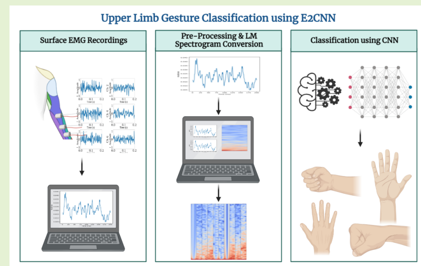
DOI:
10.1109/JSEN.2023.3255408

E2CNN: An Efficient Concatenated CNN for Classification of Surface EMG Extracted from Upper-Limb

Muhammad Farrukh Qureshi, Zohaib Mushtaq, Muhammad Zia ur Rehman, and Ernest Nlandu Kamavuako

***Abstract*—Surface electromyography are bioelectrical indicators that emerge during muscle contraction and have been widely used in a variety of clinical applications. Several prosthetic control applications can benefit from analysis based on the classification of sEMG signals. However, for the real-time application of upper-limb prosthesis, EMG-based systems need robust performance and rapid response behaviour. In this study, we propose an efficient concatenated convolutional neural network (E2CNN) for classification of sEMG extracted from upper-limb. We have tested and validated the performance of the proposed E2CNN on two datasets: a longitudinal dataset comprised of ten able-bodied (healthy) subjects as well as six transradial amputee subjects and spanned the data collected for a period of seven days; and the publicly available NinaPro DB1 dataset. The raw sEMG signals are converted to Log-Mel (LM) spectrograms. This model combines input layers with the output of each convolutional block using concatenation layers. The proposed efficient concatenated CNN (E2CNN) when applied to log-Mel spectrogram-based images, provides a good response time with high-performance accuracy of 98.31% ± 0.5% and 97.97% ± 1.41% for both able-bodied and amputee subjects. When applied to NinaPro DB1, the proposed E2CNN has attained a mean accuracy of 91.27%, an increase by 24.67% with respect to baseline CNN model. The results show that the achieved results are comparable to those obtained using SSAE and other CNN models; however, E2CNN is associated with reduced training and prediction time, making it a potential candidate for real-time classification of sEMG based on LM spectrogram images.**



***Index Terms*—surface electromyography, convolutional neural network, concatenation, log-mel spectrogram, ninapro db1,**

I. Introduction

ELECTROMYOGRAPHY, also known as EMG, is a biological signal that is frequently utilised for the purpose of recognising human motor gestures. This is an essential component of human-computer interaction systems. EMG signals have been the subject of extensive research and have been implemented as a control input for upper-limb prostheses, assistive wheel chairs, assistive humanoid robots, and meal assistive robots [1]–[4]. The control method used in the development of a prosthesis determines the device's functionality, ease of use, and acceptability. EMG-based control systems for upper-limb prosthesis have been widely studied, however implementing them in real-time with high multi-class accuracy is still a challenge. Several studies have used offline machine learning [5]–[7] and deep learning techniques [8]–[15] to classify EMG signals for upper-limb prosthesis.

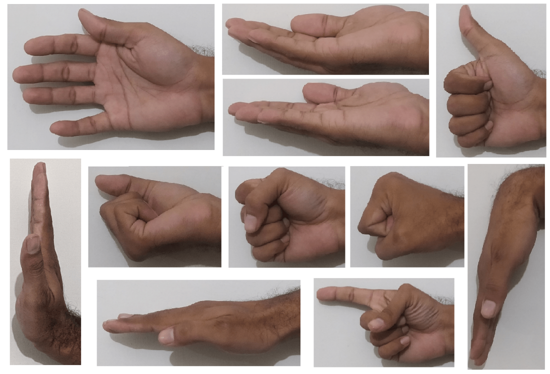
Rehman et al. [8] used a single-layered Convolutional Neural Network (CNN) on a multiday sEMG dataset that was captured for consecutive 15 days. The accuracy that was attained with CNN was 97.60%±1.99. In [13], Convolutional Neural Networks (CNN) and Multilayer Perceptron (MLP) were applied in order to categorise the 9 predetermined gestures. The average accuracy of CNN's gesture recognition is 99.47%, while the mean accuracy of MLP's gesture recognition is 98.42%. In their study, Huang and Chen [16] used the Ninapro database, which included the sEMG data of forty subjects with fifty gestures. The accuracy of their suggested technique, which included a spectrogram, a convolutional neural network (CNN), and a ling-short term memory (LSTM), was 80.929% for the basic hand movements. Using a method that is based on the Short Time Fourier Transform (STFT) representation of EMG data and convolutional neural networks (CNN), Sengur et al. [17] were able to attain an accuracy of 96.69%. In [18], the authors have suggested two streams-CNN

