This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Gesture recognition from bio-signals using hybrid deep neural networks

PLEASE CITE THE PUBLISHED VERSION

<https://doi.org/10.1109/ICAICA50127.2020.9182510>

PUBLISHER

IEEE

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

LICENCE

All Rights Reserved

REPOSITORY RECORD

Xie, Baao, James Meng, Baihua Li, and Andy Harland. 2020. "Gesture Recognition from Bio-signals Using Hybrid Deep Neural Networks". Loughborough University. <https://hdl.handle.net/2134/12964829.v1>.

Gesture Recognition from Bio-signals Using Hybrid Deep Neural Networks

Baao Xie

Loughborough
University
Loughborough, UK
B.Xie@lboro.ac.uk

James Meng

School of Clinical
Medicine
Cambridge University, UK
jfm48@cam.ac.uk

Baihua Li

Loughborough
University
Loughborough, UK
B.Li@lboro.ac.uk
*Corresponding Author

Andy Harland

Loughborough
University
Loughborough, UK
a.r.harland@lboro.ac.uk

Abstract— Surface electromyogram (sEMG) provides a promising means to develop a non-invasive prosthesis control system. In the context of transradial amputees, it allows a limited but functionally useful return of hand function that can significantly improve patients’ quality of life. In order to predict users’ motion intention, the ability to process multi-channel sEMG signals generated by muscle is required. We propose an attention-based Bidirectional Convolutional Gated Recurrent Unit (Bi-CGRU) deep neural network to analyse sEMG signals. The two key novel aspects of our work include: firstly, novel use of a bi-directional sequential GRU to focus on the inter-channel relationship between both the prior time steps and the posterior signals. This enhances the intra-channel features extracted by an initial one-dimensional CNN. Secondly, an attention component is employed at each GRU layer. This mechanism learns different intra-attention weights, enabling focus on vital parts and corresponding dependencies of the signal. This increases robustness to feature noise to further improve accuracy. The attention-based Bi-CGRU is evaluated on the Ninapro benchmark dataset of sEMG hand gestures. The electromyogram signals of 17 hand gestures from 10 subjects from the database are tested. The average accuracy achieved 88.73%, outperforming the state-of-the-art approaches on the same database. This demonstrates that the proposed attention based Bi-CGRU model provides a promising bio-control solution for robotic prostheses.

Keywords— surface electromyogram, myoelectric control, gesture and movement recognition, deep learning, attention mechanism, gated recurrent unit, signal processing

I. INTRODUCTION

In the US alone, an estimated 41,000 people live with a major loss of upper limb (i.e. excluding fingers) [1]. The loss of hand function greatly limits a person’s capabilities with a corresponding decrease in quality of life. While prostheses are available, they are limited by unreliable and/or intuitive control methods. A reliable and non-invasive bio-inferred control system is thus highly desirable. Surface electromyogram (sEMG) signals (show) promise in meeting such criteria as an effective and patient friendly input [2]. Compared to the EMG signal obtained by needled electrodes, sEMG is collected by placing non-invasive electrodes on the skin to measure muscle activation [3][4]. Current sEMG use is largely limited to lower limb control due to imprecise control methods allowing only coarse movements [5]. This work aims to develop a more accurate sEMG pattern recognition method, thereby widening the possible uses of prostheses to more precise and co-ordinated movements. This is a vital step to moving away from coarse lower limb control to the use of sEMG for hand movements. This has the potential to improve patients’ quality of life

significantly, while also contributing to the growing body of work into bio-inferred prosthesis control.

In the area of sEMG-based movement recognition, traditional machine learning approaches have been widely employed, including support vector machines, neural networks, genetic algorithms, Bayesian networks, and decision trees [6]. These methods can provide a reasonable performance on the sEMG classification but require large amounts of hand-crafted features. Such feature engineering procedures are time consuming and require high-level domain knowledge. Another limitation of traditional machine learning is lacking the ability of generalization and robustness. This means a machine learning model developed for a specific case will have difficulty transferring the training ‘experience’ to another case. In recent years, Deep Learning (DL) approaches such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have become more prominent in the fields of image and signal recognition. Compared to traditional machine learning approaches, deep learning models can extract and learn complex and high-level features from the raw signal without hand-crafted feature selection. In addition, the re-usability of the extractor in deep learning models reduces the task of generating hand-crafted features for every specific case. Therefore, we chose to develop and optimize deep learning models rather than applying traditional machine learning methods to tackle this challenge.

