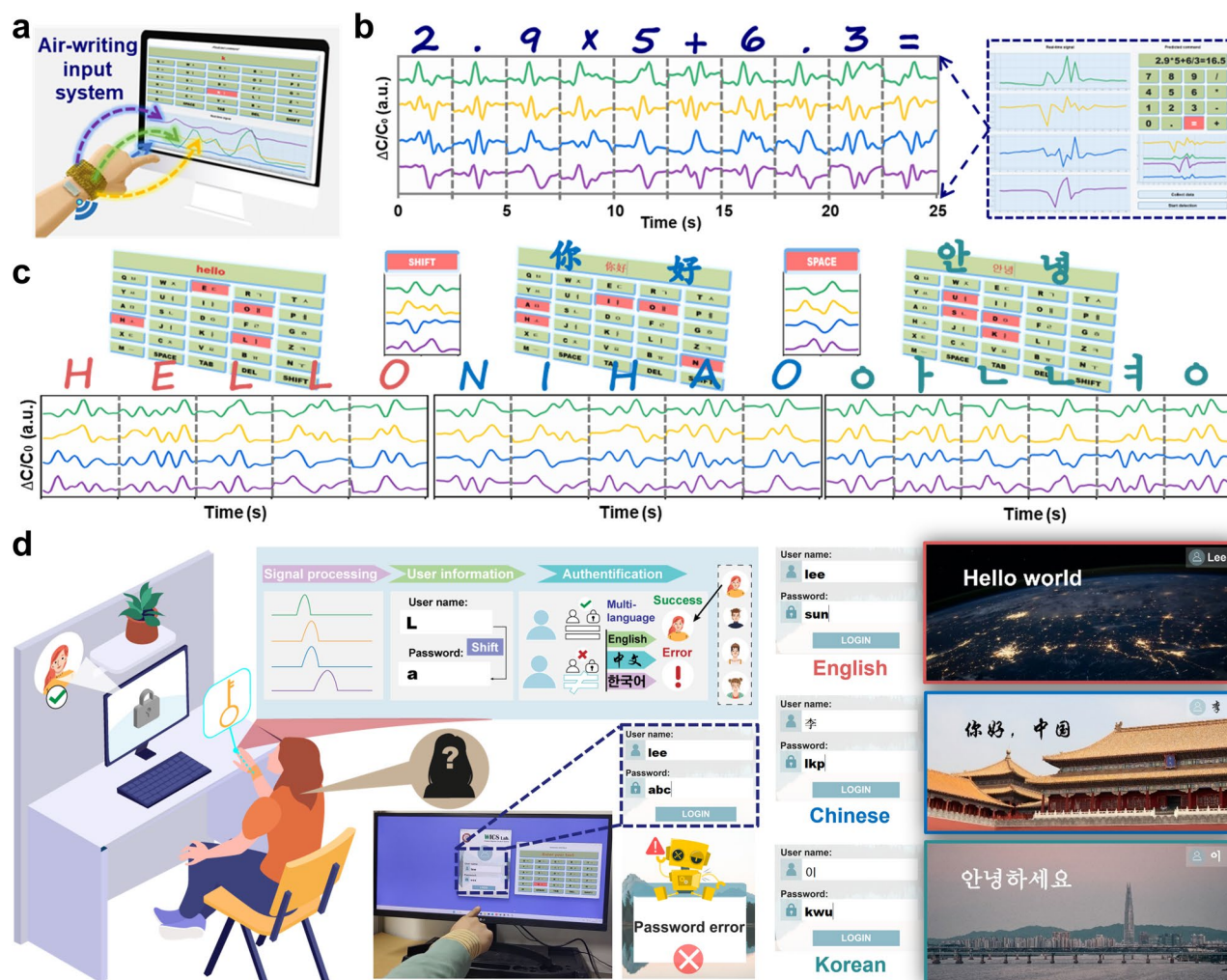


### 3.5 Demonstration of Air-Writing Input System

For the air-writing input system developed based on the wristband (Fig. 5a), signals from the wristband are transmitted to the computer through the wireless module for further processing and recognition, with the prediction results displayed on the interface in real time. The calculator system serves as a ubiquitous calculation tool in our daily lives, and Fig. 5b presents the continuous variations in the four-channel signals while air-writing the equation “ $2.9 \times 5 + 6/3 =$ ”. The corresponding characters were identified in real time, and the final result of the Eq. (16.5) can be obtained after

entering the symbol “=”. Movie S6 depicts the operational process of the calculator, demonstrating rapid and stable distinguishing between different numbers and symbols by air-writing with an index finger.

