Adaptive EMG Pattern Classification via Probabilistic Knowledge Transfer With Scale Mixture-Based Bayesian Sequential Learning

Seitaro Yoneda, *Student Member, IEEE*, and Akira Furui, *Member, IEEE*

***Abstract*— Electromyogram (EMG) signals, measured non-invasively from the skin surface, reflect human motion intentions and enable device control through pattern classification, particularly in applications such as myoelectric prostheses. However, continuous use of EMG-based interfaces remains challenging due to temporal variations in signal characteristics caused by muscle fatigue and electrode shift, leading to gradual degradation in classification accuracy. To address this limitation, we propose an adaptive method that integrates a scale mixture classification model (SMCM) with Bayesian sequential self-training (BSST), enabling probabilistic knowledge transfer across trials. In this framework, the posterior distribution of model parameters is sequentially updated through Bayesian updates using pseudo-labels assigned based on prediction confidence. This approach enables adaptation to evolving signal characteristics without storing historical data. Furthermore, SMCM represents signal variability through variance uncertainty modeling, thereby improving both the representation of EMG signal distributions and the reliability of prediction confidence estimation. We evaluated the proposed method using both short-term (within-day) and long-term (30 days) EMG datasets. The results showed that our method outperformed conventional methods in both classification accuracy and confidence estimation, while effectively mitigating accuracy degradation over time. These results demonstrate that the combination of SMCM and BSST provides effective adaptation to EMG signal variations, offering a practical solution for reliable EMG-based interfaces. Code is available at https://github.com/Yseitaro/smcm-bsst**

***Index Terms*— Electromyogram (EMG), Bayesian updating, sequential learning, adaptive pattern classification.**

I. Introduction

ELECTROMYOGRAM (EMG) signals are bioelectrical signals generated during muscle contraction that can be measured non-invasively from the skin surface. These signals contain inherent information about human movement, including muscle exertion levels and motor intent, which enables motion estimation through machine learning techniques. Leveraging these characteristics, EMG signals can serve as control interfaces for various devices, particularly in applications such as myoelectric prostheses [1], [2].

However, a major challenge in EMG-based interfaces is the gradual degradation of motion classification accuracy during prolonged use [3], [4], [5], [6]. This degradation occurs as the signal characteristics evolve over time due to multiple factors, such as muscle fatigue [7], [8] and electrode shift [9], [10], which causes significant deviation from the characteristics of the original training data. Consequently, maintaining classification accuracy over time requires sequential updates to the classification model to accommodate these evolving signal characteristics [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23].

Existing approaches to address this challenge can be broadly classified into two categories. The first category involves sequential learning using calibration data, in which additional labeled data are periodically collected from subjects to adapt the classification model [12], [13], [14], [15], [16]. However, regularly obtaining labeled calibration data imposes a significant burden on subjects and is impractical. The second category is sequential learning based on self-training [17], [18], [19], [20], [21], [22], [23]. This method performs sequential learning using predicted labels and test data, eliminating the need for labeled calibration data. Although this approach reduces subject burden, using all test data for self-training risks corrupting the model through the incorporation of incorrectly labeled samples.

Furthermore, both approaches share a common limitation: they require the retention of past data to prevent catastrophic forgetting [24], [25]. While some studies have introduced algorithms for periodic replacement of accumulated data [18], [26], these approaches heavily rely on heuristic parameter settings, such as replacement rate and data selection criteria. This requirement not only leads to significantly increased learning time as data accumulates but also raises concerns regarding privacy and memory constraints [27].

To overcome these limitations, we focus on introducing the concept of Bayesian updating into sequential self-training.

Received 22 January 2025; revised 23 May 2025; accepted 5 July 2025. Date of publication 14 July 2025; date of current version 24 July 2025. This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP23K28128. *(Corresponding authors: Seitaro Yoneda; Akira Furui.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Hiroshima University Ethics Committee under Application E-840.

The authors are with the Graduate School of Advanced Science and Engineering, Hiroshima University, Higashihiroshima 739-8527, Japan (e-mail: seitaroyoneda@hiroshima-u.ac.jp; akirafurui@hiroshima-u.ac.jp).

Digital Object Identifier 10.1109/TNSRE.2025.3587985

In this method, information about past signal characteristics is represented as a prior distribution, and the posterior distribution reflecting the current signal characteristics is sequentially updated through repeated Bayesian updates. This approach eliminates the need to accumulate past data for sequential learning while maintaining the model's ability to adapt to changing signal characteristics.

Building on this concept, we propose an adaptive classification method for EMG patterns based on Bayesian sequential self-training (BSST), which enables probabilistic knowledge transfer through sequential posterior updates. The proposed method incorporates a classification model based on scale mixture distributions, which we refer to as the scale mixture classification model (SMCM) [28], that considers the generative process of EMG signals to model changes in the distribution tails arising from signal variance uncertainty. This allows us to accurately represent the characteristics of the generative distribution of EMG signals and enhance the estimation performance of prediction confidence. Subsequently, we estimate motion class predictions and their confidences, and sequentially update the posterior distribution of the SMCM parameters using combinations of test data with high confidence and pseudo-labels. Hereinafter, we refer to this combined approach as SMCM+BSST. This proposed adaptation method is expected to contribute to reducing user burden and improving control reliability, especially in practical interfaces such as myoelectric prostheses, by eliminating the need for periodic collection of calibration data and suppressing accuracy degradation during long-term use.

A preliminary version of this work was presented at [29]. The present paper substantially extends our previous work with the following major contributions:

- Development of an enhanced Bayesian sequential learning method that employs scale mixture distributions, providing more accurate modeling of EMG signal variance fluctuations compared to conventional Gaussian distributions
- Comprehensive evaluation using both short-term and long-term EMG datasets to demonstrate the method's robustness across different temporal scales of signal characteristic changes
- Systematic analysis of the method's performance under several model update scenarios to address practical implementation considerations
- Introduction of expected calibration error (ECE) metrics to quantitatively assess the reliability of prediction confidence estimates, enabling more objective evaluation of model performance

The remainder of this paper is organized as follows: Section II discusses related work. Section III presents the details of SMCM and the algorithm for BSST. Section IV describes simulation experiments using synthetic data to visualize the sequential learning process. Section V demonstrates the effectiveness of our method through EMG pattern classification experiments using two types of EMG datasets. Finally, Section VI concludes the paper with a discussion of our findings and future directions.

II. Related Work

A. Adaptive Learning in EMG Classification

1) Adaptive Methods Using Calibration Data: Methods for adaptive EMG signal classification have been developed that maintain model accuracy by adapting to temporal changes in signal characteristics. One major approach involves collecting additional labeled calibration data from subjects to adjust model parameters to reflect current signal characteristics. Several studies have demonstrated successful adaptation using traditional statistical methods such as linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) [12], [13], [14]. These methods mitigate accuracy degradation by effectively integrating model parameters learned from previous sessions with those derived from new calibration data. Recent deep learning approaches have shown promising results: researchers have introduced session calibrators to reduce inter-session variability [15] and demonstrated adaptation through fine-tuning of pre-trained convolutional neural networks [16].

Although these calibration-based approaches, including both traditional statistical methods and deep learning techniques, can achieve high classification accuracy, they uniformly require periodic collection of labeled calibration data from subjects. This requirement becomes particularly burdensome when classifying multiple motions, limiting practical applicability.

2) Adaptive Methods Without Requiring Calibration Data: To address the burden of periodic calibration data collection, researchers have explored unsupervised adaptive learning methods that leverage the model's own predictions through self-training, thus eliminating the need for additional labeled data.

Several studies have proposed methods for sequentially updating parameters of traditional classifiers such as LDA, QDA [17], [18], [19], [20], and support vector machine (SVM) [21], [22], [23] using prediction results. To mitigate the risk of model degradation from incorrect predictions, various data selection criteria have been proposed, including likelihood estimation [19], entropy [30], prediction confidence [20], and unknown pattern rejection [31]. While using only high-confidence predictions for updates has shown promise, proper representation of uncertainty in the classification model remains a key challenge for effective self-training [20].

A common limitation across these adaptive methods is the need to retain historical data to maintain knowledge of past learning. Although some approaches attempt to manage this through periodic data replacement strategies [18], [26], these solutions heavily depend on heuristic parameter settings.

B. Classification Models and Prediction Confidence

To effectively implement self-training, it is important to appropriately evaluate the reliability of prediction results. Below, we discuss the evaluation of prediction confidence in discriminative and generative modeling approaches.

1) Approach Based on Discriminative Models: Discriminative models such as deep neural networks (DNNs) and SVMs directly learn the decision boundary $p(y|x)$ (where x and

y represent the input feature and corresponding class label, respectively) from data and generally achieve higher classification accuracy compared to generative models [32], [33]. However, these models focus solely on learning decision boundaries between classes without considering the overall probability distribution of the input space, making it difficult to appropriately evaluate prediction confidence. This limitation makes it challenging to reliably select high-confidence predictions for self-training.

2) Approach Based on Generative Models: Generative models, such as LDA and QDA, first model the data distribution as the product of the class-conditional probability $p(x|y)$ and the prior distribution $p(y)$, expressed as $p(x, y) = p(x|y)p(y)$. The posterior probability is then calculated using Bayes' theorem:

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

By modeling the input data distribution explicitly, these models can generally infer prediction confidence more reliably than discriminative models. However, their effectiveness depends heavily on how well the assumed distribution matches the true data distribution; significant mismatches can lead to inaccurate confidence estimates [34], [35].

C. Probabilistic Nature of EMG

Traditionally, reseachers have modeled the probability distribution of EMG signals using Gaussian distributions [36], [37], [38]. However, recent studies have shown that the actual distribution has heavier tails than a Gaussian distribution due to variance variations related to muscle activity [28], [39]. These distributional characteristics vary significantly with changes in muscle contraction level and fatigue, contributing substantially to signal variability both within and between recording sessions [28], [40]. Furthermore, external factors such as electrode displacement and changes in skin impedance introduce additional variability into the signal characteristics [9], [10]. Consequently, effective EMG signal classification requires probabilistic models capable of flexibly accommodating distribution shape changes, particularly in capturing the heavy-tailed nature of the signal variations.

D. Utility of Bayesian Approaches

Bayesian models, which treat parameters as probability distributions, offer potential solutions to the aforementioned challenges. Unlike traditional approaches based on maximum likelihood estimation, Bayesian models can evaluate prediction confidence while accounting for parameter uncertainty, thereby enabling effective filtering of unreliable predictions during self-training.

Bayesian approaches also provide efficient parameter updating mechanisms. By representing past learning outcomes as posterior distributions that serve as priors for new data, they enable sequential learning without direct data storage. This integration of prior knowledge and new observations occurs naturally within a theoretically consistent framework, eliminating the need for heuristic update parameters.

