

# Inter-subject Variance Transfer Learning for EMG Pattern Classification Based on Bayesian Inference

Seitaro Yoneda and Akira Furui

Graduate School of Advanced Science and Engineering, Hiroshima University, Higashi-hiroshima, Japan.

Email: {seitaroyoneda, akirafurui}@hiroshima-u.ac.jp

**Abstract**—In electromyogram (EMG)-based motion recognition, a subject-specific classifier is typically trained with sufficient labeled data. However, this process demands extensive data collection over extended periods, burdening the subject. To address this, utilizing information from pre-training on multiple subjects for the training of the target subject could be beneficial. This paper proposes an inter-subject variance transfer learning method based on a Bayesian approach. This method is founded on the simple hypothesis that while the means of EMG features vary greatly across subjects, their variances may exhibit similar patterns. Our approach transfers variance information, acquired through pre-training on multiple source subjects, to a target subject within a Bayesian updating framework, thereby allowing accurate classification using limited target calibration data. A coefficient was also introduced to adjust the amount of information transferred for efficient transfer learning. Experimental evaluations using two EMG datasets demonstrated the effectiveness of our variance transfer strategy and its superiority compared to existing methods.

## I. INTRODUCTION

Electromyogram (EMG) signals are electrical activities that can be measured from the skin surface, and their characteristics change depending on the motion performed. Machine learning leverages EMG signals in various fields, such as myoelectric prostheses, by estimating motion intentions [1], [2].

In machine learning-based EMG pattern recognition, unique classifiers are commonly trained using data collected individually from each subject [3], [4]. This is due to substantial individual differences exist in EMG signals, which can be attributed to variations in subcutaneous fat distribution, muscle fiber diameter, and the manner in which force is exerted [4], [5]. However, this approach requires the collection of extensive training data over multiple trials to ensure adequate recognition performance, which poses a heavy burden on subjects and presents a major challenge for practical applications.

Inter-subject transfer learning has been gaining attention as a method for achieving high recognition performance while reducing data collection costs for a new subject. This approach utilizes the knowledge from pre-training with multiple subjects (*source* subjects) to train a classifier for a new subject (*target* subject), anticipating that in such a scenario, only minimal labeled calibration data will be collected from the target subject. Recently, transfer learning based on deep neural networks has been widely used [6], [7]; however, issues related to computational cost and overfitting have been identified [8].

For applications on edge devices with limited computational resources, simpler, non-deep classification models are

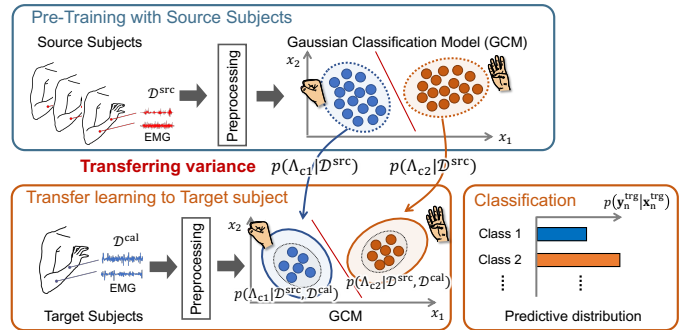


Fig. 1. Overview of the proposed method

preferable for inter-subject transfer learning. Traditionally, a framework of transfer learning is introduced for non-deep classification models [9]–[11]. For instance, in adaptive linear discriminant analysis (LDA) [9], model parameters (mean and covariance matrix) from both the target and source are combined. The weight of the combination is then empirically adjusted to determine the amount of information transferred from source to target, to achieve effective recognition.

As a non-deep strategy, the authors have explored transfer learning using simple Bayesian models in EMG classification [12]. In Bayesian transfer learning, the transfer of knowledge is interpreted as a process, whereby the prior distribution is transformed into the posterior distribution via the likelihood. The amount of transfer from source to target is determined based on the uncertainty of priors. We demonstrated the viability of this approach for intra-subject, inter-trial transfer learning [12]. However, its applicability and success in inter-subject contexts requires further investigation.

