



## City Research Online

### City, University of London Institutional Repository

---

**Citation:** Lin, Y., Palaniappan, R., De Wilde, P. & Li, L. (2023). Robust Long-Term Hand Grasp Recognition With Raw Electromyographic Signals Using Multidimensional Uncertainty-Aware Models. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31, pp. 962-971. doi: 10.1109/tnsre.2023.3236982

This is the published version of the paper.

This version of the publication may differ from the final published version.

---

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/29803/>

**Link to published version:** <https://doi.org/10.1109/tnsre.2023.3236982>

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

---

City Research Online:

<http://openaccess.city.ac.uk/>

[publications@city.ac.uk](mailto:publications@city.ac.uk)

---

# Robust Long-Term Hand Grasp Recognition With Raw Electromyographic Signals Using Multidimensional Uncertainty-Aware Models

Yuzhou Lin<sup>1b</sup>, Ramaswamy Palaniappan<sup>1b</sup>, *Senior Member, IEEE*,  
Philippe De Wilde<sup>1b</sup>, *Senior Member, IEEE*, and Ling Li<sup>1b</sup>, *Member, IEEE*

**Abstract**—Hand grasp recognition with surface electromyography (sEMG) has been used as a possible natural strategy to control hand prosthetics. However, effectively performing activities of daily living for users relies significantly on the long-term robustness of such recognition, which is still a challenging task due to confused classes and several other variabilities. We hypothesise that this challenge can be addressed by introducing uncertainty-aware models because the rejection of uncertain movements has previously been demonstrated to improve the reliability of sEMG-based hand gesture recognition. With a particular focus on a very challenging benchmark dataset (NinaPro Database 6), we propose a novel end-to-end uncertainty-aware model, an evidential convolutional neural network (ECNN), which can generate multidimensional uncertainties, including vacuity and dissonance, for robust long-term hand grasp recognition. To avoid heuristically determining the optimal rejection threshold, we examine the performance of misclassification detection in the validation set. Extensive comparisons of accuracy under the non-rejection and rejection scheme are conducted when classifying 8 hand grasps (including rest) over 8 subjects across proposed models. The proposed ECNN is shown to improve recognition performance, achieving an accuracy of 51.44% without the rejection option and 83.51% under the rejection scheme with multidimensional uncertainties, significantly improving the current state-of-the-art (SoA) by 3.71% and 13.88%, respectively. Furthermore, its overall rejection-capable recognition accuracy remains stable with only a small accuracy degradation after the last data acquisition over 3 days. These results show the potential design of a reliable classifier that yields accurate and robust recognition performance.

**Index Terms**—Hand gesture recognition, surface electromyography (sEMG), convolutional neural network, temporal variability, robustness, uncertainty, rejection.

Manuscript received 26 July 2022; revised 4 January 2023; accepted 10 January 2023. Date of publication 16 January 2023; date of current version 3 February 2023. (*Corresponding author: Yuzhou Lin.*)

Yuzhou Lin and Ramaswamy Palaniappan are with the School of Computing, University of Kent, Canterbury, CT2 7NZ Kent, U.K. (e-mail: yl339@kent.ac.uk; r.palani@kent.ac.uk).

Philippe De Wilde is with the Division of Natural Sciences, University of Kent, Canterbury, CT2 7NZ Kent, U.K. (e-mail: p.dewilde@kent.ac.uk).

Ling Li is with the Department of Engineering, School of Science and Technology, City, University of London, London, EC1V 0HB Northampton Square, U.K. (e-mail: caroline.li@city.ac.uk).

Digital Object Identifier 10.1109/TNSRE.2023.3236982

## I. INTRODUCTION

HAND gesture recognition (HGR) with surface electromyography (sEMG), which reveals neuromuscular activities by collecting electrical signals from muscles through non-invasive electrodes, has been widely acknowledged as a natural way to express intuitive intention, thus providing a control command for human-machine interaction (HMI) [1], [2]. As such, designing an accurate, robust, and reliable hand gesture classifier can help build a solid bridge between computers and humans. This is extremely valuable for transradial amputee users in controlling prosthetic limbs, allowing them to perform activities of daily living (ADL) through hand movements, including various hand grasps. Furthermore, as an independent module in HMI, sEMG-based HGR has often been studied offline with publicly available datasets [3], [4], [5].

Recently, deep learning approaches have been extensively investigated to improve the performance of sEMG-based HGR, and state-of-the-art (SoA) results have been reported on a series of benchmark data sets achieved by different deep learning models, particularly convolutional neural networks (CNNs) [6], [7]. However, the challenge remains to obtain the long-term robustness of the sEMG-based HGR. As many have reported, the recognition accuracy of a well-trained sEMG-based hand gesture classifier may vary or even decrease significantly over time [5], [8], [9]. Various possible factors responsible for the poor long-term robustness of HGR have been studied and could be summarised as muscle fatigue, skin conductivity, limb position, electrode displacement, and signal variation over days [10], [11], [12], [13], [14]. To address these issues, we believe that rejecting uncertain predictions [15], [16], [17] is the most straightforward approach compared to others such as multimodal approaches [6], [18], [19], adaptive learning [20], [21], [22], [23], [24], and alternative training protocols [24], [25], [26], [27].

Our previous work [35] has investigated the reliability (i.e., the quality of uncertainty measures) of the uncertainty-aware models that were proposed for the recognition of finger movement with raw sEMG. Furthermore, we have shown that a classifier can be considered more reliable if it knows what it does not know. The promising results encouraged us to pursue

TABLE I  
SUMMARY OF RESEARCH WORK ON NINAPRO DB6 IN THE LITERATURE

Paper	Model	Window+Step	Features	Accuracy (%)			
				Inter-subject	Intra-session	Inter-session within-day	Inter-session between-day
Palermo <i>et al.</i> (2017) [28]	Random Forest	200ms+10ms	Waveform Length	-	52.40 <sup>a</sup>	25.40 <sup>b</sup>	-
Wei <i>et al.</i> (2019) [6]	MV-CNN <sup>1</sup>	200ms+10ms	Raw sEMG <sup>1</sup>	-	62.90 <sup>c</sup>	-	-
Cene <i>et al.</i> (2019) [29]	R-RELM <sup>2</sup>	200ms+10ms	MAV+wavelength	-	69.83 <sup>a</sup>	41.75 <sup>b</sup>	-
	GAN			-	66.10 <sup>c</sup>	-	-
	GengNet [31]			-	56.40 <sup>c</sup>	-	-
Hu <i>et al.</i> (2019) [30]	DuNet [32]	200ms+N.A.	sEMG images <sup>3</sup>	-	56.80 <sup>c</sup>	-	-
	HuNet [33]			-	58.00 <sup>c</sup>	-	-
	TCN	150ms+15ms	Raw sEMG	-	-	-	49.60 <sup>d</sup>
Marano <i>et al.</i> (2021) [34]	Random Forest	200ms+10ms	ave(mDWT, MAV,	44.00 <sup>e</sup>	-	-	-
			VAR) <sup>4</sup>	75.40 <sup>f</sup>	75.80 <sup>g</sup>	-	-
Bao <i>et al.</i> (2022) [17]	CNN	150ms+25ms	FFT <sup>5</sup>	-	-	-	33.33 /73.33 <sup>h</sup>

<sup>1</sup> Even though the raw sEMG was claimed to be the input of multi-view CNN (MV-CNN), each feature set would be computed as a view of the sEMG. This means that the input of MV-CNN is, in fact, the extracted features of the raw sEMG. Note that the sampling frequency has been down-sampled to 100 Hz in this work.

<sup>2</sup> R-RELM: Reliable-Regularized Extreme Learning Machines

<sup>3</sup> The sEMG images were computed using three traditional feature sets.

<sup>4</sup> The input of the model was the average of the marginal discrete wavelet transform (mDWT), the mean absolute value (MAV), and the variance (VAR). Note that only the first eight electrodes were used in this work.