Several studies have demonstrated the effectiveness of Bayesian sequential learning in biomedical applications [41], [42]. For instance, a Bayesian model based on a Gaussian mixture distribution has successfully adapted to temporal changes in EEG signals [42]. In our preliminary work [29], we proposed a Gaussian-based Bayesian model for EMG adaptation. However, given the heavy-tailed nature of EMG signals discussed earlier, more sophisticated probabilistic models are needed to accurately represent their statistical properties.

E. Our Contribution

The challenges of existing methods can be summarized in three key aspects: the representation of prediction confidence, the stability of parameter updates, and the storage cost of past data. To address these challenges, we propose an unsupervised adaptive learning method that integrates Bayesian sequential learning with scale mixture distributions.

The proposed approach formulates scale mixture distributions in a Bayesian framework, effectively representing the heavy-tailed characteristics of EMG signals and enabling reliable evaluation of prediction confidence. This reliable confidence estimation allows us to selectively utilize high-confidence predictions for model updates, maintaining stability during adaptation. Furthermore, by encoding past signal characteristics in prior distributions rather than storing raw data, we achieve efficient sequential learning without the storage costs associated with conventional methods.

III. Proposed Method

Fig. 1 shows an overview of the proposed method. In this method, features are first extracted from EMG signals acquired during each trial t and then input to a trained classification model. The classification model estimates both the class label and prediction confidence for the input data. Subsequently, BSST is performed using only test data with prediction confidence above a specified threshold, which updates the posterior distribution of the model parameters. By repeating the above process for each trial t, adaptive motion classification is achieved.

A. Scale Mixture Classification Model (SMCM)

Fig. 2 illustrates the relationship between the EMG signal $\mathbf{x}_n \in \mathbb{R}^D$ measured from D electrodes at data point n and its corresponding motion class label $\mathbf{y}_n \in \{0, 1\}^C$ (C: number of classes) using a graphical model. In this model, EMG signal variability is represented as uncertainty in the variance by allowing the variance of $\mathbf{x}_n$ to depend on the random variable u_{nc}.

For a certain class c, the observation model of the EMG signal $\mathbf{x}_n$ is expressed via the following scale mixture distribution:

$$p(\mathbf{x}_n|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c, \nu_c) = \int p(\mathbf{x}_n|u_{nc}, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)p(u_{nc}|\nu_c)du_{nc}. \quad (1)$$

The conditional distribution $p(\mathbf{x}_n|u_{nc}, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$ represents the EMG signal $\mathbf{x}_n$ given the scale parameter u_{nc}, which we model using a Gaussian distribution:

$$p(\mathbf{x}_n|u_{nc}, \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) = \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_c, u_{nc}\boldsymbol{\Sigma}_c), \quad (2)$$

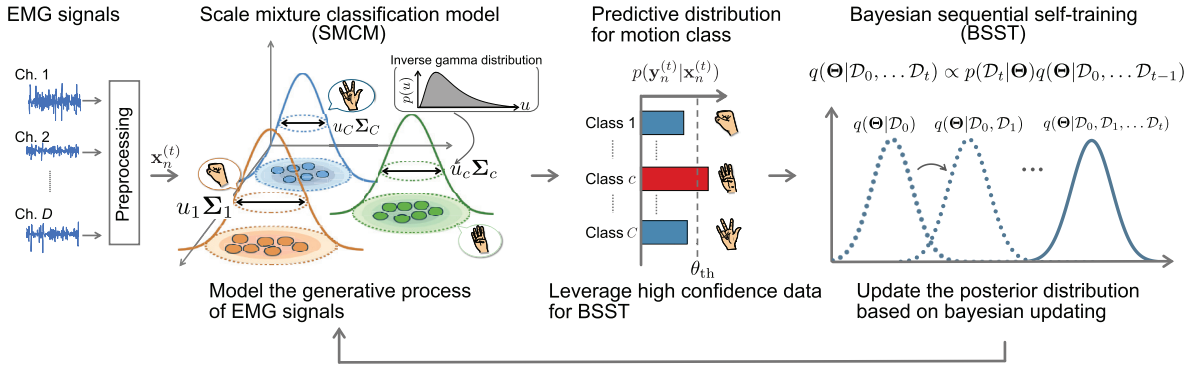

Fig. 1. Overview of the proposed adaptive EMG pattern classification method. Raw EMG signals from multiple channels are preprocessed through rectification and smoothing. The processed signals are then fed into a SMCM, which models the generative process of EMG signals. The model calculates posterior probabilities for each motion class to determine the predicted labels and their confidence levels. For samples exceeding a confidence threshold θ_{th}, BSST updates the posterior distribution of the model parameters to adapt to temporal changes in signal characteristics.

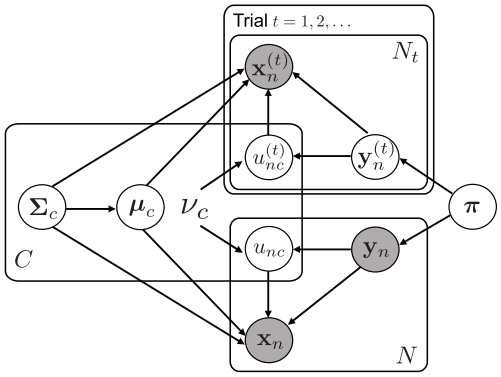

Fig. 2. Graphical model representation of SMCM. Gray nodes indicate observed variables, white nodes represent unobserved variables and model parameters. The plate notation indicates repeated variables, with N samples for initial training and N_t samples for trial t. The EMG signal $\mathbf{x}_n$ follows a scale mixture distribution parameterized by class-specific means $\boldsymbol{\mu}_c$, covariance matrices $\boldsymbol{\Sigma}_c$, and degrees of freedom ν_c, with mixing proportions $\boldsymbol{\pi}$.

where $\boldsymbol{\mu}_c \in \mathbb{R}^D$ and $\boldsymbol{\Sigma}_c \in \mathbb{R}^{D\times D}$ are the mean vector and covariance matrix corresponding to class c, respectively. To account for the uncertainty in signal variance, we model the scale parameter u_{nc} using an inverse gamma distribution:

$$p(u_{nc}|\nu_c) = \text{IG}\left(u_{nc} \,\middle|\, \frac{\nu_c}{2}, \frac{\nu_c}{2}\right), \tag{3}$$

where ν_c is the degrees of freedom parameter that controls the level of uncertainty in the variance. This distribution is a conjugate prior for the precision (inverse variance) of the Gaussian distribution, and we adopted it because it offers the computational advantage of enabling analytical update formulas in variational Bayesian inference.

By combining these distributions, we obtain a closed-form expression for the scale mixture distribution:

$$\begin{aligned} &p(\mathbf{x}_n|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c, \nu_c) \\ &\quad= \int \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_c, u_{nc}\boldsymbol{\Sigma}_c)\,\text{IG}\left(u_{nc} \,\middle|\, \frac{\nu_c}{2}, \frac{\nu_c}{2}\right) du_{nc}, \\ &\quad= \frac{\Gamma\left(\frac{\nu_c+D}{2}\right)}{\Gamma\left(\frac{\nu_c}{2}\right)} \frac{|\boldsymbol{\Sigma}_c|^{-\frac{1}{2}}}{(\pi\nu_c)^{\frac{D}{2}}} \left(1+\Delta_{nc}^2\right)^{-\frac{\nu_c+D}{2}}, \end{aligned} \tag{4}$$

where $\Delta_{nc}^2 = (\mathbf{x}_n - \boldsymbol{\mu}_c)^\top \boldsymbol{\Sigma}_c^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_c)$.

To assign the C observation models to each data point, we introduce a one-hot encoded class label representation $\mathbf{y}_n = \{y_{nc}\}_{c=1}^C$, where $y_{nc} = 1$ indicates assignment to the c-th class. The class label $\mathbf{y}_n$ is expressed using the following categorical distribution:

$$p(\mathbf{y}_n|\boldsymbol{\pi}) = \text{Cat}(\mathbf{y}_n|\boldsymbol{\pi}) = \prod_{c=1}^C \pi_c^{y_{nc}}, \tag{5}$$

where $\boldsymbol{\pi} = \{\pi_c\}$ represents the mixture proportions, with $\sum_{c=1}^C \pi_c = 1$.

Using equations (4) and (5), we obtain the complete observation model given $\mathbf{y}_n$ as follows:

$$p(\mathbf{x}_n|\mathbf{y}_n, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_c) = \prod_{c=1}^C p(\mathbf{x}_n|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c, \nu_c)^{y_{nc}}, \tag{6}$$

where $\boldsymbol{\mu} = \{\boldsymbol{\mu}_c\}$ and $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_c\}$. This model selects the appropriate scale mixture distribution corresponding to the class label as the observation model.

To treat SMCM as a Bayesian probabilistic model, we assign conjugate prior distributions to the model parameters. The mean vectors $\boldsymbol{\mu}_c$ and covariance matrices $\boldsymbol{\Sigma}_c$ follow a Gaussian-inverse Wishart distribution, while the class mixture proportions $\boldsymbol{\pi}$ follow a Dirichlet distribution:

$$p(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) = \mathcal{N}(\boldsymbol{\mu}_c|\mathbf{m}_0, \beta_0^{-1}\boldsymbol{\Sigma}_c)\mathcal{IW}(\boldsymbol{\Sigma}_c|\mathbf{W}_0, \eta_0), \tag{7}$$

$$p(\boldsymbol{\pi}) = \text{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha}), \tag{8}$$

where $\mathbf{m}_0 \in \mathbb{R}^D$, $\beta_c \in \mathbb{R}^+$, $\eta_0 > D-1$, $\mathbf{W}_0 \in \mathbb{R}^{D\times D}$, and $\boldsymbol{\alpha}_0 = \{\alpha_0\}_{c=1}^C$ are hyperparameters. We do not specify a prior for the degrees-of-freedom parameter ν_c because it lacks a conjugate prior distribution.

B. Bayesian Sequential Self-Training (BSST)

In the proposed method, we introduce BSST to adapt to changes in the signal characteristics over time. This approach unifies initial training and sequential self-training within a single Bayesian updating framework. In the initial training phase, the posterior distributions of the model parameters are inferred using training data. In the sequential self-training phase,

we assign pseudo-labels to incoming test data and recursively update the posterior distributions using these labeled samples. This sequential updating mechanism enables continuous model adaptation as new data arrives.

