Multi-Grasp Classification for the Control of Robot Hands Employing Transformers and Lightmyography Signals

Ricardo V. Godoy[1], Bonnie Guan[1], Anany Dwivedi[2], Mojtaba Shahmohammadi[1], Mahonri Owen[3], and Minas Liarokapis[1]

***Abstract*— The increasing use of smart technical devices in our everyday lives has necessitated the use of muscle-machine interfaces (MuMI) that are intuitive and that can facilitate immersive interactions with these devices. The most common method to develop MuMIs is using Electromyography (EMG) based signals. However, due to several drawbacks of EMG-based interfaces, alternative methods to develop MuMI are being explored. In our previous work, we presented a new MuMI called Lightmyography (LMG), which achieved outstanding results compared to a classic EMG-based interface in a five-gesture classification task. In this study, we extend our previous work experimentally validating the efficiency of the LMG armband in classifying thirty-two different gestures from six participants using a deep learning technique called Temporal Multi-Channel Vision Transformers (TMC-ViT). The efficiency of the proposed model was assessed using accuracy. Moreover, two different undersampling techniques are compared. The proposed thirty-two-gesture classifiers achieve accuracies as high as 92%. Finally, we employ the LMG interface in the real-time control of a robotic hand using ten different gestures, successfully reproducing several grasp types from taxonomy grasps presented in the literature.**

I. Introduction

With the advancements in technology over the past decades, intelligent electronic and robotic devices are becoming ubiquitous. Such devices have integrated into our lives seamlessly, be it a mobile phone to connect with the world, a small vacuum cleaning robot to clean our houses, or autonomous cars to travel from one place to another [1], [2]. Robotic devices in particular have even found applications in search and rescue, monitoring, maintenance, and disaster response in hard to reach and dangerous environments including applications in deep-space and deep-sea explorations [3], [4], [5]. With the increased use of such technical devices in our everyday life, we need improved interfaces which facilitate a more intuitive interaction. The traditional interfaces (e.g., keyboards, mice, joysticks, vision-based systems etc.) can be bulky, fatiguing, and difficult to use [6]. while simple to use, vision-based systems may suffer from poor lighting conditions or occlusions.

[1]Ricardo V. Godoy, Bonnie Guan, Mojtaba Shahmohammadi, and Minas Liarokapis are with the New Dexterity Research Group, Department of Mechanical and Mechatronics Engineering, The University of Auckland, Auckland, New Zealand {rdeg264, bgua324, msha851}@aucklanduni.ac.nz, minas.liarokapis@auckland.ac.nz

[2]Anany Dwivedi is with the Chair of Autonomous Systems and Mechatronics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany anany.dwivedi@fau.de

[3]Mahonri Owen is with the School of Engineering, The University of Waikato, Hamilton, New Zealand mahonri.owen@waikato.ac.nz

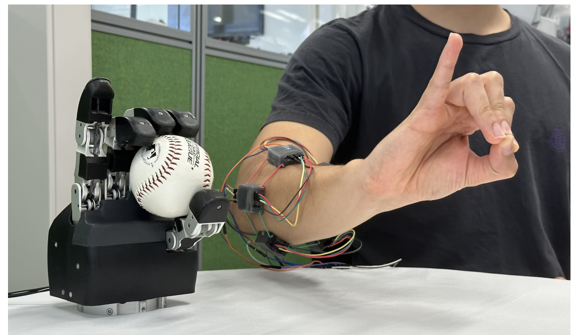

Fig. 1. Real-time robot hand teleoperation using the LMG armband and machine learning techniques to classify the gesture and grasp an object.

These issues can be addressed by using Muscle Machine Interfaces (MuMIs). MuMIs decode human intentions by utilizing data capturing the activity of the user's muscles employing different modalities, and they allow for immersive interaction with technical devices. The most common method for developing MuMIs is employing Electromyography (EMG) based methods for decoding motion, using interfaces that are non-invasive in nature [7]. Such interfaces have found applications in various fields, including but not limited to teleoperation of robotic arm-hand systems [8], [9], rehabilitation using robotic exoskeletons [10], entertainment (myo-games) [11], and muscle computer interfaces [12].