<sup>5</sup> The fast Fourier transform (FFT) was applied to compute the spectrum of sEMG.

<sup>a, b</sup> All odd trials collected on the morning of each day were used to train the model for each individual, which was tested on all even trials of the same morning data<sup>a</sup> and the same afternoon data<sup>b</sup>.

<sup>c</sup> All odd trials collected in the morning and afternoon of 5 days were used to train the model for each individual, and the accuracy reported was calculated intrasession (and intrasubject) for the remaining trials.

<sup>d</sup> The incremental training protocol was used and this result was not purely cross-day since session 5, as part of the training data, was recorded on the morning of the 3<sup>rd</sup> day, while session 6, as part of the testing data, was recorded on the afternoon of the same day.

<sup>e</sup> This is the inter-subject baseline accuracy achieved with ‘Source only’ method, i.e., no data from the testing subject was used.

<sup>f</sup> When allowing the use of 8 repetitions of the target subject to train the model, the best recognition accuracy was reported as 75.4%.

<sup>g</sup> Here,  $k$ -fold cross-validation was used, where  $k$  is the number of repetitions from other sessions of the same subject.

<sup>h</sup> This CNN was trained using the data for the day 1 and tested on the data for the day 3 – 5 for each individual. Higher recognition accuracy was reported under the rejection scheme.

further research on the long-term robustness of hand grasp recognition with the same design of uncertainty-aware models. In this study, we focus on the very challenging NinaPro Database 6, as it

- 1) was first released for repeatability analysis of hand grasp recognition with sEMG collected in data acquisitions across days;
- 2) was found to require more research compared to other benchmark data sets;
- 3) was ideal for validating the long-term robustness of the proposed uncertainty-aware models on sEMG-based hand grasp recognition.

We first design an end-to-end 3D CNN as a baseline model for the specific task of classifying hand grasps. An uncertainty-aware model, that is, evidential CNN (ECNN), is then proposed by integrating it with evidential deep learning (EDL). Despite the fact that the potential of ECNN can be investigated by reliability analysis without optimal rejection threshold determination, one may be eager to know how the performance of the HGR can be improved under rejection schemes in practical use. Therefore, the primary objective of this study is to present a practical way to determine a rejection threshold using validation sets only and to provide a comprehensive analysis of the long-term performance of rejection-capable hand grasp recognition with raw sEMG. Furthermore, ECNN could generate multidimensional uncertainties such as vacuity

and dissonance as a result of the nature of EDL [35]. A secondary objective of this study is to investigate its potential to improve the rejection-capable performance with multidimensional uncertainties by comparing it with a single uncertainty.

## II. RELATED WORK

The Non-Invasive Adaptive Hand Prosthetics Database 6 (NinaPro DB6) [28] is an undervalued benchmark data set to test the repeatability of hand grasp recognition, where repeatability was defined as the variation in repeated measurements made on consecutive days by the same subject under identical conditions. For each individual of 10 intact healthy subjects, 10 sessions were recorded in the morning (AM) and afternoon (PM) on 5 days, where each session involved 12 repetitions of 7 hand grasps. Furthermore, each subject was required to sit in front of a table with the forearm leaning on it to grasp an object for approximately 4 seconds on one repetition and rest for approximately 4 seconds afterward. Muscle activity was measured with 14 Delsys Trigno Wireless electrodes at a sampling frequency of 2 kHz, attached to the upper half of the forearm in two rows with equal space.

This section outlines two key findings from the summary of the research work on NinaPro DB6, which is presented in Table I. One is that NinaPro DB6 has been underexplored in the literature due to the difficulty of improving its recognition accuracy, as evidenced by the observation of generally poor

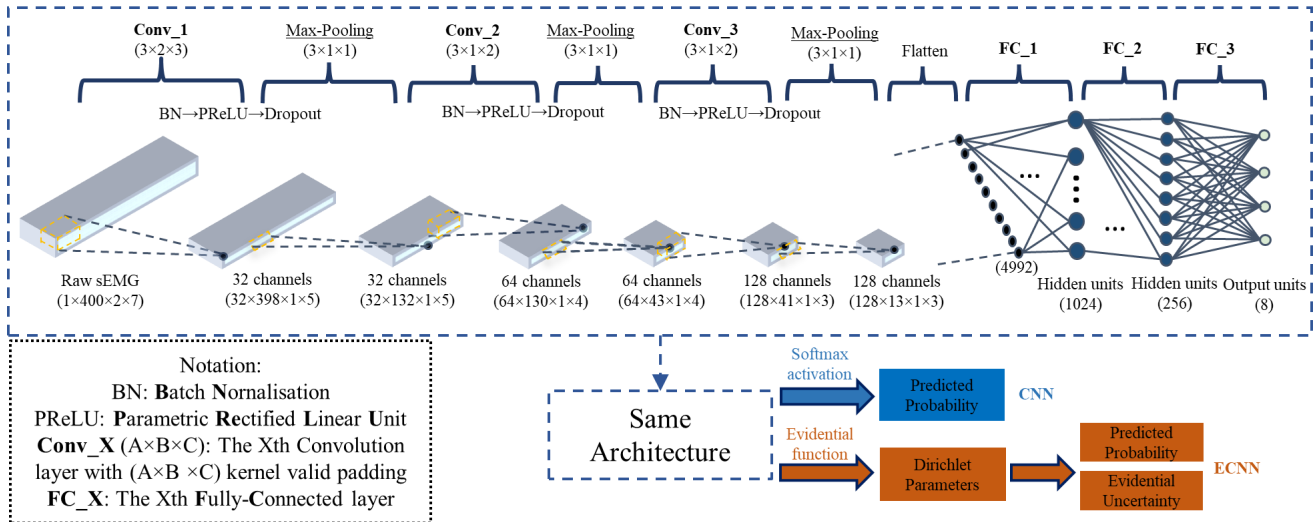


Fig. 1. The detailed illustrations of the proposed 3D Evidential Convolutional Networks (ECNN) and its conventional version (i.e., CNN).

recognition accuracy. As evaluated by Chang et al. [36], the signal quality of this data set was acceptable. Despite the fact that some signals suffer from low signal-to-noise ratio (SNR) values and incorrect labelling, we can argue that poor accuracy was due to highly confused hand grasps, together with large and various variabilities. In addition to the variability between steady and transient states and the temporal variability, at least two more variabilities must be taken into account in NinaPro DB6, namely the variability between data acquisitions and the variability between objects for each hand grasp. Note that electrodes were not required to be attached at the exact same position on each data acquisition, and two objects were used in turn when each individual performed the same hand grasp. Most studies reported intra-session accuracy in NinaPro DB6 to validate their proposed algorithms. We argue that the focus on NinaPro DB6 should be on improving the inter-session recognition accuracy only. Recall that this database was first released for the analysis of repeatability in hand grasp recognition with sEMG collected, with the aim of improving the long-term robustness of hand prostheses control systems. More importantly, Table I shows that the previous SoA inter-session accuracy was achieved as 49.6% using a proposed Temporal CNN (TCN) with multi-session training [27]. Recently, Bao et al. [17] proposed a new confidence estimate used for a CNN to improve the long-term robustness of recognition and reported an SoA inter-session between-day accuracy of 73.33% under the rejection scheme.

### III. MULTIDIMENSIONAL UNCERTAINTY-AWARE MODELS

The proposed multidimensional uncertainty-aware models are shown in Fig. 1. We first proposed a simple but efficient end-to-end 3D convolutional neural network for sEMG-based hand grasp recognition as a baseline model in this project. To improve its long-term robustness, we then constructed three multidimensional uncertainty-aware models, each with slightly different learning preferences. In this work, two dimensions of uncertainty were mainly considered: vacuity and dissonance. The uncertainty of vacuity (or known as belief vacuity) is

due to insufficient or unreliable information received from sources, while the uncertainty of dissonance (or known as belief dissonance) reflects the situation where a model holds simultaneous contradicting beliefs about a given prediction, which is usually caused by valid but conflicting evidence derived from the model output [37].

