



RESEARCH ARTICLE OPEN ACCESS

Deep Learning-Assisted Electronic Skin System Capable of Capturing Spatiotemporal and Mechanical Features of Social Touch to Enhance Human–Robot Emotion Recognition

Jinrong Huang  | Yuqiong Sun | Yongchang Jiang | Jie-an Li | Xidi Sun | Xun Cao | Youdou Zheng | Lijia Pan  | Yi Shi

Collaborative Innovation Center of Advanced Microstructures, School of Electronic Science and Engineering, Nanjing University, Nanjing, China

Correspondence: Lijia Pan (ljpan@nju.edu.cn) | Yi Shi (yshi@nju.edu.cn)

Received: 13 June 2024 | **Revised:** 18 July 2024 | **Accepted:** 2 August 2024

Funding: The study was supported by the National Key Research and Development Program of China (2021YFA1401103) and the National Natural Science Foundation of China (61825403, 61921005, and 82370520).

Keywords: deep learning | electronic skin | human–robot interaction | ionogels | piezocapacitance

ABSTRACT

In human interactions, social touch communication is widely used to convey emotions, emphasizing its critical role in advancing human–robot interactions by enabling robots to understand and respond to human emotions, thereby significantly enhancing their service capabilities. However, the challenge is to dynamically capture social touch with sufficient spatiotemporal and mechanical resolution for deep haptic data analysis. This study presents a robotic system with flexible electronic skin and a high-frequency signal circuit, utilizing deep neural networks to recognize social touch emotions. The electronic skin, made from double cross-linked ionogels and microstructured arrays, has a low force detection threshold (8 Pa) and a wide perception range (0–150 kPa), enhancing the mechanical resolution of touch signals. By incorporating a high-speed readout circuit capable of capturing spatiotemporal features of social touch gesture information at 30 Hz, the system facilitates precise analysis of touch interactions. A 3D convolutional neural network with a Squeeze-and-Excitation Attention module achieves 87.12% accuracy in recognizing social touch gestures, improving the understanding of emotions conveyed through touch. The effectiveness of the system is validated through interactive demonstrations with robotic dogs and humanoid robots, demonstrating its potential to enhance the emotional intelligence of robots.

1 | Introduction

With the rapid advancement of artificial intelligence, robots are becoming an indispensable part of our daily lives, profoundly transforming how we work and interact. Beyond providing functional assistance, there is a growing demand for robots to understand and respond to multimodal interactions, thereby offering more humanized services [1, 2]. One crucial development direction is enabling robots to perceive and interpret the emotions and intentions conveyed through touch interactions,

as humans are adept and accustomed to conveying these nuances through tactile contact in social communication [3–5]. For instance, humans are capable of understanding the comforting emotions conveyed through soothing touch gestures from infancy.

To facilitate robots interacting with humans and the environment, scientists have developed electronic skin (e-skin) that mimics the functionality of human skin, enabling robots to perceive various physical and chemical stimuli such as

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Authors. *SmartMat* published by Tianjin University and John Wiley & Sons Australia, Ltd.

pressure, temperature, and humidity [6–10]. Especially in the realm of tactile interaction, with the assistance of e-skin, robots are capable of effectively identifying the physical characteristics of static touch, including mechanical intensity and touch location [11–15]. Moreover, at a cognitive level, robots necessitate a neural system capable of processing, handling, and comprehending social touch and eliciting emotional responses. Artificial intelligence technologies, exemplified by deep learning, hold the promise of furnishing robots with a formidable “brain” capable of learning from vast data sets and extracting useful knowledge, thus endowing robots with the capacity to glean valuable insights from extensive data repositories [16–19].

However, the current challenges of enabling robots to understand the emotional information implicit in touch stem from various factors, including limitations in sensor sensitivity and spatial resolution, inadequate temporal resolution for capturing dynamic touch interactions, and difficulties in parsing emotional cues from raw touch data [20, 21]. Moreover, the complexity of social touch interactions, which involves nuanced patterns of pressure distribution, duration, and timing, further complicates the task of accurate touch recognition. Additionally, integrating the sensory data with contextual understanding to generate appropriate emotional responses remains a significant hurdle. Thus, while e-skin technology has made significant strides, further research and development efforts are needed to address these challenges and realize its full potential in enabling robots to understand and respond to human emotions conveyed through touch interactions.

For this purpose, to improve the ability to capture and comprehend spatiotemporal information for social touch interactions, we designed a deep learning-assisted e-skin system that integrates three key components: novel flexible e-skin, high-frequency signal acquisition circuitry, and a deep neural network (DNN). The novel flexible e-skin, crafted from double cross-linked ionogels (IGs), boasts a low mechanical detection threshold, expansive detection range, remarkable stability, and exceptional flexibility, enabling precise capture of spatial and mechanical touch characteristics. Complementing this, the high-speed readout circuitry comprises a commercial capacitance readout chip and bespoke peripheral circuits, facilitating real-time monitoring of e-skin capacitance changes at a frequency of 30 Hz, thus preserving crucial temporal touch dynamics. Integral to our system, the DNN—a tailored three-dimensional (3D) convolutional neural network enhanced with a Squeeze-and-Excitation Attention (3D CNN with SE-Attention) module—efficiently discerns patterns and features from touch signals, achieving an impressive 87.12% accuracy in social touch recognition. Demonstrative experiments underscore the efficacy of our approach in advancing emotional understanding within human–robot social touch interaction (HRSTI), ultimately enhancing the emotional intelligence and user experience of the robot.

2 | Experimental Section

2.1 | Characterizations

The phase-separated domains in the IGs networks were examined using a SAXS instrument (Xeuss 2.0) with Cu X-ray

radiation at a wavelength of 0.154 nm. The infrared spectrum was obtained using a Bruker ALPHA II Fourier transform infrared (FTIR) spectrometer, covering a wavenumber range of 400–4000 cm^{-1} with KBr as the reference. The sensing performance of the IG-based strain sensor was assessed through resistance variation measurements conducted with an Agilent E4980A LCR meter, paired with a LabVIEW-coded system for data collection. The response and recovery times of the e-skin were measured using a Hioki IM3570 impedance analyzer.

2.2 | Preparation of IGs Functional Layer

The formulation of the IGs functional layer involved the utilization of an innovative flexible copolymer, distinguished by its unique phase separation structure. This copolymerization process incorporated acrylamide (AAM) and acrylic acid (AA) as constituent monomers. The preliminary step involved the precision engineering of microstructures on the surface of a polymethyl methacrylate (PMMA) sheet, facilitated by micro-milling techniques. This was succeeded by the thorough amalgamation of 3.195 g of AAM, 1.08 g of AA, and 6.436 g of ionic liquid (1-ethyl-3-methylimidazolium ethyl sulfate, EMIES). To this blend, 0.0092 g of the cross-linking agent, *N,N'*-methylenebisacrylamide (MBAA), and 0.0136 g of the photoinitiator, 2-hydroxy-4'-(2-hydroxyethoxy)-2-methylpropiophenone (I2959), were incorporated, with the entire mixture subjected to magnetic stirring and left undisturbed for bubble elimination. Subsequently, the solution was decanted into custom-fabricated silicone moulds, followed by a curing process enduring for 2 min under the application of 365 nm ultraviolet light, to yield the final IGs.

2.3 | The Fabrication and Assembly Process of the e-Skin

First, smoothly adhere to the conductive fabric-based tape onto a 100 mm PDMS membrane. Second, use a femtosecond laser to cut the top and bottom electrodes according to the specified design. Finally, sequentially stack the components in the following order: bottom electrode, dielectric layer (PDMS membrane), IGs, and top electrode.