Even though EMG-based interfaces are very popular, Electromyography has several drawbacks. EMG signals are known to be non-stationary, prone to cross-talk between different neighboring muscles, sensitive to electrode shifts during their use, sweating, fatigue, and electro-magnetic noise [13]. Moreover, they can only be used to measure the contractions of surface muscles, since to measure the activity of the deep-seated muscles, the invasive variant of EMG is needed [14] that is much more tricky and difficult to use. To overcome these issues, researchers have started exploring alternate myography methods (e.g., mechanomyography, ultrasonography, near infra-red spectroscopy etc.) [15]. However, these methods have their own drawbacks - mechanomyography-based signals are affected by ambient acoustic/vibrational noise [16], ultrasonography-based interfaces are generally bulky and expensive, and their probe needs to be frequently gelled for proper functioning, while near infra-red spectroscopy-based signals have a delayed response to the motion of the user, leading to a low temporal resolution [17].

In our previous work, we proposed a novel method to develop MuMIs called Lightmyography (LMG) [18]. LMG offers a low-cost, portable, easy-to-use alternative for MuMIs. For decoding human intention and motion, the LMG modules were assembled in an armband (see Fig. 1). The armband LMG signals were employed by three machine learning techniques to decode five gestures identified by Bullock et al. [19] as frequent grasps in daily life to validate the performance of the armband. The performance of the classifiers was evaluated, and the use of LMG signals as input proved to be able to train machine learning models to decode five different hand gestures with high performance.

In this paper, we further assess the efficiency of the LMG armband in providing appropriate signals for classifying thirty-two hand gestures employing deep learning techniques. To achieve this, a deep learning technique called Temporal Multi-Channel Vision Transformer (TMC-ViT) [20] was employed to develop classification models. The results were validated using a 5-fold cross-validation method, and the decoding performance of each fold is presented. Confusion matrices are employed to understand the decoding models better while predicting each gesture. After the offline analysis, these trained models were employed to decode the LMG signals to teleoperate a robotic hand during real-time operation. In this experiment, we decoded ten of the thirty-two gestures, which were translated to a robotic hand, reproducing most of the grasps of the taxonomy proposed by Feix et al. [21].

The rest of the paper is organized as follows: Section II presents the methods and apparatus used to collect data and teleoperate the robotic hand, Section III explains how the experiments were conducted and how the models were trained and evaluated. Section IV presents the results of this study and Section V concludes the paper.

II. Methods

A. Detection of Gestures with Lightmyography

LMG utilizes the propagation of light waves through silicone mediums and into the skin where it gets reflected back. Light of different wavelengths are used. The reflected light waves are detected with a photodiode and based on its intensity, the displacement of the skin and deformation of tissue can be used for the decoding of human motion or intention. In our previous study [18], high prediction accuracies that were comparable to Electromyography based predictions were observed using an LMG armband to classify five gestures. This armband consisted of five sensing modules placed around the upper forearm where a majority of the muscles responsible for the movements of the finger digits are located (extensor digitorum, flexor digitorum superficialis, flexor digitorum profundus, and flexor pollicis longus) [22]. Each module uses two wavelengths of light to achieve different levels of skin penetration.

In this study, to test the capabilities of the armband, thirty-two gestures were performed by each participant. The gestures include every five-finger combination, where each finger can either be closed or open. These thirty-two gestures were chosen for easy mapping onto the NDX-A* hand [23]. LMG data was collected while the participants executed each gesture. Then, the data was used to train one deep learning model for each participant. The accuracy achieved by these models in the task of classifying thirty-two gestures is shown in Section IV.

B. Online Robotic Hand Teleoperation

In order to assess the real-time capabilities of the LMG armband, ten of the thirty-two grasps previously collected were used to train a deep learning model to control the NDX-A* hand. These ten grasps, shown in Fig. 2, were selected so the hand could perform most of the taxonomy grasps proposed by Feix et al, as explained in Table I. A single participant used the trained model for real-time teleoperation of the NDX-A* robot hand. A flowchart detailing this process is shown in Fig. 3. The NDX-A* hand uses six motors - one for each of the five fingers, and one for thumb opposition. For simplicity, thumb opposition was not considered. The trained model used a sliding window of thirteen samples and provided as an output the likelihood of each of the ten gestures being the one performed by the user. A maximum function was applied to find the predicted gesture. Finally, a moving average filter with a window of ten samples was used to determine which gesture would be sent to the robot hand controller which would lead to a light delay in response time.