#### A. Baseline Model

We proposed an end-to-end 3D CNN for this specific hand grasp recognition task as a baseline. Note that the classifiers in this task are required to output predictive probabilities of 8 classes (including the rest posture). To learn characteristics that take into account spatial and temporal information, the 14 channels of sEMG signals were presented in a  $2 \times 7$  matrix. As such, the size of each frame of the raw sEMG signals is  $400 \times 2 \times 7$ . Taking into account the trade-off between model efficiency and training computation cost, there are three convolutional layers and three fully connected layers, which were appended for high-level reasoning. In this article, we refer to this baseline model simply as CNN.

#### B. Uncertainty-Aware Models

We constructed uncertainty-aware models by integrating the baseline model with evidential deep learning (EDL) [38]. For the sake of consistency, they will be referred to as Evidential Convolutional Networks (ECNN), the same as in our previous study [35]. Initially, EDL based on the framework of Subjective Logic (SL) was proposed to help explicitly train a model that can make a prediction along with the quantified uncertainty of it, i.e., vacuity in the context of SL, indicating whether there is sufficient evidence to support model predictions [38]. Due to the SL framework, ECNN is capable of inferring dissonance, which is another type of explainable uncertainty derived from conflicting evidence [37]. They can be regarded as evidential uncertainty, and more details are presented later in Sec. III-C.

Let  $Y = (Y_1, Y_2, \dots, Y_K)$  be a discrete variable in a domain  $\mathbb{Y}$  that represents the class label, where  $K$  is the number of

classes. A multinomial opinion on  $Y$  in SL is then defined as an ordered triplet  $w_Y = (\mathbf{b}_Y, u_Y, \mathbf{a}_Y)$ , where  $\mathbf{b}_Y$  refers to a *belief mass distribution* over  $\mathbb{Y}$ ;  $u_Y$  is the *uncertainty mass* that expresses the vacuity of evidence;  $\mathbf{a}_Y$  represents a *base rate distribution* over  $\mathbb{Y}$ , which is known as *prior probability* in classic Bayesian theory. Note that each element of  $\mathbf{a}_Y$  equals  $1/K$  if no additional information is provided. More importantly, its additivity requirement can be seen as follows.

$$\sum \mathbf{b}_Y + u_Y = 1. \quad (1)$$

It is clear that  $u_Y$  refers to vacuity because it is inversely proportional to the total belief masses. For example, when  $u_Y$  reaches its upper limit, that is, 1, it means that no belief mass can be found in any of the classes. By replacing the *softmax* layer with an activation layer such as *ReLU*, EDL terms the nonnegative output as the evidence vector  $\mathbf{e}$ , expressing the amount of evidence collected to support the classification of samples into a specific class. The belief mass distribution  $\mathbf{b}_Y$  can then be calculated by normalising  $\mathbf{e}$ , as

$$\mathbf{b}_Y = \frac{\mathbf{e}_Y}{\sum \mathbf{e}_Y + K}. \quad (2)$$

Furthermore, the Dirichlet distribution of order  $K$  with parameter vector  $\boldsymbol{\alpha}$  could be calculated from the evidence observed by  $\mathbf{e} + 1$ . Therefore, the predicted probability for each class is the expectation of the corresponding Dirichlet distribution, i.e.,

$$\mathbb{E}_Y = \frac{\boldsymbol{\alpha}}{\sum \boldsymbol{\alpha}}. \quad (3)$$

In summary, the main difference between CNN and ECNN is that ECNN finds a way to turn the value of output units into observed evidence, thus helping the model form a multinomial opinion for the recognition task to further enrich uncertainty representation.

Previously, we investigated the reliability of three variants of ECNN, which employed different loss functions for training. To further explore their long-term robustness, we followed the same procedure to build three multidimensional uncertainty-aware models, i.e., ECNN-A, ECNN-B, and ECNN-C. Given a sample  $i$  and let  $\mathbf{y}_i$  be a one-hot encoding of the ground truth class of it with  $y_{ij} = 1$  and  $y_{im} = 0$  for all  $j \neq m$  where  $j$  and  $m$  are class labels. The loss function used for ECNN-A can be seen in (4), as

$$\mathcal{L}(\mathbf{e}_i, \mathbf{y}_i) = \sum_{j=1}^K (y_{ij}^2 - 2y_{ij}\mathbb{E}[p_{ij}] + \mathbb{E}[p_{ij}^2]), \quad (4)$$

where  $p_{ij}$  is the predicted probability that the sample  $i$  is classified as the ground truth class  $j$ . This is used to obtain more evidence when the samples are predicted correctly and to remove excessive misleading evidence to avoid misclassification. To encourage the ECNN to generate relatively high vacuity when it is likely to make wrong predictions due to outliers, a Kullback-Leibler (KL) divergence term,

$$\lambda \mathbb{E}_{(\mathbf{x}_i, \mathbf{y}_i) \sim \mathcal{D}} [KL [\text{Dir}(\mathbf{p}_{-j}; \boldsymbol{\alpha}_{-j}) \parallel \text{Dir}(\mathbf{p}_{-j}; \mathbf{1})]], \quad (5)$$

is incorporated into (4) as a regularisation term to help reduce the total evidence of a misclassified sample to zero.

With respect to ECNN-B, the impact of the KL divergence term is expected to increase gradually. This is controlled by introducing a hyperparameter, annealing step  $s$ , i.e.,  $\lambda = \min(1.0, t/s)$  where  $t$  stands for the current training epoch number. Unlike ECNN-B,  $\lambda$  in (5) is a constant and will be considered a hyperparameter directly when training ECNN-C.

### C. Multidimensional Uncertainty

As introduced previously, ECNN can generate evidential uncertainty, including vacuity and dissonance. Vacuity denotes uncertainty due to insufficient evidence or knowledge and can be calculated by

$$u_{vac} = u_Y = \frac{K}{\sum \mathbf{e}_Y + K}, \quad (6)$$

where  $K$  is the number of classes and  $\mathbf{e}_Y$  is the vector of observed evidence for each class. Dissonance represents the uncertainty due to conflicting evidence, derived from a sufficient number of conflicting evidence by comparing each two singleton belief masses [39]:

$$u_{diss} = \sum_{j=1}^K \left( \frac{b_j \sum_{m=1, m \neq j}^K \mathbf{Bal}(b_j, b_m)}{\sum_{m=1, m \neq j}^K b_m} \right), \quad (7)$$

where  $K$  is the number of classes and  $\mathbf{Bal}(b_j, b_m)$  represents the relative mass balance between a pair of belief masses  $b_j$  and  $b_m$  for the sample  $i$ , equals 0 when  $b_j + b_m = 0$  and  $1 - \frac{|b_j - b_m|}{b_j + b_m}$  otherwise.

Furthermore, two uncertainty measures, including the entropy of the probability vector and the maximum posterior probability, were often used as traditional confidence scores [17], [35], [40] for comparison. They are denoted by  $u_{nEntropy}$  and  $u_{nnmp}$  and calculated in (8), representing *normalised entropy* and *normalised negative maximum probability*, respectively.

$$\begin{aligned} u_{nEntropy} &= \frac{-\sum \mathbf{p} \ln \mathbf{p}}{\ln(K)}, \\ u_{nnmp} &= 1 - \max(\mathbf{p}), \end{aligned} \quad (8)$$

where  $\mathbf{p}$  is the predicted probability vector for the  $K$  classes.

## IV. EXPERIMENTS

The versions of PyTorch and Python used in this study are 1.10.2 and 3.9.7, respectively. The experimental sequences were constructed by data loading, data segmentation, hyperparameter determination, model training, and model testing. Due to the stochastic nature of deep learning algorithms, all models were trained 10 times with the determined hyperparameters. The source code for this study is available on GitHub (<https://github.com/YuzhouLin/3dECNN-NinaProDB6>).

