New Synthetic Goldmine: Hand Joint Angle-Driven EMG Data Generation Framework for Micro-Gesture Recognition

Nana Wang[1,2], Gen Li[1], Pengfei Ren[3], Hao Su[4], Suli Wang[5*]

[1]School of Computer Science and Engineering, Beihang University, China
[2]NoBarriers.ai Technology, Hangzhou, China
[3]Beijing University of Posts and Telecommunications, China
[4]Zhengzhou University, China
[5]Department of Computer Science, Technische Universität Darmstadt, Germany

Abstract

Electromyography (EMG)-based gesture recognition has emerged as a promising approach for human-computer interaction. However, its performance is often limited by the scarcity of labeled EMG data, significant cross-user variability, and poor generalization to unseen gestures. To address these challenges, we propose SeqEMG-GAN, a conditional, sequence-driven generative framework that synthesizes high-fidelity EMG signals from hand joint angle sequences. Our method introduces a context-aware architecture composed of an angle encoder, a dual-layer context encoder featuring the novel Ang2Gist unit, a deep convolutional EMG generator, and a discriminator, all jointly optimized via adversarial learning. By conditioning on joint kinematic trajectories, SeqEMG-GAN is capable of generating semantically consistent EMG sequences, even for previously unseen gestures, thereby enhancing data diversity and physiological plausibility. Experimental results show that classifiers trained solely on synthetic data experience only a slight accuracy drop (from 57.77% to 55.71%). In contrast, training with a combination of real and synthetic data significantly improves accuracy to 60.53%, outperforming real-only training by 2.76%. These findings demonstrate the effectiveness of our framework,also achieves the state-of-art performance in augmenting EMG datasets and enhancing gesture recognition performance for applications such as neural robotic hand control, AI/AR glasses, and gesture-based virtual gaming systems.

Introduction

Electromyography (EMG)-based gesture recognition has become a key modality for human-computer interaction (HCI), owing to its noninvasive sensing and ability to capture subtle gesture movement. It has shown great promise in applications such as neural prosthetics, AR/VR interfaces, and wearable assistive technologies. However, practical deployment remains constrained by several critical challenges: the scarcity of large-scale and diverse EMG datasets, limited generalization across users and tasks, and the difficulty in modeling temporal and semantic consistency of EMG patterns (Eddy, Scheme, and Bateman 2023).

A major bottleneck lies in the scarcity and structural limitations of existing EMG datasets. Publicly available datasets such as Ninapro (Atzori et al. 2012), CapgMyo (Dai et al. 2021), CSL-HDEMG (Amma et al. 2015), and SEEDS (Matran-Fernandez et al. 2019) are predominantly collected under predefined and isolated gesture protocols, offering limited degrees of freedom (DoFs) and failing to support natural transitions or combinations of DoFs commonly required in real-world control scenarios. This lack of flexibility restricts the datasets' utility for modeling intuitive, user-driven interactions. Although HIT-SimCo (Yang and Liu 2021) attempts to address this issue by recording randomized combinations of wrist DoFs, it is constrained to single-session recordings, thereby limiting its applicability to studies involving cross-day variability, learning adaptation, and long-term stability. The recently introduced *emg2pose* dataset (Salter et al. 2024) significantly advances the state of EMG data collection by offering large-scale, multi-user, and multi-session recordings with synchronized hand pose annotations. It comprises over 370 hours of data from 193 subjects, covering 29 gesture stages under varied sensor placements and session conditions. Compared to earlier datasets, *emg2pose* provides high-resolution joint-angle trajectories across the entire hand and introduces benchmark splits to evaluate generalization between users, between sessions and between tasks. An additional concern is the significant imbalance between dynamic and isometric contractions, which can lead to biased model performance and reduced generalizability across different contraction types.

On the algorithmic side, effective EMG pattern recognition remains highly dependent on capturing the complex temporal and semantic relationships inherent in muscle signals. While earlier studies employed traditional machine learning with handcrafted features (Phinyomark and Scheme 2018; Oskoei and Hu 2007, 2008), such methods often suffer from suboptimal feature representations and poor scalability. Although recent deep learning approaches (Gallon et al. 2024; Xiong et al. 2021; Chen et al. 2020) have improved recognition performance through end-to-end learning, several critical issues persist. Despite recent advances, EMG-based gesture modeling continues to face three interrelated challenges. First, many existing models struggle to capture the intricate temporal dependencies between joint kinemat-

ics and EMG activation, resulting in decoded or generated signals that lack semantic coherence and physiological plausibility (Tacca et al. 2024; Sitole and Sup 2023; Zhang et al. 2022). Second, the scarcity and uneven distribution of gesture data hinder model robustness, while commonly adopted augmentation strategies—such as random noise injection (Fatayer, Gao, and Fu 2022) or label-conditioned generation (Xiong et al. 2024)—lack the precision and coverage required to synthesize high-quality samples in underrepresented regions of the EMG feature space (Nasrallah et al. 2023; Jiang et al. 2021). Third, current discriminator-based evaluation frameworks rely on narrow, signal-level criteria and are ill-equipped to assess the fidelity of generated EMG signals across multiple dimensions, including amplitude continuity, temporal alignment, and waveform morphology (Farago et al. 2022; Zhao et al. 2024; Tsinganos et al. 2020).

These challenges are further amplified in the context of EMG signal generation. Unlike image or audio synthesis, generating realistic and semantically meaningful EMG signals is fundamentally constrained by the physiological complexity, non-stationary dynamics, and user-specific variability of muscle activation. Existing generative models—often based on label-conditioned or noise-driven inputs—fail to provide fine-grained control and struggle to maintain semantic consistency and physiological plausibility, particularly in gesture-sparse regions. Moreover, the lack of comprehensive, physiology-aware evaluation mechanisms exacerbates the difficulty of training reliable EMG generators. These limitations highlight the need for a controllable, physiologically grounded EMG generation framework that ensures both semantic consistency and distributional completeness.

