Random Channel Ablation for Robust Hand Gesture Classification with Multimodal Biosignals

Bimbraw, Keshav; Liu, Jing; Wang, Ye; Koike-Akino, Toshiaki

TR2024-103   July 26, 2024

Abstract

Biosignal-based hand gesture classification is an important component of effective human-machine interaction. For multimodal biosignal sensing, the modalities often face data loss due to missing channels in the data which can adversely affect the gesture classification performance. To make the classifiers robust to missing channels in the data, this paper proposes using Random Channel Ablation (RChA) during the training process. Ultrasound and force myography (FMG) data were acquired from the forearm for 12 hand gestures over 2 subjects. The resulting multimodal data had 16 total channels, 8 for each modality. The proposed method was applied to convolutional neural network architecture, and compared with baseline, imputation, and oracle methods. Using 5-fold cross-validation for the two subjects, on average, 12.2% and 24.5% improvement was observed for gesture classification with up to 4 and 8 missing channels respectively compared to the baseline. Notably, the proposed method is also robust to an increase in the number of missing channels compared to other methods. These results show the efficacy of using random channel ablation to improve classifier robustness for multimodal and multi-channel biosignal- based hand gesture classification.

International Conference of the IEEE Engineering in Medicine and Biology Society 2024

Random Channel Ablation for Robust Hand Gesture Classification with Multimodal Biosignals

Keshav Bimbraw*†, Jing Liu†, Ye Wang†, and Toshiaki Koike-Akino†
*Worcester Polytechnic Institute, Worcester, MA 01605, USA
†Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA
kbimbraw@wpi.edu, {jiliu, yewang, koike}@merl.com

Abstract—Biosignal-based hand gesture classification is an important component of effective human-machine interaction. For multimodal biosignal sensing, the modalities often face data loss due to missing channels in the data which can adversely affect the gesture classification performance. To make the classifiers robust to missing channels in the data, this paper proposes using Random Channel Ablation (RChA) during the training process. Ultrasound and force myography (FMG) data were acquired from the forearm for 12 hand gestures over 2 subjects. The resulting multimodal data had 16 total channels, 8 for each modality. The proposed method was applied to convolutional neural network architecture, and compared with baseline, imputation, and oracle methods. Using 5-fold cross-validation for the two subjects, on average, 12.2% and 24.5% improvement was observed for gesture classification with up to 4 and 8 missing channels respectively compared to the baseline. Notably, the proposed method is also robust to an increase in the number of missing channels compared to other methods. These results show the efficacy of using random channel ablation to improve classifier robustness for multimodal and multi-channel biosignal-based hand gesture classification.

Index Terms—Biomedical Signal Processing, Biomedical Imaging, Image Processing, Biomedical Sensors, Deep Learning, Robustness, Multimodal, Multi-Channel, Ultrasound, FMG

I. INTRODUCTION

Hand gesture recognition using biosignals has been a focus of research for designing pipelines for effective human-machine interfaces. For this, several biosignal modalities have been explored such as surface electromyography (sEMG) [1], ultrasound [2], force myography (FMG) [3], photoplethysmography (PPG) [4], mechanomyography [5], inertial measurement unit (IMU) [6], and electrical impedance tomography (EIT) [7]. Forearm ultrasound can help with visualization of the musculature which can be used to estimate hand gestures [8]. Recent research has shown that it can be used to estimate different hand configurations [9], finer finger movements [2] and forces [10]. With FMG, piezoelectric sensors around the forearm are used to acquire pressure information corresponding to muscle contraction, from which hand gestures can be inferred. Ha et al. showed an average classification accuracy of over 80% for estimating hand gestures and prosthetic hand control [11]. Multimodal biosignal-based gesture classification has been explored primarily with sEMG [12], in addition to modalities like ultrasound [13], FMG [14] among others. Multimodal data based systems have been shown to improve the gesture classification performance compared to using just one modality [15]. We explore using ultrasound and FMG in a multimodal fashion for gesture recognition in this paper.