For the time-series signals, Long Short-Term Memory (LSTM) model and Gated Recurrent Unit (GRU) have shown their ability to generate classification results that consider the dependencies between time steps of sequential data. Compared to LSTM, GRU has a less complex structure, enabling more rapid training and reduced computational and memory loads. This is particularly important for real-time embedded systems, such as those we envisage using for sEMG prosthesis control. Further, GRU requires less training data. This benefits both initial model training and also allows rapid refinement to produce a patient-unique response model. Meanwhile, the rapid nature of bio-signals means the ability of LSTM to remember long term data is less important.

However, in both LSTM and GRU, each cell only gathers a state from previous cells. In gesture recognition, the posterior time steps of the signal also contain useful information, which is meaningful for the current cell classification. To overcome this restriction in unidirectional RNN, we propose an advanced bidirectional GRU model. This structure allows the model to possess both backward and forward information at each time step. In addition, we

applied a novel attention mechanism which enables the model to concentrate on the vital information in features to reduce noisy and redundant data.

In our own previous work, we have developed several deep learning models for gesture recognition using wearable sEMG sensors [7]. These models obtained a good performance compared to traditional machine learning models. However, these approaches take raw signals as inputs. This decreases the accuracy due to the inherent muscle crosstalk noise and signal interference. To overcome this problem, we propose a novel attention-based Bi-CGRU model using feature-based signals as input. The model is evaluated on the Ninapro sEMG database [12], containing signals of 17 different hand gestures from 10 subjects. The result of classification has been compared with our previous models and existing approaches in this field.

The main contribution and innovation of our work:

- Apply deep learning technology for sEMG based bio-controlled prostheses.
- Introduce an efficient bidirectional GRU to obtain information both forward and backward in time.
- Implement an attention mechanism to extract the intra-attention weights and thus further increase accuracy and robustness.
- Use feature-based signals rather than the raw data to reduce feature noise.

II. RELATED WORK

In the early stages of sEMG-based gesture recognition frameworks, machine learning models were developed extensively with hand-crafted feature extraction. Ahsan et al. [8] built an artificial neural network (ANN) to classify sEMG movement signals. Features in time-domain and frequency-domain such as Root Mean Square (RMS) and Zero Crossings (ZC) were extracted in the training procedure. Oskoei et al. [9] developed a support vector machine (SVM) classifier based on the sEMG signal and applied a hyperparameter tuning strategy. However, each aforementioned machine learning approach depends on high-level domain knowledge for feature extraction. Also, it is hard to transform the features for a specific problem to another case.

Due to the limitations of machine learning methodology for signal recognition, many researchers turn to use deep learning models. In this area, Côté-Allard et al. [10] applied a transfer learning model on aggregated data from multiple users on three different databases. The model reached 68.98% for 18 gestures over 10 participants using the raw sEMG signals in NinaproDB5. Wan and Han [11] designed a time-domain sEMG-graph to represent multi-channel sEMG signals as an image, then proposed an end-to-end classifier based on the CNN to classify the gestures. The model achieved an accuracy of 82% on 14 movements from NinaproDB5. They reported overperforming traditional machine learning models by 13%.

Recently the algorithmic improvements and new hardware architectures have allowed researchers to develop more complex and advanced architectures for specific tasks. The deep learning models of CNN-RNN have shown competence in sEMG-based gesture recognition. Following this tendency, the attention mechanism and bidirectional GRU technologies have been proposed for Natural Language Processing (NLP) and signal. Previously, we developed several deep learning models to classify different hand

gestures based on raw sEMG signals. The hybrid 3+3 C-RNN model achieved the best performance [7]. This hybrid model includes three convolutional layers as feature extractors followed by three recurrent layers as the classifier. The model has been trained on three datasets including the UC Irvine (UCI) Human Activity Recognition (HAR) dataset, NinaproDB2 and NinaproDB5. For the NinaproDB5, it attained the highest accuracy of 83.6% [7]. In this work, pre-processed signals are employed as input instead of using raw signals, and propose a novel attention based Bi-CGRU model.