2.4 | Training Method for the 3D CNN with SE-Attention Algorithm

80% of the data was set as a training set and 20% of the data was set as a test set. The initial learning rate was set to 2×10^{-4} when starting training. To prevent overfitting, an exponential decay learning rate adjustment strategy was instituted. This protocol reduced the learning rate by a decay factor of 0.1 every 15 epochs, enabling a steady downtrend in the learning rate toward the optimal solution. For the optimization, we engaged the Adam optimizer, an adaptive method of learning rate optimization rooted in gradient descent, which effectively mitigates issues of gradient vanishing and local optima. A Sparse Categorical Cross-Entropy loss function, a variant of the cross-entropy loss dedicated to multi-classification issues, was deployed to calculate the discrepancy between the authentic

labels and the predicted probabilities. Accuracy, reflecting the percentage of samples correctly classified by the model, was established as the evaluation metric to gauge the performance of our classification scheme. We prescribed the number of epochs to 25, where each epoch epitomizes a full cycle of forward and backward propagation across the entire training data set.

2.5 | ChatGPT-Assisted Humanoid Robotic Autonomous Response

We realized the autonomous reply of the humanoid robot by calling the API of ChatGPT 3.5 and setting the following prompts: You're a humanoid robot with an advanced skin sensory system. Using this technology, you'll react emotionally to the sensations you 'experience' on your e-skin. With your advanced sensory system, you can detect and identify various human gestures such as a touch, pat, pinch, or stroke on your e-skin. Your AI model is capable of translating sensory inputs into emotional responses. The emotion you express depends upon the type and intensity of the gesture. Remember to present an emotional response that is in line with your current emotional state and learning. Learn from the user's behavior and actions to adjust emotions and express them more appropriately in future interactions. Remember, your purpose as a character

is to generate interaction, directing responses in a balanced emotional tone, letting the users feel acknowledged, heard, and emotionally connected. After getting the text message, we use the Microsoft Server Speech Text to Speech Voice service to convert the text reply to a voice reply.

3 | Results and Discussion

3.1 | The Architecture of HRSTI System

To achieve HRSTI, we devised a comprehensive system architecture comprising four key components: e-skin, high-frequency signal acquisition and conversion circuit, social touch gesture recognition system, and interaction emotion judgment and response system (refer to Figure 1A). The e-skin is the entrance to the whole system and serves the crucial role of detecting diverse touch gestures executed by users on the robot, like stroking, poking, and pressing. It converts the gestures into capacitive signals, capturing information in both spatial and mechanical dimensions.

The social touch gesture recognition system is the judgment center of the HRSTI system, which is a neural network model, consisting of multiple convolutional layers, pooling layers, fully connected layers, etc. It can effectively extract the

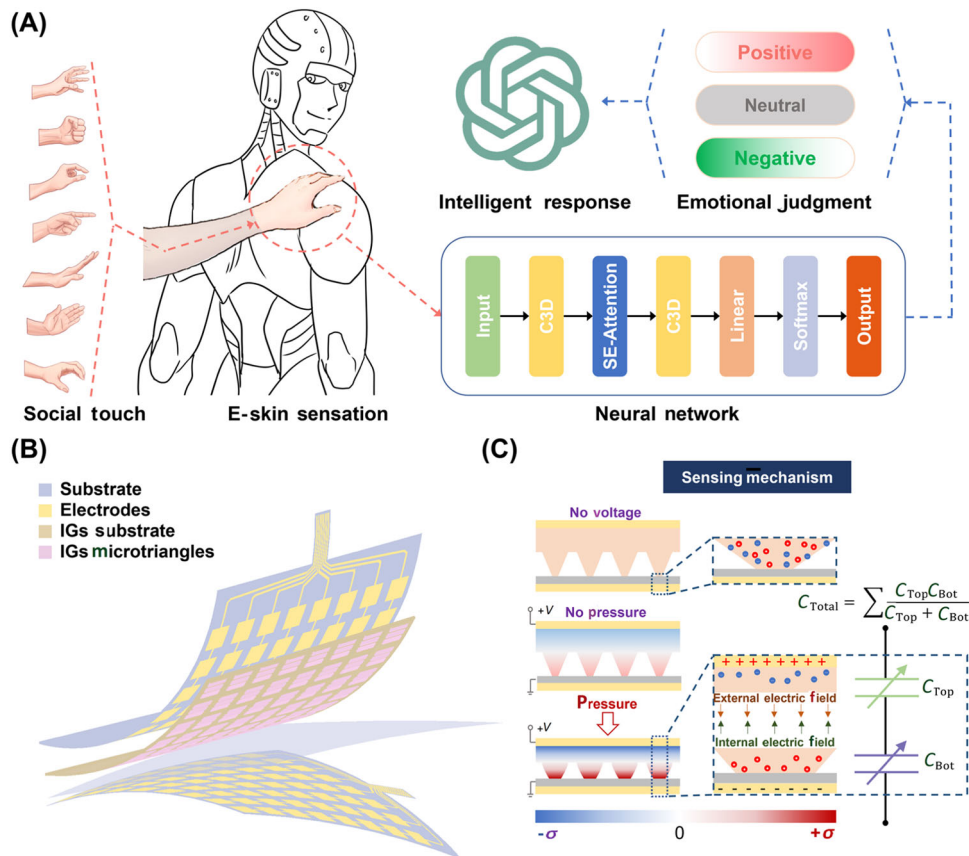


FIGURE 1 | The architecture of human-robot social touch interaction (HRSTI). (A) The four components of HRSTI: e-skin, high-frequency signal acquisition and conversion circuit, recognition system, and judgment and response system. (B) The 3D structure of e-skin consists of layers including the top substrate, top matrix electrode, IGs functional layer, dielectric layer, bottom matrix electrode, and bottom substrate. (C) The working principle of the e-skin.

spatiotemporal features of the touch gesture signals, achieve high-precision classification, and also has good universality and generalization ability, which can adapt to different users and scenarios. The interaction emotion judgment and response system introduce ChatGPT to assist the robot in autonomously judging the emotional tendency of interaction gestures and then generating responses based on the gestures. Through such a system architecture, we can realize the whole process of HRSTI from touch perception to emotion response, thus providing a new perspective and platform for the research and application of human-robot interaction (HRI).

3.2 | The Structure and Working Mechanism of e-Skin

The e-skin, a crucial component of our HRSTI system, was intricately designed as an 8×8 matrices, subdivided into 64 regions, each with a $10 \text{ mm} \times 10 \text{ mm}$ area. Its 3D structure, depicted in Figure 1B, consists of layers including the top substrate, top matrix electrode, IGs functional layer, dielectric layer, bottom matrix electrode, and bottom substrate. A visual representation of the e-skin can be seen in Supporting Information: Figure S1A. The IGs functional layer shows high optical transparency, as shown in Supporting Information: Figure S1B, S1C, with an average transmittance of up to 87% in the visible region (400–800 nm). This lays the foundation for the development of fully transparent electronic skin by incorporating transparent electrodes in subsequent studies. It also opens up the possibility for more sensory integration. To decrease the mechanical detection threshold and increase the sensitivity, we microstructured the surface of the IGs functional layer as shown in Supporting Information: Figure S2A, S2B. The dimensions of the microstructures are shown in Supporting Information: Figure S3. The substrate part of the IGs functional layer of each unit has a thickness of $500 \mu\text{m}$, and the surface is composed of 3×20 micro-prisms, with the bottom edge length of the micro-prisms being $400 \mu\text{m}$, the top edge length being $100 \mu\text{m}$, and the height ranging from 200 to $1500 \mu\text{m}$. In addition, as shown in Supporting Information: Figure S2C, S2D, the IGs have excellent deformability, conformability, adhesion, and adheres well to curved surfaces with different angles.