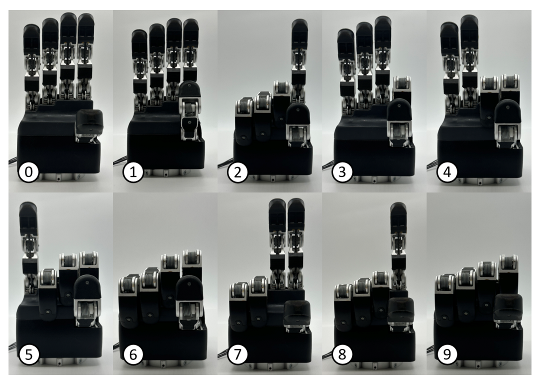

Fig. 2. Ten gestures shown on the NDX-A* robot hand that were selected to perform most of the taxonomy grasps proposed by Feix et al. The grasp #6, for example, is the same combination of fingers opened and closed used for both large-diameter and small-diameter grasps, depending on the object characteristics such as size and shape.

III. Experiments

In order to experimentally validate the performance of the LMG armband and the proposed transformer-based classifiers, we performed a set of experiments to collect data from six participants, where each participant was asked to perform thirty-two different gestures. We then performed an offline LMG-based gesture classification using the acquired data. Furthermore, the offline decoding accuracy of ten gestures based on grasp taxonomy proposed by Feix et al. [21] was also assessed. This was later employed in real-time teleoperation of the NDX-A* robot hand.

TABLE I

THE TEN GESTURES COLLECTED AND THE RESPECTIVE TAXONOMY GRASPS THE ROBOTIC HAND CAN REPRODUCE BY USING THEM.

Gesture	Taxonomy grasp
0	Rest
1	Parallel extension
2	Ventral
3	Ring, palmar pinch, tip pinch, inferior pincer
4	Sphere 3 finger, tripod variation. prismatic 2 finger, tripod, writing tripod, lateral tripod
5	Sphere 4 finger, prismatic three finger, quadpod
6	Large diameter, small diameter, medium wrap, power disk, power sphere, extension type, distal type, prismatic 4 fingers, precision disk, precision sphere
7	Adduction grip
8	Index finger extension
9	Adducted thumb, light tool, fixed hook, palmar, lateral, stick

A. Data collection

The data was collected from six able-bodied participants, with each using their dominant hand. The details regarding the participants can be found in Table II. The study was approved by the University of Auckland Human Participants Ethics Committee (UAHPEC), reference number #019043. All experiments were performed in accordance with relevant guidelines and regulations. Prior to the study, participants provided a written and informed consent to the experimental procedures.

All the participants were asked to perform thirty-two different gestures in a single session so that the LMG module placement was the same between all gestures. For each gesture, participants started with 15 seconds of rest, during which the hand was completely relaxed, followed by 15 seconds of gesture execution. This procedure was repeated four more times for each gesture. Visual cues were provided to the participants as a three-second counter on a computer screen to switch between the gesture state and the rest state. This included a software trigger to label the two states for creating the ground truth data for supervised learning algorithms. As the participant performs these gestures, data is collected using the LMG band at a frequency of 66Hz. Due to the long sessions, participants were asked between each gesture if they were fatigued. An appropriate break was given, with the armband on, if any fatigue was experienced. Additionally, where fingers were coupled, especially with the pinky and ring fingers, participants were asked not to strain their hands while keeping their fingers open. Instead, participants were asked to open those fingers to the best of their ability within comfort.