This paper presents a Bayesian approach to inter-subject transfer learning in EMG classification, employing a simple Gaussian classification model (GCM). Fig. 1 presents an overview of the proposed method. The method initially computes the posterior distributions of GCM parameters using EMG signals from multiple source subjects. These distributions are then updated using calibration data from a target subject. The proposed method shares the precision matrix among subjects, thereby transferring only variance information from source to target subjects. The amount of transferred information is determined based on the uncertainty of the pre-trained posterior distribution on the source subjects.

## II. PROPOSED METHOD

### A. Problem Setting

Consider the  $D$ -dimensional EMG feature vector at data point  $n$ , denoted as  $\mathbf{x}_n \in \mathbb{R}^D$ , and its corresponding motion class label  $\mathbf{y}_n \in \{0, 1\}^C$ , where  $C$  is the number of classes. In inter-subject transfer learning for EMG signals, the information learned from the data of multiple source subjects is utilized during the training of a classification model for a target subject, to enhance classification performance for the target. Let us consider the pre-training of a model with a labeled training dataset obtained from a source subject  $s$  as  $\mathcal{D}^s = \{(\mathbf{x}_n^s, \mathbf{y}_n^s)\}_{n=1}^{N^s}$ , where  $N^s$  is the number of data points in the source subject. The complete dataset across all source subjects is  $\mathcal{D}^{\text{src}} = \{\mathcal{D}^s\}_{s=1}^S$ , with a total sample size of  $N^{\text{src}} = \sum_s N^s$ , which is assumed to be sufficiently large. The target subject provides a small set of labeled data (referred to as calibration data), denoted as  $\mathcal{D}^{\text{cal}} = \{(\mathbf{x}_n^{\text{cal}}, \mathbf{y}_n^{\text{cal}})\}_{n=1}^{N^{\text{cal}}}$ . Through transfer learning, the ultimate goal is to maximize the classification performance on the test data  $\mathbf{x}_n^{\text{trg}}$  for the target subject.

### B. Gaussian Classification Model (GCM)

A GCM represents the relationship between EMG signals and the corresponding class labels. The observed model of the EMG signal  $\mathbf{x}_n$  for class  $c \in \{1, \dots, C\}$  is expressed via the following Gaussian distribution:

$$p(\mathbf{x}_n | \boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c) = \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c^{-1}), \quad (1)$$

where  $\boldsymbol{\mu}_c \in \mathbb{R}^D$  and  $\boldsymbol{\Lambda}_c \in \mathbb{R}^{D \times D}$  represent the mean vector and precision matrix (the inverse of the covariance matrix) of each class, respectively. We use one-hot representation for the motion class label  $\mathbf{y}_n = \{y_{n,c}\}$ , where  $y_{n,c} = 1$  for a specific  $c$  indicates the selection of the  $c$ -th class. Hence, the distribution of  $\mathbf{y}_n$  can be represented using the following categorical distribution:

$$p(\mathbf{y}_n | \boldsymbol{\pi}) = \text{Cat}(\mathbf{y}_n | \boldsymbol{\pi}) = \prod_{c=1}^C \pi_c^{y_{n,c}}, \quad (2)$$

where  $\boldsymbol{\pi} = \{\pi_c\}$  denotes the mixing coefficients ( $\pi_c \in [0, 1]$  and  $\sum_{c=1}^C \pi_c = 1$ ).

A prior distribution is introduced for each model parameter for Bayesian treatment. For computational convenience, we set the Gaussian-Wishart and Dirichlet distributions, which are conjugate priors for  $\{\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c\}$ , and  $\boldsymbol{\pi}$ , as follows:

$$p(\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c) = p(\boldsymbol{\mu}_c | \boldsymbol{\Lambda}_c) p(\boldsymbol{\Lambda}_c) \quad (3)$$

$$= \mathcal{N}(\boldsymbol{\mu}_c | \mathbf{m}_0, (\beta_0 \boldsymbol{\Lambda}_c)^{-1}) \mathcal{W}(\boldsymbol{\Lambda}_c | \nu_0, \mathbf{W}_0), \quad (4)$$

