

Hybrid Generative-Discriminative Learning of Scale Mixture Models for EMG Classification

Shunya Fukuda

*Graduate School of Advanced
Science and Engineering
Hiroshima University
Higashi-hiroshima, Japan
shunya-fukuda@hiroshima-u.ac.jp*

Toshio Tsuji

*Graduate School of Advanced
Science and Engineering
Hiroshima University
Higashi-hiroshima, Japan
toshiotsuji@hiroshima-u.ac.jp*

Akira Furui

*Graduate School of Advanced
Science and Engineering
Hiroshima University
Higashi-hiroshima, Japan
akirafurui@hiroshima-u.ac.jp*

Abstract—Electromyogram (EMG) classification plays an important role in prosthetic hands/limbs control and rehabilitation support. The scale mixture model (SMM), which captures the non-Gaussian characteristics of EMG signals, offers robustness against inter-trial distribution shifts and favorable confidence calibration properties. However, conventional methods estimate SMM parameters using maximum likelihood estimation, which is not necessarily optimal for classification. This paper implements the SMM within an automatic differentiation framework, enabling parameter optimization with respect to classification performance. Specifically, we introduce hybrid learning that combines negative log-likelihood loss (generative learning) and cross-entropy loss (discriminative learning). Experiments on three public EMG datasets demonstrate that the proposed method achieves higher classification accuracy than conventional approaches while maintaining the SMM’s robustness to distribution shifts and confidence calibration properties. These results suggest that the proposed approach enables both high-accuracy EMG classification in real-world environments prone to distribution shifts and safe prosthetic control based on well-calibrated confidence.

Index Terms—Electromyogram (EMG), Generative learning, Discriminative learning, Stochastic model, Scale mixture model

I. INTRODUCTION

Various biological phenomena in humans are accompanied by electrical activities, and the electrical signals measured from these activities are referred to as biosignals. One such biosignal is the electromyogram (EMG), which records the electrical activity of muscles. Classifying EMG signals using machine learning enables various applications, such as the control of prosthetic hands/limbs [1] and rehabilitation support [2].

Machine learning models for EMG classification are broadly categorized into discriminative and generative models. Discriminative models characterize the conditional probability $p(y|\mathbf{x})$ of a class label y given input data $\mathbf{x}$; thus, they tend to exhibit high classification accuracy. In contrast, generative models characterize the data distribution $p(\mathbf{x})$ and compute the posterior probability using Bayes’ theorem. Generative models offer several advantages, including the potential to outperform discriminative models when appropriate distributions are assumed for the input data [3], robustness against distribution shifts between trials, and favorable confidence calibration properties. In EMG classification, these advantages

are practically significant because distribution shifts frequently occur across trials due to changes in measurement conditions, and applications such as prosthetic control require safe decision-making based on well-calibrated confidence.

Conventionally, Gaussian distributions have been widely assumed for EMG classification based on generative-models [4], [5]. However, recent studies have reported that the amplitude distribution of EMG signals exhibits heavier tails than a Gaussian distribution [6]. As a model to capture such non-Gaussianity, the scale mixture model (SMM) has been proposed [7]. By weighting the variance of a Gaussian distribution with a latent variable, the SMM can represent variance uncertainty and model heavy-tailed distributions. Classification methods using this model have achieved higher accuracy than both traditional Gaussian-based methods and discriminative models, while also outputting confidence levels that closely align with actual accuracy [7], [8].

However, in conventional SMM-based classification methods, model parameters are determined using maximum likelihood estimation (MLE). Since MLE aims to accurately capture the distribution of the input data rather than directly optimizing classification boundaries, the obtained parameters are not necessarily optimal from a classification perspective. To address this limitation in generative models, previous studies have proposed learning rules that directly minimize classification error in addition to likelihood maximization, such as discriminative learning for Gaussian mixture models [9] and hybrid learning combining generative and discriminative approaches [10]. Based on these findings, applying a similar approach to SMM is expected to improve classification performance. To achieve this, we must implement SMM in a differentiable form, since the SMM density function involves the integration of latent variables and traditional EM algorithms rely on analytical update equations that cannot accommodate arbitrary loss functions.