Manuscript received ....; revised .....; accepted .... (Corresponding author: Muhammad Farrukh Qureshi).

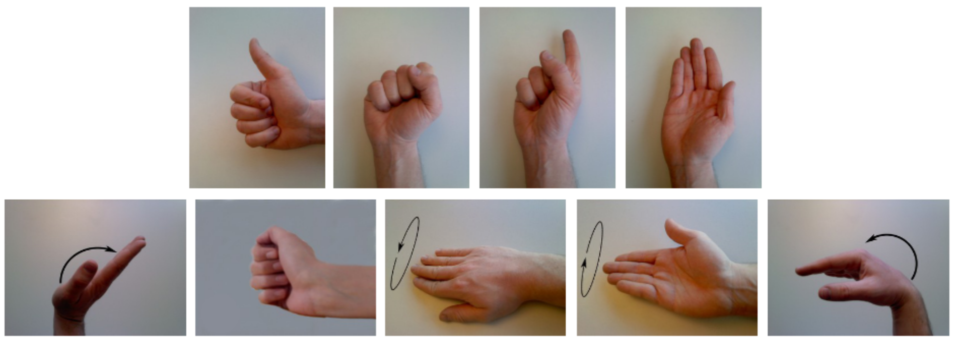
Muhammad Farrukh Qureshi and Zohaib Mushtaq are with the Department of Electrical Engineering, Riphah International University, Islamabad, Pakistan (e-mail: muhammad.farrukh@riphah.edu.pk; zohaib.mushtaq@riphah.edu.pk).

Muhammad Zia ur Rehman is with the Department of Biomedical Engineering, Riphah International University, Islamabad, Pakistan. (e-mail: ziaur.rehman@riphah.edu.pk).

Ernest Nlandu Kamavuako is with the Centre for Robotics Research, Department of Engineering, King's College London, London WC2G 4BG, UK (e-mail: ernest.kamavuako@kcl.ac.uk).



(a)



(b)

Fig. 1. Gestures used in this study (a) Hand Gestures utilized from first dataset: Rest, Opened Hand, Closed Hand, Pronation, Supination, Fine grasp, Side grip, Flex hand, Extend hand, Thumb up, and Pointer; (b) Hand gestures utilized from NinaPro DB1: Thumb up, Hand close, Hand open, Pointer, Supination, Pronation, Flexion, Extension, and Wrist extension with closed hand.

(TS-CNN) to acquire significant features from raw EMG data using multiple scales, as well as estimate the motion that is created during elbow flexion and extension. The results that were obtained were 81%±0.06, 71%±0.06, and 80%±0.13 for the estimate of the joint angle, and 78%±0.05, 79%±0.07, and 71%±0.13 for the estimation of the velocity, during isotonic contractions, isokinetic contractions, and dynamic contractions, respectively. In [19], the authors have tested the performance of Deep Learning based Pattern Recognition (DLPR) framework on features extracted from publicly available dataset NinaPro databases and achieved an accuracy of 92.18%, 91.56%, and 84.98% on DB1, DB2, and DB3 respectively. The same work also experimented on Short-Term Fourier Transform (STFT) spectrogram images from publicly available dataset NinaPro databases and achieved an accuracy of 70.14%, 74.89%, and 65.87% on DB1, DB2, and DB3 respectively. In [20], the authors have identified four hundred EMG spectrograms using hybrid deep-learning approaches that were based on transfer learning models such as AlexNet, GoogleNet, and ResNet18. The accuracy for hybrid classification using AlexNet, GoogleNet, and ResNet18 were 99.17%, 95.83%, and 93.33%, respectively.