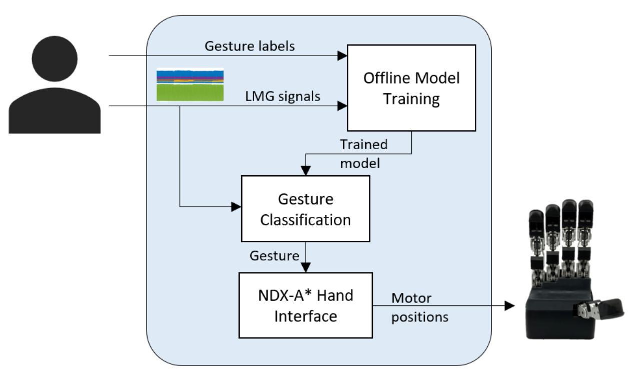

Fig. 3. Block diagram illustrating the real-time control of the robotic hand. The LMG data is collected from the participant during gesture execution. This data is used to train a deep learning model, which will be used later on to control the robotic hand in real-time using the raw LMG signals of the user as input.

TABLE II

INFORMATION OF THE PARTICIPANTS.

Subj.	**1**	**2**	**3**	**4**	**5**	**6**
Age	23	30	27	32	34	25
Sex	female	male	male	female	male	male
All participants were right hand dominant						

B. Data processing

In this section, we present the data preprocessing method employed for the LMG-based gesture classification.

1) Sliding window: A sliding window of 200 ms with a stride of 20 ms was employed to feed the training data to the supervised deep learning models. The sliding window was chosen to be larger than 125 ms to avoid high biases and variance [24], and smaller than 300 ms due to real-time constraints of prosthetic control systems [25].

2) Data balance: As each repetition of a gesture was followed by the rest gesture, after thirty-two gestures the rest class was predominant at the end of the data collection. Therefore, the data was balanced to guarantee that we have the same number of samples for each gesture to avoid bias toward any particular class. To do so, we employed two different techniques and compared their performances. The selected undersampling techniques were random undersampling and the near-miss undersampling technique based on [26]. One model was trained using each technique. The method that presents the best classification accuracy is selected and shown in Section IV.

3) Data type: The deep learning models were trained using raw LMG data without any feature engineering. LMG signals can be used with minimal preprocessing, unlike EMG which needs signal filtering and feature extration steps. Moreover, we rely on the ability of deep learning methods to automatically learn discriminative features from raw data, even with noisy signals. The use of batch normalization layers within the deep learning models incorporates normalization into the architecture, increasing training speed and eliminating the need for normalization during preprocessing steps [27].

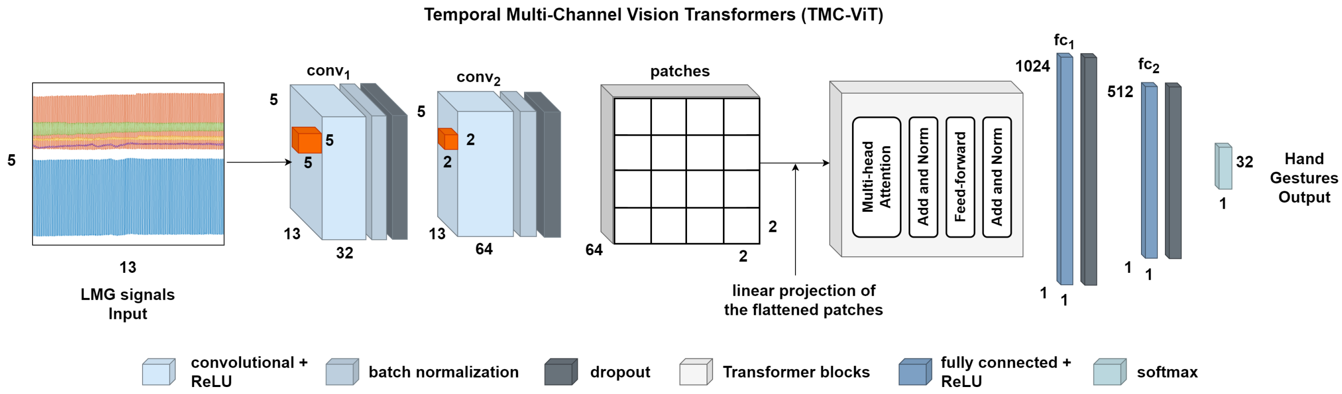

Fig. 4. TMC-ViT Model for raw LMG data. Two convolution layers extract the embeddings with an embedding size of sixty-four, which are further split into patches of 2×2 and provided to the Transformer block. A final softmax layer with thirty-two neurons outputs the decoded gesture.