With the advancements in machine learning, there is a growing interest in using biosignals for hand gesture classification not just in academia but also in industry [16]. The robustness of the machine learning models is of prime importance for effective hand gesture classification. Though combining multiple modalities and multiple channels can often lead to more accurate predictions, it also increases the fragility of the system, as in practice there are often missing channel and missing modality issues due to the unreliability of the sensors and communications, as well as partial occlusions and other disruptions. A common practice is to impute the missing values with zeros or the mean [17]. However, such ad-hoc approaches have no robustness guarantee and still suffer from significant performance degradation when the number of missing channels increases. In this paper, we aim to improve the robustness of multimodal and multi-channel signal processing applied to biosignal based hand gesture classification. We leverage the Randomized Ablation technique [18] from the Adversarial Machine Learning literature, which was proposed to defend against sparse adversarial attacks, where some pixels of an RGB image are adversarially perturbed.

In this paper, we generalize Randomized Ablation to multi-channel and multimodal settings, which we call Random Channel Ablation (RChA). We apply this to biosignal processing, with a focus on robust hand gesture prediction in the presence of missing channels. We use multimodal ultrasound and FMG data for the classification of 12 hand gestures with a convolutional neural network (CNN). Section II describes the data acquisition, model, gesture classification workflow, and random channel ablation technique. Section III describes the experimental setup for data acquisition, processing, experimental design, and evaluation metrics. Finally, the results and discussion sections describe the obtained results and major takeaways in addition to future research directions stemming from this work.

II. METHODS AND EXPERIMENTAL DESIGN

Ultrasound and FMG data were simultaneously acquired from 2 subjects. The data acquisition procedure and the 12 hand gestures were explained to the subjects. The study

---

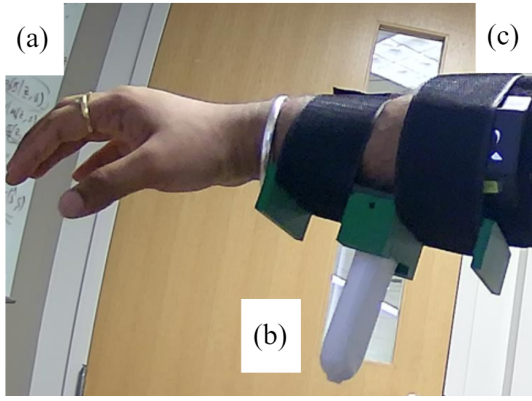*This work was performed while K.B. was an intern at MERL.

Fig. 1. Data acquisition: (a) Hand gesture; (b) Linear ultrasound probe strapped to the forearm using a custom-designed attachment; (c) BioX FMG armband strapped to the forearm.

was approved by the institutional research ethics committee at Mitsubishi Electric Research Laboratories (IRB reference number 23001), and written informed consent was given by the subjects before the data acquisition.

A. Data Acquisition

The ultrasound data was acquired using a Sonostar 4L linear palm Doppler ultrasound probe. A custom-designed 3D-printed wearable was strapped onto the subject's forearm. The data from the probe was streamed to a Windows system over Wi-Fi, and screenshots of the ultrasound images were captured using a custom Python script. The 4L linear probe has 80 channels of ultrasound data, and the post-processed beamformed B-mode data is obtained, from which 350×350-pixel image data is acquired. The FMG data was acquired using a BioX AAL-Band 2.0. The data were streamed to the Windows system over Bluetooth, and saved along with the ultrasound frames using the Python script to ensure that the ultrasound and FMG data were synchronized. The FMG band gives 8 channels of FMG data. Fig. 1 shows the data acquisition from the ultrasound and FMG.

The data were acquired for 12 hand gestures selected based on activities of daily living, as described in detail in [19]. These hand gestures consist of 4 finger flexions, 4 pinch configurations, hand wolf, fist, hook, and open hand gestures. 100 frames of data were acquired while the subject had a fixed hand gesture Per subject, 24,000 frames of data were used for the study which resulted from 20 sessions of data acquisition (20 sessions, 12 hand gestures, 100 frames per gesture: 24,000 frames). The average frame rate of data acquisition was 18 Hz. The experiments were run on a desktop with an Intel i7-13700K CPU, 128GB RAM, and an NVIDIA GeForce RTX 4090 GPU.