In this paper, we propose a rapid, responsive Deep Neural Network (DNN) that has been applied to surface EMG (sEMG) spectrogram images. This DNN is based on the Convolutional Neural Network (CNN) that takes Log-Mel (LM) spectrogram images (extracted from EMG signals) as an input. The significant contributions of this study are as follows:

1) A custom non-sequential concatenated convolutional neural network is proposed for classification of upper limb.
2) Utilization of Log-Mel Spectrogram images for sEMG signals for upper limb gesture classification.
3) Analysis of the proposed E2CNN on a longitudinal sEMG dataset comprised of ten able-bodied and six amputee subjects.
4) Validation of proposed technique on publicly available NinaPro DB1 dataset.
5) Performance comparison of proposed E2CNN on previous studies implemented on both datasets.

The rest of the paper is organised as follows: we present the methodology containing the description of datasets, preprocessing steps, and details of proposed DNN applied for classification in Sec. II, discuss the results in Sec. III, and finally provide the conclusion and future work in Sec. V.

II. Methodology

A. Experimental Dataset and Setup Description

We have used two datasets in this study: (i) First dataset based on longitudinal dataset from previous work [5]; and (ii) Second dataset is publicly available NinaPro Dataset 1 (DB1). Details of both datasets are discuss in next subsections:

1) First Dataset: The first dataset we utilize is from previous work [5]. That dataset consists of 10 able-bodied (healthy) and

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2023.3255408

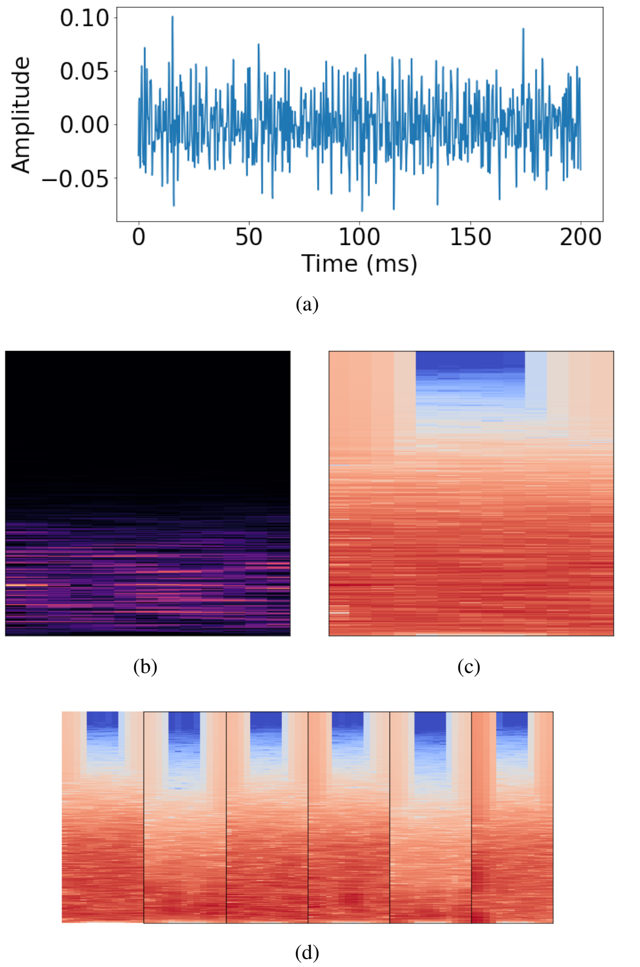


Fig. 2. Conversion of signals from first dataset to LM spectrogram images: (a) A windowed signal from one channel (b) Short-Term Fourier Transform (STFT) of the windowed signal (c) Logarithmic-Mel (LM) Spectrogram (d) A combined image of all six channels as LM spectrogram image