### A. Data Preprocessing

To satisfy the constraints of real-time control of prosthetic limbs, a sliding window of 250 ms ( $< 300ms$  [41]) was selected with an increment of 25ms. The overlap was set to as

TABLE II  
THE DETERMINED OPTIMAL HYPERPARAMETERS OF MODELS ON *NinaPro DB6*

Models	HPs	Subjects							
		1	3	4	5	6	7	8	10
CNN	batch size	64	256	64	64	256	128	256	32
	learning rate	0.0005340	0.0008067	0.0001597	0.0006327	0.0005146	0.0003064	0.0006283	0.0003267
ECNN-A	batch size	64	32	1024	32	1024	32	32	1024
	learning rate	0.0010544	0.0001190	0.0001867	0.0000352	0.0006483	0.0000894	0.0000463	0.0004348
	evidence func	<i>ReLU</i>	<i>SoftPlus</i>	<i>ReLU</i>	<i>ReLU</i>	<i>ReLU</i>	<i>ReLU</i>	<i>ReLU</i>	<i>SoftPlus</i>
ECNN-B	batch size	128	1024	32	32	256	1024	1024	128
	learning rate	0.0007813	0.0034648	0.0005427	0.0002708	0.0002021	0.0003872	0.0003869	0.0010605
	evidence func	<i>Exp</i>	<i>SoftPlus</i>	<i>SoftPlus</i>	<i>SoftPlus</i>	<i>ReLU</i>	<i>ReLU</i>	<i>ReLU</i>	<i>SoftPlus</i>
	annealing step	35	40	50	25	50	15	55	60
ECNN-C	batch size	512	256	1024	32	512	1024	64	512
	learning rate	0.0017647	0.0008522	0.0005555	0.0000297	0.0001381	0.0006240	0.0003930	0.0019805
	evidence func	<i>ReLU</i>	<i>Exp</i>	<i>ReLU</i>	<i>SoftPlus</i>	<i>SoftPlus</i>	<i>ReLU</i>	<i>SoftPlus</i>	<i>SoftPlus</i>
	lambda	0.0609794	0.1522302	0.0119255	0.0394299	0.0748643	0.0304412	0.0112908	0.0926919

high as 90% to increase the decision density and reduce the prediction delay. For fair consideration, the rest movement was treated identically to that of the other hand grasps. Specifically, each trial of ‘rest’ was taken after the first 0.5s to avoid incorrect labelling and only about 1/7 duration of its previous hand grasp to guarantee a balance in data for all classes. Note that the signals in *NinaPro DB6* were preprocessed data that had been filtered by a bandpass filter (20–450Hz) and a notch filter (50 Hz). Therefore, no additional signal preprocessing was required, and each segmented sEMG would be taken as a model input directly. Furthermore, the sEMG signals used in training and testing included both steady and transient states.

Since the main objective of this study was to investigate the long-term performance of sEMG-based hand grasp recognition, the data collected on day 1 and day 2 for each individual were used as training and validation sets, while the rest were used as test sets. To capture more generalised features, the last two cycles of each acquisition were used as validation sets. Specifically, the 11<sup>th</sup> and 12<sup>th</sup> cycles recorded at AM and PM of the first two days were used as validation sets. The ratio of training data to validation data was 5 : 1. For consistency, the 2<sup>th</sup> and 9<sup>th</sup> subjects would not be studied in this project, and the reasons are listed below.

- Only two labels of sEMG signals were found instead of 8 on the morning of day 2 for the subject 2. This may be caused by label noise or a mistake in data acquisition.
- Only 13 valid channels were found instead of 14 on the morning of day 1 for the subject 9. Specifically, you can only observe 0 on channel 7 (that is, the 8<sup>th</sup> channel).

### B. Model Training

To reduce the computational burden, we used the Tree-structured Parzen Estimator (TPE) [42], [43], which is one of the SoA hyperparameter optimisation (HPO) algorithms, to search for optimised hyperparameters. Taking advantage of sequential model-based global optimisation algorithms [42], [44], the available values of each hyperparameter

could be determined based on the previous search results because the hyperparameters have been organised into a tree-like space. The hyperparameters for each model were determined using Optuna [45], which is a powerful HPO framework. Each HPO search stopped after running 5 study trials, where each trial refers to each evaluation of an objective function to minimise validation loss. In addition, early stop was used to avoid overfitting in each training. When no improvement was found in the validation set after waiting for 10 epochs or when the training epoch reached 200, the model stopped training. Furthermore, we used ADAM as an optimiser for training and the ReduceLRonPlateau schedule with patience for 5 epochs and the decay factor of 0.8 to decrease the learning rate during training. The search result for HPO is shown in Table II.

### C. Rejection-Capable Performance Evaluation

Assume there is an uncertainty threshold  $\delta$ , a prediction can be considered uncertain when its quantified uncertainty is greater than or equal to  $\delta$  and certain otherwise. By determining  $\delta$ , a model is capable of making only confident predictions by rejecting the uncertain ones. When  $\delta = 1$ , it simply refers to standard recognition since no rejections will be made. Here, the rejection threshold of a model for each individual in each run was determined by validating its performance in misclassification detection in the validation set. In more detail, a rejection threshold was selected when it achieved the best  $F_{\beta}$ -Score,

$$F_{\beta}\text{-Score} = (1 + \beta^2) \frac{\text{Precision} \cdot \text{Recall}}{(\beta^2 \cdot \text{Precision}) + \text{Recall}}, \quad (9)$$

where  $\beta$  was chosen as 2 in this study, which means that recall was considered 2 times as important as precision. As shown in Fig. 2, the purpose of this setting was to maximise TAR as much as possible in testing data sets with a determined rejection threshold. When the number of TP and TN remains the same, a higher recall can only be achieved with fewer FNs,

		Predicted Condition		
		Positive (Reject)	Negative (Accept)	
Actual Condition	Positive (Wrong)	True Positive (TP)	False Negative (FN)	Recall $\frac{TP}{TP+FN} = \frac{\text{blue}}{\text{blue}+\text{white}}$
	Negative (Correct)	False Positive (FP)	True Negative (TN)	
		Precision $\frac{TP}{TP+FP} = \frac{\text{blue}}{\text{blue}+\text{white}}$	True Acceptance Rate (TAR) $\frac{TN}{TN+FN} = \frac{\text{blue}}{\text{blue}+\text{white}}$	Accuracy $\frac{TP+TN}{TP+FN+FP+TN}$

Fig. 2. The confusion matrix of misclassification detection. TAR can be considered the accuracy of rejection-capable sEMG-based hand grasp recognition and Precision is actually the True Rejection Rate (TRR) defined in this paper.

thus indirectly producing a higher TAR. Furthermore, to avoid setting extremely low thresholds so that no active predictions would be made in the testing phase, the determined threshold had to be checked first on the validation sets. If the number of active predictions was less than the 10% of the total number of validation samples in each acquisition, the threshold was selected again on the value which has the second best  $F_{\beta}$ -Score and et cetera.

## V. RESULTS

In all experiments, the results would be reported as an average of 10 runs, and the Wilcoxon signed-rank test was used to compare the performance of CNN (baseline model) and ECNN variants. There was a significant difference in the results between the two models when the p-value < 0.05, that is, the null hypothesis, which assumed that two related paired samples come from the same distribution, was rejected. The calculations of all the evaluation metrics used in this section can be found in Fig. 2.

### A. Recognition Performance - No Rejections

To investigate the long-term robustness of sEMG-based hand grasp recognition with the proposed uncertainty-aware models, their recognition accuracies were first presented when there was no rejection option. Recall that the proposed CNN was considered a baseline model so that its performance could be compared with the ECNN variants. Table III shows that the recognition accuracy of all models increases slightly (about 1.7%) on day 4 but decreases greatly (about 6.9%) on day 5. The best average accuracy was observed in the ECNN-A as 51.4%. Although the difference between it and CNN on each day was not statistically significant, it outperformed CNN on average with a small but statistically significant difference of 0.6%.