B. Model Architecture, Parameters and Metrics

A CNN was used in this paper designed using PyTorch based on a network used to classify hand gestures from ultrasound images from the forearm [2], [19]. The network is tailored for image compression, and its architecture is designed
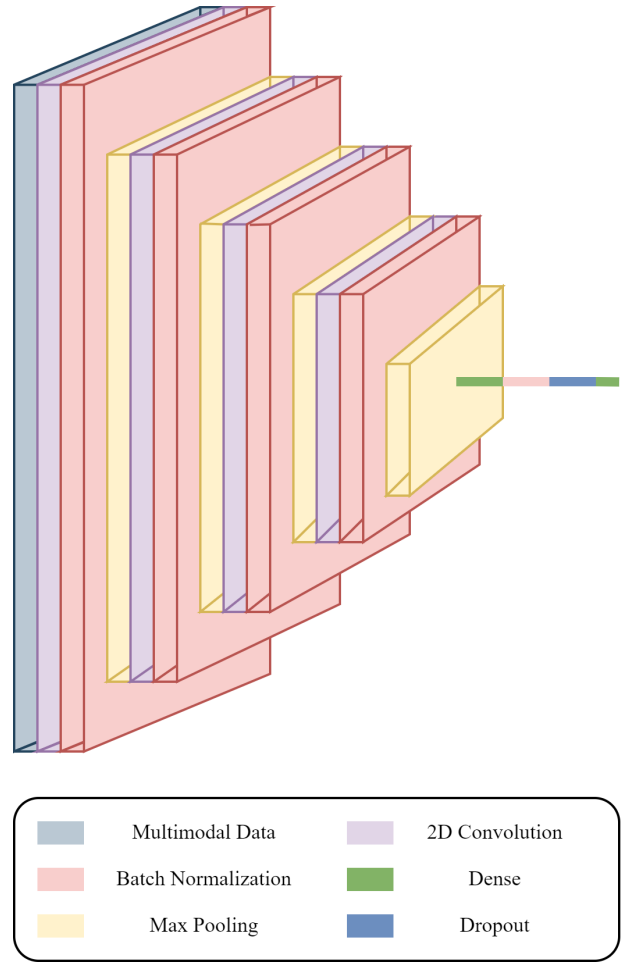

Fig. 2. The model architecture to classify hand gestures from multimodal data.

to capture hierarchical features through convolutional and pooling operations. The network has 4 cascaded convolution layers followed by batch normalization and pooling layers compared to the referenced architecture which has 5 cascaded convolution layers and a different number of parameters per layer because of different input dimensions. The 2-dimensional convolution layers have 16 channels as the output channels, and a kernel size of (3, 3), with a stride of 1. These are followed by a fully connected linear layer which flattens out the output from the fourth max-pooling layer. After another batch normalization, a dropout operation is applied with dropout probability of 0.5. This is followed by a final fully connected layer with a softmax activation which leads to the probabilities of the 12 hand gestures based on the input data. Each convolution layer has a rectified linear unit (ReLU) activation. The first fully connected (dense) layer has a ReLU activation, while the output layer has a softmax activation. The network is visualized in Fig. 2.