1) Initial Training: Using the training dataset $\mathcal{D}_0 = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^N$, we calculate the posterior distribution $p(\mathbf{U}, \boldsymbol{\Theta}|\mathcal{D}_0)$ of latent variables $\mathbf{U} = \{u_{nc}|n = 1, \ldots N; c = 1, \ldots C\}$ and model parameters $\boldsymbol{\Theta} = \{\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$. Since the marginal likelihood $p(\mathcal{D}_0)$ cannot be computed analytically, we employ variational inference. We introduce a variational posterior distribution $q(\mathbf{U}, \boldsymbol{\Theta}|\mathcal{D}_0)$ to approximate the true posterior $p(\mathbf{U}, \boldsymbol{\Theta}|\mathcal{D}_0)$ by minimizing the Kullback-Leibler (KL) divergence between these distributions [43]. Following the mean-field approximation, we assume the following independence in the variational posterior distribution:

$$q(\mathbf{U}, \boldsymbol{\Theta}|\mathcal{D}_0) \approx q(\mathbf{U}|\mathcal{D}_0)q(\boldsymbol{\Theta}|\mathcal{D}_0), \tag{9}$$

Based on this decomposition, we alternately optimize the variational posterior distribution of the latent variables and the model parameters.

First, the logarithm of the variational posterior distribution of the latent variables is given as follows:

$$\begin{aligned}\ln q(\mathbf{u}_n|\mathcal{D}_0) = \langle \ln p(\mathbf{x}_n|\mathbf{y}_n, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{u}_n)\rangle_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})} \\ + \ln p(\mathbf{u}_n|\mathbf{y}_n) + \text{const.}\end{aligned} \tag{10}$$

Here, $\langle\cdot\rangle_{q(\boldsymbol{\mu},\boldsymbol{\Sigma})}$ denotes the expectation with respect to the variational posterior distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and $\mathbf{u}_n = \{u_{nc}\}_{c=1}^C$. By calculating the above expression, the variational posterior distribution of the latent variable u_{nc} can be represented by the following inverse gamma distribution:

$$q(u_{nc}|\mathcal{D}_0) = \text{IG}(u_{nc}|a_{nc}, b_{nc}). \tag{11}$$

The parameters a_{nc} and b_{nc} of the inverse gamma distribution are defined as follows:

$$a_{nc} = \frac{\nu_c + D}{2}, \quad b_{nc} = \frac{\nu_c + \langle\Delta_{nc}^2\rangle}{2}, \tag{12}$$

where $\langle\Delta_{nc}^2\rangle$ is expressed as:

$$\langle\Delta_{nc}^2\rangle = D\beta_c^{-1} + \eta_c(\mathbf{x}_n - \mathbf{m}_c)^\top \mathbf{W}_c^{-1}(\mathbf{x}_n - \mathbf{m}_c). \tag{13}$$

Next, we determine the variational posterior distribution $q(\boldsymbol{\Theta}|\mathcal{D}_0)$. To simplify the computation, we define the following variables:

$$N_c = \sum_{n=1}^N y_{nc}, \tag{14}$$

$$\omega_c = \sum_{n=1}^N y_{nc}\langle u_{nc}^{-1}\rangle, \tag{15}$$

$$\bar{\mathbf{x}}_c = \omega_c^{-1}\sum_{n=1}^N y_{nc}\langle u_{nc}^{-1}\rangle \mathbf{x}_n, \tag{16}$$

$$\mathbf{S}_c = \omega_c^{-1}\sum_{n=1}^N y_{nc}\langle u_{nc}^{-1}\rangle (\mathbf{x}_n - \bar{\mathbf{x}}_c)(\mathbf{x}_n - \bar{\mathbf{x}}_c)^\top, \tag{17}$$

where $\langle u_{nc}^{-1}\rangle$ is defined as follows:

$$\langle u_{nc}^{-1}\rangle = \frac{a_{nc}}{b_{nc}}. \tag{18}$$

The logarithm of the variational posterior distribution $q(\boldsymbol{\Theta}|\mathcal{D}_0)$ is given as follows:

$$\begin{aligned}\ln q(\boldsymbol{\Theta}|\mathcal{D}_0) = \langle \ln p(\mathbf{X}|\mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{U})\rangle_{q(\mathbf{U})} + \ln p(\mathbf{Y}|\boldsymbol{\pi}) \\ + \ln p(\boldsymbol{\pi}) + \ln p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) + \text{const},\end{aligned} \tag{19}$$

where $\mathbf{X} = \{\mathbf{x}_n\}$ and $\mathbf{Y} = \{\mathbf{y}_n\}$ collect all observations and their labels. The resulting variational posterior distributions $q(\boldsymbol{\pi})$ and $q(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$ take the form of Gaussian-inverse Wishart and Dirichlet distributions, respectively:

$$q(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c|\mathcal{D}_0) = \mathcal{N}(\boldsymbol{\mu}_c|\mathbf{m}_c, \beta_c^{-1}\boldsymbol{\Sigma}_c)\mathcal{IW}(\boldsymbol{\Sigma}_c|\mathbf{W}_c, \eta_c), \tag{20}$$

$$q(\boldsymbol{\pi}|\mathcal{D}_0) = \text{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha}). \tag{21}$$

Here, $\mathbf{m}_c$, β_c, ν_c, $\mathbf{W}_c$, and $\boldsymbol{\alpha}$ are hyperparameters of the variational posterior distribution, defined as follows:

$$\beta_c = \beta_0 + \omega_c, \tag{22}$$

$$\mathbf{m}_c = \frac{1}{\beta_c}(\omega_c\bar{\mathbf{x}}_c + \beta_0\mathbf{m}_0), \tag{23}$$

$$\mathbf{W}_c = \mathbf{W}_0 + \omega_c\mathbf{S}_c + \frac{\beta_0\omega_c}{\beta_c}(\bar{\mathbf{x}}_c - \mathbf{m}_0)(\bar{\mathbf{x}}_c - \mathbf{m}_0)^\top, \tag{24}$$

$$\eta_c = \eta_0 + \sum_{n=1}^N y_{nc}, \tag{25}$$

$$\alpha_c = \alpha_0 + \sum_{n=1}^N y_{nc}. \tag{26}$$

The degrees of freedom parameter ν_c is updated by maximizing the following log-likelihood using stochastic gradient methods:

$$\begin{aligned}\tilde{\nu}_c &= \arg\max_{\nu_c} \ln p(\mathbf{X}|\mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_c), \\ &= \arg\max_{\nu_c} \ln \prod_{n=1}^N\prod_{c=1}^C p(\mathbf{x}_n|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c, \nu_c)^{y_{nc}}.\end{aligned} \tag{27}$$

The variational inference procedure described above iteratively updates the latent variables and model parameters. Starting with initialized prior hyperparameters β_c, η_c, $\mathbf{m}_c$, and $\mathbf{W}_c$, we alternately compute the variational posterior distributions $q(\mathbf{U}|\mathcal{D}_0)$ and $q(\boldsymbol{\Theta}|\mathcal{D}_0)$ until convergence.

2) Sequential Self-Training: We perform sequential self-training using the dataset $\mathcal{D}_t = \{(\mathbf{x}_n^{(t)}, \hat{\mathbf{y}}_n^{(t)})\}_{n=1}^{N_t}$ from test trial $t \in \{1, 2, \ldots\}$. Here, $\hat{\mathbf{y}}_n^{(t)}$ represents the pseudo-label assigned based on predictions from the model trained in trial $t-1$. The variational posterior distributions of the model parameters at trial t are computed by recursively applying Bayes' theorem:

$$\begin{aligned}&q(\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c|\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_t) \\ &\quad\propto p(\mathcal{D}_t|\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c)\, q(\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c|\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_{t-1}),\end{aligned} \tag{28}$$

$$\begin{aligned}&q(\boldsymbol{\pi}|\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_t) \\ &\quad\propto p(\mathcal{D}_t|\boldsymbol{\pi})\, q(\boldsymbol{\pi}|\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_{t-1}),\end{aligned} \tag{29}$$

$$q(\mathbf{u}_n^{(t)}|\mathcal{D}_t) \propto p(\mathcal{D}_t|\mathbf{u}_n^{(t)})\, p(\mathbf{u}_n^{(t)}). \tag{30}$$

We fix ν to its value from the initial training, assuming that the degree of uncertainty remains relatively stable across trials.

Due to the conjugacy of the prior distributions, the variational posterior distributions take the following form:

$$q(\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c|\mathcal{D}_0,\mathcal{D}_1,\ldots,\mathcal{D}_t) = \mathcal{NIW}(\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c|\mathbf{m}_c^{(t)}, \beta_c^{(t)}, \eta_c^{(t)}, \mathbf{W}_c^{(t)}), \tag{31}$$

$$q(\boldsymbol{\pi}|\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_t) = \mathrm{Dir}(\boldsymbol{\pi}|\boldsymbol{\alpha}^{(t)}). \tag{32}$$

These posterior distributions are computed using the same update equations as the initial training with two key modifications: we replace the true labels $\mathbf{y}_n$ with the pseudo-labels $\hat{\mathbf{y}}_n$, and use the posterior hyperparameters from trial $t-1$ instead of the prior hyperparameters in (22)–(26). By sequentially performing these updates for $t = 1, 2, \ldots$, the model continuously adapts to changes in the EMG signal characteristics through the recursive Bayesian updates.

C. Motion Class Prediction and Assignment of Pseudo-Labels

For motion class label assignment to the test data $\mathbf{x}_n^{(t)}$ obtained during trial t, we compute the predictive distribution:

$$p(\mathbf{y}_n^{(t)}|\mathbf{x}_n^{(t)}, \mathcal{D}_0, \ldots, \mathcal{D}_{t-1}) \propto p(\mathbf{x}_n^{(t)}|\mathbf{y}_n^{(t)}, \mathcal{D}_0, \ldots, \mathcal{D}_{t-1}) p(\mathbf{y}_n^{(t)}|\mathcal{D}_0, \ldots, \mathcal{D}_{t-1}). \tag{33}$$

The two terms in the second equation above can be calculated as follows:

$$p(\mathbf{x}_n^{(t)}|\mathbf{y}_n^{(t)}, \mathcal{D}_0, \ldots, \mathcal{D}_{t-1}) = \int\int p(\mathbf{x}_n^{(t)}|\mathbf{y}_n^{(t)}, \boldsymbol{\Theta}, \mathbf{U}) q(\boldsymbol{\Theta}, \mathbf{U}|\mathcal{D}_0, \ldots, \mathcal{D}_{t-1}) d\boldsymbol{\Theta} d\mathbf{U}, \tag{34}$$

$$p(\mathbf{y}_n^{(t)}|\mathcal{D}_0, \ldots, \mathcal{D}_{t-1}) = \mathrm{Cat}(\mathbf{y}_n^{(t)}|\hat{\boldsymbol{\pi}}), \tag{35}$$

where $\hat{\boldsymbol{\pi}} = \{\pi_c\}$, $\pi_c = \alpha_c / \sum_{c=1}^{C} \alpha_c$. Since the integral in (34) cannot be solved analytically, we approximate it for a certain class c as follows:

$$p(\mathbf{x}_n^{(t)}|y_{nc}^{(t)} = 1, \mathcal{D}_0, \ldots \mathcal{D}_{t-1}) \approx p(\mathbf{x}_n^{(t)}|\langle\boldsymbol{\mu}_c\rangle, \langle\boldsymbol{\Sigma}_c\rangle, \nu_c), \tag{36}$$

where $\langle\boldsymbol{\mu}_c\rangle = \mathbf{m}_c$, $\langle\boldsymbol{\Sigma}_c\rangle = (\eta_c - D - 1)^{-1}\mathbf{W}_c$.