### B. Recognition Performance - With Rejections

To better investigate the rejection-capable performance of sEMG-based hand grasp recognition with respect to all

TABLE III  
INTER-SESSION CROSS-DAY STANDARD RECOGNITION ACCURACY OF CONVNETS WITH COMPARISONS

	Day3	Day4	Day5	Ave.
CNN	52.0±13.5	52.2±11.6	48.3±11.7	50.8±12.3
ECNN-A	52.4±12.8*	53.1±10.1*	48.8±11.4*	51.4±11.6
ECNN-B	47.4±10.2	49.2±08.5	44.3±09.7	47.0±09.5
ECNN-C	51.0±13.2	51.8±10.7	47.4±11.7	50.1±11.9

<sup>1</sup> All results were reported as the mean and standard deviation of classification accuracy for 10 runs over 8 subjects.

<sup>2</sup> The Wilcoxon signed rank test is applied to compare CNN with the ECNN variants, and the null hypothesis (H0) assumes that the median of the differences between two distributions is zero, which will be rejected when the p-value < 0.05. For conciseness, the asterisk (\*) is used when H0 is not rejected here.

proposed models, we employed three evaluation metrics, including the true acceptance rate (TAR), the true rejection rate (TRR), and the rejection rate (RR). The rejection-capable performances of models with different uncertainty scores are summarised in Table IV. Recall that TAR can be considered the recognition accuracy under the rejection scheme. First, the same commonly seen uncertainty estimate was used for a fair comparison between the CNN and ECNN variants. It can be seen that the ECNN variants significantly outperformed CNN with  $u_{nEntropy}$  or  $u_{nnmp}$  by presenting a higher TAR. ECNN-B had achieved the highest TAR of 76.77%, which was 7.07% higher than CNN when using  $u_{nnmp}$  as an uncertainty estimate. Taking into account the use of evidential uncertainty, it also achieved the best performance with  $u_{vac}$  compared to other variants of ECNN. Note that although the difference between ECNN-B and ECNN-A in TAR was not statistically significant, ECNN-B made fewer incorrect rejections, as evidenced by the observation of an 8.56% higher TRR and a 4.35% lower RR. When using  $u_{diss}$  as the uncertainty score, it is important to note that ECNN-B achieved the lowest RR, which was only about 50%. With more predictions accepted, it produced the lowest but acceptable TAR of 65.83%. Special attention was paid to the use of multidimensional uncertainties (that is, predictions were accepted when their  $u_{vac}$  and  $u_{diss}$  were found to be smaller than their respective predetermined thresholds simultaneously). In summary, ECNN-A achieved the highest TAR (83.51%) while ECNN-B obtained the best comprehensive performance by presenting a lower TAR (80.32%) but with 4.79% more active predictions.

To investigate the repeatability of sEMG-based hand grasp recognition with the proposed ECNN models with regard to multidimensional uncertainty, the grouped violin plot was drawn to visualise their TAR differences in several data acquisitions over days. A violin plot [46] is a hybrid of a box plot and a kernel density plot, which can be used to compare the distributions of several groups. Fig. 3 shows that ECNN-C consistently achieved lower performance on each acquisition across 3 days compared with the other two ECNN variants. In other words, both ECNN-A and ECNN-B obtained robust long-term sEMG-based hand grasp recognition performance, and ECNN-A was a little more robust than ECNN-B from



TABLE IV  
OVERALL REJECTION-CAPABLE PERFORMANCE COMPARISON OF CONVNETS

Scores	Models	True Acceptance Rate $\uparrow$		True Rejection Rate $\uparrow$		Rejection Rate $\downarrow$	
		M(%) $\pm$ SD(%)	H0 ( $p$ )	M(%) $\pm$ SD(%)	H0 ( $p$ )	M(%) $\pm$ SD(%)	H0 ( $p$ )
$u_{nEntropy}$	CNN	72.28 $\pm$ 15.62	-	62.81 $\pm$ 9.28	-	60.81 $\pm$ 9.57	-
	ECNN-A	75.50 $\pm$ 14.75	0 (3.7e-26)	61.64 $\pm$ 9.19	0 (5.7e-09)	64.62 $\pm$ 8.55	0 (2.6e-30)
	ECNN-B	76.71 $\pm$ 15.20	0 (1.1e-13)	65.83 $\pm$ 7.48	0 (5.1e-19)	69.23 $\pm$ 10.80	0 (7.6e-53)
	ECNN-C	73.84 $\pm$ 16.44	0 (1.5e-07)	63.84 $\pm$ 9.97	0 (2.6e-07)	62.55 $\pm$ 9.50	0 (1.0e-08)
$u_{nmp}$	CNN	71.70 $\pm$ 15.60	-	62.95 $\pm$ 9.09	-	60.00 $\pm$ 9.68	-
	ECNN-A	75.49 $\pm$ 14.63	0 (7.3e-31)	61.47 $\pm$ 9.12	0 (5.6e-14)	64.90 $\pm$ 8.80	0 (2.6e-39)
	ECNN-B	76.77 $\pm$ 15.17	0 (1.4e-18)	65.84 $\pm$ 7.46	0 (2.7e-18)	69.28 $\pm$ 10.75	0 (1.2e-58)
	ECNN-C	73.82 $\pm$ 16.38	0 (1.3e-10)	63.76 $\pm$ 9.85	0 (4.5e-05)	62.77 $\pm$ 9.35	0 (6.9e-14)
$u_{vac}$	ECNN-A	76.11 $\pm$ 15.13	-	57.23 $\pm$ 11.66	-	73.63 $\pm$ 10.26	-
	ECNN-B	76.63 $\pm$ 15.13	1 (8.7e-01)	65.79 $\pm$ 7.46	0 (3.1e-50)	69.28 $\pm$ 10.86	0 (1.3e-15)
	ECNN-C	73.77 $\pm$ 17.18	0 (1.8e-09)	63.26 $\pm$ 11.31	0 (1.4e-68)	63.03 $\pm$ 11.42	0 (3.9e-59)
$u_{diss}$	ECNN-A	75.06 $\pm$ 14.71	-	59.83 $\pm$ 9.59	-	67.70 $\pm$ 9.07	-
	ECNN-B	65.83 $\pm$ 20.22	0 (6.7e-26)	60.98 $\pm$ 9.93	0 (9.4e-05)	52.56 $\pm$ 32.27	0 (5.3e-10)
	ECNN-C	70.96 $\pm$ 15.52	0 (1.4e-24)	63.01 $\pm$ 9.66	0 (4.8e-45)	61.22 $\pm$ 12.02	0 (1.8e-31)
$u_{vac} \& u_{diss}^*$	ECNN-A	83.51 $\pm$ 15.43	-	55.88 $\pm$ 11.35	-	81.43 $\pm$ 7.65	-
	ECNN-B	80.32 $\pm$ 16.00	0 (1.8e-13)	62.82 $\pm$ 9.79	0 (7.3e-44)	76.64 $\pm$ 10.20	0 (1.7e-17)
	ECNN-C	77.35 $\pm$ 16.53	0 (9.8e-57)	61.30 $\pm$ 10.61	0 (1.1e-69)	70.23 $\pm$ 9.08	0 (2.9e-78)

<sup>1</sup> An upward arrow indicates that a higher value is desirable and vice versa.

<sup>2</sup> The M and SD refer to the mean and standard deviation of each evaluation metric for 10 runs over 8 subjects.

<sup>3</sup> The Wilcoxon signed rank test is applied to compare CNN with ECNN variants with commonly used uncertainty measures and to compare ECNN-A with other ECNN variants with evidential uncertainty measures. The null hypothesis (H0) assumes that the median of differences between two distributions is zero, which will be rejected when the p-value  $<$  0.05.

\* This multidimensional uncertainties-based rejection performance is achieved by rejecting predictions with high vacuity or dissonance. Specifically, a prediction will be accepted only when its evidential uncertainties are less than predetermined thresholds, i.e.,  $u_{vac} < \delta_{vac}$  and  $u_{diss} < \delta_{diss}$ .