Cross Entropy Loss was used for evaluating the loss and the Adam optimizer was used for optimizing the network. A learning rate of 0.0001 and a batch size of 10 were used. The
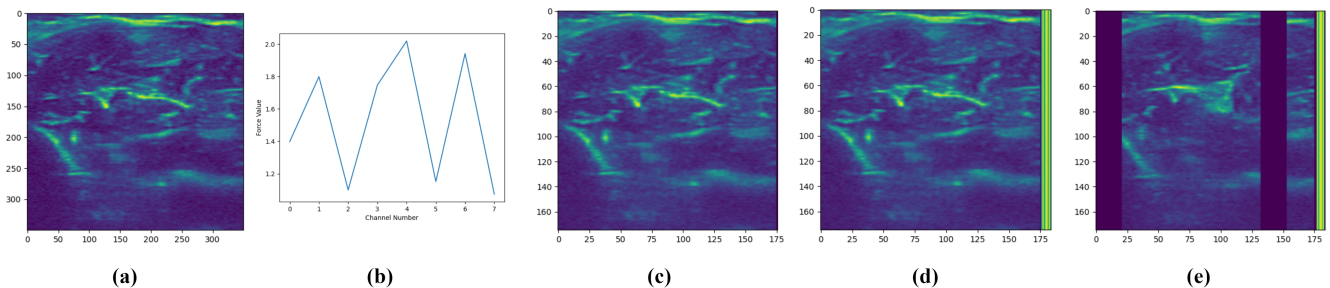
Fig. 3. Data pre-processing and ablation. (a) The 350×350-pixel forearm ultrasound image was acquired using a linear ultrasound probe. (b) 8-channel FMG data was simultaneously acquired. (c) The ultrasound image was downsized by a factor of 2. The new image dimensions were 175×176. An additional column of zeros was added to make the column width divisible by 8. (d) Ultrasound and FMG data were independently normalized and then appended per frame. The FMG data was expanded to span the rows of the ultrasound image. The dimensions of each sample were 175×184 with the last 8 columns belonging to FMG and the remaining to the downsized ultrasound image. (e) The output after random channel ablation with channels 0, 6, 8, and 15 ablated. First, the number of channels to be ablated was chosen randomly which turned out to be 4. Then, 4 channels were randomly chosen and ablated from the 16 channels.

evaluation was done for three different random seed values: 0, 1, and 11, and the results were averaged over these seeds. Classification accuracy percentage (Acc) was used as the accuracy metric, which is the fraction of correct classifications over the total number of classifications multiplied by 100.

5-fold cross-validation was used to assess the robustness and generalization performance of our method. The dataset was partitioned into five subsets, training the model on four of these subsets, and evaluating its performance on the remaining subset. This meant, that per subject and fold, the training data had 19,200 samples and the testing data had 4,800 samples. This process was repeated five times, each time using a different subset for evaluation. Throughout training and evaluation, careful attention was given to maintaining temporal independence and preventing any overlap between the train and test sets. The results were averaged to provide a more reliable estimate of the model's performance and reduce the impact of variability in a single train-test split.

C. Data Preprocessing

Fig. 3 shows the data pre-processing for the ultrasound and FMG data. The 350×350-pixel ultrasound data and 1×8 FMG data are obtained per frame. The ultrasound image is downsized by a factor of 2. Let I be the original image with dimensions $d \times d$. The downsized image, $I_{\text{downsized}}$, with dimensions $\frac{d}{2} \times \frac{d}{2}$, is expressed as follows:

$$I_{\text{downsized}}(i,j) = I(2i, 2j). \tag{1}$$

Here, i and j are pixel indices in the downsized image, and $2i$ and $2j$ are the corresponding pixel indices in the original image. The downsized image is shown in Fig. 3(c). The dimension of the downsized ultrasound image is 175×175 pixels. A zeros column was added to the downsized image to make the columns perfectly divisible by 8, with the updated dimensions being 175×176 pixels.

To have consistency in the FMG and ultrasound data, the data from both modalities were independently normalized. Following this, the FMG data was expanded to have the same number of rows as the ultrasound data and then appended

to the ultrasound data as shown in Fig. 3(d). The dimension of this image is 175×184 pixels with the last 8 columns representing data obtained from FMG and the remaining 176 columns belonging to the ultrasound image.

D. Robust Training for Hand Gesture Classification

It has been previously shown that randomly ablating image pixels can help make a classifier robust to sparse adversarial perturbations that only corrupt very few number of image pixels [18]. This is because the fraction of corrupted pixels is very small and the vast majority of the pixels are benign. There is a high probability that the randomly retained (non-ablated) pixels are not corrupted. Further, it is possible to recognize images with partial observation (i.e., retained pixels). If a classifier is trained deliberately with randomly ablated information (i.e., partial observation), using the Randomized Ablation strategy can be certifiably robust to such sparse adversarial perturbations. In our work, instead of an adversary, we only consider randomly missing channels, and our robust classifier is trained on randomly observed/retained channels.