For each class c, we compute posterior probability using (33) and determine the most probable class $\hat{c}_n$ as follows:

$$\hat{c}_n = \arg\max_c \; p\left(y_{nc}^{(t)} = 1|\mathbf{x}_n^{(t)}, \mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_{t-1}\right). \tag{37}$$

In the proposed BSST, we construct pseudo-labels $\hat{\mathbf{y}}_n^{(t)}$ based on $\hat{c}_n$. Using all predicted class labels for self-training could degrade model performance through the incorporation of incorrect predictions. To mitigate this risk, we select predictions based on their confidence values. The prediction confidence for each data point in trial t is calculated as:

$$\mathrm{conf}\left[\mathbf{y}_n^{(t)}\right] = \max_c \; p\left(y_{nc}^{(t)} = 1|\mathbf{x}_n^{(t)}, \mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_{t-1}\right). \tag{38}$$

We then assign pseudo-labels only to data points whose confidence exceeds a threshold θ_{th}, effectively excluding ambiguous predictions from the sequential self-training process.

D. Visualization of Sequential Updates

To demonstrate the adaptive behavior of the proposed method under sequential distribution changes, we use two-dimensional synthetic data with two classes. The data for class 1 and class 2 were generated from scale mixture distributions following $p(\mathbf{x}|\boldsymbol{\mu} = [-5 + t, 3]^\top, \boldsymbol{\Sigma} = 3.0\mathbf{I}, \nu = 2)$ and $p(\mathbf{x}|\boldsymbol{\mu} = [5 - t, 3]^\top, \boldsymbol{\Sigma} = 3.0\mathbf{I}, \nu = 2)$, where $\mathbf{I}$ denotes the identity matrix, and t represents the trial. The mean value of each distribution changes depending on the trial $t = 1, 2, \ldots, 10$, simulating a gradual shift in the data distribution. Note that the class distribution shift in this simulation is intentionally exaggerated to visually demonstrate the adaptive behavior of the proposed method, and may differ from the actual scale of EMG signal variations observed in practice.

For each trial t, we generated 300 data points for each class, constituting a dataset $\mathcal{D}_t$. The dataset $\mathcal{D}_1$ of trial 1 was used for the initial training, and subsequent datasets $\mathcal{D}_t$ were sequentially input into the model to demonstrate the BSST process. For comparison, we also show the results of a Gaussian classification model (GCM), which assumes Gaussian distributions instead of scale mixture distributions.

Fig. 3 visualizes the adaptation of the proposed method and GCM to distribution shifts. Circles and stars indicate data points and their mean points for each class, respectively. The colored regions represent prediction probabilities, with red and blue hues indicating the probability levels for each class.

A key advantage of the proposed method is its ability to appropriately represent prediction uncertainty. As shown in Fig. 3(a), prediction confidence decreases with distance from class means, reflecting reduced reliability in sparse data regions. In contrast, GCM shows uniformly high confidence even in uncertain regions (Fig. 3(b)). As the class distributions shift over trials, the proposed method maintains appropriate uncertainty estimates while GCM produces overconfident predictions, demonstrating the effectiveness of our approach in handling both distribution shifts and uncertainty estimation.

IV. Experiments

A. Datasets and Processing

To verify the effectiveness of the proposed method, we conducted a comprehensive evaluation using two EMG datasets collected over different time periods. The details of the datasets are as follows.

- Dataset I (Short-term) [29]: EMG signals were recorded from six healthy male participants (mean age: 22.6 years) performing six motions ($C = 6$): hand opening/closing, wrist flexion/extension, and forearm pronation/supination. During the approximately 30-minutes experiment, each participant performed 20 trials while seated with their right elbow resting on a desk. The signals were recorded using four electrodes ($D = 4$) placed equidistantly around the forearm using a wireless measurement system (Delsys Trigno) at 2000 Hz.
- Dataset II (Long-term) [44]: EMG recordings were collected from five participants (mean age: 25.4 years) over 30 days, with each participant performing eight

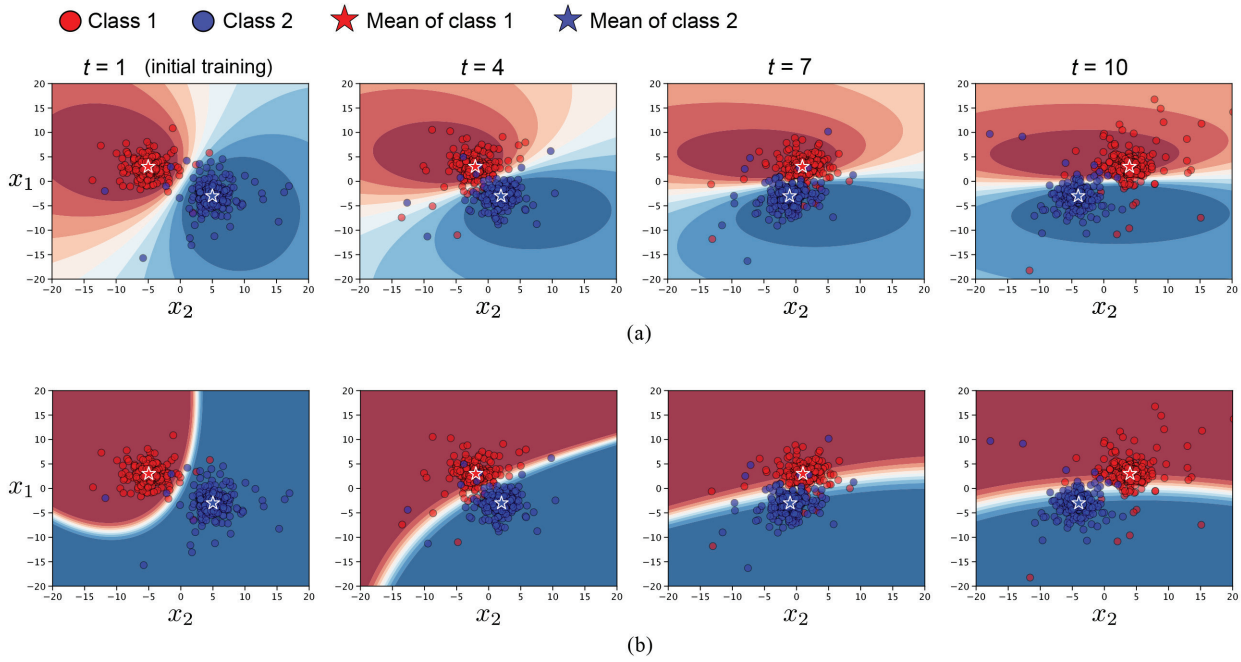

Fig. 3. Visualization of sequential model adaptation for simulated two-class data across different trials. (a) SMCM+BSST. (b) GCM+BSST. Scatter plots represent data points (circles) and class means (stars) for each class. The colored regions indicate class prediction probabilities, with red and blue hues representing the probability of belonging to class 1 and class 2, respectively. The progression from $t = 1$ (initial training) to $t = 10$ demonstrates how each model adapts to the gradually shifting class distributions.

single-DoF motions ($C = 8$) including hand opening/closing, wrist flexion/extension, radial/ulnar deviation, and forearm pronation/supination. Each participant completed four daily trials, resulting in 120 total trials (4×30). The signals were recorded using a Myo Gesture Control Armband with eight electrodes ($D = 8$) placed around the forearm at 200 Hz. This dataset is publicly available on GitHub.[1]

- Dataset III (Amputee) [45]: EMG signals were recorded from nine amputee participants (seven traumatic and two congenital) performing six motions ($C = 6$): thumb/index flexion, fine pinch, tripod/hook/spherical grip. Each participant performed each motion at three force levels for 5 to 8 trials, with eight electrodes ($D = 8$) placed around the stump. Signals were recorded at 2000 Hz using LABVIEW (National Instruments, USA). This dataset is available on Dr. Khushaba's webpage.[2] For this study, only medium-force data from the first five trials were used for all participants to ensure consistency and avoid potential confounding effects from force variation.

Dataset I was collected by our research group following approval from the Hiroshima University Ethics Committee (approval number: E-840). All participants provided written informed consent before participating in the experiment.

For feature extraction, all EMG signals underwent full-wave rectification followed by smoothing using a second-order Butterworth low-pass filter (cutoff frequency: 2.0 Hz). For both datasets, the first two trials were used for training, while the remaining served as test data.

[1]https://github.com/Suguru55/Wearable_Sensor_Long-term_sEMG_Dataset

[2]https://www.rami-khushaba.com/biosignals-repository

B. Model Configuration and Training Parameters

The hyperparameters for the prior distributions in the proposed method were initialized to make a weakly informative prior as follows: $\beta_0 = 1$, $\eta_0 = D+1$, $\alpha_0 = 0.001$, and $\mathbf{m}_0 = \mathbf{0}$. Since $\mathbf{W}_0$ is highly influenced by the data scale, we initialize it conservatively using the training data for each class as follows: $\mathbf{W}_0 = (\eta_0 + D + 1)^{-1}\mathrm{diag}(\mathrm{Cov}(\mathbf{X}_c))$ where $\mathrm{Cov}(\mathbf{X}_c)$ is the covariance matrix of training data $\mathbf{X}_c$ for class c. As demonstrated in prior work [46], we employed a common value $\nu_c = \nu$ across all classes rather than class-specific degrees of freedom, preventing potential overrepresentation of uncertainty that could adversely affect generalization performance.

The threshold for prediction confidence in BSST was set to $\theta_{\mathrm{th}} = 0.5$. While higher thresholds could potentially provide more reliable updates, we chose this moderate threshold to achieve sufficient adaptability to the expected distribution shifts between trials while maintaining reasonable selectivity in the update process.

C. Comparative Evaluation

We conducted comparative experiments against both baseline variants and conventional methods. We adopted a trial-wise evaluation approach, following standard experimental protocols from previous studies. In this process, we first performed initial training of the model using the training data.

Then, we fed the test data into the model one trial at a time, and performed BSST using test data and pseudo-labels.

To isolate the impact of different components, we established three baseline comparisons:

- SMCM: The proposed method without BSST mechanism, serving as a baseline for evaluating the contribution of adaptive learning.
- SMCM+BSST (optimal): An idealized version of the proposed method that uses true labels (normally unavailable in practice) instead of pseudo-labels, representing the theoretical upper bound of performance.
- GCM+BSST [29]: A variant using Gaussian distributions instead of using scale mixture distributions, allowing us to evaluate the specific contribution of the scale mixture modeling.