Joint angle–driven conditioning is a promising direction: joint angles summarize underlying activations, and conditioning on them renders EMG synthesis more plausible under inverse-dynamics constraints on feasible muscle activity. Here we introduce SeqEMG-GAN, a sequence-driven conditional generative model that synthesizes high-fidelity EMG signals based on hand joint kinematic trajectories. Our key motivation is to leverage joint angle sequences—an interpretable and compact representation of movement—as a prior control signal to guide the generation of temporally aligned, physiologically consistent EMG sequences. The proposed framework comprises four core components: (1) a joint angle encoder to extract movement embeddings, (2) a latent context encoder incorporating a novel Ang2Gist unit to model fine-grained temporal dependencies between motion and muscle activation, (3) a deep convolutional generator to produce raw EMG waveforms, and (4) a multi-perspective discriminator that evaluates signal fidelity across amplitude, timing, and morphological perspectives. This design enables both targeted synthesis in data-scarce regions and robust generalization across gestures and users.

The main contributions of this work are as follows:

- **A latent context encoder** incorporating a novel *Ang2Gist* unit to capture temporal dependencies between joint movement and EMG activation, ensuring semantic and physiological coherence in generated signals;

- **A distribution-aware conditional augmentation strategy** that utilizes joint-angle conditioning to guide sample generation in underrepresented regions of the EMG feature space, improving data diversity and coverage;

- **A multi-perspective discriminator** that jointly evaluates the fidelity of synthetic signals across amplitude profiles, temporal alignment, and waveform morphology, enhancing the realism of generated EMG data.

Together, these components form a controllable and physiologically grounded EMG generation pipeline, offering an effective solution for data augmentation under limited or imbalanced data regimes, and opening new directions for EMG-based gesture recognition in neural interface applications.

Related Work

Traditional EMG Data Augmentation and Generation Methods

To augment scarce EMG data, early attempts such as temporal warping (Gu et al. 2018) and noise injection (Stachaczyk, Atashzar, and Farina 2020) have been explored, but they lack semantic consistency. Transfer learning (Ameri et al. 2019; Wu, Yang, and Sawan 2023) alleviates inter-user variability, yet it relies on large-scale source data. Earlier biomechanical models (Kiguchi et al. 2008) map EMG to joint motion via physical equations, but their complexity and reliance on expert knowledge limit scalability. Statistical models like GMM and HMM (Yang, Pan, and Li 2017) impose strong assumptions, reducing generative diversity and generalization across poses and users.

These limitations motivate the need for data-driven generative approaches that can capture both temporal dynamics and individual variability.

Generative Methods for EMG

Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion-based models have emerged as foundational tools to address EMG-specific challenges such as data scarcity, signal degradation, and inter-subject variability.

GAN-Based Methods: (Mendez, Lhoste, and Micera 2022) applied GANs to generate synthetic EMG signals for hand grasp classification, leading to improved classifier performance in low-resource settings. (Chen et al. 2022) employed DCGANs to augment multi-channel EMG data represented as grayscale images, achieving marginal but consistent gains in recognition accuracy. (Lin et al. 2023) proposed RoHDE, a GAN-based framework that introduces perturbation patterns into HD-EMG to enhance classifier robustness against real-world signal artifacts. (Lin et al. 2023) introduced a style-transfer GAN to disentangle identity-specific signal features from gesture-related information, highlighting privacy vulnerabilities and EMG anonymization opportunities.

Privacy-Preserving Synthesis: Beyond task-driven synthesis, (Hazra and Byun 2020) designed SynSigGAN to generate synthetic biomedical signals, including EMG, achiev-
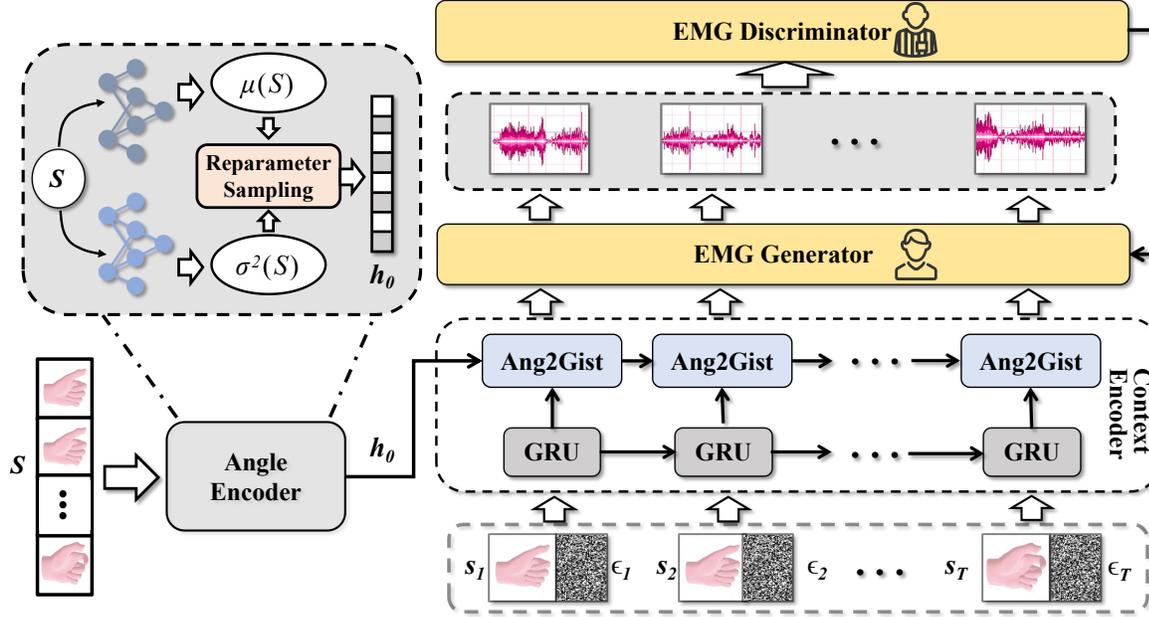
Figure 1: Overview of the SeqEMG-GAN framework. Joint angle sequences are encoded into a global latent variable and contextual hidden states, which guide the generator in synthesizing EMG signals. A discriminator jointly evaluates temporal and semantic alignment.

ing high signal correlation and enabling data sharing under privacy constraints.