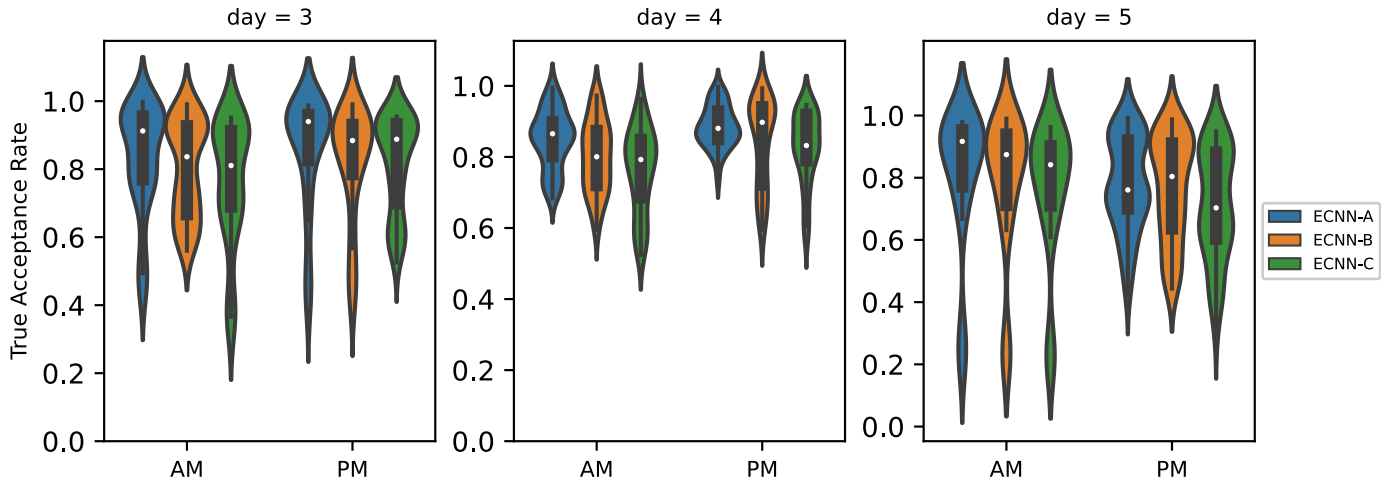


Fig. 3. Multidimensional uncertainty-based rejection-capable accuracy in several data acquisitions across days.

observation of its fewer position shifts of wider sections as time goes on. Note that a wider section of the violin plot represents a higher probability that a model will show on the given TAR. There are more details in the confusion matrix results of the ECNN-A and ECNN-B, as presented in Fig. 4 and will be discussed in Sec. VI.

### C. Comparison With SoA

The comparison of the inter-session cross-day recognition accuracy in terms of non-rejection and rejection with the previous study was presented in Table V. It is observed that our

proposed uncertainty-aware models outperformed SoA models under all conditions. The best performance was achieved by ECNN-A, which significantly improves the recognition accuracy by 3.71% and 13.88% under the non-rejection and rejection schemes, respectively.

## VI. DISCUSSION

NinaPro DB6 as a valuable benchmark dataset deserves more investigation because it was built to be challenging by including variability from data acquisitions over days and aimed at improving the long-term robustness of hand

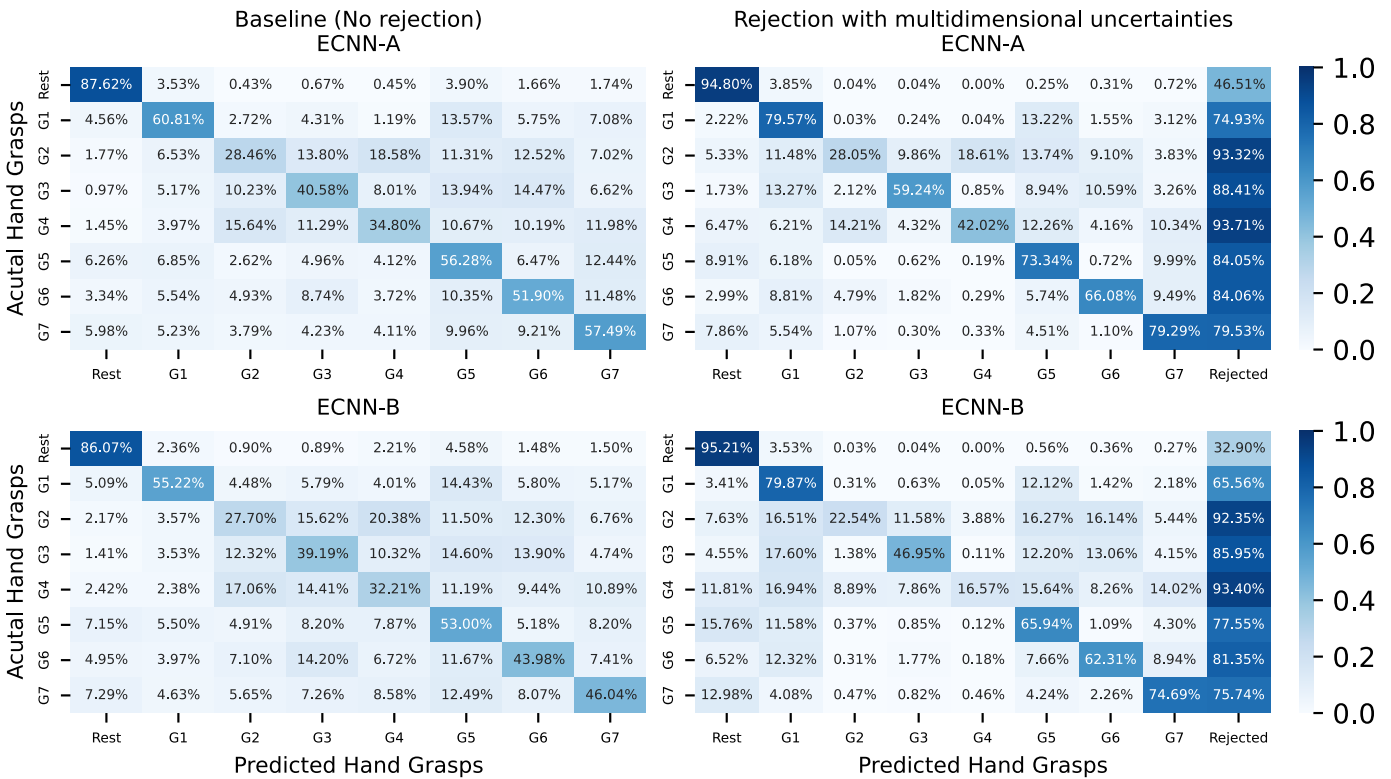


Fig. 4. The normalised confusion matrices of ECNN-A and ECNN-B averaged for 10 runs over 8 subjects. For the sake of conciseness, G1-G7 represent 'Large diameter grasp', 'Adducted Thumb grasp', 'Index finger extension grasp', 'Medium wrap', 'Writing tripod grasp', 'Power sphere grasp', and 'Precision sphere grasp'. Note that an extra x-axis was introduced when presenting a confusion matrix with rejection, which should be considered a separate part from the standard one where each cell refers to the rejection rate for a class.

TABLE V  
COMPARISON OF THE PROPOSED MODELS WITH THE SOA RESULTS ON THE INTER-SESSION CROSS-DAY ACCURACY

Model	Window + Step	Features	Inter-session cross-day accuracy (%)	
			Without rejection	With Rejection
TCN [27]	150ms + 15ms	raw sEMG	49.60 <sup>1</sup>	-
CNN [17]	150ms + 25ms	the spectrum of sEMG computed by FFT	33.33	73.33 <sup>2</sup>
ECNN-A	200ms + 20ms	raw sEMG	<b>51.44</b>	<b>83.51</b> *
ECNN-B	200ms + 20ms	raw sEMG	46.97	80.32*
ECNN-C	200ms + 20ms	raw sEMG	50.07	77.35*

<sup>1</sup> The incremental training protocol was used and this result were not purely cross-day, since the session 5, as part of the training data, was recorded on the morning of day 3, while the session 6, as part of the testing data, was recorded on the afternoon of the same day. Hence, assuming the strict intersession cross-day training protocol was used, the actual accuracy would be lower than the reported one.