As depicted in Fig. 5c, the signal variations when air-writing greetings in three languages (English, Chinese, Korean), are switched using the function key “SHIFT.” The word “HELLO” is initially air-written with the wristband, and waveforms of responses are displayed on screen in real time, with corresponding words recognized. The “SHIFT” and “SPACE” functions are represented by specific strokes (Fig. S12), and their waveforms differ from those of the



**Fig. 5** Illustration of the calculator, keyboard input, and login system based on air-writing. **a** Schematic of air-writing input system. **b** Waveforms of four signals during the calculation operation process, and a screenshot of the calculator interface. **c** Screenshots of keyboard input interface, and waveforms of four signals of air-writing words “hello”, “你好” and “안녕”. **d** Schematic of multi-language login system based on the wristband. Real-time prediction of inputting username and password, and screenshots of login interface



other letters. Upon switching the input method, “你好” and “안녕” continue to be entered and are successfully recognized, as demonstrated in Movie S7. A further developed login system that supports multi-language input is illustrated in Fig. 5d, and users can access various language systems by switching their handwritten username inputs. In the scenario where the user enters the password “abc” corresponding to the username “lee” in the login system, which does not match the preset correct password “sun,” the interface will automatically prompt “Password error.” Upon entering the correct username and password, users can freely switch to the login system in different languages following successful authentication. Each interface is shown in Fig. 5d, and the overall process of handwritten input, password verification, and interface display is shown in Movie S8. The capability to convert air-written characters into signals comprehensible to computers, without the need for learning proprietary touch symbols, minimizes user burden, and it is particularly useful for individuals with visual impairments.

## 4 Conclusions

Leveraging the innovative capabilities of TS-VFC learning, we introduced a wearable wristband designed to conform to wrist anatomy (see Fig. S33 for the breathability experiment of the wristband material), facilitating rapid adaptation to diverse scenarios of dynamic gesture tracking. As indicated in Table S6, compared with other representative studies of wristband systems, our innovative approach enables transfer learning utilizing unlabeled data. Each iontronic sensing device within the wristband, featuring hierarchical microcones and ultrathin flexible electrodes, has demonstrated exceptional performance, boasting a high sensitivity of  $353 \text{ kPa}^{-1}$  and a board response range of 150 kPa. Moreover, we designed a compact circuit with a Wi-Fi module for the wireless acquisition of four-channel signals from the array. Following the formation of the LTS during the pre-training stage in the model, the wristband system has proven effective in handling various tasks across multiple scenarios. These include the prediction of precise eight-direction commands and air-writing of numbers and letters. Despite the encoder in the model comprising only three layers of one-dimensional convolutions, it can adapt to new gesture recognition tasks without requiring architectural redesigning or extensive training for

specific tasks. With minimal labeled data, the model can be fine-tuned for generalization to new tasks, even those not originally included in the training data. Moreover, practical applications, such as game control, calculator operation, and login systems, were demonstrated, highlighting the feasibility and potential of human–machine interaction. This self-supervised wristband system seamlessly integrates with a user, offering an intuitive means of communication and control of digital interfaces through gestures that align with everyday habits. It has the potential to provide a more natural, immersive, efficient, and personalized digital experience in the future.

**Acknowledgements** This work was supported by the Research Grant Fund from Kwangwoon University in 2023, the National Natural Science Foundation of China under Grant (62311540155), the Taishan Scholars Project Special Funds (tsqn202312035), and the open research foundation of State Key Laboratory of Integrated Chips and Systems.

**Authors' Contribution** YG, KL, and WY contributed equally to this work. YG, KL, YL, GS, and JL conceived the ideas and designed the research. YG and WY performed the sensor experiment and analyzed the sensing properties. KL handled the software development and programming tasks. YG designed the signal acquisition circuit. YG and KL performed applications with contrastive learning. YG, KL, and WY drafted and wrote the manuscript. NK, YL, GS, and JL reviewed and edited the manuscript.

### Declarations

**Conflict of interest** The authors declare no interest conflict. They have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40820-024-01545-8>.