C. Classification models

In order to decode the hand gestures, we trained classifiers using a novel deep learning technique called TMC-ViT. This technique was chosen based on evidence from our previous works which showed that the TMC-ViT outperformed the well-established machine learning techniques such as convolutional neural networks and random forests [18], [28]. The TMC-ViT is a Transformer-based model that adapts the Vision Transformer [29] to process temporal data with multiple channels, e.g., LMG signals as input, employing convolutional to extract its embeddings.

Two convolutional layers are used before the data is supplied to a ViT that extracts 2×2 patches and provides the output to a Transformer encoder composed of four Multi-head Attention layers [30] with four heads each. The ViT processes the multi-channel signals with a linear projection of the flattened patches, whose components indicate low-dimensional correlations in the patches, and the Multi-Head Attention mechanism aggregates signal information across all layers. The TMC-ViT is shown in Fig. 4.

D. Training and evaluation

Our models were developed in Python using Tensorflow and Keras. The models were trained and evaluated on an Alienware computer equipped with an i9-11900F processor, 126GiB memory, and NVIDIA GeForce RTX 3090 GPU. The subject-specific classifiers were trained and validated using the 5-fold cross-validation method. During the training of each deep learning model, the loss function was the sparse categorical cross-entropy. Optimization was performed using Adam [31], with the efficiency of the trained model being assessed using accuracy. One model was trained for each participant.

IV. Results

In this section, we discuss the findings from the experiments conducted with the LMG armband. More precisely, we present the offline gesture decoding accuracies of the models trained with the data collected by the LMG armband for a thirty-two gesture set and a ten-gesture set for six participants. We also evaluate the performance of the trained models in decoding a variety of hand gestures in real-time to teleoperate a robotic hand.

A. Hand gesture decoding using LMG

The gesture classification accuracies of the subject-specific models trained with the data collected with the LMG armband for decoding thirty-two gestures are presented in Table III.

TABLE III
DECODING ACCURACY FOR THE THIRTY-TWO GESTURES. *SD* STANDS FOR STANDARD DEVIATION.

Subj.	Undersampling Technique	Accuracy (%)						
		Fold					AVG	SD
		1	2	3	4	5		
1	Near-miss	84	97	94	93	90	**92**	**5**
2	Random	90	95	94	91	89	**92**	**3**
3	Random	86	95	94	90	78	**89**	**7**
4	Random	73	92	86	90	86	**85**	**7**
5	Near-miss	75	90	88	87	74	**83**	**8**
6	Random	85	93	95	90	89	**90**	**4**

Table III shows that the deep learning models can decode the thirty-two gestures with high overall accuracy for all the participants, with an average accuracy as high as 92%. It can be noticed that the first fold presents a lower accuracy, which may be attributed to the fact that the participants needed some trials to adapt to the data collection procedure, while the last repetition tends to have lower accuracy due to an onset of muscle fatigue. The average accuracy between the tested participants was 88%, and all the participants presented a low standard deviation (SD), indicating good performance of the TMC-ViT technique in decoding the hand gestures. For the thirty-two gesture classifiers, two models presented better results when the near-miss undersampling technique was used. For example, with participant 5, using near-miss instead of random undersampling led to an increase of 30% in terms of decoding accuracy for the rest gesture, i.e. the class that was downsampled. The rest gesture