$$p(\boldsymbol{\pi}) = \text{Dir}(\boldsymbol{\pi} | \boldsymbol{\alpha}_0), \quad (5)$$

where  $\mathbf{m}_0 \in \mathbb{R}^D$ ,  $\beta_0 \in \mathbb{R}^+$ ,  $\nu > D - 1$ ,  $\mathbf{W}_0 \in \mathbb{R}^{D \times D}$ , and  $\boldsymbol{\alpha}_0 = \{\alpha_0\}^C$  are the prior hyperparameters.

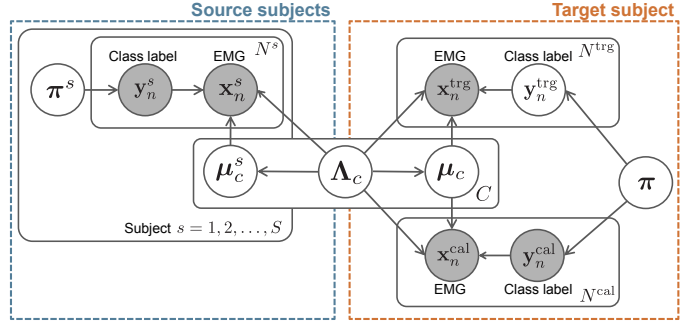


Fig. 2. Graphical model of GCM for inter-subject classification. The gray and white nodes represent the observed and unobserved variables, respectively.

### C. Inter-Subject Transfer Learning

In transfer learning based on structured Bayesian models, the issue reduces to whether to share parameters among subjects or to treat them independently. EMG feature vectors vary significantly across subjects due to individual biological differences, thereby affecting the mean values of the features. By contrast, feature variance (the variability over time) may display common trends across subjects. With these simple assumptions, we set subject-specific means and a shared variance in GCM. Fig. 2 shows the graphical model of the extended GCM. The training steps of the model are divided into pre-training for source subjects and transfer learning for the target subject.

1) *Pre-training with source subjects:* The posterior distributions of the source subject  $s$  for each class can be computed by applying Bayes' theorem as follows:

$$p(\boldsymbol{\mu}_c^s, \boldsymbol{\Lambda}_c | \mathcal{D}^{\text{src}}) \propto p(\mathcal{D}^{\text{src}} | \boldsymbol{\mu}_c^s, \boldsymbol{\Lambda}_c) p(\boldsymbol{\mu}_c^s | \boldsymbol{\Lambda}_c) p(\boldsymbol{\Lambda}_c), \quad (6)$$

$$p(\boldsymbol{\pi}^s | \mathcal{D}^{\text{src}}) \propto p(\mathcal{D}^{\text{src}} | \boldsymbol{\pi}^s) p(\boldsymbol{\pi}^s). \quad (7)$$

From conjugacy, each posterior distribution is attributed to the following Gaussian-Wishart and Dirichlet distributions, which are of the same type as the priors:

$$p(\boldsymbol{\mu}_c^s, \boldsymbol{\Lambda}_c | \mathcal{D}^{\text{src}}) = \mathcal{N}(\boldsymbol{\mu}_c^s | \mathbf{m}_c^s, (\beta_c^s \boldsymbol{\Lambda}_c)^{-1}) \mathcal{W}(\boldsymbol{\Lambda}_c | \nu_c^{\text{src}}, \mathbf{W}_c^{\text{src}}), \quad (8)$$

$$p(\boldsymbol{\pi}^s | \mathcal{D}^{\text{src}}) = \text{Dir}(\boldsymbol{\pi}^s | \boldsymbol{\alpha}^s), \quad (9)$$

where  $\beta_c^s$ ,  $\mathbf{m}_c^s$ ,  $\nu_c^{\text{src}}$ ,  $\mathbf{W}_c^{\text{src}}$ , and  $\boldsymbol{\alpha}^s = \{\alpha_c^s\}_{c=1}^C$  are the posterior hyperparameters for the source subjects. Note that the hyperparameters for the subject-specific mean carry a subject index  $s$ , whereas for the shared variance (precision) across subjects, the superscript "src" denotes its commonality.