In this paper, we implement the SMM in a differentiable format and introduce hybrid learning that combines generative and discriminative approaches. This enables parameter optimization for arbitrary loss functions and improves classification performance compared to conventional MLE-based methods. The primary contributions of this study are

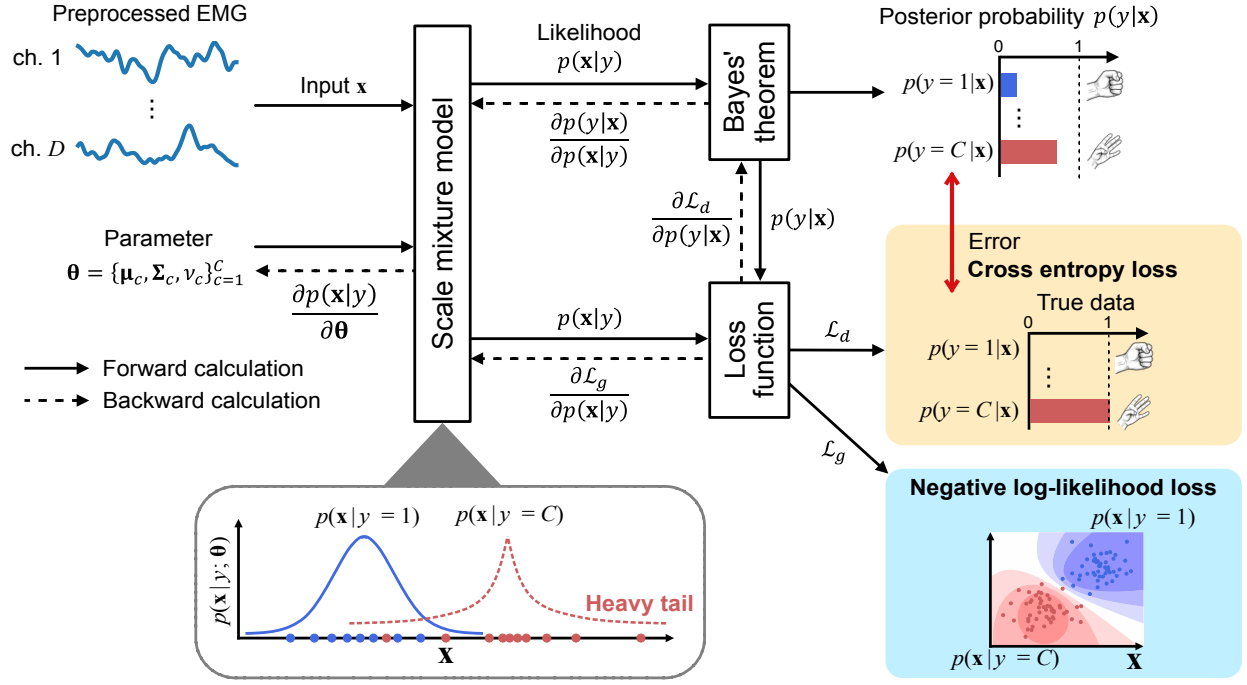


Fig. 1. Overview of the proposed method. The SMM computes class-conditional likelihoods, which are converted to posterior probabilities via Bayes' theorem. Parameters are optimized by backpropagating gradients from the loss function.

as follows:

- We implemented the SMM as a differentiable module within an automatic differentiation framework. This allows for parameter optimization with any differentiable loss function, which was difficult to achieve with conventional EM algorithms.
- We applied hybrid learning, combining generative and discriminative approaches, to the SMM for the first time. This improved classification performance while maintaining the SMM's inherent robustness to distribution shifts and its calibration properties.
- Experiments using three public EMG datasets demonstrated that the proposed method achieves superior classification accuracy compared to conventional methods. Furthermore, we analyzed the characteristics of the optimal mixing coefficient α in hybrid learning.

II. METHODS

This section describes a method for estimating the parameters of the SMM from a classification perspective. We formulate parameter estimation as an optimization problem with a differentiable loss function and estimate the parameters using gradient-based optimization. Fig. 1 shows a schematic diagram of the proposed method.