## References

1. A. Moin, A. Zhou, A. Rahimi, A. Menon, S. Benatti et al., A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition. *Nat. Electron.* **4**, 54 (2020). <https://doi.org/10.1038/s41928-020-00510-8>
2. M. Wang, Z. Yan, T. Wang, P. Cai, S. Gao et al., Gesture recognition using a bioinspired learning architecture that integrates visual data with somatosensory data from stretchable sensors. *Nat. Electron.* **3**, 563 (2020). <https://doi.org/10.1038/s41928-020-0422-z>
3. Z. Sun, M. Zhu, X. Shan, C. Lee, Augmented tactile-perception and haptic-feedback rings as human-machine interfaces aiming for immersive interactions. *Nat. Commun.* **13**, 5224 (2022). <https://doi.org/10.1038/s41467-022-32745-8>
4. X. Liang, H. Li, W. Wang, Y. Liu, R. Ghannam et al., Fusion of wearable and contactless sensors for intelligent gesture recognition. *Adv. Intell. Syst.* **1**, 1900088 (2019). <https://doi.org/10.1002/aisy.201900088>
5. T. Watanabe, Md. Maniruzzaman, Md.A.M. Hasan, H.-S. Lee, S.-W. Jang et al., 2D camera-based air-writing recognition using hand pose estimation and hybrid deep learning model. *Electronics* **12**, 995 (2023). <https://doi.org/10.3390/electronic12040995>
6. S. Ahmed, W. Kim, J. Park, S.H. Cho, Radar-based air-writing gesture recognition using a novel multistream cnn approach. *IEEE Internet Things J.* **9**, 23869 (2022). <https://doi.org/10.1109/JIOT.2022.3189395>
7. Y. Guo, F. Yin, Y. Li, G. Shen, J. Lee, Incorporating wireless strategies to wearable devices enabled by a photocurable hydrogel for monitoring pressure information. *Adv. Mater.* **35**, 2300855 (2023). <https://doi.org/10.1002/adma.202300855>
8. W. Yue, E.-S. Kim, E. Ganbold, B.-H. Zhu, B. Oh et al., A miniature and reusable radiofrequency biosensor combining microfluidic and integrated passive technology for glucose detection. *Sens. Actuators B Chem.* **392**, 134108 (2023). <https://doi.org/10.1016/j.snb.2023.134108>
9. Y. Li, Q. Lin, T. Sun, M. Qin, W. Yue, S. Gao, A perceptual and interactive integration strategy toward telemedicine healthcare based on electroluminescent display and triboelectric sensing 3D stacked device. *Adv. Funct. Mater.* (2024). <https://doi.org/10.1002/adfm.202402356>
10. H. Niu, H. Li, S. Gao, Y. Li, X. Wei et al., Perception-to-cognition tactile sensing based on artificial-intelligence-motivated human full-skin bionic electronic skin. *Adv. Mater.* **34**, 2202622 (2022). <https://doi.org/10.1002/adma.202202622>
11. Y. Liu, C. Yiu, Z. Song, Y. Huang, K. Yao et al., Electronic skin as wireless human-machine interfaces for robotic VR. *Sci. Adv.* **8**, eab16700 (2022). <https://doi.org/10.1126/sciadv.ab16700>
12. J. Huang, R. Zhou, Z. Chen, Y. Wang, Z. Li et al., Highly stable and reliable capacitive strain sensor for wearable electronics based on anti-dry hydrogel electrode. *Mater. Today Phys.* **35**, 101123 (2023). <https://doi.org/10.1016/j.mtphys.2023.101123>
13. K.K. Kim, I. Ha, M. Kim, J. Choi, P. Won et al., A deep-learned skin sensor decoding the epicentral human motions. *Nat. Commun.* **11**, 2149 (2020). <https://doi.org/10.1038/s41467-020-16040-y>