Diffusion-Based Methods: Recent diffusion-based approaches have shown superior modeling of complex EMG distributions. (Xiong et al. 2024) introduced PatchEMG, which employs patch-wise diffusion for few-shot gesture recognition, demonstrating high fidelity synthesis with limited data. (Liu et al. 2024) proposed SDEMG, a score-based diffusion model that effectively removes ECG artifacts from diaphragmatic EMG. (Qing et al. 2024) utilized a diffusion model to synthesize realistic EMG for ergonomic studies, preserving utility while mitigating privacy concerns.

VAE-Based Methods: Latent-variable models offer calibrated uncertainty that is valuable for open-set scenarios. (Olsson et al. 2021) propose a VAE-based framework for open-set gesture recognition using sparse multichannel EMG signals. Their method enables reliable classification of known gestures while rejecting unknown ones, addressing a key need in real-world prosthetic control. Unlike closed-set models, the VAE captures uncertainty, allowing effective separation of seen and unseen gestures.

Together, these contributions demonstrate the versatility of generative models in advancing EMG-based applications across signal enhancement, data augmentation, privacy, and human-computer interaction.

Methodology

We propose **SeqEMG-GAN**, a sequence-driven conditional generative framework that synthesizes electromyographic (EMG) signal sequences from hand joint kinematic inputs.

The architecture is designed to model complex temporal dependencies and generate physiologically plausible EMG signals, even for underrepresented gestures. As shown in Figure 1, SeqEMG-GAN consists of four main modules: Angle Encoder, Context Encoder, EMG Generator, and EMG Discriminator.

Overall Framework

Given a joint angle sequence $S = \{s_1, s_2, \ldots, s_T\}$, the model generates a corresponding EMG sequence $\hat{X} = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_T\}$. The generation process is governed by a global latent context vector $\mathbf{h}_0$, which summarizes high-level motion semantics. To enable stochastic generation and improve generalization, we apply the *reparameterization trick* (Kingma, Salimans, and Welling 2015) to sample $\mathbf{h}_0$ from a learned posterior distribution:

$$\mathbf{h}_0 = \mu(S) + \Sigma(S)^{1/2} \odot \epsilon_s, \quad \epsilon_s \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (1)$$

Here, $\mu(S)$ and $\Sigma(S)$ are the outputs of neural networks conditioned on S. This stochastic sampling introduces variability into the generated signals while maintaining differentiability during training.

Context Encoding via Ang2Gist

The **Context Encoder** models sequential dependencies between joint movement and muscle activation. It is implemented as a two-layer recurrent structure:

- **Lower Layer:** Standard Gated Recurrent Unit (GRU) units encode motion dynamics.

- **Upper Layer:** A novel Angle-to-Gist (Ang2Gist) unit refines contextual representation.

At each timestep t, we compute:

$$\mathbf{i}_t, \mathbf{g}_t = \text{GRU}(\mathbf{s}_t \| \epsilon_t, \mathbf{g}_{t-1}), \qquad (2)$$

where $\epsilon_t \sim \mathcal{N}(0, I)$ is noise and $\|$ denotes concatenation. The output $\mathbf{i}_t$ is passed to the Ang2Gist unit:

$$\mathbf{o}_t, \mathbf{h}_t = \text{Ang2Gist}(\mathbf{i}_t, \mathbf{h}_{t-1}), \qquad (3)$$

producing a Gist vector $\mathbf{o}_t$ for EMG synthesis and updating the hidden state $\mathbf{h}_t$. The Ang2Gist computations are defined as:

$$
\begin{aligned}
\mathbf{z}_t &= \sigma_z(\mathbf{W}_z \mathbf{i}_t + \mathbf{U}_z \mathbf{h}_{t-1} + \mathbf{b}_z), \\
\mathbf{r}_t &= \sigma_r(\mathbf{W}_r \mathbf{i}_t + \mathbf{U}_r \mathbf{h}_{t-1} + \mathbf{b}_r), \\
\mathbf{h}_t &= (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \sigma_h(\mathbf{W}_h \mathbf{i}_t + \mathbf{U}_h \mathbf{h}_{t-1} + \mathbf{b}_h), \\
\mathbf{o}_t &= \text{Filter}(\mathbf{i}_t) * \mathbf{h}_t.
\end{aligned}
$$
$$(4)$$

Here, $\sigma_{\{z,r,h\}}$ are nonlinear activations; $\mathbf{W}_*$, $\mathbf{U}_*$, and $\mathbf{b}_*$ are trainable parameters; $\odot$ and $*$ denote element-wise product and filtering, respectively. Filter($\cdot$) is a 1D depthwise-separable convolution along time (kernel size = 5, stride = 1, padding = 2) applied to the angle embedding i_t; its parameters are learned jointly with the generator (Xavier-uniform initialization, SGD with momentum), and each filter channel is shared across gesture dimensions to enforce temporal consistency at low cost.