Algorithm 1 describes the robust training procedure of the proposed Random Channel Ablation method. The pre-processed data is first combined for both modalities. The maximum number of ablated channels K, the batch size M for training, and the learning rate η are specified. Each batch $X_{\text{batch}}, \boldsymbol{y}_{\text{batch}}$ in $X_{\text{train}}, \boldsymbol{y}_{\text{train}}$ is then used for training the neural network and its parameters $\boldsymbol{\theta}$ are then updated. The parameters are updated based on the gradient $\nabla_{\boldsymbol{\theta}}$ of the loss function w.r.t. the model parameters $\boldsymbol{\theta}$. The loss function $\ell(f_{\boldsymbol{\theta}}(X_{\text{ablate}}), \boldsymbol{y}_{\text{batch}})$ is based on the loss obtained from the predictions $f_{\boldsymbol{\theta}}(X_{\text{ablate}})$ and the true labels $\boldsymbol{y}_{batch}$.

Fig. 3(e) shows the multimodal data sample after randomly ablating 4 channels. First, the number of channels to be ablated was set as 4. Then, 4 channels were randomly chosen and ablated from the total 16 channels. For FMG, each column corresponds to one channel, and for ultrasound, each channel is 22-pixel wide column. This was done so that the ablation was consistent across the modalities.

Algorithm 1 Robust training via Random Channel Ablation of T epochs. The neural network f is parameterized by $\boldsymbol{\theta}$.

1: Input $X_{\text{train}} = \text{append}(X_{\text{us}}, X_{\text{fmg}})$, $\boldsymbol{y}_{\text{train}}$, max # of ablated channels K, batch size M, learning rate η
2: **for** $t = 1, \ldots, T$ **do**
3:     **for** $\{X_{\text{batch}}, \boldsymbol{y}_{\text{batch}}\}$ in $\{X_{\text{train}}, \boldsymbol{y}_{\text{train}}\}$ **do**
4:         $X_{\text{ablate}} \leftarrow \text{ablate}(X_{\text{batch}}, K)$
5:         $\boldsymbol{\theta} = \boldsymbol{\theta} - \eta \nabla_{\boldsymbol{\theta}} \ell(f_{\boldsymbol{\theta}}(X_{\text{ablate}}), \boldsymbol{y}_{\text{batch}})$
6:     **end for**
7: **end for**
8: **return** neural network parameters $\boldsymbol{\theta}$
9:
10: **Subroutine** $X_{\text{ablate}} \leftarrow \text{ablate}(X_{\text{batch}}, K)$:
11: Initialize $X_{\text{ablate}} = X_{\text{batch}}$
12: Randomly choose # of ablated channels $k \in [0, \ldots, K]$
13: **if** $k > 0$, **then** randomly choose k channels of X_{ablate} and set them to zero
14: **end if**

TABLE I
CLASSIFICATION ACCURACY FOR DIFFERENT METHODS WHEN SAMPLES ARE RANDOMLY MISSING UP TO 25% CHANNELS

	Baseline	Imputation	Proposed	Oracle
Subj 1	70.0%	81.9%	**86.2%**	87.3%
Subj 2	87.3%	**95.6%**	95.5%	97.1%
Average	78.7%	88.8%	**90.9%**	92.2%

E. Evaluation

During testing, the index of missing channels is chosen uniformly at random for each sample. For missing channels, the proposed method fills zero to the missing channels, and then applies robust classifier trained in Algorithm 1 for prediction.

The proposed method is compared with the Baseline method, Imputation method, and Oracle, where the classifier of these 3 methods is trained on all channels. The Oracle has all channels available during testing as well, and its performance can be viewed as an upper bound for the missing-channel case. To deal with missing channels at testing, the Baseline method simply fills in with zero values. While for the Imputation method, the missing channel is imputed using the overall sample mean of that channel from the training samples.