A. Posterior Probability Calculation based on SMM

We consider the classification problem of predicting a class label $y \in \{1, \dots, C\}$ from a D -dimensional biosignal $\mathbf{x} \in$

$\mathbb{R}^D$. The conditional probability $p(y = c|\mathbf{x})$ that $\mathbf{x}$ belongs to class c is derived using Bayes' theorem as follows:

$$p(y = c|\mathbf{x}) = \frac{p(\mathbf{x}|y = c)p(y = c)}{\sum_{c'=1}^C p(\mathbf{x}|y = c')p(y = c')}, \quad (1)$$

where $p(y)$ represents the prior probability of the class, and $p(\mathbf{x}|y)$ is the likelihood calculated based on the density function of the SMM [5]. In the SMM, the class-conditional density $p(\mathbf{x}|y = c)$ is modeled as:

$$p(\mathbf{x}|y = c) = \int \mathcal{N}(\mathbf{x}|\mu_c, u_c \Sigma_c) \text{IG}(u_c|\nu_c/2, \nu_c/2) du_c, \quad (2)$$

where $\mathcal{N}(\cdot)$ denotes the Gaussian distribution, $\text{IG}(\cdot)$ is the inverse gamma distribution, $u_c \in \mathbb{R}^+$ is a latent variable, $\mu_c \in \mathbb{R}^D$ is the mean vector, $\Sigma_c \in \mathbb{R}^{D \times D}$ is a symmetric scale matrix, and $\nu_c \in \mathbb{R}^+$ represents the degrees of freedom. The covariance matrix of the Gaussian distribution is determined by weighting the scale matrix Σ_c with the random latent variable u_c . By modeling variance uncertainty through this structure, the SMM can represent distributions with heavier tails than the Gaussian distribution. Furthermore, since the inverse gamma distribution serves as a conjugate prior for the Gaussian variance, Eq. (2) can be calculated analytically.

Previous studies have reported that sharing the degrees of freedom ν_c across classes improves classification performance [7]. This improvement is attributed to a regularization effect that suppresses overfitting by assuming a common tail behavior across all classes. In conventional studies, the shared degrees of freedom ν were treated as a hyperparameter. The proposed gradient-based framework also enables learning ν

TABLE I
SUMMARY OF THE THREE EMG DATASETS USED IN EXPERIMENTS

Dataset	# Motions (C)	# Electrodes (D)	Sampling frequency	Recording duration	# Participants	Training trial	Validation trial	Test trials
I	10	2	4,000 Hz	5 s	10	1	2	3, 4, 5, 6
II	14	8	4,000 Hz	10 s	8	1	2	3, 4
III	15	8	4,000 Hz	20 s	8	1	2	3

as a trainable parameter. We compare both approaches in our experiments.

B. Gradient-Based Parameter Estimation

We consider learning the model parameters given a training dataset $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^N$ of sample size N . Learning is achieved by minimizing a loss function $\mathcal{L}$ that measures the discrepancy between the predicted probabilities and the ground-truth labels y_i . When $\mathcal{L}$ is the negative log-likelihood, analytical update rules based on the EM algorithm can be obtained [5]; however, deriving such updates is difficult for other loss functions. Therefore, the proposed method updates the model parameters using stochastic gradient descent (SGD):

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \gamma \frac{1}{|B|} \sum_{j \in B} \frac{\partial \mathcal{L}(\mathbf{x}_j, y_j; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \quad (3)$$

where $B \subset \{1, \dots, N\}$ is the set of indices for a mini-batch randomly sampled from the training data, $|B|$ is the batch size, and $\gamma \in \mathbb{R}^+$ is the learning rate. The model parameters are summarized as $\boldsymbol{\theta} = \{\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c, \nu_c\}_{c=1}^C$. The density function in Eq. (2) is continuously differentiable with respect to $\boldsymbol{\mu}_c$, $\boldsymbol{\Sigma}_c$, and ν_c . Consequently, the partial derivatives can be efficiently computed using automatic differentiation in standard deep learning frameworks. This enables SMM parameter optimization with any differentiable loss function.