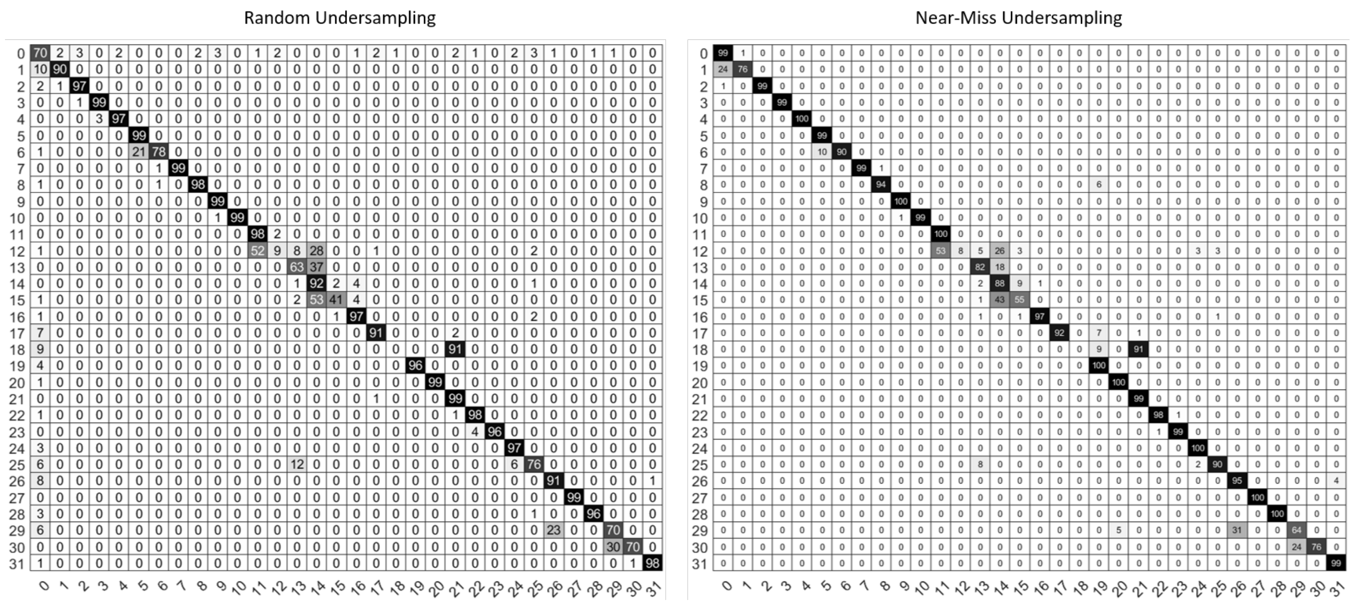

Fig. 5. Confusion matrices of one tested subject for the thirty-two-gestures classifier, using different undersampling techniques. The left confusion matrix shows the results achieved by the model using the random undersampling technique, and the right one shows the results of the near-miss technique. For the presented subject, using near-miss undersampling enhances the model's performance. By using near-miss, the performance of the most problematic gestures, e.g. gesture number 0, improved significantly compared to the results achieved by the random undersampling technique and so did the overall accuracy (for this case, changed from 85% to 91%).

classification improvement for participant 5 when using the near-miss downsampling technique can also be seen in Fig. 5.

B. *Real-time teleoperation of a robotic hand*

Before the teleoperation of the robot hand, we limited the dataset to the ten gestures that could reproduce the grasps of the taxonomy and trained one deep learning model for each subject in order to analyze the accuracy and models' behavior. The results achieved by the TMC-ViT in a ten-gesture classification task are shown in Table IV.

TABLE IV

DECODING ACCURACY FOR THE TEN GESTURES OF THE TAXONOMY. *SD* STANDS FOR STANDARD DEVIATION.

Subj.	**Undersampling Technique**	**Accuracy (%)**						
		Fold					**AVG**	**SD**
		1	**2**	**3**	**4**	**5**		
1	Random	94	96	97	96	93	**95**	**2**
2	Near-miss	89	84	93	95	77	**88**	**7**
3	Near-miss	96	99	99	95	91	**96**	**3**
4	Random	80	94	86	95	87	**88**	**6**
5	Near-miss	90	98	97	94	92	**94**	**3**
6	Near-miss	97	98	100	99	90	**97**	**4**