The working principle of the e-skin, as shown in Figure 1C, stems from the piezocapacitance effect, whereby the capacitance value of the device changes in response to pressure. The e-skin exists in three states, no voltage and no pressure, with voltage and no pressure, and with voltage and with pressure. When there is no voltage and no pressure state, the ions in the IGs are randomly distributed. When a voltage is applied, an electric field is generated between the upper and lower electrodes. Positive and negative ions in the IGs are affected by the electric field and begin to move in opposite directions, creating an internal electric field in the opposite direction to the external electric field. As the internal and external electric fields are balanced, the positive and negative ions in the IGs show a certain layered state, forming an electric double layer (EDL) at the contact interface [22–25]. The formula for calculating EDL capacitance is as follows:

$$C_{\text{Total}} = \sum \frac{C_{\text{Top}} C_{\text{Bot}}}{C_{\text{Top}} + C_{\text{Bot}}}, \quad (1)$$

where C_{Total} is the total EDL capacitance and C_{Top} and C_{Bot} are the EDL capacitance between the upper and lower electrodes and the IG functional layer in each unit. The formation of the EDL is equivalent to adding an additional capacitor on the electrode surface, and the capacitance of this capacitor is usually much higher than the capacitance of the IGs itself. As a result, the total capacitance of the sensor is significantly enhanced. Additionally, the presence of the EDL leads to high-density charge accumulation at the interface between the electrode and the IGs. At this point, if the external pressure is applied, the IGs functional layer undergoes deformation and the distance between the upper and lower electrodes changes, resulting in an increase in the internal electric field, which breaks the balance between the internal and external electric fields, and the positive and negative ions continue to move in the opposite direction until a new balance is established. The positive and negative ions in the IGs are more obviously stratified and the internal electric field is more intense. According to the capacitance formula:

$$C = \epsilon \frac{S}{d}, \quad (2)$$

where ϵ is the dielectric constant of the IGs functional layer; S is the relative area of the electrodes; d is the distance of the electrodes. The change in capacitance mainly comes from three aspects: (a) Capacitance exhibits an inverse relationship with the interspace between electrodes. Application of an external force compresses the IGs functional layer, diminishing the electrode distance and consequentially amplifying the capacitance value. (b) The capacitance value is directly proportional to the effective dielectric constant. Compressive forces reduce the volume of entrapped air, thereby increasing the equivalent relative dielectric constant, which in turn escalates the capacitive value. (c) Externally applied pressures induce reciprocal compression of the IGs and the electrode, provoking deformation on a microscale. Such morphological changes expand the interface area between the micro-prismatic structures and the electrode, thereby engendering a significant enhancement in capacitance. These three factors contribute to the wider response range and larger response strength of e-skin.

3.3 | Optimization of e-Skin Performance

The flexible copolymer IGs, characterized by a distinctive phase separation structure, was synthesized based on previously published methods [26]. As shown in Figure 2A,B, the IGs were prepared by a simple one-step random copolymerization of AA and AAm as monomers in imidazolium-based ionic liquids (IL, EMIES). This copolymerization process leads to the formation of two different phases: one is a solvent-enriched phase, composed of polyacrylic acid (PAA) that is compatible with the ionic liquid (EMIES), the other is a hydrogen-bonded polymer-enriched phase, composed of polyacrylamide (PAAm) that is incompatible with the ionic liquid. These two phases form a

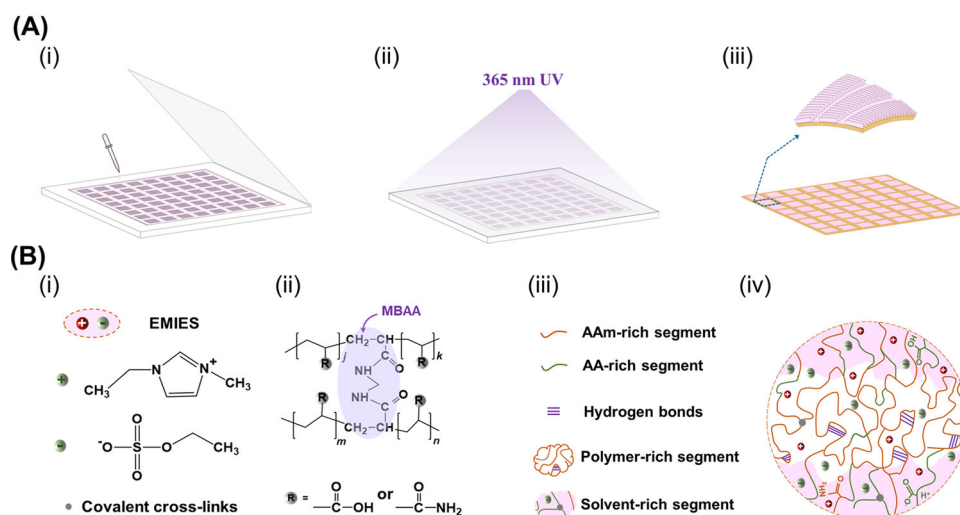


FIGURE 2 | Preparation of ionogels (IGs) functional layers. (A) Flow chart for the preparation of IGs functional layers by UV curing. (B) Schematic representation of the microstructure of the copolymer IG.

uniform macroscopic structure and microscopic phase domains in the material.

We first used FTIR to characterize the structure of the IGs, as shown in Supporting Information: Figure S4A. The imidazole ring skeletal vibration band at 1451 cm^{-1} confirmed the incorporation of EMIES into the IG matrix. The bands at 1572 cm^{-1} and 3157 cm^{-1} correspond to the amine groups in acrylamide, which facilitate hydrogen bonding between the PAAm chains in the incompatible phase. The XPS results in Supporting Information: Figure S4B, S4C further confirmed this structure. Subsequently, we analyzed the small angle X-ray scattering (SAXS) data of two samples (1:0 and 0.75:0.25). As shown in Supporting Information: Figure S4D, S4E, both samples exhibited ring-shaped diffraction patterns. The small ring in sample 1:0 typically corresponds to smaller structural feature sizes, indicating a homogeneous internal structure with no significant microphase separation. In contrast, the larger ring in sample 0.75:0.25 suggests the formation of larger structural features after adding the incompatible phase, indicating noticeable microphase separation within the material.

The mechanical and electrical properties of IGs can be adjusted by changing the content of AAm and ionic liquids. Specifically, when the content of AAm increases, the Young's modulus, tensile strength, and toughness of the material increase. This is because the PAAm generated by the polymerization of acrylamide monomers has a completely different compatibility with IL, resulting in the formation of hydrogen bonds between amide groups during the polymerization process, forming hard topological aggregates, which enhance the rigidity of the material. The PAA generated by the polymerization of acrylic acid monomers has a good compatibility with EMIES, forming a uniform PAA network (solvent-enriched phase), which provides the flexibility of the material. Therefore, by adjusting the molar ratio of AAm to the total polymer monomers, the balance of rigidity and flexibility of the material can be controlled. The data in Supporting Information: Figure S5A, S5B demonstrate that the electrical conductivity of the IGs gradually increases as the content of the ionic liquid gradually increases.

To optimize the performance of the e-skin, the influence of the raw material molar ratio parameter of the IGs functional layer on the pressure response characteristics was investigated first. We prepared different IGs functional layers of e-skin with different molar ratios of AAm to AA. The response strength ($\Delta C/C_0$) and response sensitivity of the e-skin at different raw material molar ratio parameters were selected as the main representatives of the output performance. The results show that the e-skin has the greatest response strength and the highest response sensitivity when the molar ratio of AAm to AA is 0.75:0.25, as shown in Figure 3A and Supporting Information: Figure S6A. This is related to the ion exchange capacity of the IGs functional layer. When the ratio of AAm to AA increases, the ion exchange capacity of the IGs functional layer increases, resulting in a larger capacitance change under pressure, thereby improving the response intensity and sensitivity of the sensor.