<sup>2</sup> This accuracy was calculated by 1 - its proposed error rate, which only considers the number of false acceptance samples. Hence, the actual accuracy would be lower than their reported one.

\* This best rejection-capable accuracy was reported based on multidimensional uncertainties, i.e.,  $u_{vac}$  and  $u_{diss}$ .

prostheses control systems. It is worth noting that electrodes were not required to be placed in expected locations exactly during each data acquisition, which approaches real-life scenarios and brings the challenge of remaining the long-term robustness of sEMG-based hand grasp recognition to the table. In this study, we proposed a 3D uncertainty-aware model (ECNN) to address this challenge by rejecting doubtful predictions with multidimensional uncertainties. The potential of ECNN has been explored with three training strategies implemented by slightly different loss functions. The main implication of our results is that a CNN can be easily modified

as an uncertainty-aware model by integrating with evidential deep learning. Leaving aside the training strategies used by ECNN, all ECNN variants statistically significantly outperformed CNN under the rejection scheme, even if the overall mean recognition accuracy of CNN was found to be 3.80% and 0.7% higher than those of ECNN-B and ECNN-C when there is no rejection option. It is worth noting that ECNN-B achieved the best TAR and TRR with a common uncertainty  $u_{nEntropy}$  or  $u_{nnmp}$ , although the lowest standard accuracy was achieved by it under the non-rejection scheme. This suggests that model selection based on standard recognition accuracy is not

sufficient, especially when rejection is introduced. Following our previously proposed reliability analysis [35], the mean reliability of ECNN-B in terms of the normalised area under precision-recall was calculated as 58.81% and 59.88% with  $u_{nEntropy}$  and  $u_{nnmp}$ , which was 4.16% and 5.50% higher than CNN. This matches the comparison between CNN and ECNN-B with respect to rejection-capable recognition performance, indicating that ECNN-B is more reliable than CNN. This finding is consistent with previous research showing that reliability analysis can be considered a useful supplementary measure for studying sEMG-based hand gesture recognition.

The outstanding rejection-capable hand grasp recognition performance achieved by ECNN variants with multidimensional uncertainties confirms that evidential uncertainties are not only understandable but also useful in practical terms. By rejecting predictions with either insufficient or conflicting evidence, the ECNN variants are able to produce robust long-term hand grasp recognition. This implies that model reliability is associated with model robustness. However, it may be quite difficult to determine which model yields the best rejection-capable performance, as this is actually a trade-off problem. Ideally, if a model has a standard accuracy of 52% and knows exactly what it does not know, it will obtain 100% TAR and 100% TRR with 48% RR by an optimal rejection threshold under the rejection scheme. A comprehensive analysis is required to consider the efficiency of an uncertainty-aware model by making a comparison between the ideal and practical recognition performances. For example, ECNN-B can be considered more efficient than ECNN-A because it obtained an improvement of 54.46% in recognition accuracy by making 59.67% more rejections, while ECNN-A obtained an improvement of 60.60% in recognition accuracy by making 69.65% more rejections compared to the ideal situation.

One concern shown in Fig. 3 was that ECNN-C would achieve 0% TAR on some runs because no accepted predictions were found. This indicates that the approach used in this paper to determine the rejection threshold may sometimes yield a very strict threshold. It is confident that ECNN-C can get better rejection-capable recognition with appropriate rejection thresholds. Conversely, it can be said that rejection-capable performance suffers from the limitation that it is highly dependent on optimal rejection thresholds.

Fig. 4 shows the confusion matrices of the predictions made by ECNN-A and ECNN-B under the non-rejection and rejection scheme. The first finding from this figure is that it is difficult to improve the recognition accuracy of some classes, which were confused with others, by rejections. For example, it can be observed that G2 and G4 were difficult to correctly classify. When the highest number of rejections (about 93% RR) was made in the samples labelled G4, only a limited improvement in TAR was achieved by ECNN-A and ECNN-B. The recognition accuracy of G4 even decreased by approximately 50% after making approximately 93% rejections in ECNN-B. Associated with the relatively poor rejection-capable performance achieved by the ECNN variants with  $u_{diss}$  shown in Table IV, it reveals a limitation in the training of the ECNN variants. The current loss functions

used in this paper did not directly consider penalising training samples with high dissonance because it may affect the normal convergence of training CNNs. A boosting algorithm may be a possible solution to address this problem [47]. The second finding is that there is a positive correlation between baseline performance and rejection-capable performance. For example, ECNN-B obtained the highest classification accuracy on ‘rest’ as 86.07% and 95.21% with respect to the non-rejection options and with rejection. This example also shows that, not surprisingly, ‘Rest’ is considered the easiest class. Leaving the easiest and hardest classes (‘Rest’, G2 and G4) aside, it was found that about 20% improvements in hand grasps could be obtained by allowing uncertain rejections with multidimensional uncertainties regarding both ECNN-A and ECNN-B. This implies that one can focus on improving the straightforward baseline performance first without suffering from the limitations of investigating its rejection-capable performance when designing an uncertainty-aware model integrated by evidential deep learning.

## VII. CONCLUSION

This study uses a very challenging benchmark dataset, NinaPro DB6, to demonstrate the potential of designing an uncertainty-aware model to improve the long-term robustness of hand grasp recognition with raw sEMG signals. The proposed ECNN allows us to reject predictions with high evidential uncertainty from either less supported or conflicting evidence. When there is no rejection option, it outperformed the existing SoA with a significant improvement of 3.71% in recognition accuracy. More importantly, it achieved as high recognition accuracy as 83.51%, which was found to improve the current SoA by 13.88% under the rejection scheme. Furthermore, its long-term robustness has been verified by presenting a high (> 85%) rejection-capable recognition accuracy on each of 3 days with only a small degradation observed in the afternoon of day 5. This encourages us to extend the investigation to amputee subjects in improving the real-time sEMG-based control system.

## REFERENCES

- [1] M. Hakonen, H. Piitulainen, and A. Visala, “Current state of digital signal processing in myoelectric interfaces and related applications,” *Biomed. Signal Process. Control*, vol. 18, pp. 334–359, Apr. 2015.
- [2] A. Jaramillo-Yanez, M. E. Benalcazar, and E. Mena-Maldonado, “Real-time hand gesture recognition using surface electromyography and machine learning: A systematic literature review,” *Sensors*, vol. 20, no. 9, p. 2467, 2020.
- [3] M. Atzori and H. Müller, “The ninapro database: A resource for sEMG naturally controlled robotic hand prosthetics,” in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 7151–7154.
- [4] M. Cognolato et al., “Gaze, visual, myoelectric, and inertial data of grasps for intelligent prosthetics,” *Sci. Data*, vol. 7, p. 43, Feb. 2020.
- [5] U. Côté-Allard et al., “A transferable adaptive domain adversarial neural network for virtual reality augmented EMG-based gesture recognition,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 546–555, 2021.
- [6] W. Wei, Q. Dai, Y. Wong, Y. Hu, M. Kankanhalli, and W. Geng, “Surface-electromyography-based gesture recognition by multi-view deep learning,” *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2964–2973, Oct. 2019.
- [7] U. Côté-Allard et al., “Deep learning for electromyographic hand gesture signal classification using transfer learning,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 760–771, Apr. 2019.