In Bayesian transfer learning, the amount of transferred information is determined by the uncertainty of the distribution, which is inversely related to data quantity. With more source subjects or larger samples, there is the risk of transferring excessive information to the target. Accordingly, we introduce a weight coefficient  $w^s \in \mathbb{R}^+$  to control the distribution uncertainty and adjust the impact of source data. Using this

TABLE I  
DATASET INFORMATION

Dataset	# Motions ( $C$ )	# Electrodes ( $D$ )	Sampling frequency	# Subjects	Calibration trials	Test trials
I	6	4	2000 Hz	6	1	2, 3, 4, 5, 6
II	8	8	200 Hz	25	1	2, 3, 4

coefficient, the modified posterior hyperparameters for the source subjects can be calculated as follows:

$$\beta_c^s = \sum_{n=1}^{N^s} w^s y_{n,c}^s + \beta_0, \quad (10)$$

$$\mathbf{m}_c^s = \frac{\sum_{n=1}^{N^s} w^s y_{n,c}^s \mathbf{x}_n^s + \beta_0 \mathbf{m}_0}{\beta_c^s}, \quad (11)$$

$$\nu_c^{\text{src}} = \frac{1}{S} \sum_{s=1}^S \sum_{n=1}^{N^s} w^s y_{n,c}^s + \nu_0, \quad (12)$$

$$(\mathbf{W}_c^{\text{src}})^{-1} = \frac{1}{S} \sum_{s=1}^S \left( \sum_{n=1}^{N^s} w^s y_{n,c}^s \mathbf{x}_n^s \mathbf{x}_n^{s\top} + \beta_0 \mathbf{m}_0 \mathbf{m}_0^\top - \beta_c^s \mathbf{m}_c^s \mathbf{m}_c^{s\top} \right) + \mathbf{W}_0^{-1}, \quad (13)$$

$$\alpha_c^s = \sum_{n=1}^{N^s} y_{n,c}^s + \alpha_0. \quad (14)$$

Note that a larger  $w^s$  reduces distribution uncertainty, increasing the amount of information transferred from source to target.

2) *Transfer learning to target subject*: The model for the target subject is trained with transferred information from pre-training on source subjects. The posterior distributions for the target subject are computed based on both  $\mathcal{D}^{\text{src}}$  and  $\mathcal{D}^{\text{cal}}$ , as follows:

$$p(\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c | \mathcal{D}^{\text{cal}}, \mathcal{D}^{\text{src}}) \propto p(\mathcal{D}^{\text{cal}} | \boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c) p(\boldsymbol{\mu}_c | \boldsymbol{\Lambda}_c) p(\boldsymbol{\Lambda}_c | \mathcal{D}^{\text{src}}), \quad (15)$$

$$p(\boldsymbol{\pi} | \mathcal{D}^{\text{cal}}) \propto p(\mathcal{D}^{\text{cal}} | \boldsymbol{\pi}) p(\boldsymbol{\pi}). \quad (16)$$

Note that only the precision matrix  $\boldsymbol{\Lambda}_c$  is shared between source and target, conditioned on  $\mathcal{D}^{\text{src}}$ . Similar to (8) and (9), each posterior distribution can be calculated using conjugation as follows:

$$p(\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c | \mathcal{D}^{\text{cal}}, \mathcal{D}^{\text{src}}) = \mathcal{N}(\boldsymbol{\mu}_c | \mathbf{m}_c, (\beta_c \boldsymbol{\Lambda}_c)^{-1}) \mathcal{W}(\boldsymbol{\Lambda}_c | \nu_c, \mathbf{W}_c), \quad (17)$$

$$p(\boldsymbol{\pi} | \mathcal{D}^{\text{cal}}) = \text{Dir}(\boldsymbol{\pi} | \boldsymbol{\alpha}), \quad (18)$$