When learning parameters using Eq. (3), updated values may violate parameter constraints (e.g., $\nu_c > 0$, $\boldsymbol{\Sigma}_c$ positive definite), resulting in undefined density values. To address this, we apply transformations to ensure that constraints are satisfied throughout learning. Specifically, the softplus function $f(\nu_c) = \log(1 + \exp(\nu_c))$ is applied to ν_c to ensure positivity. The mean vector $\boldsymbol{\mu}_c$ is unconstrained. To ensure positive definiteness, $\boldsymbol{\Sigma}_c$ is parameterized via Cholesky decomposition $\boldsymbol{\Sigma}_c = \mathbf{A}_c \mathbf{A}_c^T$, where $\mathbf{A}_c \in \mathbb{R}^{D \times D}$ is a lower triangular matrix. The softplus function is applied to the diagonal elements of $\mathbf{A}_c$ to ensure they are positive, while off-diagonal elements remain unconstrained.

C. Loss Functions

In the proposed method, we consider the following three loss functions $\mathcal{L}$. Let c^* denote the ground-truth class for input $\mathbf{x}$.

- **Negative log-likelihood (NLL) Loss** (generative learning): Minimizing this loss maximizes the log-likelihood of the ground-truth class.

$$\mathcal{L}_g = -\frac{1}{|B|} \sum_{i \in B} \ln p(\mathbf{x}_i | y_i = c^*). \quad (4)$$

This loss encourages accurate modeling of the class-conditional data distribution but does not directly optimize decision boundaries.

- **Cross-entropy (CE) Loss** (discriminative learning): Measures the cross-entropy between the predicted posterior $p(y|\mathbf{x})$ and the ground-truth label.

$$\mathcal{L}_d = -\frac{1}{|B|} \sum_{i \in B} \ln p(y_i = c^* | \mathbf{x}_i). \quad (5)$$

Since this loss directly optimizes the posterior probability, it is effective for learning decision boundaries.

- **Hybrid loss** [10] (hybrid learning): A weighted combination of the generative and discriminative losses.

$$\mathcal{L}_h = \alpha \mathcal{L}_g + (1 - \alpha) \mathcal{L}_d, \quad (6)$$

where $\alpha \in [0, 1]$ is a hyperparameter controlling the balance; $\alpha = 1$ corresponds to pure generative learning, while $\alpha = 0$ corresponds to pure discriminative learning.

III. EXPERIMENTS

To evaluate the effectiveness of the proposed method, we conducted classification experiments using three public EMG datasets related to upper limb movements. Table I summarizes the details of each dataset. Datasets I, II, and III were obtained from [11], [12], and [13], respectively.

Each dataset consists of multiple measurement trials for each motion. The first trial was used as training data, the second trial as validation data, and the remaining trials as test data. As preprocessing, the EMG signals were full-wave rectified and then smoothed using a second-order low-pass filter with a cutoff frequency of 2 Hz.

For the SMM-based classifier, the parameters were learned using the three loss functions described in Section II-C. The prior probability of each class was set to $p(y = c) = 1/C$, and the model parameters were initialized by sampling from a standard Gaussian distribution. Since sharing the parameter ν among classes has been reported to improve classification performance [7], we compared three configurations for ν : (1) ν_c , where each class has its own learnable degrees of freedom, (2) ν_{train} , where a single shared ν is learned across all classes, and (3) $\hat{\nu}_{\text{val}}$, where a shared ν is treated as a hyperparameter optimized on validation data.

Adam was used for optimization with a learning rate of $\gamma = 0.001$ and a batch size of $|B| = 1024$. Note that since each sample point is treated as a training sample, the number of training samples is given by the product of the sampling frequency and the recording duration. Training was conducted for 100 epochs, and the model with the minimum validation