The **Ang2Gist** unit is designed to address the challenge of capturing *temporally consistent* and *semantically aligned* representations for EMG signal generation. Standard GRU-based encoders often fail to retain local dynamics while preserving global contextual consistency. Ang2Gist explicitly fuses motion-specific features $\mathbf{i}_t$ with evolving context states $\mathbf{h}_t$, enabling the generator to condition on both instantaneous motion and accumulated semantics. This design enhances the generator's ability to produce physiologically plausible and temporally coherent EMG sequences.

EMG Generation and Discrimination

The **EMG Generator** synthesizes EMG signals at each timestep using the corresponding Gist vector:

$$\hat{\mathbf{x}}_t = \mathcal{G}(\mathbf{o}_t), \qquad (5)$$

where $\mathcal{G}$ is implemented as a multi-layer convolutional decoder with transposed convolutions for temporal upsampling.

The **EMG Discriminator** evaluates whether the generated EMG signal $\hat{\mathbf{x}}_t$ is realistic and semantically aligned with the joint angle sequence S_t, conditioned on global context $\mathbf{h}_0$:

$$m_t = \mathcal{D}(\hat{\mathbf{x}}_t, S_t \mid \mathbf{h}_0). \qquad (6)$$

Implementation

The proposed **SeqEMG-GAN** framework generates EMG signal sequences from joint kinematic inputs, maintaining temporal coherence and physiological plausibility. The model is built on a conditional GAN architecture and consists of four core components:

- **Angle Encoder:** Encodes joint angle sequences into compact latent representations.
- **Context Encoder:** Models temporal dependencies using recurrent layers and produces a global context vector h_0.
- **EMG Generator** (G)**:** Synthesizes EMG signals from joint embeddings, noise, and context.
- **EMG Discriminator** (D)**:** Evaluates the realism and motion-alignment of generated EMG sequences.

Data Representation The training data consists of paired samples $\left\{(S^{(i)}, X^{(i)})\right\}_{i=1}^{N}$, where each pair contains:

- **Joint angle sequence** $S \in \mathbb{R}^{T \times d_j}$, representing the hand gesture over T time steps
- **EMG sequence** $X \in \mathbb{R}^{T \times d_e}$, capturing the corresponding muscle activation signals

The joint angle sequence S is used as the conditioning input to guide the generation of the EMG signal X through the SeqEMG-GAN model.

Training Objective

The learning objective integrates a standard adversarial loss and a Kullback–Leibler (KL) divergence regularization:

$$\min_{\theta} \max_{\psi} \left(\alpha \, \mathcal{L}_{\text{GAN}} + \mathcal{L}_{\text{KL}}\right) \qquad (7)$$

where θ and ψ are the parameters of the generator and discriminator, respectively, and α is a balancing hyperparameter.

Adversarial Loss The adversarial loss encourages the generator to synthesize realistic EMG sequences conditioned on joint angle sequences:

$$
\begin{aligned}
\mathcal{L}_{\text{GAN}} = \ & \mathbb{E}_{(x_t, s_t)} \left[\log D(x_t, s_t, h_0)\right] \\
& + \mathbb{E}_{(\epsilon_t, s_t)} \left[\log \left(1 - D(G(\epsilon_t, s_t; \theta), s_t, h_0)\right)\right]
\end{aligned}
$$
$$(8)$$

Here, x_t denotes the real EMG signal, s_t is the joint angle input, $\epsilon_t \sim \mathcal{N}(0, I)$ is the noise input, and h_0 is the global context variable.

KL Divergence Regularization To ensure a structured latent space, a KL divergence regularization term constrains the global latent variable h_0 to follow a standard normal distribution:

$$\mathcal{L}_{\text{KL}} = D_{\text{KL}}\left(\mathcal{N}(\mu(S), \text{diag}(\sigma^2(S))) \,\|\, \mathcal{N}(0, I)\right) \qquad (9)$$

Here, $\mu(S)$ and $\sigma^2(S)$ are parameterized by neural networks and computed from the joint sequence S.

Experiments

Dataset and Experimental Setup

The experiments in this chapter utilize the latest **emg2pose** dataset proposed by Meta (Salter et al. 2024). This dataset is currently the largest publicly available high-quality dataset for wrist-based electromyography (EMG) recordings with hand gesture annotations. The **emg2pose** dataset includes
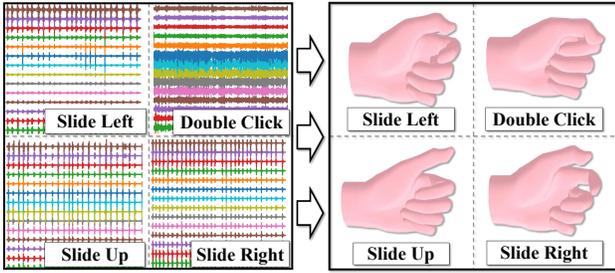
Figure 2: Synthesized EMG Gesture Data and Its Visualization Results by SeqEMG-GAN

16-channel EMG signals sampled at 2 kHz, along with precise gesture annotations captured by 26 cameras. It covers a total of 193 users, 370 hours of recording time, and 29 gesture categories, ensuring a high degree of gesture diversity.

During the experiments, the dataset is randomly split into training, validation, and test sets in an 8:1:1 ratio. Specifically, the training set contains 20,203 EMG samples, while the validation and test sets each contain 2,525 EMG samples. The scale and precision of this dataset make it an ideal benchmark for EMG-based gesture recognition research, providing a reliable data foundation for the experiments in this chapter.

Implementation Details

In the experiments, the batch size is set to 32, and training is conducted for 100 epochs. The initial learning rate is set to 0.002, with an annealing strategy applied during the pretraining phase to gradually decay the learning rate. During model fine-tuning, the learning rate is reduced by a factor of 10 at the 21st, 24th, and 27th epochs.