where  $\beta_c, \mathbf{m}_c, \nu_c, \mathbf{W}_c, \boldsymbol{\alpha} = \{\alpha_c\}$  are the posterior hyperparameters for the target subject and defined as follows:

$$\beta_c = \sum_{n=1}^{N^{\text{cal}}} y_{n,c}^{\text{cal}} + \beta_0, \quad (19)$$

$$\mathbf{m}_c = \frac{\sum_{n=1}^{N^{\text{cal}}} y_{n,c}^{\text{cal}} \mathbf{x}_n^{\text{cal}} + \beta_0 \mathbf{m}_0}{\beta_c}, \quad (20)$$

$$\nu_c = \sum_{n=1}^{N^{\text{cal}}} y_{n,c}^{\text{cal}} + \nu_c^{\text{src}}, \quad (21)$$

$$\mathbf{W}_c^{-1} = \sum_{n=1}^{N^{\text{cal}}} y_{n,c}^{\text{cal}} \mathbf{x}_n^{\text{cal}} \mathbf{x}_n^{\text{cal}\top} + \beta_0 \mathbf{m}_0 \mathbf{m}_0^\top - \beta_c \mathbf{m}_c \mathbf{m}_c^\top + (\mathbf{W}_c^{\text{src}})^{-1}, \quad (22)$$

$$\alpha_c = \sum_{n=1}^{N^{\text{cal}}} y_{n,c}^{\text{cal}} + \alpha_0. \quad (23)$$

From (21) and (22), the calculation of the posterior hyperparameters for  $\boldsymbol{\Lambda}_c$  utilizes hyperparameters determined by the source subjects. This allows transfer of variance information from source to target.

#### D. Motion class prediction

To perform motion recognition on new target subject test data  $\mathbf{x}_n^{\text{trg}}$ , we compute the following predictive distribution:

$$p(\mathbf{y}_n^{\text{trg}} | \mathbf{x}_n^{\text{trg}}, \mathcal{D}_{\text{cal}}, \mathcal{D}_{\text{src}}) = \int p(\mathbf{y}_n^{\text{trg}} | \mathbf{x}_n^{\text{trg}}, \boldsymbol{\theta}_c) p(\boldsymbol{\theta}_c | \mathcal{D}_{\text{cal}}, \mathcal{D}_{\text{src}}) d\boldsymbol{\theta}_c, \quad (24)$$

where  $\boldsymbol{\theta}_c = \{\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c, \boldsymbol{\pi}\}$  represents the set of model parameters. The class prediction probabilities defined by the predicted distribution are then calculated, and the class with the highest probability is determined to be the predicted motion class:

$$\hat{c}_n = \arg \max_c p(\mathbf{y}_{n,c}^{\text{trg}} = 1 | \mathbf{x}_n^{\text{trg}}, \mathcal{D}_{\text{cal}}, \mathcal{D}_{\text{src}}). \quad (25)$$

### III. EXPERIMENTS

#### A. Methods

To evaluate the effectiveness of the proposed inter-subject transfer learning, EMG pattern classification experiments were conducted using two EMG datasets. Table I summarizes the information for each dataset, which contains multi-channel EMG signals from several subjects over multiple trials. Datasets I and II are obtained from [12] and [10], respectively. For feature extraction, each dataset was subjected to full-wave rectification processing and smoothing using a second-order Butterworth low-pass filter with a cutoff frequency of 1 Hz. One subject was designated as the target subject, and the rest as source subjects, with validation performed on all possible combinations. All trials from source subjects were used as the pre-training set  $\mathcal{D}^{\text{src}}$ . For the target subject, only the first trial served as the calibration set  $\mathcal{D}^{\text{cal}}$ , and the remaining trials served as the test set for accuracy evaluation.