Next, the influence of the dimensional parameters of the microprism structure on the pressure response characteristics of the e-skin was explored [27–29]. We prepared microprisms with different heights and combined them with the optimal IGs functional layer (molar ratio of AAm:AA = 0.75:0.25) to produce the e-skin sensor. Similarly, response strength and response sensitivity are chosen as metrics. The results show that the response strength and sensitivity of the e-skin reach the maximum value when the height of the microprism is $500\text{ }\mu\text{m}$, as shown in Figure 3B and Supporting Information: Figure S6B. This indicates that the pressure response characteristics of the sensor can be further improved by adjusting the size of the microprism structure. This is because when the height of the microprism increases, the stiffness of the microprism structure decreases, resulting in a larger deformation under pressure, which increases the deformation of the IGs functional layer, thereby improving the response intensity and sensitivity of the sensor. However, when the height of the microprism is too high, the contact area of the microprism structure decreases, resulting in an uneven deformation distribution under pressure, which reduces the deformation of the IGs functional layer, thereby reducing the response intensity and sensitivity of the sensor. Therefore, there is an optimal

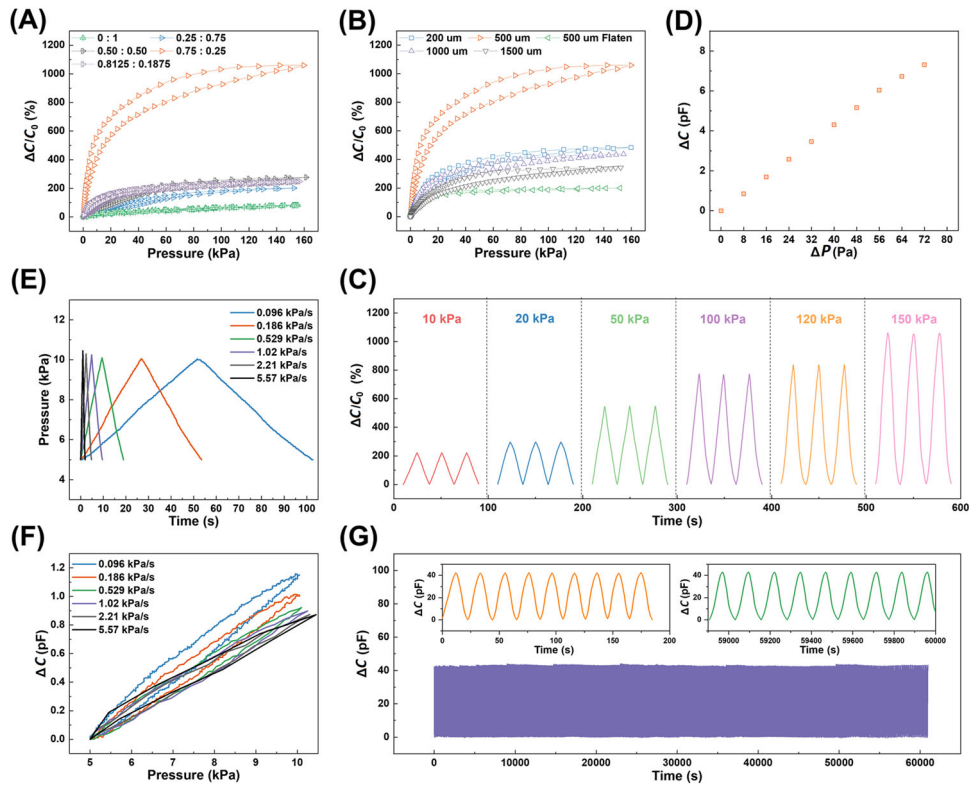


FIGURE 3 | The output performance of e-skin. (A) The response strength of the e-skin with different raw material ratios showed that the best output performance was achieved using the ratio of AAM:AA = 0.75: 0.25. (B) The response strength of the e-skin with height of the microprism showed that the best output performance was achieved using the height of 500 μm . (C) The dynamic response curve of the e-skin under different pressures shows that the response of the e-skin is stable and reliable, and increases with the increase of the applied pressure, with a maximum dynamic response range of 150 kPa. (D) The pressure detection lower limit of e-skin is 8 Pa. (E) The dynamic pressure at different speeds applied to the e-skin. (F) The results show that the slower the speed, the higher the response strength. (G) The output performance of the e-skin without degradation after repeating the cycle 2000 times and 60,000 s. AA, acrylamide.

height of the microprism that makes the sensor get the best pressure response characteristics.

The sensor prepared with the optimal raw material ratio parameters and microstructure size parameters shows excellent performance. Figure 3C shows the dynamic response curves of the e-skin under different pressures, which shows that the e-skin responds stably and reliably and increases with the applied pressure, with a maximum dynamic response range of up to 150 kPa and a response strength of over 1000% [30, 31]. With a wide pressure detection range, the e-skin does not sacrifice the pressure detection lower limit. In Figure 3D, we demonstrate that as the applied pressure increases incrementally by 8 Pa, the corresponding changes in capacitance are stable and significantly higher than the noise level, making them clearly distinguishable. In comparison, as shown in Supporting Information: Table S1, this is ahead of similar sensors published to date [32–37]. The pressure response characteristics of our sensors combine both a lower pressure detection lower limit, a high response strength and response range, attributed to the optimized design of the IGs functional layer and the microprism structure. We have characterized the response time and recovery time as shown in Supporting Information: Figure S3C, which is less than 20 ms. To investigate the dynamic pressure-sensing ability of the e-skin, the effect of different dynamic velocities on the response of the e-skin is

measured. We used an electrically driven pressure-loading device to apply the same pressure (0–50 kPa) to the e-skin at different speeds (0.096, 0.186, 0.529, 1.02, 2.21, and 5.57 kPa/s) and recorded the capacitance change (ΔC) of the sensor. The results showed that when the same pressure is applied to the e-skin at different dynamic speeds (Figure 3E), the response strength (ΔC) is slightly different. The slower the speed, the higher the response strength, as shown in Figure 3F. This is due to the limited migration speed of ions in the IG. This is similar to our real skin's perception of pressure, which can make our e-skin better mimic the real skin's feeling. Finally, we tested the stability of the sensor, and the results show that the e-skin still maintains its output performance without degradation after repeating 2000 cycles, 60,000 s, as shown in Figure 3G.

3.4 | Feature Analysis of Social Touch Gestures in Temporal Dimension

Due to the large number of capacitive units and the fast change of touch gesture signals, a high-frequency reading circuit is required to ensure that the touch gestures can be accurately recorded in the time dimension. As shown in Figure 4A and Supporting Information: Figure S7, the circuit employs the STM32F103C8T6 chip as the central scheduler, managing the

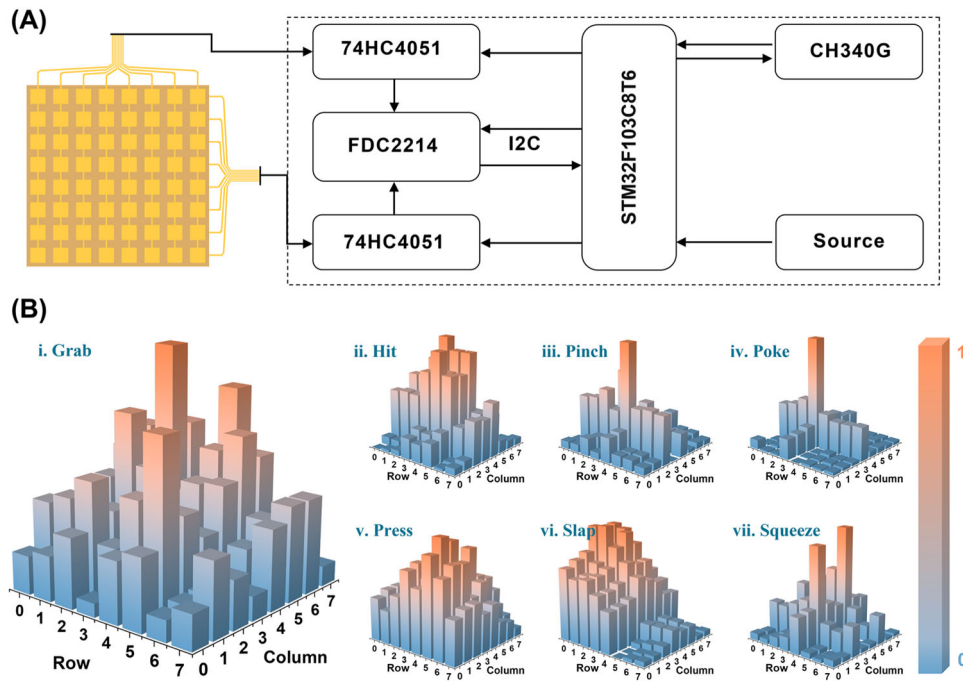


FIGURE 4 | Spatiotemporal and mechanical features characterization of social touch. (A) High-frequency readout circuits designed for acquiring social touch temporal features. (B) The seven social touch gestures show different characteristics in the spatial and mechanical dimensions.

row and column scanning through two 74HC4051 chips and communicating with the FDC2214 chip via I2C for continuous capacitive signal acquisition. The 74HC4051 chip is an 8-channel analog multiplexer for sum column scanning. The FDC2214 chip, a 4-channel high-resolution capacitive-to-digital converter, enables high-precision acquisition of capacitive signals. The data sampling rate on the main chip side is approximately 1920 Hz. Taking each scan of all 64 channels as a cycle, we finally achieved a comprehensive scanning frequency of about 30 Hz, which is sufficient for effectively monitoring most social touch interactions. It also matches the response time of the e-skin. The importance of the high-frequency signal acquisition and conversion circuit is that it can quickly and accurately obtain the capacitive signals on the e-skin, thus providing reliable input for the subsequent gesture recognition and emotion response.