III. RESULTS

Upon averaging the hand gesture classification results over the two subjects, the average Oracle case hand gesture classification accuracy was 92.2%. Table I shows the gesture classification accuracy for the Baseline, Imputation, Proposed, and Oracle methods when the fraction of missing channels is up to 25%. Since the multimodal data comprises 16 channels, this leads to a maximum of 4 channels being missing (i.e., the number of missing channels of each testing sample is randomly chosen from 1 to 4). On average, our proposed method's classification accuracy was 90.9%, compared to 88.8% for the Imputation method and 78.7% for the Baseline method.

TABLE II
CLASSIFICATION ACCURACY FOR DIFFERENT METHODS WHEN SAMPLES ARE RANDOMLY MISSING UP TO 50% CHANNELS

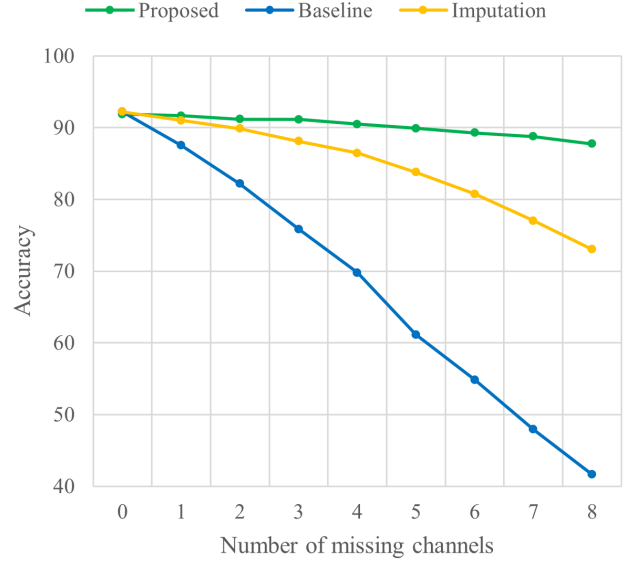|	Baseline	Imputation	Proposed	Oracle
Subj 1	56.1%	75.1%	**83.8%**	87.3%
Subj 2	73.9%	92.2%	**95.1%**	97.1%
Average	65.0%	83.7%	**89.5%**	92.2%

Fig. 4. Classification accuracy for a fixed number of channels being missing.

Table II shows the gesture classification accuracy for the Baseline, Imputation, Proposed, and Oracle methods when the fraction of missing channels is up to 50%. This leads to a maximum of 8 channels being missing. On average, our proposed method's classification accuracy was 89.5% compared to 83.7% for the Imputation method and 65.0% for the Baseline method.

Fig. 4 shows the results where a fixed number of channels are missing per sample for the different methods. For 4 missing channels, on average, our proposed method's classification accuracy was 90.5% compared to 86.5% for the Imputation method and 69.8% for the Baseline method. For 8 missing channels, on average, our proposed method's classification accuracy was 87.7% compared to 73.0% for the Imputation method and 41.6% for the Baseline method. It is also worth noting that, when there is no missing channel, our robustly trained classifier performs almost the same as the Baseline method (which is also the Oracle method since there are no missing channels).

IV. DISCUSSION

In this section, we discuss the effect of random channel ablation, and the effect of the number of missing channels on the classification performance. We also highlight the advantages of our method in making it robust to missing channels and some future research directions stemming from this work.

A. Effect of Random Channel Ablation

As can be seen in both Tables I and II, the proposed method outperforms the Baseline and the Imputation methods. For a maximum of 4 (25%) channels being missing, there is an improvement of 12.2% compared to the Baseline method, and 2.1% compared to the Imputation method. The classification accuracy is 1.3% lower than the Oracle case, which is expected since the proposed method only has partial observation of the channels during testing. Similar trends were observed with a maximum of 8 (50%) channels being missing with an improvement of 24.5% from the Baseline method and 5.8% from the Imputation method. The classification accuracy for the proposed method is only 2.7% less than the Oracle case. This shows that random channel ablation can help make the classifier robust to missing channels, while still maintaining very high utility.