TABLE II
CLASSIFICATION ACCURACY AND ECE ON THE TEST DATA

Method	Dataset I		Dataset II		Dataset III	
	Accuracy ($\uparrow$)	ECE ($\downarrow$)	Accuracy ($\uparrow$)	ECE ($\downarrow$)	Accuracy ($\uparrow$)	ECE ($\downarrow$)
<i>Baseline classifiers</i>						
LDA	0.5318 $\pm$ 0.1409	0.1933 $\pm$ 0.0754	0.7322 $\pm$ 0.1376	0.2375 $\pm$ 0.1331	0.6737 $\pm$ 0.0999	0.2163 $\pm$ 0.0705
LR	0.5696 $\pm$ 0.1561	0.2608 $\pm$ 0.0978	0.6800 $\pm$ 0.1044	0.3107 $\pm$ 0.1073	0.6619 $\pm$ 0.1043	0.2805 $\pm$ 0.0852
MLP	0.5680 $\pm$ 0.1125	0.3306 $\pm$ 0.0592	0.6641 $\pm$ 0.1187	0.2995 $\pm$ 0.1169	0.6445 $\pm$ 0.1252	0.3265 $\pm$ 0.1226
Non-linear SVM	0.5189 $\pm$ 0.1314	0.2131 $\pm$ 0.0708	0.7144 $\pm$ 0.1333	0.2027 $\pm$ 0.1105	0.6514 $\pm$ 0.1175	0.2086 $\pm$ 0.0849
<i>Gaussian model</i>						
generative learning	0.4938 $\pm$ 0.1515	0.2296 $\pm$ 0.0665	0.4885 $\pm$ 0.1346	0.4971 $\pm$ 0.1380	0.6249 $\pm$ 0.1121	0.3206 $\pm$ 0.0954
discriminative learning	0.4469 $\pm$ 0.1661	0.1508 $\pm$ 0.0779	0.5481 $\pm$ 0.0888	0.4519 $\pm$ 0.0888	0.5612 $\pm$ 0.1045	0.3128 $\pm$ 0.0983
hybrid learning	0.5711 $\pm$ 0.1264	0.2520 $\pm$ 0.0801	0.6802 $\pm$ 0.1363	0.3063 $\pm$ 0.1347	0.6904 $\pm$ 0.1273	0.2721 $\pm$ 0.1101
<i>Scale mixture model (ours)</i>						
generative learning	0.6082 $\pm$ 0.1141	0.1324 $\pm$ 0.0615	0.7313 $\pm$ 0.1189	0.1622 $\pm$ 0.1014	0.7270 $\pm$ 0.1125	0.1436 $\pm$ 0.0847
discriminative learning	0.5469 $\pm$ 0.1150	0.1846 $\pm$ 0.0935	0.6535 $\pm$ 0.1125	0.2586 $\pm$ 0.1142	0.6914 $\pm$ 0.1227	0.2141 $\pm$ 0.0883
hybrid learning	0.6169 $\pm$ 0.1227	0.1437 $\pm$ 0.0764	0.7441 $\pm$ 0.0997	0.1497 $\pm$ 0.0767	0.7270 $\pm$ 0.1297	0.1471 $\pm$ 0.0855

loss was selected. The hyperparameters α and $\hat{\nu}_{\text{val}}$ were tuned using Bayesian optimization with the Ax library, employing a Gaussian process surrogate and log expected improvement as the acquisition function. The number of optimization trials was 30, with search ranges of $\alpha \in [0.0, 1.0]$ and $\hat{\nu}_{\text{val}} \in [0.001, 5]$.

Models were trained separately for each participant. Performance was evaluated using classification accuracy and expected calibration error (ECE). The ECE measures how well predicted confidence aligns with actual accuracy: predictions are grouped into bins by confidence level, and ECE is computed as the weighted average of the absolute difference between mean confidence and accuracy within each bin. Lower ECE indicates better-calibrated predictions. For comparison, we used the following baseline methods: linear discriminant analysis (LDA), logistic regression (LR), multilayer perceptron (MLP), non-linear support vector machine (SVM) and a Gaussian-based generative classifier. The MLP had three hidden layers with 100, 50, and 25 units, respectively. The SVM used a radial basis function (RBF) kernel with L2 regularization and a regularization coefficient of 1.

IV. RESULTS

Table II presents the classification results on the test data for each dataset. For the SMM, we report the results using the hyperparameter configuration $\hat{\nu}_{\text{val}}$. The best metric values for each dataset are shown in bold. Across all datasets, the proposed classifier with hybrid learning achieved the highest classification accuracy (Dataset I: 61.69%, Dataset II: 74.41%, Dataset III: 72.70%). Comparing generative and discriminative learning for the SMM, the generative approach achieved a higher accuracy. Furthermore, the SMM-based classifiers exhibited a lower ECE than conventional methods, and this characteristic was maintained even with hybrid learning.