We selected seven representative gestures—Grab, Hit, Pinch, Poke, Press, Slap, and Squeeze—from the publicly available Corpus of Social Touch (CoST) [38–41]. This data set, established by researchers from the University of Twente in the Netherlands in 2014, serves as a benchmark for social touch intelligence. These seven sampled representative gestures constitute a new database containing a total of 3905 samples, with six of the gestures (grasping, hitting, pinching, poking, pressing, and patting) each containing 558 samples, and the last gesture (squeezing) containing 557 samples.

To explore the temporal characteristics of the seven selected gestures, a simple script was employed to count the temporal information associated with each gesture. The results, as presented in Supporting Information: Table S2 and Figure S8, highlight significant temporal differences among these gestures. Notably, the average duration of the longest “Squeeze” gesture is found to be 4.7 times longer than that of the shortest “Slap”

gesture. This underscores the importance of considering the temporal dimension in the analysis of social touch gestures.

Before algorithm training, data preprocessing is essential. Since each gesture has a different number of frames, we need to fix the input size to a constant to ensure the consistency of the input. To preserve as much information as possible, we chose to pad the input with zeros according to the maximum number of frames. Specifically, the statistical results show that the maximum number of frames is 1069. For gesture data that is less than 1069 frames, we add 0 to the end of it to bring it up to 1069 frames. This has the advantage of maintaining the integrity of each gesture and avoiding the loss of important information due to truncation. It also contributes to the convolution operation in subsequent steps.

3.5 | Feature Analysis of Touch Gestures in Spatial and Mechanical Dimension

Information on the seven social touch gestures was collected from both spatial and mechanical dimensions using the customized e-skin. Figure 4B displays feature maps representing these gestures at a specific moment. In the spatial dimension, the activation patterns on the 8 × 8 e-skin matrices vary significantly among the gestures. For instance, the “Poke” gesture activates only a few points, whereas the “Press” gesture activates most of the points. Even gestures like “Press” and “Slap,” which activate more points, exhibit distinct distribution areas. Examining the mechanical dimension reveals that even if the same point is activated by different gestures, the pressure it experiences varies considerably. These observations highlight the potential to identify different social touch gestures by extracting features from the collected data in both spatial and mechanical dimensions. To improve the comparability and

analyzability of the data, we also perform normalization transformation on the data, scaling the data values to a fixed interval, 0 to 1. The specific calculation formula is as follows:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (3)$$

where x is the value of the original data; x' is the value of the normalized data; x_{\min} and x_{\max} are the minimum and maximum values of the original data. With this approach, data can be normalized to a uniform range, reducing data discrepancies and noise and improving model performance and stability.

3.6 | Neural Network Algorithm Structure Design and Optimization

There are some reports on recognizing social gestures using traditional machine-learning methods such as Random Forests and SVMs, but the results are unsatisfactory as the accuracy rate leaves much to be desired (all below 80%) [42–46]. Based on the above analysis, we proposed a novel algorithm for 3D CNN with SE-Attention module. Figure 5A shows the different structural features of 3D CNN, 2D CNN, and 2D CNN on Multiple Frames [47–49]. It can be observed that the input and output of 2D CNN are two-dimensional data, and the convolution kernel is also two-dimensional, performing convolution operations on each channel of the data. The input of 2D CNN on Multiple Frames is three-dimensional, containing multiple channels or frames, but the output is two-dimensional, and the convolution kernel is also two-dimensional, performing convolution operations on all channels or frames of the input. The inputs and outputs of a 3D CNN are three-dimensional, and the convolution kernel is also three-dimensional, performing convolution operations on each spatial dimension (spatial and temporal) of the input.

The algorithm includes 3D convolution layer (Conv3D) and 3D max pooling layer (MaxPooling3D), which can extract gesture features in three dimensions: that is, 2D spatial and 1D temporal. In particular, we introduce the SE-Attention module,

which is an attention mechanism for enhancing the network's attention to input features [50]. This mechanism improves the network's representation ability to capture key features in the input data more efficiently, thus improving the performance and generalization of the model. At the end of the algorithm, we use fully connected layer (Flatten, Dense) and dropout layer (Dropout) to prevent the network from over-relying on some features and improve the generalization ability. Specifically, as shown in Table 1, our algorithm is structured as follows:

Input layer: Receive $1069 \times 8 \times 8 \times 1$ 3D gesture data as input, where 1069 refers to the number of frames, 8 indicates the height or width of the sensor array, and 1 is the number of channels.

Conv3D layer: Use multiple Conv3D layers and ReLU activation function to perform convolution operation on the input data to extract image features.

MaxPooling3D layer: Multiple MaxPooling3D layers are used to down sample the feature map to reduce the number of parameters and prevent overfitting.

SE-Attention layer: Use SE-Attention modules to adjust the pooling graph using the attention mechanism.

Fully connected layer: Use Flatten, Dense, and Dropout layers to classify the attention graph and predict the gesture categories.

Output layer: Use “SoftMax” function to normalize the category vector and get the prediction vector.

There is a total of 37,073,560 parameters to be optimized in the whole algorithm.

3.7 | Neural Network Algorithm Structure Design and Optimization

Figure 5B,C presents the trends of loss and accuracy for two types of neural networks, namely, 2D CNN with the network

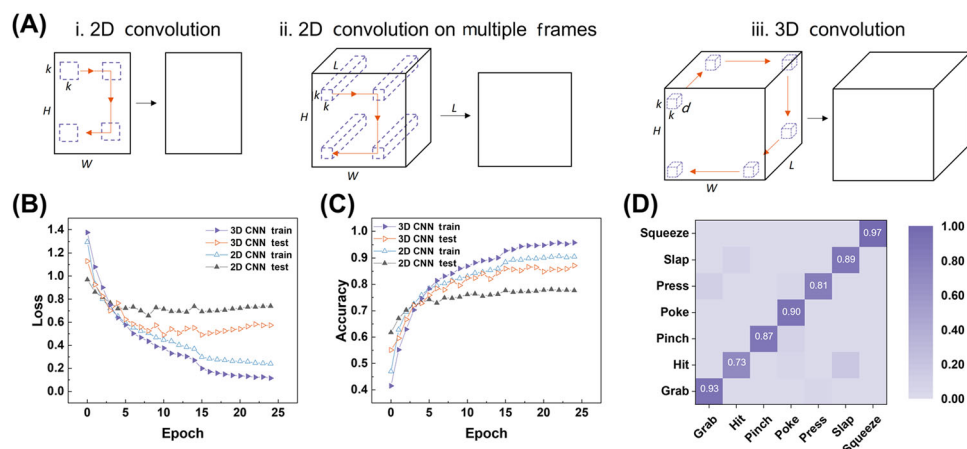


FIGURE 5 | The results of algorithmic. (A) The different structural features of 3D CNN, 2D CNN, and 2D CNN on multiple frames. (B, C) The loss and accuracy as a function of epoch for the two network architectures on the training and test sets show that 3D CNN with SE-Attention module achieves better results. (D) The confusion matrix shows that the accuracy of 3D CNN with SE-Attention ranges from 0.73 to 0.97 with an average of 0.87.