Additionally, we compared our method against six conventional EMG classification approaches. Four were traditional statistical methods: LDA, QDA, polynomial classifier (PC), and SVM. The remaining two were deep learning methods: one-dimensional convolutional neural network (1D-CNN) and multilayer perceptron (MLP).

For traditional statistical methods, we implemented sequential self-training (SST) by combining training and test data, retaining 50% of previous data while replacing reminder with new samples [18]. LDA and QDA utilized only high-confidence samples ($\theta_{\text{th}} = 0.5$) for updates, while SVM and PC used all test samples due to their inability to output probability scores. The hyperparameters of SVM were optimized through grid search with two-fold cross-validation on the training data.

For deep learning methods, we performed sequential adaptation through fine-tuning (FT). Models were initially trained for 300 epochs on the first two trials (with validation-based selection), then fine-tuned for five epochs on each subsequent trial using high-confidence predictions ($\theta_{\text{th}} = 0.5$) as pseudo-labels. The architectures used were:

- 1D-CNN: Four convolutional layers with 16, 32, 64, and 128 feature maps, respectively. Each layer used a kernel size of 3, followed by batch normalization, PReLU activation, and max pooling. To capture temporal dynamics, input data were segmented into 200 ms windows with 50 ms steps.
- MLP: Three hidden layers with 100, 50, and 25 units, respectively. After each hidden layer, batch normalization and a ReLU activation were applied.

Both networks were trained using stochastic gradient descent with a learning rate of 10^{-3} and batch size of 128.

D. Behavioral and Sensitivity Analysis

We investigated the impact of different updating strategies by comparing two BSST approaches: trial-wise updating (processing all motion classes per trial) and class-wise updating (processing one motion class at a time). Since practical applications typically lack discrete trial boundaries, we examined class-wise updating as a more implementation-feasible strategy for real-world systems. To ensure robust evaluation, we randomized the sequence of motions within each trial and repeated this process ten times with different random seeds.

To deploy sequential learning on resource-constrained platforms such as microcontrollers in myoelectric prostheses and interfaces, minimizing memory usage is essential. Accordingly, we evaluated the memory footprint of our proposed method on long-term datasets II. We compared it against two LDA-based baselines: LDA+SST and LDA+CST (cumulative sequential training), which retains all past data throughout the sequence. Memory usage was calculated by tracking the size of model parameters and stored training examples for each approach throughout 120 trials.

Next, to assess parameter sensitivity, we analyzed how the confidence threshold for pseudo-label assignment affects classification performance. We systematically varied the threshold θ_{th} from 0.2 to 0.9 in increments of 0.1 to determine the optimal threshold that maximized classification accuracy for each participant.

V. Results

A. Classification Performance and Calibration

Fig. 4 shows the classification accuracy averaged across participants: per trial for (a) dataset I and (c) dataset III, and per day for (b) dataset II. For dataset II, the first day shows the average of two test trials (as two trials were used for initial training), while subsequent days show the average of all four daily trials. Note that SMCM+BSST (optimal) represents an ideal case where the true labels of test data are known in advance. Across all datasets, the proposed SMCM+BSST achieved higher classification accuracy than SMCM in most trials, with a particularly notable improvement over GCM+BSST in dataset II.

The reliability diagrams in Fig. 5 shows the differences in calibration performance between GCM+BSST and SMCM+BSST for each dataset. The plots group predictions by confidence level and compare them with actual accuracies. For well-calibrated models, the prediction confidence should match the actual accuracy (e.g., predictions with 70% confidence should be correct 70% of the time). The expected calibration error (ECE) [47] quantifies the overall calibration performance by computing the weighted average of these discrepancies across all bins, with lower values indicating better calibration. The proposed method demonstrated the lower ECE in both datasets, indicating that it could accurately estimate the prediction confidence.

Table I summarizes the performance using four metrics: "Overall accuracy" (average across all trials), "Final accuracy" (last trial performance), "Accuracy change" (calculated by subtracting the average accuracy of the first four test trials from that of the last four trials, with negative values indicating degradation; note that for dataset III, due to its limited number of trials, this metric is calculated by subtracting the accuracy of the first test trial from that of the last trial), and ECE (all models except SVM and PC). We clarify the performance gap between the proposed method and its theoretical upper bound by including the results of SMCM+BSST (optimal), which uses true labels instead of pseudo-labels for updates. To statistically analyze performance differences, we conducted

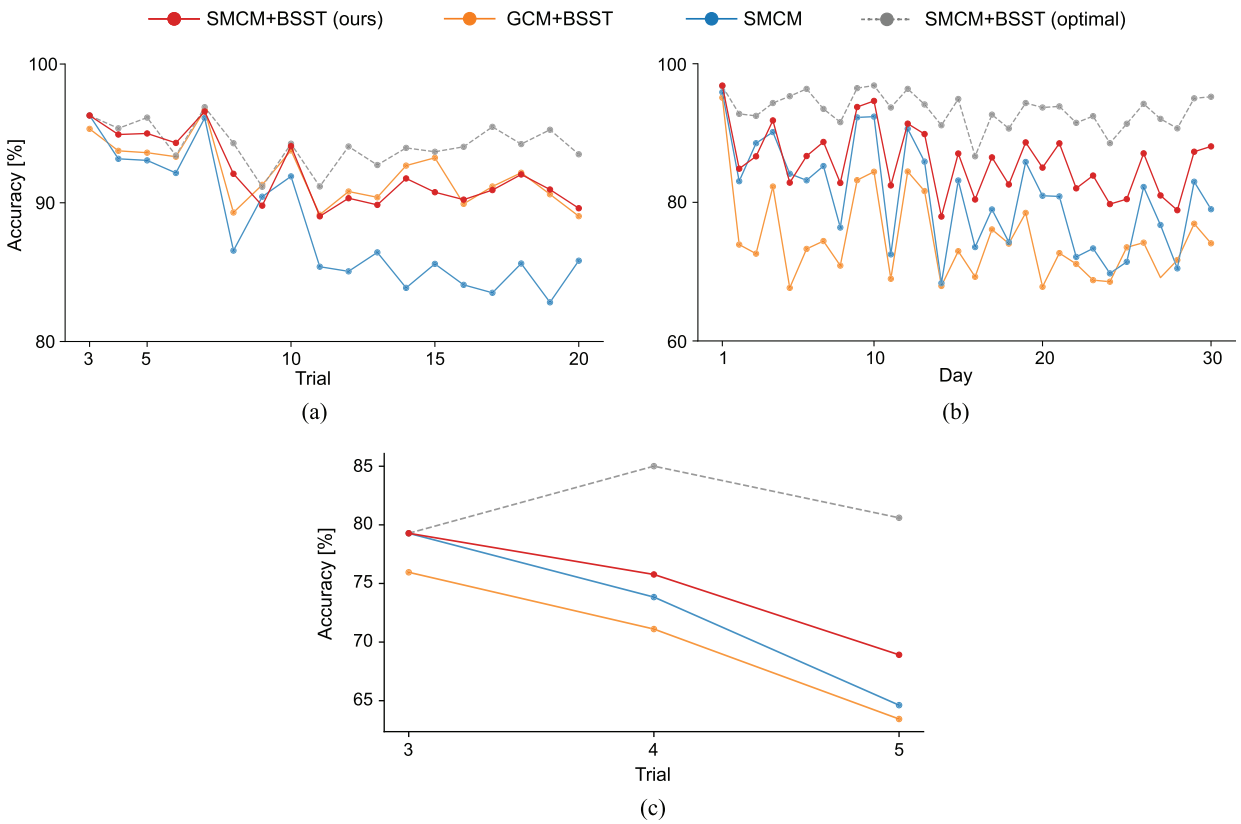

Fig. 4. Classification accuracy across trials for different models. (a) Dataset I. (b) Dataset II. (c) Dataset III. The first two trials were used for training, with remaining trials for testing. Red: proposed SMCM+BSST, orange: GCM+BSST, blue: SMCM, and gray: SMCM+BSST (optimal).

TABLE I
PERFORMANCE METRICS FOR EACH DATASET

Dataset	Classifier	Overall accuracy (↑)	Final accuracy (↑)	Accuracy change (↑)	ECE (↓)
I	LDA+SST	83.8%*	84.7%	-2.9%	5.4%
	QDA+SST	90.8%	**89.7%**	-2.3%	7.3%
	PC+SST	87.9%*	86.1%	-2.6%	N/A
	SVM+SST	89.8%	87.7%	-3.6%	N/A
	1D-CNN+FT	80.7%*	78.9%	**-1.9%**	14.7%
	MLP+FT	83.4%*	80.2%	-6.4%	13.9%
	SMCM+BSST (Ours)	**92.1%**	89.6%	-2.6%	**4.3%**
	SMCM+BSST (Optimal)†	94.2%	93.5%	-1.2%	2.2%
II	LDA+SST	78.9%*	74.6%	-7.1%	18.6%
	QDA+SST	74.5%*	77.5%	-7.7%	23.6%
	PC+SST	76.8%*	78.5%	-10.0%	N/A
	SVM+SST	83.0%*	85.3%	-7.8%	N/A
	1D-CNN+FT	78.4%*	73.0%	-19.1%	**6.4%**
	MLP+FT	83.4%*	80.2%	-6.4%	30.6%
	SMCM+BSST (Ours)	**85.8%**	**87.1%**	**-5.6%**	10.9%
	SMCM+BSST (Optimal)†	93.3%	95.3%	-0.3%	3.8%
III	LDA+SST	73.1%	70.8%	-6.7%	**13.5%**
	QDA+SST	71.6%	66.2%	-9.8%	21.4%
	PC+SST	**75.6%**	**74.7%**	**-2.9%**	N/A
	SVM+SST	71.9%	64.9%	-13.8%	N/A
	1D-CNN+FT	58.6%*	56.0%	-6.6%	28.5%
	MLP+FT	61.9%*	58.1%	-7.7%	25.4%
	SMCM+BSST (Ours)	74.7%	68.9%	-10.3%	16.9%
	SMCM+BSST (Optimal)†	81.6%	80.6%	1.3%	10.1%

* Significant difference from the proposed method by pairwise comparisons with a linear mixed-effects model ($p<0.01$).
† Theoretical upper bound performance using true labels for BSST updates, included for reference.
The **bold** and underlined values represent the best and second best performances, respectively.