14. Z. Zhou, K. Chen, X. Li, S. Zhang, Y. Wu et al., Sign-to-speech translation using machine-learning-assisted stretchable sensor arrays. *Nat. Electron.* **3**, 571 (2020). <https://doi.org/10.1038/s41928-020-0428-6>
15. T. Wang, Y. Zhao, Q. Wang, A flexible iontronic capacitive sensing array for hand gesture recognition using deep convolutional neural networks. *Soft Robot.* **10**, 443 (2023). <https://doi.org/10.1089/soro.2021.0209>
16. P. Tan, X. Han, Y. Zou, X. Qu, J. Xue et al., Self-powered gesture recognition wristband enabled by machine learning for full keyboard and multicommand input. *Adv. Mater.* **34**, 2200793 (2022). <https://doi.org/10.1002/adma.202200793>
17. H. Fang, L. Wang, Z. Fu, L. Xu, W. Guo et al., Anatomically designed triboelectric wristbands with adaptive accelerated learning for human-machine interfaces. *Adv. Sci.* **10**, 2205960 (2023). <https://doi.org/10.1002/advs.202205960>
18. X. Le, Q. Shi, Z. Sun, J. Xie, C. Lee, Noncontact human-machine interface using complementary information fusion based on MEMS and triboelectric sensors. *Adv. Sci.* **9**, 2201056 (2022). <https://doi.org/10.1002/advs.202201056>
19. A. Tashakori, Z. Jiang, A. Servati, S. Soltanian, H. Narayana et al., Capturing complex hand movements and object interactions using machine learning-powered stretchable smart textile gloves. *Nat. Mach. Intell.* **6**, 106 (2024). <https://doi.org/10.1038/s42256-023-00780-9>
20. J. Park, D. Kang, H. Chae, S.K. Ghosh, C. Jeong et al., Frequency-selective acoustic and haptic smart skin for dual-mode dynamic/static human-machine interface. *Sci. Adv.* **8**, eabj9220 (2022). <https://doi.org/10.1126/sciadv.abj9220>
21. K.K. Kim, M. Kim, K. Pyun, J. Kim, J. Min et al., A substrate-less nanomesh receptor with meta-learning for rapid hand task recognition. *Nat. Electron.* **6**, 64 (2023). <https://doi.org/10.1038/s41928-022-00888-7>
22. Q. He, Z. Feng, X. Wang, Y. Wu, J. Yang, A smart pen based on triboelectric effects for handwriting pattern tracking and biometric identification. *ACS Appl. Mater. Interfaces* **14**, 49295 (2022). <https://doi.org/10.1021/acsami.2c13714>
23. H. Lee, S. Lee, J. Kim, H. Jung, K.J. Yoon et al., Stretchable array electromyography sensor with graph neural network for static and dynamic gestures recognition system. *npj Flex. Electron.* **7**, 20 (2023). <https://doi.org/10.1038/s41528-023-00246-3>
24. L. Cui, C. Hu, W. Wang, J. Zheng, Z. Zhu et al., An adhesive, stretchable, and freeze-resistant conductive hydrogel strain sensor for handwriting recognition and depth motion monitoring. *J. Colloid Interface Sci.* **677**, 273 (2024). <https://doi.org/10.1016/j.jcis.2024.07.214>
25. K.R. Pyun, K. Kwon, M.J. Yoo, K.K. Kim, D. Gong et al., Machine-learned wearable sensors for real-time hand motion recognition: toward practical applications in reality. *Natl. Sci. Rev.* **11**, nwad298 (2024). <https://doi.org/10.1093/nsr/nwad298>