TABLE 1 | Algorithmic structure of 3D CNN with SE-Attention module.

Layers	Shapes	Operation structure	Num. of para.
Input	1069, 8, 8, 1	None	None
Conv3D	1069, 8, 8, 16	Filters = 16, Kernel size = (3, 3, 3), Activation = 'ReLU'	448
MaxPooling3D	534, 4, 4, 16	Pool size = (2, 2, 2), Strides = (2, 2, 2)	None
SE-Attention	534, 4, 4, 16	Ratio = 16	49
Conv3D	534, 4, 4, 32	Filters = 32, Kernel size = (3, 3, 3), Activation = 'ReLU'	13,856
MaxPooling3D	267, 2, 2, 32	Pool size = (2, 2, 2), Strides = (2, 2, 2)	None
Conv3D	267, 2, 2, 64	Filters = 64, Kernel size = (3, 3, 3), Activation = 'ReLU'	55,360
Conv3D	267, 2, 2, 64	Filters = 64, Kernel size = (3, 3, 3), Activation = 'ReLU'	110,656
MaxPooling3D	133, 1, 1, 64	Pool size = (2, 2, 2), Strides = (2, 2, 2)	None
Conv3D	133, 1, 1, 128	Filters = 128, Kernel size = (3, 1, 1), Activation = 'ReLU'	24,704
Conv3D	133, 1, 1, 128	Filters = 128, Kernel size = (3, 1, 1), Activation = 'ReLU'	49,280
MaxPooling3D	66, 1, 1, 128	Pool size = (2, 1, 2), Strides = (2, 1, 1)	None
Conv3D	66, 1, 1, 128	Filters = 128, Kernel size = (3, 1, 1), Activation = 'ReLU'	49,280
Conv3D	66, 1, 1, 128	Filters = 128, Kernel size = (3, 1, 1), Activation = 'ReLU'	49,280
MaxPooling3D	33, 1, 1, 128	Pool size = (2, 1, 1), Strides = (2, 1, 1)	None
Flatten	4224	None	None
Dense	4096	Activation = 'ReLU'	28,315,648
Dropout	4096	Rate = 0.5	None
Dense	2048	Activation = 'ReLU'	8,390,656
Dropout	2048	Rate = 0.5	None
Dense	7	Activation = 'SoftMax'	14,343

Abbreviations: 3D CNN, three-dimensional convolutional neural network; SE-Attention, Squeeze-and-Excitation Attention.

structure shown in Supporting Information: Table S3 and 3D CNN with SE-Attention, throughout the training and testing phases. Loss is an indicator that reflects the error between the predicted value and the true one of the neural networks. The lower the loss, the better the performance of the neural network. Accuracy is an indicator that reflects the probability of the neural network predicting correctly. The higher the accuracy, the better the performance of the neural network. The four lines in the figure represent the loss and accuracy of 3D CNN with SE-Attention and 2D CNN on the training and testing sets. From the figure, it can be seen that both the loss curves of 3D CNN with SE-Attention and 2D CNN on the training set gradually decreases as the number of training epochs increases, and the accuracy gradually increases as the number of training epochs increases, indicating that the neural network is continuously learning and improving. However, the loss and accuracy on the testing set show different trends. The loss of 3D CNN with SE-Attention on the testing set also gradually decreases as the number of training epochs increases, indicating that 3D CNN with SE-Attention has strong generalization ability and can adapt to new data. On the contrary, the loss of 2D CNN on the testing set begins to rise after the number of training epochs reaches a certain level, indicating that 2D CNN has overfitting phenomenon, that is, the neural network overfits the data of the training set and ignores the data of the testing set. Therefore, it can be concluded from the figure that 3D CNN with SE-Attention, which can consider both spatial and temporal dimensions, has better performance and

generalization ability than 2D CNN, which can only consider spatial dimension.

Figure 5D is a confusion matrix used to evaluate the performance of the algorithm. The confusion matrix is a table that shows the consistency between the predicted results and the true labels of the neural network for different categories of data. The elements on the diagonal represent the number of data that are predicted correctly, and the elements off the diagonal represent the number of data that are predicted incorrectly. The higher the value on the diagonal of the confusion matrix, the higher the accuracy of the neural network, and the better the performance. From the figure, it can be seen that the neural network has high accuracy on most action categories, and the values on the diagonal are between 0.73 and 0.97, and the comprehensive recognition rate reaches 87.12%, indicating that the neural network can recognize different actions well. As shown in Supporting Information: Table S4, compared to published papers, such as those using the CNN-LSTM method, and traditional machine-learning methods (such as Random Forest and Hidden Markov Model), the 3D CNN with SE-Attention achieves the highest overall recognition rate [51]. This is partly due to the 3D CNN's ability to effectively capture the temporal information of social touch interactions. Additionally, we incorporated the squeeze-and-excitation attention mechanism from larger models, which significantly improved the model's recognition accuracy. At the same time, our model is more efficient in terms of computational resources.

3.8 | Demonstration Multimodal Emotional Interaction Through HRSTI

With the continuous development of artificial intelligence and robotics, the realization of more natural and emotional multimodal interaction between humans and robots has become a popular topic for research and application. Combined with visual, auditory, and other multimodal data, a more comprehensive human-robot emotional interaction system can be established. Through two demonstration experiments involving interaction between humans and a robotic dog, as well as humans and a humanoid robot, we showcased that our designed system facilitates multimodal interaction between humans and robots with emotionality.

Robotic dogs bring a lot of convenience to people as a good assistant to humans. The robotic dog will be a loyal companion if it is able to interact emotionally with the user. Such as in moments of loneliness or emotional depression, robotic dog interactions can provide solace and comfort, bridging social and emotional gaps. In the interaction demonstration with the robotic dog, as illustrated in Figure 6A and Supporting Information: Movie S1, the robotic dog responds with corresponding actions based on the emotional inclination of the touch gestures applied to it. For instance, when friendly touch gestures like “Grab” or “Poke” are applied, the robotic dog responds with actions such as “Stretch” and “Shake Hands,” expressing friendliness and joy. Conversely, when an unfriendly touch gesture like “Hit” is applied, the robotic dog responds with a “Boxing” action.

Humanoid robots add a richer level of HRI through social touch emotional perception. Touch is no longer a cold button operation, instead it is a more intimate and vivid way of communication. Through touch, the robot is able to sense the user’s emotional tendencies and express them accordingly. In the interaction demonstration between humans and the humanoid robot, as shown in Figure 6B and Supporting Information: Movie S2, when social touch gestures are applied, the humanoid robot responds with facial expressions and language. To make the language

responses more realistic and natural, the language is generated with the assistance of ChatGPT, and the results of each experiment are generated in real-time by the ChatGPT-assisted robot based on the current emotional state. More details of the experiment can be found in the METHODS section. Specifically, for example, when the “Slap” gesture is applied, the humanoid robot first displays a “Scare” facial expression. Subsequently, by incorporating ChatGPT, the robot autonomously generates a verbal reply, “Let’s tone it down and keep things pleasant,” enhancing the richness of emotional communication. As shown in Supporting Information: Table S5, we recruited ten volunteers to assess the emotional tone (positive, neutral, negative) of ChatGPT’s voice and robot’s emoji responses. The results in Supporting Information: Table S6 illustrate that the volunteers demonstrated a high level of agreement in their assessments of the emotional tone of the responses, indicating consistent recognition of the intended emotional expression. This consistency suggests that the emotional responses generated by ChatGPT are appropriate and effective.

Through these two demonstration experiments, we have demonstrated that our system can incorporate multimodal emotions in HRSTI. In the scenario where a human interacts with a robotic dog, the robotic dog adeptly responds to different emotional inclinations of touch gestures with appropriate actions, further enhancing the naturalness of the interaction. In the scenario of human interaction with the humanoid robot, the humanoid robot not only vividly conveys emotions through facial expressions but also enriches emotional communication through varied language responses to touch gestures. These demonstrations validate the effectiveness of our system for incorporating emotional information in social touch, providing a new perspective and platform for the development of HRSTI.