III. METHODOLOGY

1) Model

The attention-based Bi-CGRU is shown in Fig. 1. Each channel of the raw sEMG signal is firstly divided by sequential sliding windows (section A) in the size of 128 samples (~640ms). The sliding window layer generates a dense matrix of windows for 16 channel signals. Then, the window matrices go through a pre-processing stage to obtain the featured signals, namely Mean, SD, MAD, RMS (section B). This produces a more representative output with less noise. The pre-processed feature signals will be normalized into a fixed scale (section C), followed by three one dimensional convolutional layers to extract intra-channel characteristic features. The features of each channel will then be integrated into a feature map matrix, for use as input into the GRU module (section D). The bidirectional GRU module has three identical layers, each layer composed of two parallel processes; a forward time sequence and backward time sequence. Inputs are fed simultaneously into each sequence for processing. The output of the forward sequence will also be fed as the input for the backward time sequence to produce the result with backward time state information. With this structure, each GRU cell will form weighted inter-channel features from both prior and posterior cells. A feature sequence output from the previous Bi-GRU layer is transferred to the next Bi-GRU layer through an attention layer (section E). The attention layer extracts vital Bi-GRU features and assigns attention weights for the feature cells. The attention weight for each time step is calculated based on states between cells. In the final third Bi-GRU layer, the prediction from the last cell is considered to be the classification result.

Fig. 1. Structure of the attention-based Bi-CGRU model

2) Sliding Window

The continuous sEMG signal is divided into sequential windows of a specific size. A window size too big will reduce signal resolution and thus lose the characteristic nature of high-frequency bio-signals. Further, it will increase decision delay to an unacceptable level that becomes perceptible to a prosthesis user. Equally, too small a window size will increase the computational load and limit the ability to reduce raw sEMG noise. In our work, the length of sliding windows is informed by related work by Scheme and Englehart [16]. They report that 300ms-800ms is an acceptable delay that a prosthesis user is unlikely to detect in real-time signal recognition. As a short sliding window allows the model to respond quicker to real-time predictions, a 640ms (128 samples) sliding window size was used.

3) sEMG Signal Pre-processing

In the signal recognition field, deep learning based approaches tend to directly use raw signals as inputs. However, the sEMG bio-signals are naturally noisy due to frequent muscle crosstalk, signal interference and variation in sensor attachment conditions. To avoid such adverse effects, a pre-processing stage is proposed to obtain a pre-processed feature signal. This solution is inspired by related research, which shows the improved performance of the models with salient signature features. Hu and Peng [17] used time-frequency features as the input of an RNN+CNN model for joint angle recognition. Furthermore, Shim et al. [18] developed an unsupervised pre-trained deep belief network for motion recognition. The input of the model is time-domain features generated by DeepLearnToolbox. The feature extraction strategy of my work is based on Côté-Allard et al. [10] and [19]. The former compared the performance of raw signal input with state-of-the-art feature input on several deep learning models. On the aspect of sEMG feature extracting, [19] has done much work on the Ninapro in both the time domain and frequency domain.

Table 1: the feature sets tested in the experiment. There are 6 feature sets ($S_1, S_2, \dots, S_6$) listed in the table. All the sets are experimentalized on the Bi-CGRU model with the same deep network configurations.

Time Domain Features	S_1	S_2	S_3	S_4	S_5	S_6
Root Mean Square (RMS)	●	●	●	●	●	●
Mean Value	●					●
Zero Crossings (ZC)	●	●				
Slope Sign Changes (SSC)	●		●			
Mean Absolute Deviation (MAD)	●			●	●	●
Standard Deviation (SD)	●		●			●
Waveform Length (WL)	●	●	●		●	
Willison Amplitude (WA)	●			●		

We conducted signal analysis experiments in both the temporal and frequency domains. The frequency features investigated were Mean Frequency (MF), Median Frequency (MDF) and Power Spectrum (PS) generated by Fast Fourier Transform. However, the results show that the scale of time-domain features and frequency-domain features differ significantly even with some normalization procedures. When the time-frequency feature sets are fed into the Bi-CGRU model, the model treats all the input as the same class, which leads to a termination of recognition. For this reason, we decided to employ only time-domain features in this stage. From the trials in Table 1, S_1 and S_6 achieved the best and most accurate performance on the NinaproDB5.