The prior hyperparameters for GCM were set as  $\mathbf{m}_0 = \mathbf{0}$ ,  $\beta_0 = 1$ ,  $\nu_0 = D + 1$ ,  $\alpha_0 = 1$ , and  $\mathbf{W}_0 = \mathbf{I}$  ( $\mathbf{I}$  as

the identity matrix). The weight coefficient  $w^s$  introduced during pre-training, is an important parameter for adjusting the amount of information transferred from source to target. In our experiments, we tentatively set  $w^s$  as follows:

$$w^s = \frac{N^{\text{cal}}}{N^s} \quad (26)$$

so that the ratio of the amount of transferred data is equal to that of the calibration data. This setting assumes that if the transfer amount is too high, the information from source subjects may dominate, hindering the appropriate incorporation of the target subject information into the learning process.

To investigate the impact of varying transferred information, we performed an ablation study for the proposed method. We also compared the performance of the proposed method with the following two existing methods:

- **Adaptive LDA** [9] is a classification method that introduces transfer learning into LDA. In this method, the model parameters,  $\mu_c$  and  $\Sigma$ , are determined as follows:

$$\mu_c = \tau \mu_c^{\text{cal}} + (1 - \tau) \mu_c^{\text{src}}, \quad (27)$$

$$\Sigma = \lambda \Sigma^{\text{cal}} + (1 - \lambda) \Sigma^{\text{src}}, \quad (28)$$

where  $\mu_c^{\text{src}}$  and  $\Sigma^{\text{src}}$  are the class-specific mean and shared covariance matrix calculated from the source dataset, respectively, whereas  $\mu_c^{\text{cal}}$  and  $\Sigma^{\text{cal}}$  are those from the target subject’s calibration dataset. The hyperparameters  $\{\tau, \lambda\}$  control the amount of transferred information and were tuned using  $11 \times 11$  grid search (ranging from 0 to 1) through two-fold cross-validation on the calibration dataset.

- **Adaptive quadratic discriminant analysis (Adaptive QDA)** [11] introduces transfer learning into QDA, which is an extension of LDA, where each class is given a class-specific covariance matrix  $\Sigma_c$ . Unlike LDA, QDA can learn quadratic decision boundaries. The model parameters,  $\mu_c$  and  $\Sigma_c$ , can be determined by replacing  $\Sigma^{\text{cal}}$  and  $\Sigma^{\text{src}}$  in (28) with  $\Sigma_c^{\text{cal}}$  and  $\Sigma_c^{\text{src}}$ , respectively. Tuning of the hyperparameters was performed in the same way as for the adaptive LDA.

In both the ablation study and the comparison with existing methods, we evaluated the effect of calibration data size on accuracy using the entire set and only the first 25%. Apart from the above evaluations, the sensitivity of the proposed method to the coefficient  $w^s$  was also explored.

## B. Results

In the ablation study, we calculated the classification accuracy for different transfer scenarios: transferring only variance information (ours), only mean information, both mean and variance information, and not transferring any information. The results are presented in Table II. Regardless of the dataset and the percentage of calibration data used, the proposed method of transferring only variance information demonstrated the best performance. Table III shows the results of the comparison with existing methods. The proposed method outperformed the existing methods in all cases.

TABLE II  
AVERAGE CLASSIFICATION ACCURACY (%) IN ABLATION STUDY

Dataset	Type of transfer	% of calibration data used	
		25%	100%
I	No transfer	73.27 ± 18.42	78.89 ± 12.46
	Mean & variance transfer	61.10 ± 22.99	69.64 ± 20.05
	Mean transfer	62.41 ± 16.86	75.48 ± 20.10
	Variance transfer ( <b>ours</b> )	<b>75.99 ± 10.69</b>	<b>85.51 ± 10.26</b>
II	No transfer	20.87 ± 7.25	61.82 ± 6.75
	Mean & variance transfer	47.47 ± 8.47	60.84 ± 5.51
	Mean transfer	21.27 ± 7.81	61.03 ± 6.41
	Variance transfer ( <b>ours</b> )	<b>48.49 ± 10.04</b>	<b>64.27 ± 4.70</b>