Fig. 2 illustrates the distribution of the coefficient α selected by hybrid learning (i.e., the value achieving the highest validation accuracy). These results correspond to the configuration where ν was optimized as a hyperparameter $\hat{\nu}_{\text{val}}$, and the distribution is visualized using kernel density estimation. In all datasets, the mode lies within $0 < \alpha < 1$, where

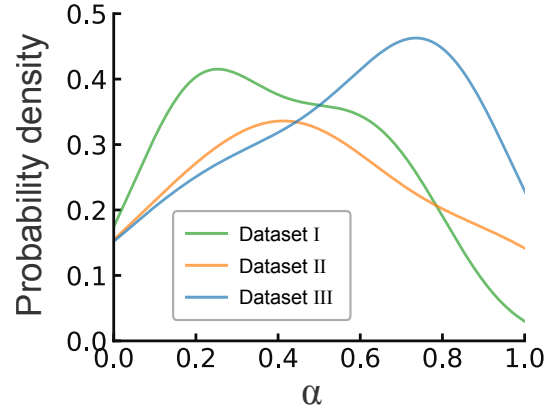


Fig. 2. Distribution of optimal α values selected based on validation accuracy for each dataset, visualized using kernel density estimation.

TABLE III
ABLATION STUDY: IMPACT OF DEGREES-OF-FREEDOM CONFIGURATION

Learning	Dataset I	Dataset II	Dataset III
<i>Generative</i>			
ν_c	0.5914 $\pm$ 0.1281	0.7355 $\pm$ 0.0989	0.6154 $\pm$ 0.1397
ν_{train}	0.5881 $\pm$ 0.1216	0.7360 $\pm$ 0.0930	0.6329 $\pm$ 0.1244
$\hat{\nu}_{\text{val}}$	0.6082 $\pm$ 0.1141	0.7313 $\pm$ 0.1189	0.7270 $\pm$ 0.1125
<i>Hybrid</i>			
ν_c	0.5875 $\pm$ 0.1310	0.7467 $\pm$ 0.1132	0.6942 $\pm$ 0.1101
ν_{train}	0.5856 $\pm$ 0.1325	0.7443 $\pm$ 0.1144	0.7248 $\pm$ 0.0981
$\hat{\nu}_{\text{val}}$	0.6169 $\pm$ 0.1227	0.7441 $\pm$ 0.0997	0.7270 $\pm$ 0.1297

generative and discriminative learning are combined. The peak was observed around $\alpha \approx 0.2$ for Dataset I and $\alpha \approx 0.7$ for Dataset III, indicating that the optimal α varies across datasets.

Table III presents the results of an ablation study on the degrees of freedom configuration. The best metric values for each dataset are shown in bold. For Datasets I and III, the $\hat{\nu}_{\text{val}}$ approach achieved the highest classification accuracy. In contrast, for Dataset II, the differences among the configurations were negligible, with all three approaches achieving similar accuracy around 73–74%.

V. DISCUSSION

We first discuss the factors contributing to the lower classification accuracy of discriminative learning compared to generative learning, as shown in Table II. In Dataset I, the training accuracy was equivalent to or lower than that achieved by generative learning, suggesting that the optimization may have converged to local optima. In Datasets II and III, although the training accuracy reached nearly 100%, the test accuracy decreased, indicating overfitting. These results suggest that pure discriminative learning is highly sensitive to the size and distribution of the training data.

In contrast, the hybrid learning approach proved to be effective. As illustrated in Fig. 2, the classification accuracy on the validation data often reached its maximum within the range of $0 < \alpha < 1$ across all datasets, demonstrating the effectiveness of combining generative and discriminative learning. Generative learning models the underlying distribution of the input data, making it robust to distribution shifts between trials, although it does not directly optimize classification boundaries. Hybrid learning combines the advantages of both approaches by preserving this robustness while directly minimizing classification errors.

Regarding ECE, SMM-based classification methods demonstrated lower values compared to conventional methods (Table II). This characteristic was preserved even under hybrid learning, which incorporates discriminative learning. This is likely because the SMM assumes a heavy-tailed distribution, which prevents the model from assigning extremely low probabilities to samples distant from the distribution center, thereby avoiding overconfident predictions. A classifier that achieves both high accuracy and low ECE is crucial for practical applications where reliability-based decision-making is required.