Five of the six models trained for the tested participants performed better when classifying ten gestures than thirty-two gestures, as expected since the number of classes is lower, and so is the model complexity. It can be noted that the decoding models for one of the participants (participant 2) performed worse with ten gestures, which may be attributed to the fact that some trials of these specific gestures may have been performed wrongly during data collection by the participant. Four of the six models achieved better performance when employing the near-miss undersampling technique, so we decided to employ that technique when training the model that was used to decode in real-time the hand gestures. It is notable the outstanding performance of the TMC-ViT model in predicting ten gestures using raw LMG signals as input, achieving accuracies as high as 97% and low standard deviation. For the online teleoperation of the robotic hand, a new set of data was collected and used to train a deep learning model. The participant successfully controlled the NDX-A* hand in real-time with minimal prediction errors. However, we noted that with prolonged use, lower accuracy of predictions was evident due to fatigue. A video of the user successfully performing some of the taxonomy grasps can be found at the following URL:

www.newdexterity.org/LMGTclassification

V. CONCLUSION

In this work, we employed a novel Transformer-based technique called TMC-ViT to process LMG signals acquired using a wearable, lightweight LMG-based armband. The deep learning models were used to decode thirty-two gestures, which include every five-finger combination that could be mapped onto a robotic hand. Data from six participants were collected, and one model was trained per subject. The results were evaluated in terms of overall accuracy, and the confusion matrices were also analyzed. Two undersampling techniques were compared in order to balance the data prior to training. The thirty-two-gesture classifiers achieved accuracies as high as 92%. We also trained and evaluated the models using only ten gestures from the collected dataset, chosen so that the robotic hand could reproduce most of the grasps proposed by the grasp taxonomy. For the reduced dataset, the TMC-ViT achieved accuracies as high as 97%. A final dataset containing only the ten selected grasps was collected for one subject, and we employed the trained model to control a robot hand in real-time successfully.

Future work will focus on integrating the training and robotic hand control into one integrated prosthetic system as well as on developing more efficient subject-generic models.

References

[1] M. Dragone, J. Saunders, and K. Dautenhahn, "On the integration of adaptive and interactive robotic smart spaces," *Paladyn, Journal of Behavioral Robotics*, vol. 6, no. 1, 2015.

[2] G. Wilson, C. Pereyda, N. Raghunath, G. de la Cruz, S. Goel, S. Nesaei, B. Minor, M. Schmitter-Edgecombe, M. E. Taylor, and D. J. Cook, "Robot-enabled support of daily activities in smart home environments," *Cognitive Systems Research*, vol. 54, pp. 258–272, 2019.

[3] S. Kawatsuma, M. Fukushima, and T. Okada, "Emergency response by robots to fukushima-daiichi accident: summary and lessons learned," *Industrial Robot: An International Journal*, 2012.

[4] M. A. Diftler, J. Mehling, M. E. Abdallah, N. A. Radford, L. B. Bridgwater, A. M. Sanders, R. S. Askew, D. M. Linn, J. D. Yamokoski, F. Permenter *et al.*, "Robonaut 2-the first humanoid robot in space," in *2011 IEEE international conference on robotics and automation*. IEEE, 2011, pp. 2178–2183.

[5] P. Ridao, M. Carreras, E. Hernandez, and N. Palomeras, "Underwater telerobotics for collaborative research," in *Advances in telerobotics*. Springer, 2007, pp. 347–359.

[6] T. S. Saponas, D. S. Tan, D. Morris, R. Balakrishnan, J. Turner, and J. A. Landay, "Enabling always-available input with muscle-computer interfaces," in *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, 2009, pp. 167–176.

[7] A. Dwivedi, "Analysis, Development, and Evaluation of Muscle Machine Interfaces for the Intuitive Control of Robotic Devices," Ph.D. dissertation, The University of Auckland, 2021.

[8] A. Dwivedi, D. Shieff, A. Turner, G. Gorjup, Y. Kwon, and M. Liarokapis, "A shared control framework for robotic telemanipulation combining electromyography based motion estimation and compliance control," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 9467–9473.

[9] R. V. Godoy, A. Dwivedi, B. Guan, A. Turner, D. Shieff, and M. Liarokapis, "On emg based dexterous robotic telemanipulation: Assessing machine learning techniques, feature extraction methods, and shared control schemes," *IEEE Access*, vol. 10, pp. 99 661–99 674, 2022.