The training process utilizes the Stochastic Gradient Descent (SGD) optimizer, with a momentum hyperparameter of 0.9 and a weight decay of 5×10^{-4}. The weight parameter λ is set to 2. All experiments are implemented using the PyTorch library and trained on two NVIDIA RTX 3090 Ti GPUs.

Evaluation Metrics

To assess the effectiveness of SeqEMG-GAN, we conducted two main experiments: *similarity analysis* and *performance validation*. Given that standard GAN evaluation metrics such as Inception Score (IS), Frechet Inception Distance (FID), and Kernel MMD (Xu et al. 2018) are primarily designed for image generation and high-dimensional feature space comparison, they are not directly applicable to EMG time-series data.

To address this, we adopt metrics more suitable for temporal biomedical signals, following prior work (Delaney, Brophy, and Ward 2019), and propose a multi-perspective evaluation scheme covering frequency-domain, time-domain, and signal morphology consistency:

Fast Fourier Transform MSE (FFT MSE) We compute the mean squared error between the FFT magnitudes of generated and real EMG signals. This metric evaluates frequency-domain similarity and reflects how well the generated signals preserve spectral characteristics essential to muscle activation patterns.

Dynamic Time Warping (DTW) To measure structural similarity in the time domain, we use DTW to non-linearly align generated and real sequences. DTW is a distance between two time series that allows non-linear time stretching/compression (lower is better). This technique is especially robust to temporal shifts and warping in EMG signals. We implement FastDTW (Salvador and Chan 2007) to reduce computational cost to $O(N)$.

EMG Envelope Cross-Correlation (CC) We evaluate global shape consistency using cross-correlation between smoothed signal envelopes. The envelope is computed by applying a moving average filter to the absolute EMG signals, and normalized CC quantifies trend alignment and tremor consistency.

Together, these metrics provide a comprehensive evaluation framework for EMG signal generation, capturing fidelity across spectral, temporal, and morphological dimensions.

Assessing the Quality of Generated Datasets

We evaluate the similarity between synthetic and real data by measuring classification accuracy and analyzing the contribution of synthetic data in data augmentation. The experimental design follows three different data partitioning strategies to compare the performance of real and synthetic data in classification tasks:

1. **Baseline Experiment**: 80% of the real data is used as the training set, while the remaining 20% of the real data is used as the test set. The classification accuracy is computed as a benchmark for comparison.

2. **Synthetic-Only Training**: The training set consists solely of synthetic data generated by DCGAN, while the test set remains the 20% real data used in the baseline experiment. This setup evaluates whether synthetic data alone can effectively support classification tasks.

3. **Hybrid Training**: The training set combines 80% real data with synthetic data, while the test set remains the 20% real data from the baseline experiment. This setup analyzes the impact of synthetic data on model performance when used for data augmentation.

By comparing the classification accuracy across these three strategies, we can quantify the similarity between synthetic and real data and assess the role of synthetic data in improving model generalization and enhancing data augmentation.

Similarity Results Specifically, we select several state-of-the-art EMG signal generation methods based on Generative Adversarial Networks (GANs) as comparison baselines. These methods are trained using the same dataset under identical experimental conditions and evaluated on the same test set using correlation analysis metrics. Table 2 presents a performance comparison of different generative models

Model	Conditional Generation	Unknown Gesture Generation	DTW ↓	FFT MSE↓	EECC↑
GAN	×	×	103.427321	19.564078	0.719453
StyleTransfer	√	×	98.532786	13.531477	0.781721
DCGAN	√	×	93.439412	9.675622	0.791930
Ours	√	√	**91.756312**	**8.764323**	**0.817492**

Table 1: Similarity Analysis Between Generated and Real Data. 'Conditional Generation' = generation controllable via joint-angle input; 'Unknown Gesture Generation' = synthesis for gestures absent from the generator's training set.

for EMG signal synthesis, including metrics such as **conditional generation**, **unknown gesture generation**, **Dynamic Time Warping (DTW)**, **Fast Fourier Transform Mean Squared Error (FFT MSE)**, and **EMG Envelope Cross-Correlation (EECC)**.

The baseline GAN model performs the worst across all metrics, failing to support conditional generation or unknown gesture generation. In 2020, Rafael Anicet Zanini et al. (Anicet Zanini and Luna Colombini 2020) proposed an EMG synthesis method by combining DCGAN with **Style Transfer**. Their approach employs a CNN-based architecture to generate EMG signals and introduces four feature pipelines (raw EMG, FFT, envelope FFT, and wavelet decomposition) to enrich the feature space, making the generated signals more realistic. While this method supports conditional generation, it lacks the ability to generate unknown gestures.

The **DCGAN** model, proposed by Yaojia Qian et al. (Wu, Chen, and Meng 2020), applies matrix transformation normalization to convert EMG signals into grayscale images before training a network using image generation techniques. Although this approach leverages advances in image synthesis, it suffers from information loss. Like the method of Zanini et al., DCGAN supports conditional generation but is incapable of generating unknown gestures.

In contrast, the **SeqEMG-GAN** model proposed in this paper (denoted as **Ours**) achieves the best performance across all evaluation metrics. It supports both **conditional generation** and **unknown gesture generation** while demonstrating superior performance in DTW (91.76), FFT MSE (8.76), and EMG Envelope Cross-Correlation (0.82). These results indicate that SeqEMG-GAN exhibits significant advantages in EMG signal quality, temporal alignment, frequency-domain similarity, and envelope similarity.