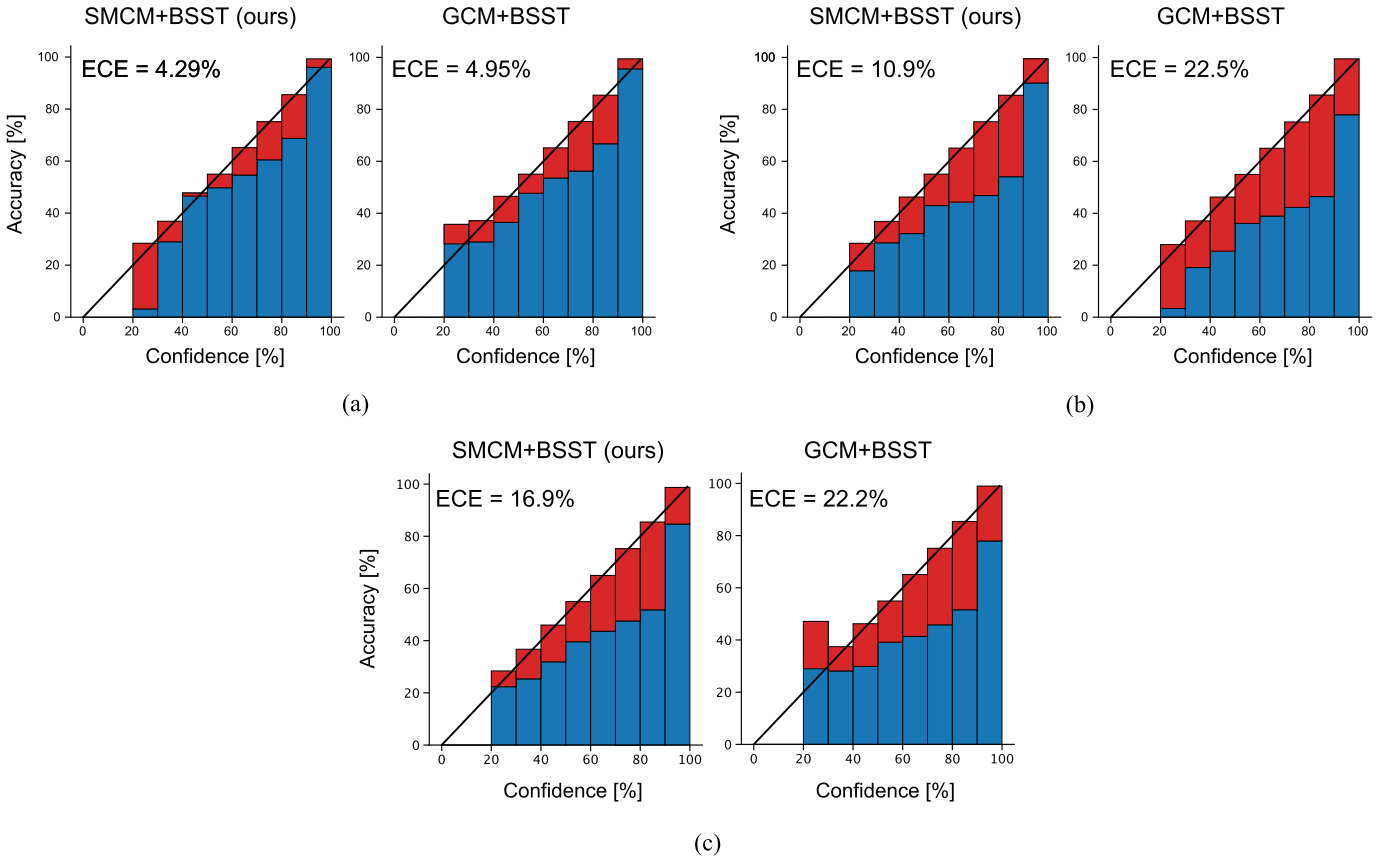

Fig. 5. Reliability diagrams for SMCM+BSST (ours) and GCM+BSST for each dataset. (a) Dataset I. (b) Dataset II. (c) Dataset III. The horizontal axis represents the prediction confidence, and the vertical axis represents the actual classification accuracy. The blue bars show the actual classification accuracy for each confidence interval, while the red bars indicate the calibration error (the difference between prediction confidence and actual accuracy). The diagonal line represents perfect calibration. The ECE, which is the weighted average of the calibration errors, is presented in each figure.

pairwise comparisons between the proposed method and each method (excluding SMCM+BSST (optimal)) using a linear mixed-effects model with classification method as fixed effect and subject as a random effect (significance level: 1%). This approach was selected to account for the within-subject design of our experiments. The analysis was performed on the overall accuracy metric, with significant differences ($p < 0.01$, Holm-adjusted for multiple comparisons) indicated with asterisks in the table.

Across the datasets, the proposed method often achieved leading 'Final accuracy' and consistently demonstrated favorable ECE values, indicating robust calibration, especially when compared to deep learning models. For instance, on datasets I and II, our method achieved the highest overall accuracies (92.1% and 85.8%), significantly outperforming most alternatives. While PC+SST registered the highest 'Overall accuracy' on the challenging dataset III (amputee subjects), our method remained competitive, particularly showing a relatively good ECE on this dataset compared to several other techniques.

B. Analysis of the Proposed Method

To evaluate the impact of incorrect pseudo-labels in BSST, we analyzed the relationship between the pseudo-label error rate at trial t and the classification accuracy at trial $t + 1$ (Fig. 6). Dataset III was excluded due to its limited number of trials. We observed moderate negative correlations in both datasets (dataset I: $r = -0.52$, $p = 0.0333$; dataset II: $r = -0.61$, $p = 3.34 \times 10^{-13}$), indicating that errors in pseudo-labels do affect subsequent performance. However, the small to moderate correlation magnitudes suggest that accumulated labeling errors are not the dominant factor in long-term accuracy degradation, indicating reasonable convergence properties of BSST.

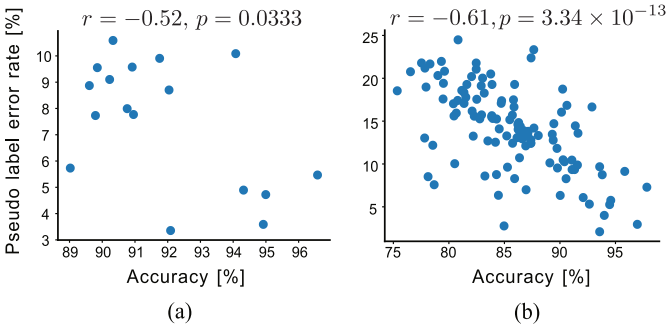

Fig. 6. Relationship between pseudo-label error rates and subsequent classification accuracy. (a) Dataset I. (b) Dataset II. The vertical axis shows the error rate of pseudo-labels at trial t, and the horizontal axis shows the classification accuracy at trial $t+1$, demonstrating the moderate negative correlation between labeling errors and subsequent performance.

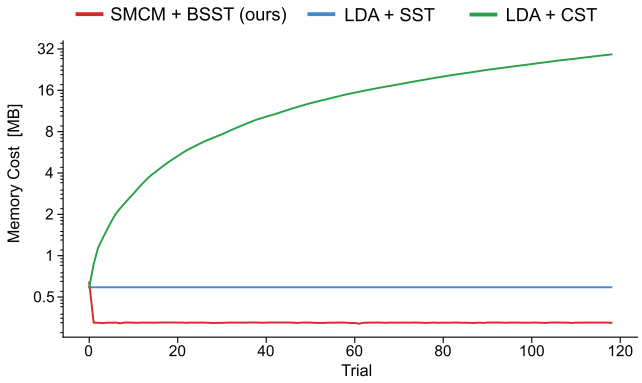

Fig. 7. Memory cost evolution for each adaptive learning approach in Dataset II, plotted on a logarithmic ($\log_2$) scale. Red: proposed SMCM+BSST (ours), blue: LDA+SST, and green: LDA+CST. Values represent averages across all participants per trial.

Fig. 7 shows the memory cost progression of our proposed method and two baselines on dataset II. The red, blue, and green curves denote the proposed SMCM+BSST, LDA+SST, and LDA+CST, respectively. The horizontal axis indicates trial index, and the vertical axis shows memory cost in megabytes (MB) on a logarithmic scale. Our method enables sequential learning with the consistently smallest memory footprint among all methods. In contrast, LDA+SST maintains a constant–but higher–memory cost across trials, while LDA+CST's memory usage grows sharply as trials accumulate.

The performance comparison between trial-wise and class-wise updating strategies is shown in Fig. 8. The red boxes represent trial-wise BSST and cyan represents the results of class-wise BSST. In all datasets, although the overall classification accuracy remained comparable between the two scenarios, the confidence inference capability tended to degrade under class-wise BSST.

The distribution of optimal threshold values across participants is presented as histograms in Fig. 9. In dataset I, most participants achieved the highest accuracy at a threshold of 0.6 or higher (Fig. 9(a)), while in dataset II and dataset III, the optimal thresholds showed a broader distribution across participants (Fig. 9(b), (c)).

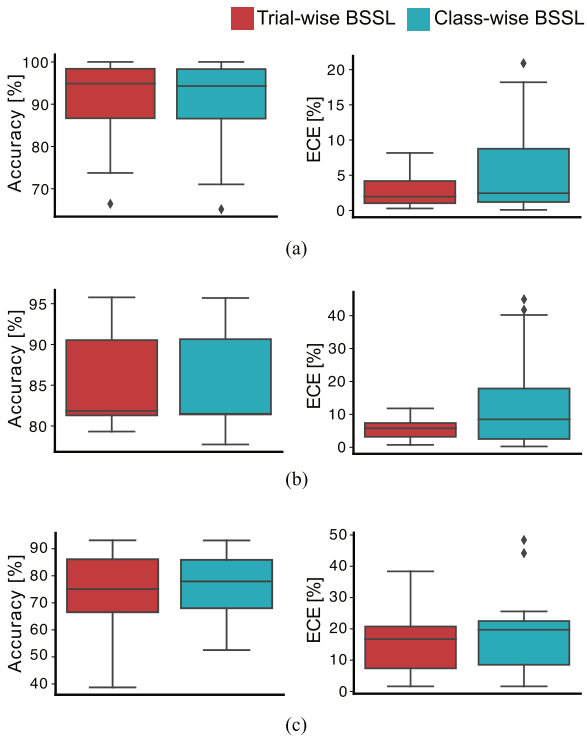

Fig. 8. Comparison of accuracy and ECE between trial-wise and class-wise BSST strategies. (a) Dataset I. (b) Dataset II. (c) Dataset III. Box plots summarize the distribution of performance metrics across all trials and participants, with red boxes representing trial-wise BSST and cyan boxes representing class-wise BSST.