[10] K. Kiguchi and Y. Hayashi, "An emg-based control for an upper-limb power-assist exoskeleton robot," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 4, pp. 1064–1071, 2012.

[11] L. Van Dijk, C. K. Van Der Sluis, H. W. Van Dijk, and R. M. Bongers, "Learning an emg controlled game: Task-specific adaptations and transfer," *PloS one*, vol. 11, no. 8, p. e0160817, 2016.

[12] T. S. Saponas, D. S. Tan, D. Morris, J. Turner, and J. A. Landay, "Making muscle-computer interfaces more practical," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2010, pp. 851–854.

[13] 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.

[14] A. Péter, E. Andersson, A. Hegyi, T. Finni, O. Tarassova, N. Cronin, H. Grundström, and A. Arndt, "Comparing surface and fine-wire electromyography activity of lower leg muscles at different walking speeds," *Frontiers in physiology*, p. 1283, 2019.

[15] A. Dwivedi, H. Groll, and P. Beckerle, "A systematic review of sensor fusion methods using peripheral bio-signals for human intention decoding," *Sensors*, vol. 22, no. 17, p. 6319, 2022.

[16] R. B. Woodward, M. J. Stokes, S. J. Shefelbine, and R. Vaidyanathan, "Segmenting mechanomyography measures of muscle activity phases using inertial data," *Scientific reports*, vol. 9, no. 1, pp. 1–10, 2019.

[17] W. Guo, X. Sheng, H. Liu, and X. Zhu, "Development of a multi-channel compact-size wireless hybrid semg/nirs sensor system for prosthetic manipulation," *IEEE Sensors Journal*, vol. 16, no. 2, pp. 447–456, 2015.

[18] M. Shahmohammadi, B. Guan, R. V. Godoy, A. Dwivedi, P. Nielsen, and M. Liarokapis, "On lightmyography based muscle-machine interfaces for the efficient decoding of human gestures and forces," *Scientific Reports*, vol. 13, no. 1, pp. 1–13, 2023.

[19] I. M. Bullock, J. Z. Zheng, S. De La Rosa, C. Guertler, and A. M. Dollar, "Grasp frequency and usage in daily household and machine shop tasks," *IEEE transactions on haptics*, vol. 6, no. 3, pp. 296–308, 2013.

[20] R. V. Godoy, G. J. G. Lahr, A. Dwivedi, T. J. S. Reis, P. H. Polegato, M. Becker, G. A. P. Caurin, and M. Liarokapis, "Electromyography-based, robust hand motion classification employing temporal multi-channel vision transformers," *IEEE Robotics and Automation Letters*, pp. 1–8, 2022.

[21] T. Feix, J. Romero, H.-B. Schmiedmayer, A. M. Dollar, and D. Kragic, "The grasp taxonomy of human grasp types," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 66–77, 2016.

[22] J. Stone and R. Stone, *EBOOK: Atlas of Skeletal Muscles*. McGraw Hill, 2011.

[23] G. Gao, A. Dwivedi, N. Elangovan, Y. Cao, L. Young, and M. Liarokapis, "The new dexterity adaptive, humanlike robot hand," in *IEEE International Conference on Robotics and Automation*, vol. 188, 2019.

[24] M. Asghari Oskoei and H. Hu, "Myoelectric control systems—a survey," *Biomedical Signal Processing and Control*, vol. 2, no. 4, 2007.

[25] K. Englehart, B. Hudgin, and P. Parker, "A wavelet-based continuous classification scheme for multifunction myoelectric control," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 3, pp. 302–311, 2001.

[26] I. Mani and I. Zhang, "knn approach to unbalanced data distributions: a case study involving information extraction," in *Proceedings of workshop on learning from imbalanced datasets*, vol. 126. ICML, 2003, pp. 1–7.

[27] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.

[28] R. V. Godoy, A. Dwivedi, M. Shahmohammadi, and M. Liarokapis, "Lightmyography based decoding of human intention using temporal multi-channel transformers," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 6087–6094.

[29] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2020. [Online]. Available: https://arxiv.org/abs/2010.11929

[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[31] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15, dec 2014.