Assessing the Effectiveness of Generated Data for Classification

To validate the effectiveness of the data generated by SeqEMG-GAN, we use classification accuracy as the evaluation metric. In particular, we select six gestures from either generated or real data and use them as the training set to train a classifier. We classify Slide Left/Right/Up/Down, Click, Double Click — a standard HCI micro-gesture set aligned with widely recognized EMG wristband interfaces, ensuring real-world applicability. The classifier's performance is then evaluated to determine whether the data contributes to model training. In this study, we select two traditional machine learning classification algorithms: **Support Vector**

Machine (SVM) and **Random Forest (RF)**, along with two deep learning-based classifiers. The experimental results are presented in Table 2.

Four classifiers are trained under identical conditions and subsequently tested on the test set to obtain classification accuracy for six-class wrist movement classification tasks. Each classifier undergoes three sets of tests:

- **RR**: The classifier is trained using 80% real data and tested on 20% real data.

- **GR**: The classifier is trained using synthetic data and tested on 20% real data.

- **MR**: The classifier is trained using mixed data (80% real data and synthetic data) and tested on 20% real data.

Dataset Partition	RR	GR	MR
SVM	37.08%	35.92%	39.88%
RF	54.88%	51.32%	58.32%
Vemg2pose	67.78%	65.73%	68.95%
NeuroPose	71.32%	69.87%	74.98%
Average Accuracy	**57.77%**	**55.71%**	**60.53%**

Table 2: Classification Accuracy Analysis

Table 2 presents the classification accuracy results for performance validation analysis. The results indicate that when classifiers are trained solely on synthetic data, the average accuracy decreases by 2.06%, dropping from 57.77% to 55.71%. However, by combining real data with synthetic data for training, the average accuracy significantly increases to 60.53%, achieving an improvement of 2.76% compared to classifiers trained solely on real data. Qualitatively, Slide gestures exhibit smoother envelope transitions than Click/Double-Click, aligning with their kinematic phases and supporting our angle-conditioned design. Taken together, these results indicate that integrating synthetic with real data effectively enhances classifier performance.

A further analysis of the experimental results in Table 2 reveals that the SVM and RF classifiers exhibit relatively lower classification accuracy. This may be attributed to the lack of feature extraction from the raw multi-channel EMG signals during the preprocessing stage, resulting in a high level of noise that negatively impacts classification performance. In contrast, deep learning-based models, such as Vemg2pose and NeuroPose, demonstrate superior classification performance.

Cross-user/session Generalization

Surface EMG varies markedly across **users** (anatomy, muscle recruitment) and **sessions** (electrode shift, skin impedance, fatigue); thus, cross-subject/session evaluations directly test robustness to these distribution shifts.

Leveraging `emg2pose` metadata (`held_out_user`, `held_out_stage`, `generalization`), we enforce disjoint train/test partitions and evaluate four protocols: sample-wise, cross-subject, cross-session, and cross user+stage. Unless noted otherwise, models are trained under REAL-ONLY and REAL+SYNTHETIC (1:1) conditions, and results are reported as mean $\pm$ SD over five runs. Mixing synthetic data yields consistent gains across all protocols.

Setting	Sample-wise	Cross-Subj	Cross-Sess	Cross User+Stage
Real-only	57.8 ± 1.0	52.6 ± 1.1	56.0 ± 0.8	48.9 ± 1.2
Real + Synthetic	60.5 ± 0.9	57.1 ± 0.9	59.8 ± 0.7	54.3 ± 1.0

Table 3: Cross-user/session generalization accuracy (%). Adding synthetic data (1:1 mix). Means $\pm$ SD over five runs.

Ablation Study

To better understand the contribution of each component in our SeqEMG-GAN framework, we conduct a series of ablation experiments. We evaluate the performance of each variant using three metrics: (1) Dynamic Time Warping (DTW) distance, which measures the temporal alignment between generated and ground-truth EMG sequences; (2) Mean Squared Error in the frequency domain (FFT MSE), which reflects the preservation of spectral characteristics; and (3) EMG Envelope Cross-Correlation, which assesses the similarity of signal envelopes and indicates semantic alignment. We report results on both seen and unseen gestures to evaluate signal quality, model generalization, and generation diversity.

We expect the following trends across the ablated variants:

- **(a) w/o Angle Encoder**: Without a dedicated angle encoder, the model loses the ability to extract rich representations from joint angles. This may lead to degraded performance in both DTW and envelope correlation, especially for complex or unseen gestures.

- **(b) w/o GRU**: Removing the recurrent backbone disrupts the modeling of temporal dependencies. We anticipate higher DTW distances and lower cross-correlation scores, indicating poor temporal alignment and weaker contextual coherence.

- **(c) w/o Ang2Gist**: Eliminating the semantic transformation blocks (Ang2Gist) is expected to reduce the model's ability to abstract gesture-level intent, resulting in lower envelope correlation and higher FFT MSE, especially for gestures involving subtle spectral variations.

- **(d) w/o Discriminator (Non-GAN)**: raining without adversarial feedback is likely to produce overly smooth or averaged signals, leading to lower realism. We expect a noticeable drop in cross-correlation and an increase in FFT MSE, while DTW may show marginal improvement due to reduced variation.

These experiments help disentangle the effects of temporal modeling, semantic abstraction, and adversarial learning on the quality, generalizability, and diversity of generated EMG signals.

Experiment	DTW $\downarrow$	FFT MSE $\downarrow$	EECC $\uparrow$
(a) w/o Angle Encoder	34.0170	3.1484	0.3262
(b) w/o GRU	50.3485	4.5006	0.4767
(c) w/o Ang2Gist	62.2033	5.7655	0.5636
(d) w/o Discriminator	65.4667	6.1234	0.6132
Ours (Full)	**91.7563**	**8.7643**	**0.8175**

Table 4: Ablation Study.