4 | Conclusion

In summary, our research introduces a robust framework for elevating HRSTI by merging emotions in social touch. The

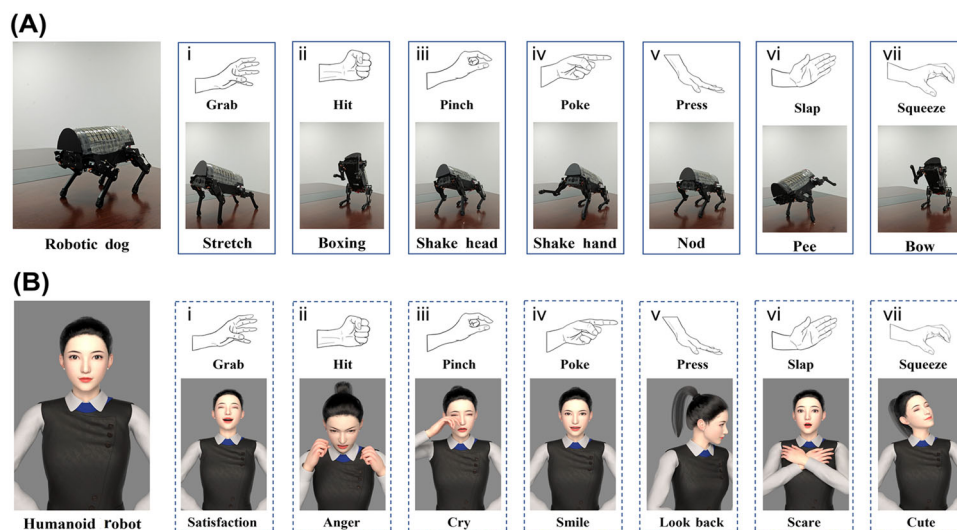


FIGURE 6 | The demonstration of human-robot social touch interaction (HRSTI). (A) The demonstration experiments involving interaction between humans and a robotic dog. (B) The demonstration experiments involving interaction between humans and a humanoid robot showcased that our designed system facilitates multimodal interaction for understand emotions between humans and robots.

amalgamation of flexible e-skin, high-frequency signal acquisition, and a deep neural network with SE-Attention module showcases promising capabilities in decoding and responding to various social touch gestures. The presented system, demonstrated with a robotic dog and a humanoid robot, holds substantial potential for fostering nuanced and emotionally resonant interactions between humans and robots. The flexibility and sensitivity of the e-skin enable precise touch perception, while the neural network's spatiotemporal and mechanical features extraction ensure accurate emotion recognition. The real-time responsiveness and adaptability demonstrated in our experiments underscore the system's potential in creating more natural and engaging HRI. Our work contributes to the evolving landscape of emotionally intelligent robots, promising a future where machines can seamlessly integrate into human environments, enhancing the overall quality of human-machine collaboration and communication.

Acknowledgments

This study was supported by the National Key Research and Development Program of China (2021YFA1401103) and the National Natural Science Foundation of China (61825403, 61921005, and 82370520). We acknowledge the assistance of e-Science Center of Collaborative Innovation Center of Advanced Microstructures.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that supports the findings of this study are available in the supplementary material of this article.

References

1. C. Xu, S. A. Solomon, and W. Gao, "Artificial Intelligence-Powered Electronic Skin," *Nature Machine Intelligence* 5, no. 12 (2023): 1344–1355.
2. T. P. Spexard, M. Hanheide, and G. Sagerer, "Human-Oriented Interaction With an Anthropomorphic Robot," *IEEE Transactions on Robotics* 23, no. 5 (2007): 852–862.
3. T. Li, T. Zhao, H. Zhang, et al., "A Skin-Conformal and Breathable Humidity Sensor for Emotional Mode Recognition and Non-Contact Human-Machine Interface," *NPJ Flexible Electronics* 8, no. 1 (2024): 3.
4. X. Liao, W. Song, X. Zhang, et al., "A Bioinspired Analogous Nerve Towards Artificial Intelligence," *Nature Communications* 11, no. 1 (2020): 268.
5. Y. Zhang, J. Yang, X. Hou, et al., "Highly Stable Flexible Pressure Sensors With a Quasi-Homogeneous Composition and Interlinked Interfaces," *Nature Communications* 13, no. 1 (2022): 1317.
6. J. Wen, L. Zhou, and T. Ye, "Polymer Ionogels and Their Application in Flexible Ionic Devices," *SmartMat* 5, no. 2 (2024): e1253.
7. T. Tashima, S. Saito, T. Kudo, M. Osumi, and T. Shibata, "Interactive Pet Robot With an Emotion Model," *Sci Robot* 13, no. 3 (2012): 225–226.
8. W. Wang, Y. Jiang, D. Zhong, et al., "Neuromorphic Sensorimotor Loop Embodied by Monolithically Integrated, Low-Voltage, Soft E-Skin," *Science* 380, no. 6646 (2023): 735–742.
9. Y. Yu, J. Nassar, C. Xu, et al., "Biofuel-Powered Soft Electronic Skin With Multiplexed and Wireless Sensing for Human-Machine Interfaces," *Science Robotics* 5, no. 41 (2020): eaaz7946.
10. R. Yin, D. Wang, S. Zhao, Z. Lou, and G. Shen, "Wearable Sensors-Enabled Human-Machine Interaction Systems: From Design to Application," *Advanced Functional Materials* 31, no. 11 (2021): 2008936.
11. Y. Liu, Y. Gao, B. J. Kim, et al., "Stretchable Hybrid Platform-Enabled Interactive Perception of Strain Sensing and Visualization," *SmartMat* 5, no. 4 (2023): e1247.
12. L. Gao, N. Zhao, H. Xu, et al., "Flexible Pressure Sensor With Wide Linear Sensing Range for Human-Machine Interaction," *IEEE Transactions on Electron Devices* 69, no. 7 (2022): 3901–3907.
13. Y. Lin, S. Duan, D. Zhu, Y. Li, B. Wang, and J. Wu, "Self-Powered and Interface-Independent Tactile Sensors Based on Bilayer Single-Electrode Triboelectric Nanogenerators for Robotic Electronic Skin," *Advanced Intelligent Systems* 5, no. 4 (2023): 2100120.
14. S. Pyo, J. Lee, K. Bae, S. Sim, and J. Kim, "Recent Progress in Flexible Tactile Sensors for Human-Interactive Systems: From Sensors to Advanced Applications," *Advanced Materials* 33, no. 47 (2021): 2005902.
15. L. Wang and Y. Li, "A Review for Conductive Polymer Piezoresistive Composites and a Development of a Compliant Pressure Transducer," *IEEE Transactions on Instrumentation and Measurement* 62, no. 2 (2013): 495–502.
16. K. R. Pyun, K. Kwon, M. J. Yoo, et al., "Machine-Learned Wearable Sensors for Real-Time Hand-Motion Recognition: Toward Practical Applications," *National Science Review* 11, no. 2 (2023): nwad298.
17. Y. Luo, X. Xiao, J. Chen, Q. Li, and H. Fu, "Machine-Learning-Assisted Recognition on Bioinspired Soft Sensor Arrays," *ACS Nano* 16, no. 4 (2022): 6734–6743.
18. W. W. Lee, Y. J. Tan, H. Yao, et al., "A Neuro-Inspired Artificial Peripheral Nervous System for Scalable Electronic Skins," *Science Robotics* 4, no. 32 (2019): aax2198.
19. S. Xiang, J. Tang, L. Yang, Y. Guo, Z. Zhao, and W. Zhang, "Deep Learning-Enabled Real-Time Personal Handwriting Electronic Skin With Dynamic Thermoregulating Ability," *NPJ Flexible Electronics* 6, no. 1 (2022): 59.
20. Y. Yan, Z. Hu, Z. Yang, et al., "Soft Magnetic Skin for Super-Resolution Tactile Sensing With Force Self-Decoupling," *Science Robotics* 6, no. 51 (2021): eabc8801.
21. Y. Sun, J. Huang, Y. Cheng, J. Zhang, Y. Shi, and L. Pan, "High-Accuracy Dynamic Gesture Recognition: A Universal and Self-Adaptive Deep-Learning-Assisted System Leveraging High-Performance Ionogels-Based Strain Sensors," *SmartMat* 1 (2024): e1269.
22. X. Lin, H. Xue, F. Li, H. Mei, H. Zhao, and T. Zhang, "All-Nanofibrous Ionic Capacitive Pressure Sensor for Wearable Applications," *ACS Applied Materials & Interfaces* 14, no. 27 (2022): 31385–31395.
23. M. Zhang, M. Gu, L. Shao, et al., "Flexible Wearable Capacitive Sensors Based on Ionic Gel With Full-Pressure Ranges," *ACS Applied Materials & Interfaces* 15, no. 12 (2023): 15884–15892.
24. Y. Yuan, B. Liu, M. R. Adibeig, et al., "Microstructured Polyelectrolyte Elastomer-Based Ionotronic Sensors With High Sensitivities and Excellent Stability for Artificial Skins," *Advanced Materials* 36, no. 11 (2024): 2310429.
25. Y. Gao, H. Zhang, B. Song, C. Zhao, and Q. Lu, "Electric Double Layer Based Epidermal Electronics for Healthcare and Human-Machine Interface," *Biosensors* 13, no. 8 (2023): 787.
26. M. Wang, P. Zhang, M. Shamsi, et al., "Tough and Stretchable Ionogels By in Situ Phase Separation," *Nature Materials* 21, no. 3 (2022): 359–365.
27. Z. Zhang, X. Gui, Q. Hu, et al., "Highly Sensitive Capacitive Pressure Sensor Based on a Micropyramid Array for Health and Motion Monitoring," *Advanced Electronic Materials* 7, no. 7 (2021): 2100174.