Based on the results in Table III, we analyze the impact of the degrees-of-freedom configuration on classification performance. In both generative and hybrid learning, the $\hat{\nu}_{\text{val}}$ approach tended to yield better results than ν_{train} . The parameter ν determines the heaviness of the distribution's tails and is related to the model's robustness. While ν_{train} is estimated from the training data, $\hat{\nu}_{\text{val}}$ is optimized using validation data from a different trial. When distribution shifts occur between trials, $\hat{\nu}_{\text{val}}$ is more effective for unseen data because it accounts for such shifts. However, since $\hat{\nu}_{\text{val}}$ was explored using Bayesian optimization in this experiment, computational overhead remains a limitation. An iterative approach that alternates between estimating ν and other parameters may address this problem.

VI. CONCLUSION

In this paper, we proposed a method for optimizing the parameters of SMM from a classification perspective by implementing the SMM within an automatic differentiation framework, which enables optimization with arbitrary differentiable loss functions. This implementation enables hybrid learning, which combines generative learning for likelihood maximization with discriminative learning for direct minimization of

classification error. To verify its effectiveness, we conducted classification experiments on three public EMG datasets using three types of loss functions. While pure discriminative learning was prone to local optima and overfitting, the hybrid learning approach achieved higher classification accuracy and lower ECE, benefiting from the regularization effect of generative learning. These findings suggest that hybrid learning provides a robust alternative for EMG classification, especially in the presence of distribution shifts between trials. Additionally, our ablation study revealed that treating the degrees of freedom ν as a hyperparameter optimized on validation data outperformed learning it as a trainable parameter, particularly when distribution shifts exist between training and test trials.

In future work, we will further exploit the flexibility of the proposed method, which enables optimization with any differentiable loss function. Specifically, we plan to develop learning strategies tailored to distribution shifts and investigate the integration of the SMM-based approach with deep neural networks.

REFERENCES

- [1] A. Krasoulis, S. Vijayakumar, and K. Nazarpour, "Multi-grip classification-based prosthesis control with two emg-imu sensors," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 508–518, 2020.
- [2] M. Lin, J. Huang, J. Fu, Y. Sun, and Q. Fang, "A vr-based motor imagery training system with emg-based real-time feedback for post-stroke rehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1–10, 2023.
- [3] A. Ng and M. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes," *Adv. Neural Inf. Process. Syst.*, vol. 14, 2001.
- [4] Y. Huang, K. Englehart, B. Hudgins, and A. Chan, "A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 11, pp. 1801–1811, 2005.
- [5] 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, 2021.
- [6] K. Nazarpour, A. H. Al-Timemy, G. Bugmann, and A. Jackson, "A note on the probability distribution function of the surface electromyogram signal," *Brain Res. Bull.*, vol. 90, pp. 88–91, 2013.
- [7] 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, p. 115644, 2021.
- [8] A. Furui, "Evaluating classifier confidence for surface emg pattern recognition," in *Proc. 45th Ann. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2023, pp. 1–4.
- [9] A. Klautau, N. Jevtic, and A. Orlitsky, "Discriminative Gaussian mixture models: A comparison with kernel classifiers," in *Proc. 20th Int. Conf. Mach. Learn.*, 2003, pp. 353–360.
- [10] J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West, "Generative or discriminative? getting the best of both worlds," *Bayesian Stat.*, vol. 8, no. 3, pp. 3–24, 2007.
- [11] R. N. Khushaba, S. Kodagoda, M. Takruri, and G. Dissanayake, "Toward improved control of prosthetic fingers using surface electromyogram (EMG) signals," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 10731–10738, 2012.
- [12] R. N. Khushaba, S. Kodagoda, D. Liu, and G. Dissanayake, "Muscle computer interfaces for driver distraction reduction," *Comput. Methods Programs Biomed.*, vol. 110, no. 2, pp. 137–149, 2013.
- [13] R. N. Khushaba and S. Kodagoda, "Electromyogram (EMG) feature reduction using mutual components analysis for multifunction prosthetic fingers control," in *Proc. 12th Int. Conf. Control Autom. Robot. Vision.*, 2012, pp. 1534–1539.