Table 4 summarizes the results of our ablation study. We report three metrics—DTW ($\downarrow$), FFT MSE ($\downarrow$), and EECC($\uparrow$)—averaged over both seen and unseen gestures. The full model achieves the best performance across all metrics, confirming the effectiveness of its modular design.

Removing the Angle Encoder (a) leads to a substantial drop in envelope correlation and an increase in DTW, highlighting the importance of high-level joint representation. Disabling the GRU module (b) causes a pronounced degradation in DTW and correlation, suggesting temporal modeling is essential for sequential consistency. Excluding the Ang2Gist blocks (c) reduces semantic alignment, as reflected in lower envelope correlation and worse frequency-domain similarity. Finally, in the non-GAN baseline (d), DTW decreases slightly without the discriminator, but spectral and envelope consistency degrade markedly; overall fidelity is therefore highest for the full model.

These findings validate that each component contributes to improving generation quality, generalization, and diversity in a complementary manner.

Conclusion and Future Work

We propose **SeqEMG-GAN**, an electromyographic (EMG) signal generation method based on Generative Adversarial Networks (GANs), designed to address the challenges posed by the scarcity of surface EMG data, individual physiological variability, and posture variations that affect the generalization capability of gesture recognition models. Traditional EMG data acquisition is constrained by high collection costs, significant differences in individual physiological characteristics, and complex motion states, making it challenging to construct large-scale, high-quality datasets. We visualize our synthesized EMG data in Figure 2, which shows that this approach can effectively expand existing datasets and improve the generalization ability of gesture recognition models. Future work will explore improved physiological modeling, integration of multimodal control inputs, and deployment in real-world HCI and robotics systems.

References

Ameri, A.; Akhaee, M. A.; Scheme, E.; and Englehart, K. 2019. A deep transfer learning approach to reducing the effect of electrode shift in EMG pattern recognition-based control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(2): 370–379.

Amma, C.; Krings, T.; Böer, J.; and Schultz, T. 2015. Advancing Muscle-Computer Interfaces with High-Density Electromyography. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, 929–938. New York, NY, USA: ACM. ISBN 978-1-4503-3145-6.

Anicet Zanini, R.; and Luna Colombini, E. 2020. Parkinson's disease EMG data augmentation and simulation with DCGANs and style transfer. *Sensors*, 20(9): 2605.

Atzori, M.; Gijsberts, A.; Heynen, S.; Hager, A.-G. M.; Deriaz, O.; Van Der Smagt, P.; Castellini, C.; Caputo, B.; and Müller, H. 2012. Building the Ninapro database: A resource for the biorobotics community. In *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 1258–1265. IEEE.

Chen, H.; Zhang, Y.; Li, G.; Fang, Y.; and Liu, H. 2020. Surface electromyography feature extraction via convolutional neural network. *International Journal of Machine Learning and Cybernetics*, 11(1): 185–196.

Chen, Z.; Qian, Y.; Wang, Y.; and Fang, Y. 2022. Deep convolutional generative adversarial network-based emg data enhancement for hand motion classification. *Frontiers in Bioengineering and Biotechnology*, 10: 909653.

Dai, Q.; Li, X.; Geng, W.; Jin, W.; and Liang, X. 2021. CAPG-MYO: a muscle-computer interface supporting user-defined gesture recognition. In *Proceedings of the 9th international conference on computer and communications management*, 52–58.

Delaney, A. M.; Brophy, E.; and Ward, T. E. 2019. Synthesis of realistic ECG using generative adversarial networks. *arXiv preprint arXiv:1909.09150*.

Eddy, E.; Scheme, E. J.; and Bateman, S. 2023. A framework and call to action for the future development of EMG-based input in HCI. In *Proceedings of the 2023 CHI conference on human factors in computing systems*, 1–23.

Farago, E.; MacIsaac, D.; Suk, M.; and Chan, A. D. 2022. A review of techniques for surface electromyography signal quality analysis. *IEEE Reviews in Biomedical Engineering*, 16: 472–486.

Fatayer, A.; Gao, W.; and Fu, Y. 2022. sEMG-based gesture recognition using deep learning from noisy labels. *IEEE Journal of Biomedical and Health Informatics*, 26(9): 4462–4473.

Gallon, V. M.; Velez, S. M.; Ramirez, J.; and Bolanos, F. 2024. Comparison of machine learning algorithms and feature extraction techniques for the automatic detection of surface EMG activation timing. *Biomedical Signal Processing and Control*, 94: 106266.

Gu, Y.; Yang, D.; Huang, Q.; Yang, W.; and Liu, H. 2018. Robust EMG pattern recognition in the presence of confounding factors: features, classifiers and adaptive learning. *Expert Systems with Applications*, 96: 208–217.

Hazra, D.; and Byun, Y.-C. 2020. SynSigGAN: Generative adversarial networks for synthetic biomedical signal generation. *Biology*, 9(12): 441.

Jiang, X.; Bardizbanian, B.; Dai, C.; Chen, W.; and Clancy, E. A. 2021. Data management for transfer learning approaches to elbow EMG-torque modeling. *IEEE Transactions on Biomedical Engineering*, 68(8): 2592–2601.

Kiguchi, K.; Rahman, M. H.; Sasaki, M.; and Teramoto, K. 2008. Development of a 3DOF mobile exoskeleton robot for human upper-limb motion assist. *Robotics and Autonomous systems*, 56(8): 678–691.

Kingma, D. P.; Salimans, T.; and Welling, M. 2015. Variational dropout and the local reparameterization trick. *Advances in neural information processing systems*, 28.

Lin, Z.; Liang, P.; Zhang, X.; and Qin, Z. 2023. Toward robust high-density EMG pattern recognition using generative adversarial network and convolutional neural network. In *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER)*, 1–5. IEEE.