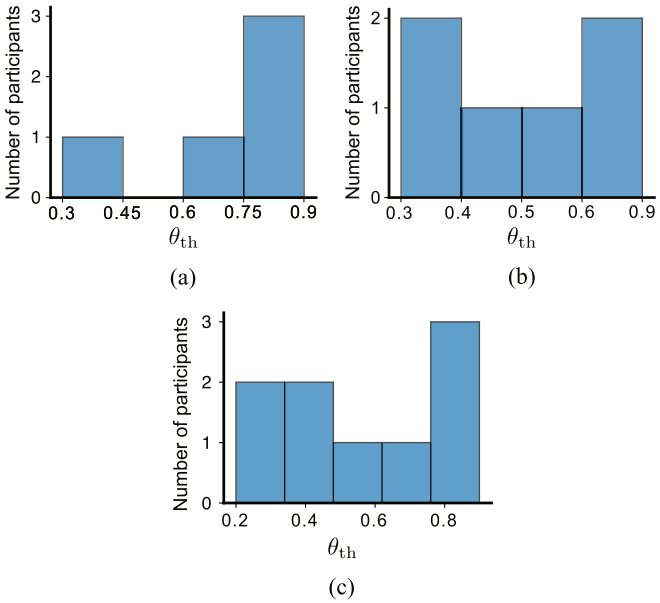

Fig. 9. Distribution of optimal confidence thresholds θ_{th} for BSST across participants. (a) Dataset I. (b) Dataset II. (c) Dataset III. The histograms indicate the threshold values that yielded the highest classification accuracy for each participant.

VI. Discussion

A. Principal Findings

The classification accuracy gradually decreases without BSST as the trials progress (Fig. 4). This degradation can be attributed to variations in signal characteristics caused by multiple factors such as muscle fatigue and electrode shift, resulting in a substantial shift in the distribution characteristics of the test data. In contrast, the proposed method maintains classification accuracies of 92.1%, 85.8% and 74.7% in the respective datasets, effectively mitigating the decrease in accuracy. These results demonstrate that the proposed method is capable of adaptive classification to changes in EMG signal characteristics.

The effectiveness of our pseudo-labeling approach varies across different data conditions while outperforming existing methods. For dataset I, the proposed method achieved accuracy close to SMCM+BSST (optimal), indicating effective pseudo-label assignment without calibration data. However, for datasets II and III, we observed larger performance gaps. These discrepancies can be attributed to their challenging

characteristics—dataset II contains measurements across multiple days with large distribution shifts, while dataset III includes amputee data with altered muscle anatomy and scar tissue that introduce signal variability. The higher ECE values in these datasets compared to dataset I suggest that reduced calibration quality affected pseudo-label selection accuracy. Nevertheless, our proposed method tended to outperform other conventional adaptive approaches across all datasets (Table I). These findings suggest that while effective for sequential learning without calibration data, there remains room for improvement in pseudo-labeling accuracy under challenging conditions.

B. Comparison With Existing Methods

The classification accuracy of GCM+BSST is substantially lower than the proposed method in dataset II (Fig. 4(b)). This degradation can be attributed to the high ECE value (22.5%) of GCM+BSST (Fig. 5(b)), which demonstrates that the Gaussian distribution fails to adequately represent the underlying distribution of EMG signal distributions. While Gaussian distributions have lighter tails with rapidly decreasing probability density away from the mean, the scale mixture distributions used in SMCM have heavier tails that can represent a broader range of probability distributions. Consequently, GCM tends to underestimate the uncertainty inherent in signal variability. Notably, dataset II contains several days of data, resulting in greater variability and uncertainty in the EMG signals compared to dataset I. This higher variability causes GCM to become overconfident in its predictions, leading to incorrectly predicted labels being included in the BSST and ultimately decreasing classification accuracy.

In contrast, the proposed SMCM+BSST demonstrated superior performance in both datasets through its sophisticated modeling of EMG signal characteristics. By representing signal uncertainty as stochastic variations in variance and controlling distribution tail heaviness through parameter ν, SMCM accurately captures the underlying EMG signal distributions. This appropriate modeling led to both improved classification accuracy (Fig. 4) and reliable confidence estimation (Fig. 5), resulting in sequential learning that outperformed conventional methods (Table I).

Deep-learning methods (1D-CNN+FT and MLP+FT) underperformed the proposed method in our experiments, despite their theoretical capacity to model complex patterns. This can be attributed to several factors inherent to deep learning in this adaptive context: (1) Sensitivity to hyperparameters and risk of overfitting, particularly when fine-tuning with limited pseudo-labels and without validation data for optimal epoch selection [48]. In sequential learning scenarios where labeled validation data is unavailable, deep networks often struggle to balance adaptation and stability. (2) Poorly calibrated confidence estimates (Table I), which leads to lower quality pseudo-labels and undermines the self-training process. The high ECE values observed in deep learning methods (up to 30.6% for MLP+FT) indicate significant overconfidence in their predictions. In contrast, the robust confidence estimation in our SMCM+BSST proved more effective for sequential self-training based on prediction confidence. Additionally, once the initial model is trained, our approach requires no further hyperparameter tuning during the adaptation process, unlike deep learning methods that need careful optimization of learning rates and training durations for each adaptation step.

C. Practical Implications

The efficiency of the sequential learning process further enhances our approach. The proposed BSST framework offers a more efficient solution compared to conventional SST methods. By encoding historical information through Bayesian updating of prior distributions, BSST eliminates the need for explicit data storage, resulting in a low and nearly constant memory footprint (Fig. 7). This difference is particularly significant when dealing with long-term datasets, where the storage and management of accumulated samples becomes increasingly challenging. Despite this more compact knowledge representation, BSST achieves comparable or better accuracy performance in suppressing accuracy degradation compared to SST methods (Table I), suggesting that representing past knowledge through probabilistic distributions can be effective while reducing storage overhead.

The long-term accuracy maintenance shown in Fig. 4 and Table I demonstrates that BSST exhibits resilience against error accumulation. This resilience stems from both the reliable confidence estimation by SMCM and the constraints from prior distributions, which together limit the impact of incorrect labels. Our correlation analysis (Fig. 6) further suggests that distribution shifts between trials, rather than accumulated errors, are the primary driver of accuracy fluctuations. Nevertheless, the risk of performance degradation remains under extreme conditions, such as consecutive major distribution shifts or when initial model accuracy is low (as in dataset III). In such cases, future work could explore implementing change-point detection, confidence recalibration, or model reset mechanisms to detect and correct potential degradation. These mechanisms would complement BSST's inherent stability by providing explicit safeguards against pathological error accumulation scenarios.

We compared trial-wise and class-wise updating strategies for BSST. While both strategies achieved comparable classification accuracy (Fig. 8), class-wise BSST tended to be higher ECE values in all datasets. This degradation in confidence estimation occurs because class-wise updates can create temporary imbalances in the probability structure when individual class parameters are updated separately. In contrast, trial-wise updating maintains better balance by processing all classes simultaneously, leading to more reliable probability estimates. However, the similar classification accuracy between strategies suggests that the choice of update strategy has limited impact on overall performance. For practical online implementations in prosthetic devices and interfaces that lack discrete trial boundaries, class-wise updating could be implemented using fixed time windows or sample thresholds, providing a feasible approach for continuous use scenarios.

The observed variation in optimal confidence thresholds across datasets provides insights into adaptation requirements for different EMG scenarios (Fig. 9). In dataset I (short-term recordings), the higher optimal thresholds (> 0.6) indicate

relatively stable signal characteristics within single sessions. Conversely, the broader threshold distributions in dataset II (long-term recordings) and dataset III (amputee data) reflect the challenges of temporal signal drift and inherent variability in residual muscle signals, respectively. These findings highlight that adaptation strategies must be context-sensitive. A key limitation of our current approach is that, with sequentially arriving unlabeled data, traditional threshold optimization methods like cross-validation cannot be applied. Therefore, developing an intelligent threshold adaptation mechanism that responds to changing signal quality and individual characteristics without requiring ground truth labels presents a critical direction for practical myoelectric interfaces. Potential approaches might include Bayesian optimization frameworks that estimate threshold performance based on model uncertainty metrics or adaptive techniques that adjust thresholds based on detected distribution shifts in the incoming signal.

D. Limitations and Future Work

Our proposed method has several limitations that warrant further investigation. While effective with gradual signal changes, the approach may struggle with abrupt variations in electrode positions or fatigue levels. Additionally, complex signal patterns from multi-degree-of-freedom movements or amputee data could challenge the model's adaptation capabilities. Future research should focus on enhancing robustness through improved modeling techniques and validation under varied conditions, including different exercise types, electrode configurations, and controlled fatigue levels.

We observed that using class-wise updates, though maintaining accuracy, reduced confidence estimation reliability, indicating that updating model components separately may disrupt its probabilistic structure. Furthermore, the optimal confidence threshold varied between subjects, suggesting the need for personalized adaptation mechanisms. Future work should focus on developing automatic threshold adjustment methods within the Bayesian framework and implementing efficient update triggering mechanisms for continuous use scenarios, such as adaptive time windows that consider both prediction confidence and elapsed time.

The BSST framework could potentially extend beyond EMG classification to other biosignals. Previous studies have successfully applied probabilistic models to EEG and ECG signals [49], [50], [51], [52]. Validating our approach across diverse biosignals would further establish its universality and practical value for adaptive human-machine interfaces.

VII. Conclusion

This paper proposes an adaptive EMG pattern classification method that combines SMCM and BSST. The proposed method adapts to signal variations by updating the posterior distribution of model parameters through BSST, which represents past signal characteristics as prior distributions without requiring direct data storage. To enhance confidence estimation accuracy, we incorporate SMCM, which accurately captures EMG signal characteristics through sophisticated probability modeling.

Experimental results show that the proposed method achieved high average discrimination accuracy across three datasets (dataset I: 92.1%, dataset II: 85.8%, dataset III: 74.7%), demonstrating its effectiveness in suppressing accuracy degradation. Furthermore, memory-cost experiments assuming long-term interface usage indicate the proposed method can sustain sequential learning over extended periods with minimal memory consumption. These results present an adaptive EMG pattern classification method that both requires no additional labeled data and runs on limited computational resources. Such a lightweight, self-trained approach is an important step toward practical next-generation myoelectric interfaces–especially prosthetic controls that demand long-term stability.

In the future, we should focus on improving robustness under more diverse real-world environments and complex motion conditions. We should also verify and enhance its applicability to challenging tasks, such as amputee data. Additionally, we plan to advance the optimization of the method itself, including the dynamic adjustment of confidence thresholds.

References

[1] A. Tsoli and O. C. Jenkins, "Robot grasping for prosthetic applications," in *Robotics Research*. Berlin, Germany: Springer, 2011, pp. 1–12.

[2] A. Fougner, Ø. Stavdahl, P. J. Kyberd, Y. G. Losier, and P. A. Parker, "Control of upper limb prostheses: Terminology and proportional myoelectric control—A review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 5, pp. 663–677, Sep. 2012.

[3] H. Zhang, D. Yang, C. Shi, L. Jiang, and H. Liu, "Robust EMG pattern recognition with electrode donning/doffing and multiple confounding factors," in *Proc. Int. Conf. Intel. Robot. Appl. (ICRA)*, Jan. 2017, pp. 413–424.