- [8] J. He, D. Zhang, N. Jiang, X. Sheng, D. Farina, and X. Zhu, "User adaptation in long-term, open-loop myoelectric training: Implications for EMG pattern recognition in prosthesis control," *J. Neural Eng.*, vol. 12, no. 4, 2015, Art. no. 046005.
- [9] M. Zia ur Rehman et al., "Multiday EMG-based classification of hand motions with deep learning techniques," *Sensors*, vol. 18, no. 8, p. 2497, 2018.
- [10] L. Hargrove, K. Englehart, and B. Hudgins, "The effect of electrode displacements on pattern recognition based myoelectric control," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2006, pp. 2203–2206.
- [11] P. Kaufmann, K. Englehart, and M. Platzner, "Fluctuating EMG signals: Investigating long-term effects of pattern matching algorithms," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug. 2010, pp. 6357–6360.
- [12] E. Scheme, A. Fougner, Ø. Stavadahl, A. D. C. Chan, and K. Englehart, "Examining the adverse effects of limb position on pattern recognition based myoelectric control," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug. 2010, pp. 6337–6340.
- [13] E. J. Scheme and K. B. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," *J. Rehabil. Res. Develop.*, vol. 48, no. 6, pp. 643–659, 2011.
- [14] A. J. Young, L. J. Hargrove, and T. A. Kuiken, "Improving myoelectric pattern recognition robustness to electrode shift by changing interelectrode distance and electrode configuration," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 3, pp. 645–652, Mar. 2012.
- [15] J. W. Robertson, E. J. Scheme, and K. B. Englehart, "Effects of confidence-based rejection on usability and error in pattern recognition-based myoelectric control," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 5, pp. 2002–2008, Sep. 2019.
- [16] L. Wu, X. Zhang, X. Zhang, X. Chen, and X. Chen, "Metric learning for novel motion rejection in high-density myoelectric pattern recognition," *Knowl.-Based Syst.*, vol. 227, Sep. 2021, Art. no. 107165.
- [17] T. Bao, S. A. R. Zaidi, S. Q. Xie, P. Yang, and Z.-Q. Zhang, "CNN confidence estimation for rejection-based hand gesture classification in myoelectric control," *IEEE Trans. Human-Mach. Syst.*, vol. 52, no. 1, pp. 99–109, Feb. 2022.
- [18] A. Krasoulis, I. Kyranou, M. S. Erden, K. Nazarpour, and S. Vijayakumar, "Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements," *J. Neuroeng. Rehabil.*, vol. 14, no. 1, p. 71, Jul. 2017.
- [19] M. Cognolato, M. Atzori, R. Gassert, and H. Müller, "Improving robotic hand prosthesis control with eye tracking and computer vision: A multimodal approach based on the visuomotor behavior of grasping," *Frontiers Artif. Intell.*, vol. 4, p. 199, Jan. 2022.
- [20] T. Tommasi, F. Orabona, C. Castellini, and B. Caputo, "Improving control of dexterous hand prostheses using adaptive learning," *IEEE Trans. Robot.*, vol. 29, no. 1, pp. 207–219, Feb. 2013.
- [21] M. M.-C. Vidovic, H.-J. Hwang, S. Amsüss, J. M. Hahne, D. Farina, and K.-R. Müller, "Improving the robustness of myoelectric pattern recognition for upper limb prostheses by covariate shift adaptation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 9, pp. 961–970, Sep. 2016.
- [22] X. Zhu, J. Liu, D. Zhang, X. Sheng, and N. Jiang, "Cascaded adaptation framework for fast calibration of myoelectric control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 3, pp. 254–264, Mar. 2017.
- [23] J. L. Betthausen et al., "Limb position tolerant pattern recognition for myoelectric prosthesis control with adaptive sparse representations from extreme learning," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 4, pp. 770–778, Apr. 2018.
- [24] U. Côté-Allard et al., "Unsupervised domain adversarial self-calibration for electromyography-based gesture recognition," *IEEE Access*, vol. 8, pp. 177941–177955, 2020.
- [25] R. J. Beaulieu et al., "Multi-position training improves robustness of pattern recognition and reduces limb-position effect in prosthetic control," *J. Prosthetics Orthotics*, vol. 29, no. 2, pp. 54–62, Apr. 2017.
- [26] D. Yang, Y. Gu, L. Jiang, L. Osborn, and H. Liu, "Dynamic training protocol improves the robustness of PR-based myoelectric control," *Biomed. Signal Process. Control*, vol. 31, pp. 249–256, Jan. 2017.
- [27] M. Zanghieri, S. Benatti, A. Burrello, V. Kartsch, F. Conti, and L. Benini, "Robust real-time embedded EMG recognition framework using temporal convolutional networks on a multicore IoT processor," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 2, pp. 244–256, Apr. 2020.
- [28] F. Palermo, M. Cognolato, A. Gijssberts, H. Müller, B. Caputo, and M. Atzori, "Repeatability of grasp recognition for robotic hand prosthesis control based on sEMG data," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2017, pp. 1154–1159.
- [29] V. Cene, M. Tosin, J. Machado, and A. Balbinot, "Open database for accurate upper-limb intent detection using electromyography and reliable extreme learning machines," *Sensors*, vol. 19, no. 8, p. 1864, Apr. 2019.
- [30] Y. Hu, Y. Wong, Q. Dai, M. Kankanhalli, W. Geng, and X. Li, "sEMG-based gesture recognition with embedded virtual hand poses and adversarial learning," *IEEE Access*, vol. 7, pp. 104108–104120, 2019.
- [31] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, no. 1, p. 36571, Nov. 2016.
- [32] Y. Du et al., "Semi-supervised learning for surface EMG-based gesture recognition," in *Proc. IJCAI*, 2017, pp. 1624–1630.
- [33] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W. Geng, "A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition," *PLoS ONE*, vol. 13, no. 10, Oct. 2018, Art. no. e0206049.
- [34] G. Marano, C. Brambilla, R. M. Mira, A. Scano, H. Müller, and M. Atzori, "Questioning domain adaptation in myoelectric hand prostheses control: An inter- and intra-subject study," *Sensors*, vol. 21, no. 22, p. 7500, Nov. 2021.
- [35] Y. Lin, R. Palaniappan, P. De Wilde, and L. Li, "Reliability analysis for finger movement recognition with raw electromyographic signal by evidential convolutional networks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 96–107, 2022.
- [36] J. Chang, A. Phinyomark, and E. Scheme, "Assessment of EMG benchmark data for gesture recognition using the NinaPro database," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 3339–3342.
- [37] A. Josang, J.-H. Cho, and F. Chen, "Uncertainty characteristics of subjective opinions," in *Proc. 21st Int. Conf. Inf. Fusion (FUSION)*, Jul. 2018, pp. 1998–2005.
- [38] M. Sensoy, L. M. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," in *Proc. NeurIPS*, 2018, pp. 3183–3193.
- [39] X. Zhao, F. Chen, S. Hu, and J. Cho, "Uncertainty aware semi-supervised learning on graph data," in *Proc. NeurIPS*, vol. 33, 2020, pp. 12827–12836.
- [40] E. Scheme and K. Englehart, "A comparison of classification based confidence metrics for use in the design of myoelectric control systems," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 7278–7283.
- [41] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, Jan. 1993.
- [42] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyperparameter optimization," in *Proc. NIPS*, 2011, pp. 2546–2554.
- [43] J. Bergstra, D. Yamins, and D. D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures," in *Proc. ICML*, vol. 28, 2013, pp. 115–123.
- [44] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," in *Proc. LION (Lecture Notes in Computer Science)*, vol. 6683. Berlin, Germany: Springer, 2011, pp. 507–523.
- [45] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. KDD*, 2019, pp. 2623–2631.
- [46] J. L. Hintze and R. D. Nelson, "Violin plots: A box plot-density trace synergism," *Amer. Statistician*, vol. 52, no. 2, pp. 181–184, 1998.
- [47] Z. Li, B. Wang, C. Yang, Q. Xie, and C.-Y. Su, "Boosting-based EMG patterns classification scheme for robustness enhancement," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 3, pp. 545–552, May 2013.