TABLE III  
AVERAGE CLASSIFICATION ACCURACY (%) FOR EACH METHOD

Dataset	Method	% of calibration data used	
		25%	100%
I	Adaptive LDA	73.72 ± 6.43	84.49 ± 4.92
	Adaptive QDA	52.62 ± 13.29	63.62 ± 19.66
	<b>Ours</b>	<b>75.99 ± 10.69</b>	<b>85.51 ± 10.26</b>
II	Adaptive LDA	31.53 ± 13.88	60.50 ± 3.88
	Adaptive QDA	22.98 ± 9.16	57.86 ± 6.87
	<b>Ours</b>	<b>48.49 ± 10.04</b>	<b>64.27 ± 4.70</b>

The effect of modulating the transfer amount by varying the coefficient  $w^s$  on performance was also explored. As the number of source subjects,  $S$ , varied across datasets, we expressed  $w^s$  as a ratio  $r$  of the volumes of source to calibration data, as follows:

$$r = \frac{\sum_{s=1}^S w^s N^s}{S N^{\text{cal}}}. \quad (29)$$

Performance was assessed at different  $r$  values ( $r = 0, 0.5, 1, 2, 5, 10, 20, 50, 100$ ), and the results are shown in Fig. 3. Note that in previous ablation and comparative studies, the default  $w^s$  corresponded to  $r = 1.0$ . In both datasets, average accuracy peaked around 0.5–1.0 and then gradually decreased as  $r$  increased.

## C. Discussion

The ablation study showed that the proposed method (variance transfer) outperforms other scenarios (Table II). With limited calibration data, the computed variance tends to be underestimated compared to the actual variability during prediction. The proposed method can better represent the spread of the target data distribution by transferring the variance from source subjects containing sufficient variability, to improve classification accuracy. In contrast, transferring the mean leads to lower accuracy than no-transfer scenario. This could be because EMG feature means vary largely among individuals due to biomechanical and physiological factors, and thus sharing mean information may negatively impact model training.

In both datasets, the existing methods, adaptive LDA and adaptive QDA, exhibited lower accuracy compared to the proposed method (Table III). The discrepancy in accuracy

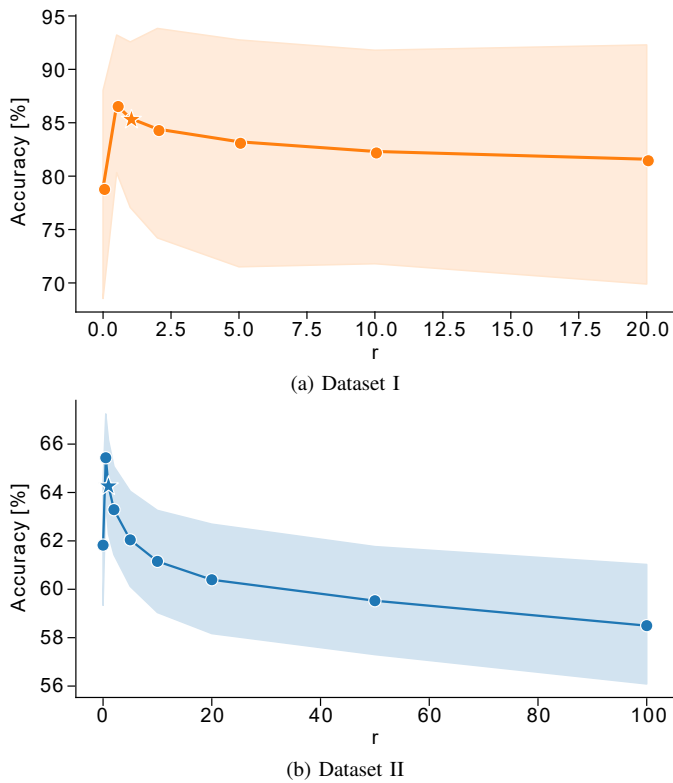


Fig. 3. Average classification accuracy with changing  $r$  for each dataset. Note that the maximum value of  $r$  varies across datasets due to differences in the number of source subjects and data length of them. Each point and the shaded areas represent average accuracy and 95% confidence interval respectively, for all participants. Stars indicate an accuracy corresponding to  $r = 1$ .

was especially pronounced when using 25% of the calibration data, suggesting that the existing methods might overfit the calibration data. In particular, QDA, being a nonlinear classifier, is more prone to overfitting than LDA [13]. By contrast, the proposed Bayesian method inherently includes a form of regularization through the prior distribution. Additionally, the introduction of weight coefficient  $w^s$  allows for adjusting the uncertainty of the distributions (and thus the transferred information) based on the amount of calibration data, leading to higher classification accuracy.