However, S_1 with eight features costs additional resources and increases the computation time. Hence, S_6 is employed as the feature set for model learning. It contains four widely used time-domain features in sEMG-based recognition, where $\{x_1, x_2, \dots, x_N\}$ represents the signal sample values, N is the total number of samples:

- Mean value (M): as shown in Eq.1

$$M = \frac{1}{N} \left(\sum_{i=1}^N x_i \right) \quad (1)$$

- Standard Deviation (SD): the amount value of variation or dispersion of signal values, as shown in Eq.2

$$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - M)^2} \quad (2)$$

- Mean Absolute Deviation (MAD): the mean absolute difference of the signal sample and the mean value, as shown in Eq.3

$$MAD = \frac{1}{N} \sum_{i=1}^N |x_i - M| \quad (3)$$

- Root Mean Square (RMS): an indicator of average signal value. It cancels out the negative values by squaring these values to calculate a mean value, as shown in Eq.4

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (4)$$

In the pre-processing stage, a single set of four time-domain features [RMS, Mean, MAD, SD] is generated from 16 raw sEMG samples, with 75% sample overlap. Each 16 samples are treated as a sub-window within the sliding window of size 128 samples. Therefore, the length of the sliding window output will be the same as the input, which is 128 in this case. These features provide salient information about muscle activities while also maintaining the characteristic of the raw signal. Based on our experiments, incorporating a pre-processing stage increases the accuracy, compared to directly using the raw sEMG signal.

4) Batch Normalization

Batch normalization is a methodology intended to mitigate internal covariate shift [20]. This problem is defined as a large, non-uniform change in the distribution of each layer's inputs during network training. The change in input distribution is due to the continuous updating of network parameters. This constant covariate shift could have a detrimental effect on the training process.

To standardize the distribution of each layer's input, we adopted a two-step process in the developed Bi-CGRU model. Firstly, after the pre-processing stage (as shown in Fig.1), feature scaling is applied to each signal channel independently through a linear transformation, then the output of each channel is standardized into a same scale for the entire model. Secondly, batch normalization is employed to normalise the mean and variance of activation functions at each one dimensional CNN and GRU layers. Therefore, the input of a subsequent layer is adjusted towards a more stable distribution, that is less sensitive to the changing parameters in subsequent layers. The batch normalization improves the input distribution of each layer to a similar scale level.

5) Inter- and Intra- Deep Feature Generation by CNN and Bidirectional GRU

Three layers of a one-dimensional CNN is applied to each sEMG channel to extract sequential intra-channel dependencies. The first two layers of CNN have a filter size of 4 and a stride of 2, while the final CNN layer has a size of 2 and stride of 1. The decrease in filter size and stride across the layers is a better discrimination strategy to learn features, from coarse to fine. Each convolutional layer is followed by a dropout layer, with dropout rates of 0.3, 0.2 and 0.2 respectively. This aims to solve the problem of overfitting in deep networks by randomly dropping cells from the neural network. It means some cells are ignored in the current layer, but the information of these cells will still be delivered to the next layer to keep the integrity of data. The produced features from multi-channel sEMG channels (16 in our work) will then be integrated into a multi-channel feature map, as input for the bidirectional GRU module.

The GRU is a branch of RNN, which was raised to solve the vanishing gradient problem, similar to the LSTM model. GRU applies an update gate to control the states of cells, whereas LSTM employs the more complex strategy in this procedure, including state selection, addition and removal (forgetting). Therefore, the GRU reduces the model complexity and parameters, computational time and memory cost is correspondingly reduced. Furthermore, this more efficient GRU learning structure requires less training data. This is essential to guarantee model performance when only smaller, more difficult to obtain datasets, are available, such as with the limited sEMG bio-signals in our work.

In signal classification, traditional LSTM and GRU models only consider the information from previous cells to predict the result. Bidirectional RNNs are an improvement on such unidirectional deep networks. The bidirectional GRU was initially designed for NLP, where information from the future (posterior) is equally essential as information from the past (prior). We use the Bi-GRU to learn the relationship between both the prior cells and posterior cells in the data sequence. The states of posterior cells can be used to provide feedback for the model to calculate the state of the current cell. The approach achieves this by connecting two layers of opposite directions and updates each cell's state with the information from these layers.

Motivated by this architecture, three bidirectional GRU layers are proposed as shown in Fig. 1. Each GRU layer has 128 cells, followed a dropout layer with dropout rates of 0.25, 0.2 and 0.15 respectively. The input of the first Bi-GRU layer is the multi-channel feature map extracted from the CNN layers. Each GRU cell utilises information from both prior and posterior samples to generate the output. The produced feature sequence output from the current Bi-GRU layer is then transferred to the next Bi-GRU layer through the attention layer. In the final Bi-GRU layer, the prediction from the last cell produces the classification result for the corresponding sliding window.