B. Effect of Number of Missing Channels

With a higher number of missing channels in the test set, it is expected to see a significant drop in classification accuracy, as can be seen for the Baseline method in Fig. 4. For 4 channels missing, the proposed method's accuracy is 20.7% better than the Baseline method and 4.0% better than the Imputation method. For 8 channels missing, the proposed method's accuracy is 46.1% better than the Baseline method and 14.7% better than the Imputation method. Overall, the drop in the classification accuracy from 0 channels missing to 8 channels missing is a mere 4.1% for the proposed method compared to 19.2% for the Imputation method and 50.6% for the Baseline method, demonstrating clearly superior performance compared to the other methods.

C. A Universal All-in-One Classifier

For the proposed random channel ablation method, only one universal classifier needs to be trained, and it can be applied to all situations with up to K out of n channels missing. In contrast, one could train a specific classifier for the specific missing-channel situation. For example, one could train a specific classifier for the situation where only the second channel is missing. However, this would scale poorly when the number of total channels n and the maximum number of missing channels K is large, since one would have to train $\sum_{k=0}^{K} \binom{n}{k}$ classifiers in total, which would not be practical.

D. Future Research

This work focused on showcasing the effectiveness of our random channel ablation technique for multimodal and multi-channel biosignals through ultrasound and FMG modalities. For future work, we aim to diversify the data acquisition by incorporating additional modalities, such as surface electromyography, inertial measurement unit gloves, and cameras, among others. Considering the temporal aspects of the input data can be used to improve the classification of hand gestures, and it can also help leverage some additional modalities in a multimodal setting such as sEMG which has been proven to

deliver good gesture classification performance [20]. For merging data from multimodal ultrasound and FMG information, alternative learning based approaches can be explored such as in [21] and [22]. Within the scope of the system described in the paper, these approaches would entail learning a low-dimensional embedding of the ultrasound data which would then be fused with the 1D FMG signal.

Additionally, the current study focused on data acquired from 2 subjects. To ensure the generalizability and applicability of multimodal and multi-channel biosignal based hand gesture recognition using our proposed random channel ablation approach, future work will focus on acquiring data from multiple subjects. The expanded set of subjects will enable an evaluation of the proposed approach for subjects with varying characteristics, making it more versatile.

V. Conclusion

In this study, we introduced our random channel ablation technique aimed at enhancing the robustness of deep learning classifiers used for hand gesture recognition when faced with missing channels in multimodal and multi-channel biosignal data. Our approach demonstrated notable resilience to up to 50% missing channels, surpassing the performance of the Baseline and Imputation methods. We observed minimal degradation in classifier performance compared to the Oracle case. The efficacy of our method unveils promising prospects for its application in diverse multimodal and multi-channel biosignal training scenarios. This technique holds practical implications for real-world data acquisition and deep learning model performance, particularly in scenarios where certain channels may be absent. We illustrated the capability to bolster model robustness which can be utilized for several applications and across several modalities.

VI. Acknowledgments

References

[1] M. Zheng, M. S. Crouch, and M. S. Eggleston, "Surface electromyography as a natural human–machine interface: A review," *IEEE Sensors Journal*, vol. 22, no. 10, pp. 9198–9214, 2022.

[2] K. Bimbraw, C. J. Nycz, M. J. Schueler, Z. Zhang, and H. K. Zhang, "Prediction of metacarpophalangeal joint angles and classification of hand configurations based on ultrasound imaging of the forearm," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 91–97.

[3] X. Jiang, L.-K. Merhi, and C. Menon, "Force exertion affects grasp classification using force myography," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 2, pp. 219–226, 2018.

[4] D. Li, P. Kang, K. Zhu, J. Li, and P. B. Shull, "Feasibility of wearable PPG for simultaneous hand gesture and force level classification," *IEEE Sensors Journal*, vol. 23, no. 6, pp. 6008–6017, 2023.

[5] M.-K. Liu, Y.-T. Lin, Z.-W. Qiu, C.-K. Kuo, and C.-K. Wu, "Hand gesture recognition by a MMG-based wearable device," *IEEE Sensors Journal*, vol. 20, no. 24, pp. 14 703–14 712, 2020.