We found that there is an optimal value for the amount of transferred information controlled by the coefficient  $w^s$  (Fig. 3); setting it too high leads to excessive transfer of source subject information, resulting in reduced accuracy. In the performance comparison experiments, we tentatively set  $w^s$  such that  $r = 1$ , which, coincidentally, was close to the average optimal value. However, as indicated by the error range in the figure, the accuracy variation trends with  $r$  significantly differ among individuals. Future improvements in accuracy can be achieved by introducing theoretical or computational methods for an efficient determination of the optimal  $r$ .

#### IV. CONCLUSION

In this paper, we proposed an inter-subject GCM-based transfer learning method. The proposed method can achieve

high accuracy with limited labeled training or calibration data by transferring the variance information acquired from the pre-training of source subjects to the target subject. In addition, by introducing a coefficient, the uncertainty of the pre-trained posterior distributions was adjusted to control the amount of transferred information, thereby preventing accuracy decline from excessive transfer.

The efficacy of the proposed method was evaluated using two EMG datasets. The results indicate that the proposed method outperforms other methods in both datasets, demonstrating the effectiveness of transferring variance information in inter-subject transfer learning. In the future, we will introduce a framework to adjust the optimal amount of transferred information for each subject and improve accuracy through more complex model structures.

#### ACKNOWLEDGMENT

This work was partially supported by JSPS KAKENHI Grant Number JP23H03438 and JP23K28128.

#### REFERENCES

- [1] A. Tsoli and O. C. Jenkins, "Robot grasping for prosthetic applications," in *Robotics Research*. Springer Berlin Heidelberg, 2011, pp. 1–12.
- [2] I. M. Rezazadeh, M. Firoozabadi, H. Hu, and S. M. R. Golpayegani, "Co-adaptive and affective human-machine interface for improving training performances of virtual myoelectric forearm prosthesis," *IEEE Trans. Affect. Comput.*, vol. 3, no. 3, pp. 285–297, 2012.
- [3] X. Zhang *et al.*, "Domain adaptation with self-guided adaptive sampling strategy: Feature alignment for cross-user myoelectric pattern recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1374–1383, May 2022.
- [4] 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. Biomim. (ROBIO)*, Dec. 2011, pp. 390–395.
- [5] 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, Jul. 2023.
- [6] Y. Zou and L. Cheng, "A transfer learning model for gesture recognition based on the deep features extracted by CNN," *IEEE Trans. Artif. Intell.*, vol. 2, no. 5, pp. 447–458, Oct. 2021.
- [7] K.-T. Kim, C. Guan, and S.-W. Lee, "A subject-transfer framework based on single-trial EMG analysis using convolutional neural networks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 94–103, Jan. 2020.
- [8] T. Bao, S. A. R. Zaidi, S. Xie, P. Yang, and Z.-Q. Zhang, "Inter-subject domain adaptation for CNN-based wrist kinematics estimation using sEMG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1068–1078, Jun. 2021.
- [9] M. M.-C. Vidovic *et al.*, "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.
- [10] 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, p. 102817, Jul. 2021.
- [11] M. M.-C. Vidovic *et al.*, "Covariate shift adaptation in EMG pattern recognition for prosthetic device control," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2014, pp. 4370–4373.
- [12] 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)*, 2023, pp. 3310–3315.
- [13] A. Krasoulis, K. Nazarpour, and S. Vijayakumar, "Use of regularized discriminant analysis improves myoelectric hand movement classification," in *Proc. 8th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, May 2017, pp. 395–398.