6) Attention Mechanism

The GRU/LSTM based sequential models can achieve good results when features are relatively equal in their importance and discriminative value. However, when the model has a long structure of cells and input sequences, the salient features tend to be lost due to data noise and averaging of attention, making it difficult to learn accurate

cell weights. When the input sequence is long, the model can be confused by bad inferences due to noisy pieces of data. This is particularly problematic when the input signal lacks a dominant component, or when the vital features are hidden due to limited training data.

To tackle this problem, we employ an attention mechanism, as shown in Fig. 1. The attention mechanism aims to relate different positions of an input sequence. It allows the model to mine the entire sequence and extract the critical weights for each cell. In the proposed attention-based Bi-CGRU model, each bidirectional GRU layer is followed with an attention layer to re-assign the cell weights of the GRU inputs. The input of the attention layer is the 128-dimensional feature map generated by the previous Bi-GRU layer, and its output has the same feature length but with re-assigned cell weights.

At each attention layer, the attention score is calculated across the current hidden states of the Bi-GRU cells. New weights are passed to the Sigmoid activation function to evaluate. Resulting attention scores will then be fed into a standard Softmax classifier. The probability is calculated depending on both the current cell hidden state and related cell hidden states. Finally, the probability score of the attention layer is evaluated using Cross-entropy loss.

In summary, a feature sequence of 128, equal to the window size, is inputted into the attention layer. The layer re-assigns weights according to attention score, producing an output of corresponding attention weights for these features. This further reduces training loss and limits noise interference. The model can thus effectively capture the importance of salient features from essential muscle activations.

Large weights in a complex deep neural network are likely to quickly overfit a training dataset with limited examples. To reduce the overfitting problem, each attention layer in the model is followed by a dropout layer with a dropout rate of 0.2. This means that a random selection of 20% of cells is set to an activation value of 0. Therefore, they are ignored in the current layer, but these cells' states and relationships with other cells will still be delivered to the next layer to keep the integrity of data. Probabilistically dropping out cells in the network is a simple, computationally cheap and effective regularization method to reduce overfitting and improve model generalization.

IV. EXPERIMENTAL RESULTS

A. Database

The well-established public sEMG dataset, Ninapro [12], was used to train and test the proposed model. The Ninapro database contains recordings of muscle activity using sEMG sensors from different subjects. Those datasets are targeted to benefit research on advanced hand myoelectric prosthetics. The sub-dataset 5 (Ninapro DB5) was recorded with two Myo armbands with 16 electrodes, providing the up-sampled sEMG signals at 200 Hz from 10 intact subjects. This sub-dataset is built to benchmark sEMG gesture recognition frameworks [13]. For our study of gesture recognition, 17 gestures in the NinaproDB5 are selected as shown in Fig. 2.

Fig. 2. 17 hand gestures in NinaproDB5 from 10 subjects [13]. (a) Gestures 1 to 8 contain basic finger movements (flexions and extensions). (b) The gesture 9 to 17 contain the basic movements of the wrist. Volunteers are allowed short breaks between gestures to avoid muscle fatigue.

During signal collection, the armbands are fixed closely to the elbow based on the Ninapro standards. Each subject performs 17 gestures 6 times. Each gesture lasts for 5 seconds and is followed by 3 seconds of rest. For each gesture in the database, the signals of repeat 1,3,4,6 are employed as the training set and repeat 2, 5 as the testing set. Moreover, in the training set, 30% of data is separated for use as the validation set. The aim of dividing the dataset is to avoid the unbalanced allocation and overfitting problem.

B. Training and Evaluation

The training and testing were carried out on a computer with GPU of GTX 1080ti with RAM size of 16 GB and CPU of Intel(R) Core(TM) i7-7700k @ 4.20Ghz. Besides, the model runs on the solid-state drive to reduce the time cost. The attention bi-CGRU model was implemented using Keras and TensorFlow.

The signals of repeat 1, 3, 4, 6, of each gesture are employed as the training set and the signals of repeat 2, 5, as the testing set. The hyperparameters in the networks have been tuned through a series of trials. Through these experiments, the learning rate was chosen to be 0.001 and the epoch size was 500. An early stopping mechanism was used to avoid overfitting by terminating the training progress when the validation accuracy stopped increasing.