26. X. Wei, H. Li, W. Yue, S. Gao, Z. Chen et al., A high-accuracy, real-time, intelligent material perception system with a machine-learning-motivated pressure-sensitive electronic skin. *Matter* **5**, 1481 (2022). <https://doi.org/10.1016/j.matt.2022.02.016>
27. Y. Li, Z. Qiu, H. Kan, Y. Yang, J. Liu et al., A human-computer interaction strategy for an FPGA platform boosted integrated “perception-memory” system based on electronic tattoos and memristors. *Adv. Sci.* (2024). <https://doi.org/10.1002/advs.202402582>
28. H. Niu, F. Yin, E. Kim, W. Wang, D. Yoon et al., Advances in flexible sensors for intelligent perception system enhanced by artificial intelligence. *InfoMat* **5**, e12412 (2023). <https://doi.org/10.1002/inf2.12412>
29. L. Jin, Z. Li, Z. Liu, B. Richardson, Y. Zheng et al., Flexible unimodal strain sensors for human motion detection and differentiation. *npj Flex. Electron.* **6**, 74 (2022). <https://doi.org/10.1038/s41528-022-00205-4>
30. W. Yang, H. Kan, G. Shen, Y. Li, A network intrusion detection system with broadband  $\text{WO}_{3-x}/\text{WO}_{3-x}\text{-Ag}/\text{WO}_{3-x}$  optoelectronic memristor. *Adv. Funct. Mater.* **34**, 2312885 (2024). <https://doi.org/10.1002/adfm.202312885>
31. T. Kim, Y. Shin, K. Kang, K. Kim, G. Kim et al., Ultrathin crystalline-silicon-based strain gauges with deep learning algorithms for silent speech interfaces. *Nat. Commun.* **13**, 5815 (2022). <https://doi.org/10.1038/s41467-022-33457-9>
32. H. Zhang, H. Li, Y. Li, Biomimetic electronic skin for robots aiming at superior dynamic-static perception and material cognition based on triboelectric-piezoresistive effects. *Nano Lett.* **24**, 4002 (2024). <https://doi.org/10.1021/acs.nanolett.4c00623>
33. F. Wen, Z. Zhang, T. He, C. Lee, AI enabled sign language recognition and VR space bidirectional communication using triboelectric smart glove. *Nat. Commun.* **12**, 5378 (2021). <https://doi.org/10.1038/s41467-021-25637-w>
34. M.L. Hammock, A. Chortos, B.C.-K. Tee, J.B.-H. Tok, Z. Bao, 25th Anniversary article: the evolution of electronic skin (e-skin): a brief history, design considerations, and recent progress. *Adv. Mater.* **25**, 5997 (2013). <https://doi.org/10.1002/adma.201302240>
35. T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in *Proceedings of the 37th International Conference on Machine Learning (PMLR)* (2020), pp. 1597–1607
36. J.-B. Grill, F. Strub, F. Altché, C. Tallec, P.H. Richemond et al., Bootstrap your own latent—a new approach to self-supervised learning. *arXiv*, 2006.07733. <https://doi.org/10.48550/arXiv.2006.07733>
37. E. Eldele, M. Ragab, Z. Chen, M. Wu, C.-K. Kwoh et al., Self-supervised contrastive representation learning for semi-supervised time-series classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 15604 (2023). <https://doi.org/10.1109/TPAMI.2023.3308189>
38. B.K. Iwana, S. Uchida, Time series data augmentation for neural networks by time warping with a discriminative teacher, in *2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy* (2021), pp. 3558–3565
39. M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski et al., Unsupervised learning of visual features by contrasting cluster assignments, in *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)* (2020), pp. 9912–9924
40. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv*, 1810.04805 (2018). <https://doi.org/10.48550/arXiv.1810.04805>
41. Y. Chang, L. Wang, R. Li, Z. Zhang, Q. Wang et al., First decade of interfacial iontronic sensing: from droplet sensors to artificial skins. *Adv. Mater.* **33**, 2003464 (2021). <https://doi.org/10.1002/adma.202003464>
42. J. Shi, Y. Dai, Y. Cheng, S. Xie, G. Li et al., Embedment of sensing elements for robust, highly sensitive, and cross-talk-free iontronic skins for robotics applications. *Sci. Adv.* **9**, eadf8831 (2023). <https://doi.org/10.1126/sciadv.adf8831>
43. W. Cheng, J. Wang, Z. Ma, K. Yan, Y. Wang et al., Flexible pressure sensor with high sensitivity and low hysteresis based on a hierarchically microstructured electrode. *IEEE Electron Device Lett.* **39**, 288 (2018). <https://doi.org/10.1109/LED.2017.2784538>
44. B. Ji, Q. Zhou, M. Lei, S. Ding, Q. Song et al., Gradient architecture-enabled capacitive tactile sensor with high sensitivity and ultrabroad linearity range. *Small* **17**, 2103312 (2021). <https://doi.org/10.1002/sml.202103312>
45. Z. Wang, W. Yan, T. Oates, Time series classification from scratch with deep neural networks: a strong baseline, in *2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA* (2017), pp. 1578–1585. <https://doi.org/10.1109/IJCNN.2017.7966039>
46. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones et al., Attention is all you need, in *Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, United States* (2017)
47. Q. Wang, B. Li, T. Xiao, J. Zhu, C. Li et al., Learning deep transformer models for machine translation, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy* (2019), pp. 1810–1822. <https://doi.org/10.18653/v1/P19-1176>
48. G.S. Dhillon, P. Chaudhari, A. Ravichandran, S. Soatto, A baseline for few-shot image classification. *arXiv*, 1909.02729 (2019). <https://doi.org/10.48550/arXiv.1909.02729>
49. W.-Y. Chen, Y.-C. Liu, Z. Kira, Y.-C. F. Wang, J.-B. Huang, A closer look at few-shot classification. *arXiv*, 1904.04232 (2019). <https://doi.org/10.48550/arXiv.1904.04232>
50. J. Soni, N. Prabakar, H. Upadhyay, Visualizing high-dimensional data using t-distributed stochastic neighbor embedding algorithm, in *Principles of Data Science* (Springer, Cham, 2020), pp. 189–206. [https://doi.org/10.1007/978-3-030-43981-1\\_9](https://doi.org/10.1007/978-3-030-43981-1_9)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.