[6] O. Makaussov, M. Krassavin, M. Zhabinets, and S. Fazli, "A low-cost, IMU-based real-time on device gesture recognition glove," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2020, pp. 3346–3351.

[7] B. Ben Atitallah, Z. Hu, D. Bouchaala, M. A. Hussain, A. Ismail, N. Derbel, and O. Kanoun, "Hand sign recognition system based on EIT imaging and robust CNN classification," *IEEE Sensors Journal*, vol. 22, no. 2, pp. 1729–1737, 2022.

[8] K. Bimbraw, J. Rothenberg, and H. Zhang, "Leveraging ultrasound sensing for virtual object manipulation in immersive environments," in *2023 IEEE 19th International Conference on Body Sensor Networks (BSN)*, 2023, pp. 1–4.

[9] K. Bimbraw, E. Fox, G. Weinberg, and F. L. Hammond, "Towards sonomyography-based real-time control of powered prosthesis grasp synergies," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 4753–4757.

[10] K. Bimbraw and H. K. Zhang, "Estimating force exerted by the fingers based on forearm ultrasound," in *2023 IEEE International Ultrasonics Symposium (IUS)*, 2023, pp. 1–4.

[11] N. Ha, G. P. Withanachchi, and Y. Yihun, "Performance of forearm FMG for estimating hand gestures and prosthetic hand control," *Journal of Bionic Engineering*, vol. 16, pp. 88–98, 2019.

[12] Q. Gao, J. Liu, and Z. Ju, "Hand gesture recognition using multimodal data fusion and multiscale parallel convolutional neural network for human–robot interaction," *Expert Systems*, vol. 38, no. 5, p. e12490, 2021.

[13] S. Wei, Y. Zhang, and H. Liu, "A multimodal multilevel converged attention network for hand gesture recognition with hybrid sEMG and A-mode ultrasound sensing," *IEEE Transactions on Cybernetics*, 2022.

[14] S. Jiang, Q. Gao, H. Liu, and P. B. Shull, "A novel, co-located EMG-FMG-sensing wearable armband for hand gesture recognition," *Sensors and Actuators A: Physical*, vol. 301, p. 111738, 2020.

[15] M. Abavisani, H. R. V. Joze, and V. M. Patel, "Improving the performance of unimodal dynamic hand-gesture recognition with multimodal training," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 1165–1174.

[16] R. Tchantchane, H. Zhou, S. Zhang, and G. Alici, "A review of hand gesture recognition systems based on noninvasive wearable sensors," *Advanced Intelligent Systems*, vol. 5, no. 10, p. 2300207, 2023.

[17] A. R. T. Donders, G. J. van der Heijden, T. Stijnen, and K. G. Moons, "Review: A gentle introduction to imputation of missing values," *Journal of Clinical Epidemiology*, vol. 59, no. 10, pp. 1087–1091, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0895435606001971

[18] A. Levine and S. Feizi, "Robustness certificates for sparse adversarial attacks by randomized ablation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 4585–4593.

[19] K. Bimbraw, C. J. Nycz, M. Schueler, Z. Zhang, and H. K. Zhang, "Simultaneous estimation of hand configurations and finger joint angles using forearm ultrasound," *IEEE Transactions on Medical Robotics and Bionics*, vol. 5, no. 1, pp. 120–132, 2023.

[20] M. Simão, P. Neto, and O. Gibaru, "Emg-based online classification of gestures with recurrent neural networks," *Pattern Recognition Letters*, vol. 128, pp. 45–51, 2019.

[21] C. Li, Y. Hou, W. Li, Z. Ding, and P. Wang, "Dfn: A deep fusion network for flexible single and multi-modal action recognition," *Expert Systems with Applications*, vol. 245, p. 123145, 2024.

[22] T. Yang, T. Xu, Y. Cheng, Z. Tang, S. Su, and Y. Cao, "A fusion method based on 1d vibration signals and 2d images for detection of railway surface defects," in *2023 3rd International Conference on Neural Networks, Information and Communication Engineering (NNICE)*. IEEE, 2023, pp. 282–286.