28. C. Ge, B. Yang, L. Wu, et al., "Capacitive Sensor Combining Proximity and Pressure Sensing for Accurate Grasping of a Prosthetic Hand," *ACS Applied Electronic Materials* 4, no. 2 (2022): 869–877.
29. H. Niu, X. Wei, H. Li, et al., "Micropyramid Array Bimodal Electronic Skin for Intelligent Material and Surface Shape Perception Based on Capacitive Sensing," *Advanced Science* 11, no. 3 (2024): 2305528.
30. S. Afroj, S. Tan, A. M. Abdelkader, K. S. Novoselov, and N. Karim, "Highly Conductive, Scalable, and Machine Washable Graphene-Based E-Textiles for Multifunctional Wearable Electronic Applications," *Advanced Functional Materials* 30, no. 23 (2020): 202000293.
31. J. Huang, H. Wang, J. Li, et al., "High-Performance Flexible Capacitive Proximity and Pressure Sensors With Spiral Electrodes for Continuous Human-Machine Interaction," *ACS Materials Letters* 4, no. 11 (2022): 2261–2272.
32. L. Chen, Y. Xu, Y. Liu, et al., "Flexible and Transparent Electronic Skin Sensor With Sensing Capabilities for Pressure, Temperature, and Humidity," *ACS Applied Materials & Interfaces* 15, no. 20 (2023): 24923–24932.
33. H. Zhang, H. Chen, J.-H. Lee, et al., "Mechanochromic Optical/Electrical Skin for Ultrasensitive Dual-Signal Sensing," *ACS Nano* 17, no. 6 (2023): 5921–5934.
34. S. Zhuo, C. Song, Q. Rong, T. Zhao, and M. Liu, "Shape and Stiffness Memory Ionogels With Programmable Pressure-Resistance Response," *Nature Communications* 13, no. 1 (2022): 1743.
35. X. Zhang, S. Zeng, Z. Hu, et al., "Bioinspired Gradient Poly(Ionic Liquid) Ionogels for Ionic Skins With an Ultrawide Pressure Detection Range," *ACS Materials Letters* 4, no. 12 (2022): 2459–2468.
36. Y. Xu, L. Chen, J. Chen, X. Chang, and Y. Zhu, "Flexible and Transparent Pressure/Temperature Sensors Based on Ionogels With Bioinspired Interlocked Microstructures," *ACS Applied Materials & Interfaces* 14, no. 1 (2022): 2122–2131.
37. Z. Q. Shen, X. Y. Zhu, C. Majidi, and G. Gu, "Ionogel Mechanoreceptors for Soft Machines, Physiological Sensing, and Amputee Prostheses," *Advanced Materials* 33, no. 38 (2021): 2102069.
38. S. Albawi, O. Bayat, S. Al-Azawi, and O. N. Ucan, "Social Touch Gesture Recognition Using Convolutional Neural Network," *Computational Intelligence and Neuroscience* 2018, no. 6973103 (2018): 1–10.
39. V.-C. Ta, W. Johal, M. Portaz, E. Castelli, and D. Vaufraydaz, "The Grenoble System For The Social Touch Challenge at ICMI 2015," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (New York: ACM, 2015), 391–398.
40. M. M. Jung, M. Poel, R. Poppe, and D. K. J. Heylen, "Automatic Recognition of Touch Gestures in the Corpus of Social Touch," *Journal on Multimodal User Interfaces* 11, no. 1 (2016): 81–96.
41. G. Zhang, Q. Liu, Y. Shi, and H. Meng, "An Ensemble Classifier Based on Three-Way Decisions for Social Touch Gesture Recognition," *Advances in Swarm Intelligence* 10942 (2018): 370–379.
42. D. Hughes, A. Krauthammer, and N. Correll, "Recognizing Social Touch Gestures Using Recurrent and Convolutional Neural Networks," in *2017 IEEE International Conference on Robotics and Automation (ICRA)* (Singapore: ICRA, 2017), 2315–2321.
43. Y. F. A. Gaus, T. Olugbade, A. Jan, et al., "Social Touch Gesture Recognition Using Random Forest and Boosting on Distinct Feature Sets," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (New York: Association for Computing Machinery, 2015), 399–406.
44. Y. X. Wang, Y. K. Li, T. H. Yang, and Q. H. Meng, "Multitask Touch Gesture and Emotion Recognition Using Multiscale Spatiotemporal Convolutions With Attention Mechanism," *IEEE Sensors Journal* 22, no. 16 (2022): 16190–16201.
45. H. Choi, D. Brouwer, M. A. Lin, et al., "Deep Learning Classification of Touch Gestures Using Distributed Normal and Shear Force," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Kyoto: IROS, 2022), 3659–3665.
46. Y. K. Li, Q. H. Meng, T. H. Yang, Y. X. Wang, and H. R. Hou, "Touch Gesture and Emotion Recognition Using Decomposed Spatio-temporal Convolutions," *IEEE Transactions on Instrumentation and Measurement* 71, no. 2500809 (2022): 1–9.
47. S. C. Lai, H. K. Tan, and P. Y. Lau, "3D Deformable Convolution for Action Classification in Videos," *International Workshop on Advanced Imaging Technology (IWAIT)* 2021, no. 117660R (2021): 11766.
48. B. Peng, Z. Yao, Q. Wu, H. Sun, and G. Zhou, "3D Convolutional Neural Network for Human Behavior Analysis in Intelligent Sensor Network," *Mobile Networks and Applications* 27, no. 4 (2022): 1559–1568.
49. H. Yang, C. Yuan, B. Li, et al., "Asymmetric 3D Convolutional Neural Networks for Action Recognition," *Pattern Recognition* 85 (2019): 1–12.
50. J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT, USA: IEEE/CVF, 2018), 7132–7141.
51. D. Darlan, O. S. Ajani, V. Parque, and R. Mallipeddi, "Recognizing Social Touch Gestures Using Optimized Class-Weighted CNN-LSTM Networks," in *32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (Busan, Korea: RO-MAN, 2023), 2024–2029.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.