Liu, Y.-T.; Wang, K.-C.; Liu, K.-C.; Peng, S.-Y.; and Tsao, Y. 2024. Sdemg: Score-based diffusion model for surface electromyographic signal denoising. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1736–1740. IEEE.

Matran-Fernandez, A.; Rodríguez Martínez, I. J.; Poli, R.; Cipriani, C.; and Citi, L. 2019. SEEDS, simultaneous recordings of high-density EMG and finger joint angles during multiple hand movements. *Scientific data*, 6(1): 186.

Mendez, V.; Lhoste, C.; and Micera, S. 2022. EMG data augmentation for grasp classification using generative adversarial networks. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 3619–3622. IEEE.

Nasrallah, C.; Boudaoud, S.; Laforet, J.; Chazard, E.; Beuscart, J.-B.; and Istrate, D. 2023. SEMG signal generation for data augmentation using time series transformer based conditional GAN. In *2023 Seventh International Conference on Advances in Biomedical Engineering (ICABME)*, 137–141. IEEE.

Olsson, A. E.; Malešević, N.; Björkman, A.; and Antfolk, C. 2021. Can Deep Synthesis of EMG Overcome the Geometric Growth of Training Data Required to Recognize Multiarticulate Motions? In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 6380–6383. IEEE.

Oskoei, M. A.; and Hu, H. 2007. Myoelectric control systems—A survey. *Biomedical signal processing and control*, 2(4): 275–294.

Oskoei, M. A.; and Hu, H. 2008. Support vector machine-based classification scheme for myoelectric control applied to upper limb. *IEEE transactions on biomedical engineering*, 55(8): 1956–1965.

Phinyomark, A.; and Scheme, E. 2018. EMG pattern recognition in the era of big data and deep learning. *Big Data and Cognitive Computing*, 2(3): 21.

Qing, L.; Jung, S.; Su, B.; Lu, L.; and Xu, X. 2024. Enhancing Privacy Protection for Time-Series Signals in Ergonomics Studies via Data Synthesis. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 68, 1465–1468. SAGE Publications Sage CA: Los Angeles, CA.

Salter, S.; Warren, R.; Schlager, C.; Spurr, A.; Han, S.; Bhasin, R.; Cai, Y.; Walkington, P.; Bolarinwa, A.; Wang, R. J.; et al. 2024. emg2pose: A large and diverse benchmark for surface electromyographic hand pose estimation. *Advances in Neural Information Processing Systems*, 37: 55703–55728.

Salvador, S.; and Chan, P. 2007. Toward accurate dynamic time warping in linear time and space. *Intelligent data analysis*, 11(5): 561–580.

Sitole, S. P.; and Sup, F. C. 2023. Continuous prediction of human joint mechanics using emg signals: A review of model-based and model-free approaches. *IEEE Transactions on Medical Robotics and Bionics*, 5(3): 528–546.

Stachaczyk, M.; Atashzar, S. F.; and Farina, D. 2020. Adaptive spatial filtering of high-density EMG for reducing the influence of noise and artefacts in myoelectric control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(7): 1511–1517.

Tacca, N.; Dunlap, C.; Donegan, S. P.; Hardin, J. O.; Meyers, E.; Darrow, M. J.; Colachis IV, S.; Gillman, A.; and Friedenberg, D. A. 2024. Wearable high-density EMG sleeve for complex hand gesture classification and continuous joint angle estimation. *Scientific Reports*, 14(1): 18564.

Tsinganos, P.; Cornelis, B.; Cornelis, J.; Jansen, B.; and Skodras, A. 2020. Data augmentation of surface electromyography for hand gesture recognition. *Sensors*, 20(17): 4892.

Wu, D.; Yang, J.; and Sawan, M. 2023. Transfer learning on electromyography (EMG) tasks: Approaches and beyond. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31: 3015–3034.

Wu, Q.; Chen, Y.; and Meng, J. 2020. DCGAN-based data augmentation for tomato leaf disease identification. *IEEE access*, 8: 98716–98728.

Xiong, B.; Chen, W.; Li, H.; Niu, Y.; Zeng, N.; Gan, Z.; and Xu, Y. 2024. Patchemg: Few-shot emg signal generation with diffusion models for data augmentation to improve classification performance. *IEEE Transactions on Instrumentation and Measurement*.

Xiong, D.; Zhang, D.; Zhao, X.; and Zhao, Y. 2021. Deep learning for EMG-based human-machine interaction: A review. *IEEE/CAA Journal of Automatica Sinica*, 8(3): 512–533.

Xu, J.; Ren, X.; Lin, J.; and Sun, X. 2018. Diversity-promoting GAN: A cross-entropy based generative adversarial network for diversified text generation. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, 3940–3949.

Yang, D.; and Liu, H. 2021. An EMG-based deep learning approach for multi-DOF wrist movement decoding. *IEEE Transactions on Industrial Electronics*, 69(7): 7099–7108.

Yang, J.; Pan, J.; and Li, J. 2017. sEMG-based continuous hand gesture recognition using GMM-HMM and threshold model. In *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 1509–1514. IEEE.

Zhang, L.; Zhu, X.; Gutierrez-Farewik, E. M.; and Wang, R. 2022. Ankle joint torque prediction using an NMS solver informed-ANN model and transfer learning. *IEEE journal of biomedical and health informatics*, 26(12): 5895–5906.

Zhao, Y.; Jing, S.; Wu, H.; Li, H.; and Todoh, M. 2024. E-trgan: A novel transformer generative adversarial network for high-density surface electromyography signal reconstruction. *IEEE Transactions on Instrumentation and Measurement*.