During the testing, the average accuracy of each model is measured by testing set with the loss function of multiclass cross-entropy as shown in Eq.5,

$$-\sum_{c=1}^M y_{o,c} \log(p_{o,c}) \quad (5)$$

where M is the number of classes, y is the binary indicator, p is the predicted probability, o is the observation of the class c . Cross-entropy loss increases as the predicted probability diverges from the actual class.

C. Performance

The developed attention-based Bi-CGRU model has achieved competitive recognition accuracy of 88.73% on 17 hand gestures on the NinaproDB5. We compared our results with state-of-the-art machine learning and deep learning based models. As shown in Table 1, Pizzolato et al. [15] achieved accuracy of $69.04\% \pm 5.24\%$ on classifying 41 hand gestures using the NinaproDB5. They applied the multivariate Discrete Wavelet Transform (mDWT) as the feature extractor, and the Support Vector Machine (SVM) as the classifier. Compared to our method, they used raw signals as the input of the model. However, this accuracy was achieved using a window size of 200 samples with 50% overlapping, which is almost three times our window size (128 samples with 75% overlapping). This means the online classification delay is around 500ms of every 100 samples [15], while our model can achieve 160ms of every 32 samples. Our model facilitates a smaller detection window

and quicker response, which is particularly important for deploying the model for “real-time” use.

Table 2. comparisons between different methods on NinaproDB5

Model	Accuracy	Description
Multivariate Discrete Wavelet + Support Vector Machine [15]	69.04%	41 selected movements from the dataset with twin Myo armbands.
CWT + ConvNet + transfer learning [10]	68.98%	17 selected movements from the dataset with single Myo armband.
Residual Convolutional Neural Network [11]	82%	14 selected movements from the dataset with twin Myo armbands. Instead of using the sliding window technique, the model regards the whole signal as an image.
Hybrid 3+3 CNN-RNN (3 Conv layers and 3 Recurrent layers) [7]	83.61%	17 selected movements of the datasets with twin Myo armbands. These models were developed in our previous work with the same experiment configurations.
Attention-based Bi-CGRU	88.73%	17 selected movements with twin Myo armbands.

The Continuous Wavelet Transform (CWT) based ConvNet [10] reached 68.98% accuracy on the same 17 selected gestures as in our experiments. They extracted complex features (e.g. RMS, marginal Discrete Wavelet Transform mDWT and EMG Histogram) as the input of the ConvNet. All trial data from 9 subjects in NinaproDB5 were used for training. They then fine-tuned the learning using 10 subjects on the pre-trained 9-subject model for four cycles. That means 90% of the validation and testing sets are known to the model. Their explanation was that no large enough dataset was available for the NinaproDB5. Compared to this research, we achieved a higher average accuracy of 88.73%, evaluated on data from unknown trials (6 total trials for each subject and gesture; 2 unknown used for evaluation, 4 for training). Meanwhile, the Residual CNN from Wan and Han obtained an accuracy of 82.15% for 14 selected hand gestures [11]. Unlike other works, they transformed the raw sEMG signal from single armband into an image, then applied a Residual CNN to classify these images. Such an approach can be challenging to apply to real-time applications as the model needs a segmented input covering the whole gesture period.

In our own previous work, several deep learning models were evaluated on the NinaproDB5, including a hybrid C-RNN model and a hybrid 3CNN+3RNN model. The hybrid 3CNN+3RNN model achieved the best accuracy of 83.61% on the NinaproDB5 [7]. Compared to this 3CNN+3RNN model, Bi-CGRU gained 5% accuracy by introducing the bi-directional GRU and attention mechanism.

V. CONCLUSION

An attention-based Bi-CGRU deep neural network has been developed to recognise hand gestures, using sEMG signals. The model utilises a three layer, one-dimensional CNN and a three layer bi-directional GRU to extract intra- and inter-channel features respectively. An attention

mechanism is employed to enhance the bi-GRU module and re-assigns salient feature weights through attention learning.

The proposed model was tested on the NinaproDB5 dataset of 17 hand gestures. It achieved a recognition accuracy of 88.73%, which is higher than the state-of-art work and our previous attempts. In addition, our approach was tailored for use in real-time applications by employing a sliding window of 640ms. This allows online classification of each time window at an acceptable delay to a potential prosthesis user.

This work has shown the feasibility of our novel approach and its success is demonstrated by the high accuracy with NinaproDB5. In future work, we hope to expand our proof of concept to include more sEMG hand gesture data from a wider range of subjects. To deal with limited availability of human bio-signal training data, data augmentation technology or optimal hyperparameter tuning would be explored. We thus envisage a baseline model that requires only limited data from an individual patient to fine-tune. Therefore, the model can rapidly adapt to become tailored to the subtleties of an individual's movements, ultimately allowing robust and personalised bio-control.