[4] N. Jiang, S. Dosen, K.-R. Müller, and D. Farina, "Myoelectric control of artificial limbs—Is there a need to change focus? [in the spotlight]," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 150–152, Sep. 2012.

[5] P. Kaufmann, K. Englehart, and M. Platzner, "Fluctuating EMG signals: Investigating long-term effects of pattern matching algorithms," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug. 2010, pp. 6357–6360.

[6] J. He, D. Zhang, X. Sheng, and X. Zhu, "Effects of long-term myoelectric signals on pattern recognition," in *Proc. Int. Conf. Intell. Robot. Appl. (ICRA)*, Jan. 2013, pp. 396–404.

[7] P. K. Artemiadis and K. J. Kyriakopoulos, "An EMG-based robot control scheme robust to time-varying EMG signal features," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 3, pp. 582–588, May 2010.

[8] J. H. T. Viitasalo and P. V. Komi, "Signal characteristics of EMG during fatigue," *Eur. J. Appl. Physiol. Occupational Physiol.*, vol. 37, no. 2, pp. 111–121, 1977.

[9] O. W. Samuel et al., "Resolving the adverse impact of mobility on myoelectric pattern recognition in upper-limb multifunctional prostheses," *Comput. Biol. Med.*, vol. 90, pp. 76–87, Nov. 2017.

[10] S. Kanoga, A. Kanemura, and H. Asoh, "Are armband sEMG devices dense enough for long-term use—Sensor placement shifts cause significant reduction in recognition accuracy," *Biomed. Signal Process. Control*, vol. 60, Jul. 2020, Art. no. 101981.

[11] T. Matsubara, S.-H. Hyon, and J. Morimoto, "Learning and adaptation of a stylistic myoelectric interface: EMG-based robotic control with individual user differences," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, Dec. 2011, pp. 390–395.

[12] J. Liu, X. Sheng, D. Zhang, J. He, and X. Zhu, "Reduced daily recalibration of myoelectric prosthesis classifiers based on domain adaptation," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 1, pp. 166–176, Jan. 2016.

[13] M. M.-C. Vidovic, H.-J. Hwang, S. Amsüss, J. M. Hahne, D. Farina, and K.-R. Müller, "Improving the robustness of myoelectric pattern recognition for upper limb prostheses by covariate shift adaptation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 9, pp. 961–970, Sep. 2016.

[14] X. Zhu, J. Liu, D. Zhang, X. Sheng, and N. Jiang, "Cascaded adaptation framework for fast calibration of myoelectric control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 3, pp. 254–264, Mar. 2017.

[15] M. Karrenbach, P. Preechayasomboon, P. Sauer, D. Boe, and E. Rombokas, "Deep learning and session-specific rapid recalibration for dynamic hand gesture recognition from EMG," *Frontiers Bioeng. Biotechnol.*, vol. 10, Dec. 2022, Art. no. 1034672.

[16] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "A deep transfer learning approach to reducing the effect of electrode shift in EMG pattern recognition-based control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 370–379, Feb. 2020.

[17] X. Chen, D. Zhang, and X. Zhu, "Application of a self-enhancing classification method to electromyography pattern recognition for multifunctional prosthesis control," *J. NeuroEng. Rehabil.*, vol. 10, no. 1, p. 44, Dec. 2013.

[18] H. Zhang, Y. Zhao, F. Yao, L. Xu, P. Shang, and G. Li, "An adaptation strategy of using LDA classifier for EMG pattern recognition," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 4267–4270.

[19] S. Amsüss, P. M. Goebel, N. Jiang, B. Graimann, L. Paredes, and D. Farina, "Self-correcting pattern recognition system of surface EMG signals for upper limb prosthesis control," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1167–1176, Apr. 2014.

[20] J. W. Sensinger, B. A. Lock, and T. A. Kuiken, "Adaptive pattern recognition of myoelectric signals: Exploration of conceptual framework and practical algorithms," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 3, pp. 270–278, Jun. 2009.

[21] Y. Okawa, S. Kanoga, T. Hoshino, and T. Nitta, "Sequential learning on sEMGs in short- and long-term situations via self-training semi-supervised support vector machine," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2022, pp. 3183–3186.

[22] D. Yang, L. Jiang, R. Liu, and H. Liu, "Adaptive learning of multi-finger motion recognition based on support vector machine," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2013, pp. 2231–2238.

[23] J. Liu, "Adaptive myoelectric pattern recognition toward improved multifunctional prosthesis control," *Med. Eng. Phys.*, vol. 37, no. 4, pp. 424–430, Apr. 2015.

[24] S. Hua, C. Wang, H. K. Lam, and S. Wen, "An incremental learning method with hybrid data over/down-sampling for sEMG-based gesture classification," *Biomed. Signal Process. Control*, vol. 83, May 2023, Art. no. 104613.

[25] Y. D. Kwon, J. Chauhan, A. Kumar, P. H. HKUST, and C. Mascolo, "Exploring system performance of continual learning for mobile and embedded sensing applications," in *Proc. IEEE/ACM Symp. Edge Comput. (SEC)*, Dec. 2021, pp. 319–332.

[26] X. Yao, X. Wang, Y. Liu, and W. Zhu, "Continual recognition with adaptive memory update," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 19, no. 3s, pp. 1–15, Oct. 2023.

[27] S. Kanoga, R. Karakida, T. Hoshino, Y. Okawa, and M. Tada, "Deep generative replay-based class-incremental continual learning in sEMG-based pattern recognition," in *Proc. 46th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2024, pp. 1–4.

[28] A. Furui, H. Hayashi, and T. Tsuji, "A scale mixture-based stochastic model of surface EMG signals with variable variances," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2780–2788, Oct. 2019.

[29] S. Yoneda and A. Furui, "Bayesian approach for adaptive EMG pattern classification via semi-supervised sequential learning," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2023, pp. 3310–3315.

[30] O. Fukuda, T. Tsuji, M. Kaneko, and A. Otsuka, "A human-assisting manipulator teleoperated by EMG signals and arm motions," *IEEE Trans. Robot. Autom.*, vol. 19, no. 2, pp. 210–222, Apr. 2003.

[31] E. J. Scheme, K. B. Englehart, and B. S. Hudgins, "Selective classification for improved robustness of myoelectric control under nonideal conditions," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1698–1705, Jun. 2011.

[32] T. Minka, "Discriminative models, not discriminative training," Microsoft Research, Redmond, WA, USA, Tech. Rep. MSR-TR-2005-144, 2005.

[33] J. A. Lasserre, C. M. Bishop, and T. P. Minka, "Principled hybrids of generative and discriminative models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2006, pp. 87–94.

[34] A. Y. Ng and M. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes," in *Proc. Neural. Inf. Process. Syst.*, vol. 14, Jan. 2001, pp. 841–848.

[35] C. M. Bishop and J. Lasserre, "Generative or discriminative? Getting the best of both worlds," in *Bayesian Statistics 8*. Oxford, U.K.: Oxford Univ. Press, pp. 13–34.

[36] V. Der Bilt and V. Der Glas, "Detection of onset and termination of muscle activity in surface electromyograms," *J. Oral Rehabil.*, vol. 25, no. 5, pp. 365–369, May 1998.

[37] N. Hogan and R. W. Mann, "Myoelectric signal processing: Optimal estimation applied to electromyography—Part I: Derivation of the optimal myoprocessor," *IEEE Trans. Biomed. Eng.*, vol. BME-27, no. 7, pp. 382–395, Jul. 1980.

[38] P. A. Parker, J. A. Stuller, and R. N. Scott, "Signal processing for the multistate myoelectric channel," *Proc. IEEE*, vol. 65, no. 5, pp. 662–674, May 1977.

[39] N. Messaoudi, R. E. Bekka, P. Ravier, and R. Harba, "Assessment of the non-gaussianity and non-linearity levels of simulated sEMG signals on stationary segments," *J. Electromyogr. Kinesiol.*, vol. 32, pp. 70–82, Feb. 2017.

[40] A. Furui and T. Tsuji, "Muscle fatigue analysis by using a scale mixture-based stochastic model of surface EMG signals," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 1948–1951.

[41] Z. Oravecz, M. Huentelman, and J. Vandekerckhove, "Sequential Bayesian updating for big data," in *Big Data in Cognitive Science*, vol. 373. East Sussex, U.K.: Psychology Press, 2017, pp. 13–33.

[42] S. Sun, Y. Lu, and Y. Chen, "The stochastic approximation method for adaptive Bayesian classifiers: Towards online brain–computer interfaces," *Neural Comput. Appl.*, vol. 20, no. 1, pp. 31–40, Feb. 2011.

[43] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, vol. 4, no. 4. Cham, Switzerland: Springer, 2006.

[44] S. Kanoga, T. Hoshino, and H. Asoh, "Semi-supervised style transfer mapping-based framework for sEMG-based pattern recognition with 1- or 2-DoF forearm motions," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102817.

[45] A. H. Al-Timemy, R. N. Khushaba, G. Bugmann, and J. Escudero, "Improving the performance against force variation of EMG controlled multifunctional upper-limb prostheses for transradial amputees," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 6, pp. 650–661, Jun. 2016.

[46] A. Furui, T. Igaue, and T. Tsuji, "EMG pattern recognition via Bayesian inference with scale mixture-based stochastic generative models," *Expert Syst. Appl.*, vol. 185, Dec. 2021, Art. no. 115644. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095741742101037X

[47] M. P. Naeini, G. Cooper, and M. Hauskrecht, "Obtaining well calibrated probabilities using Bayesian binning," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, vol. 29, no. 1, Feb. 2015, pp. 2901–2907.

[48] D. Wu, J. Yang, and M. Sawan, "Transfer learning on electromyography (EMG) tasks: Approaches and beyond," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 3015–3034, 2023.

[49] A. Furui, R. Onishi, A. Takeuchi, T. Akiyama, and T. Tsuji, "Non-gaussianity detection of EEG signals based on a multivariate scale mixture model for diagnosis of epileptic seizures," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 2, pp. 515–525, Feb. 2021.

[50] A. Furui, T. Akiyama, and T. Tsuji, "A time-series scale mixture model of EEG with a hidden Markov structure for epileptic seizure detection," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 5832–5836.

[51] R. J. Martis, C. Chakraborty, and A. K. Ray, "A two-stage mechanism for registration and classification of ECG using Gaussian mixture model," *Pattern Recognit.*, vol. 42, no. 11, pp. 2979–2988, Nov. 2009.

[52] P.-C. Chang, J.-J. Lin, J.-C. Hsieh, and J. Weng, "Myocardial infarction classification with multi-lead ECG using hidden Markov models and Gaussian mixture models," *Appl. Soft Comput.*, vol. 12, no. 10, pp. 3165–3175, Oct. 2012.