REFERENCES

- [1] Atzori, M. & Müller, H. (2015) Control Capabilities of Myoelectric Robotic Prostheses by Hand Amputees: A Scientific Research and Market Overview, *Frontiers in Systems Neuroscience*, Vol 9:162
- [2] D. Farina et al., "The Extraction of Neural Information from the Surface EMG for the Control of Upper-Limb Prostheses: Emerging Avenues and Challenges," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 797-809, July 2014.
- [3] J. Wang, Y. Chen, S. Hao, X. Peng, X and L. Hu, "Deep Learning for Sensor-based Activity Recognition: A Survey," *Pattern Recognition Letters*, vol. 103, pp. 1-9, Feb. 2018.
- [4] Ma, R., Kim, D., McCormick, M., Coleman, T.P., & Rogers, J.A., "A stretchable electrode array for non-invasive, skin-mounted measurement of electrocardiography (ECG), electromyography (EMG) and electroencephalography (EEG)." 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 6405-6408, 2010.
- [5] W. Chen, Z. Wang, H. Xie and W. Yu, "Characterization of Surface EMG Signal Based on Fuzzy Entropy," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 2, pp. 266-272, June 2007.
- [6] O. D. Lara, M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Communications Surveys & Tutorials* 15, vol. 3, pp. 1192-1209, Nov. 2012.
- [7] B. Xie, B. Li, A. Harland, "Movement and Gesture Recognition Using Deep Learning and Wearable-sensor Technology", in 2018 International Conference on Artificial Intelligence and Pattern Recognition. ACM, pp.26-31, 2018.
- [8] M. Ahsan, M. Ibrahimy, and O. Khalifa, "Electromyography (EMG) signal-based hand gesture recognition using artificial neural network," in *Int. Conf on Mechatronics*. (Kuala Lumpur, Malaysia), pp. 1-6, IEEE, 2011.
- [9] Oskoei M. A. Oskoei and H. Hu, "Support vector machine-based classification scheme for myoelectric control applied to upper limb," *IEEE Transactions on biomedical engineering*, vol. 55, pp. 1956-1965, Mar. 2008.
- [10] U. Côté-Allard, C L. Fall, A. Drouin, et al. "Deep learning for electromyographic hand gesture signal classification using transfer learning." *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, IEEE, pp.760-771, 2019.
- [11] Y. Wan, Z. Han, J. Zhong, G. Chen. "Pattern recognition and bionic manipulator driving by surface electromyography signals using convolutional neural network." *International Journal of Advanced Robotic Systems*, PLOS ONE, 2018.
- [12] M. Atzori, A. Gijsberts, and C. Castellini, et al. "Electromyography data for non-invasive naturally-controlled robotic hand prostheses." *Scientific data*, 2014.
- [13] M. Atzori, A. Gijsberts, and C. Castellini, et al. "Characterization of a benchmark database for myoelectric movement classification.", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23(1): p73-83, 2014.
- [14] S. Ioffe, C. Szegedy, 'Batch normalization: Accelerating deep network training by reducing internal covariate shift.' *arXiv preprint*, 2015.
- [15] S. Pizzolato et al., "Comparison of six electromyography acquisition setups on hand movement classification tasks," *PloS one*, vol. 12, no. 10, p. e0186132, 2017
- [16] S. Erik and E. Kevin, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use", *Journal of Rehabilitation Research and Development* 48, 6 (2011), 643–660.
- [17] P. Xia, J. Hu, Y. Peng, "EMG-Based Estimation of Limb Movement Using Deep Learning With Recurrent Convolutional Neural Networks," *Artif. Organs* 2018, 42, E67–E77.
- [18] H.M. Shim, S. Lee, "Multi-Channel Electromyography Pattern Classification Using Deep Belief Networks for Enhanced User Experience," *J. Cent. South Univ.* 2015, 22, p1801–1808.
- [19] C. P. Robinson, B. Li, Q. Meng., & M. Pain, "Effectiveness of surface electromyography in pattern classification for upper limb amputees." 2018 International Conference on Artificial Intelligence and Pattern Recognition pp. 107-112, 2018.
- [20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *arXiv preprint arXiv:1502.03167*, 2015