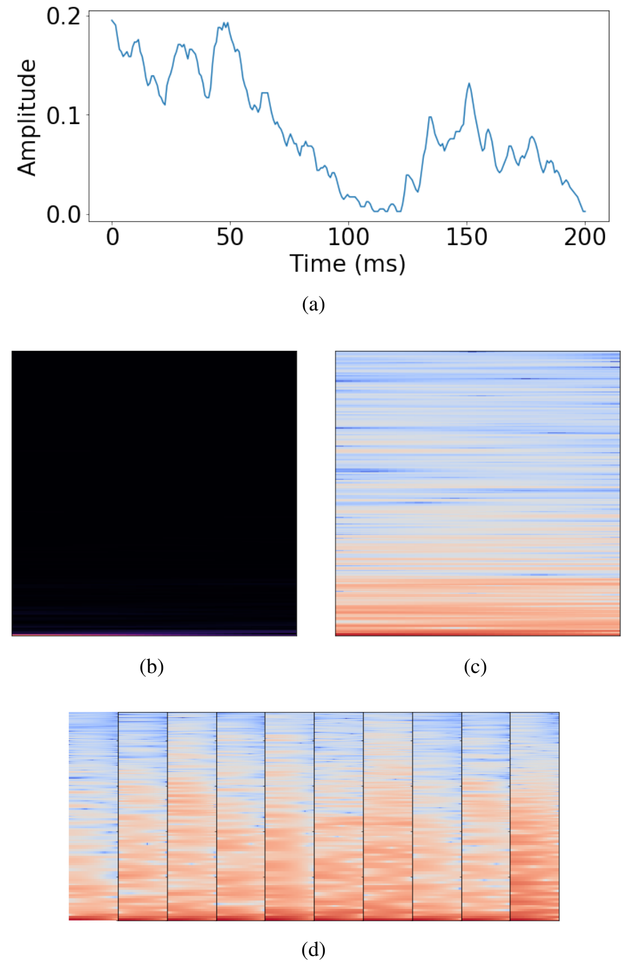


Fig. 3. Conversion of NinaPro DB1 signals to LM spectrogram images: (a) A windowed signal from one channel (b) Short-Term Fourier Transform (STFT) of the windowed signal (c) Logarithmic-Mel (LM) Spectrogram (d) A combined image of all ten channels as LM spectrogram image

6 trans-radial amputee subjects. All the subjects were male and their mean ages were 24.5 ± 0.22 and 34.8±0.32 years for able-bodied and amputee subjects, respectively. An ethical approval was taken prior to data collection (Approval No.: Ref No. Riphah/RCRS/REC/000121/20012016). The authors simultaneously recorded the surface as well as intramuscular EMG. Six surface and six intramuscular electrodes were used. An 8 kHz sample rate is used to obtain data from six channelled surface EMG signals. Eleven distinct gestures (including rest) were performed by all subjects with four repetition of each movement. The data was collected for seven consecutive days. The hand gestures performed by participants are: rest, open hand, closed hand, pronation, supination, fine grasp, side grip, flex hand, extend hand, thumb up, and pointer; and are illustrated in Fig. 1 (a). For the purpose of this study, we have only utilized the surface EMG dataset.

2) Second Dataset: The first NinaPro database (DB1) includes 27 intact subjects [21]. This dataset includes 52 distinct movements carried out by 27 subjects, with each movement having been performed ten times. These movements are categorised into three distinct types of exercises: movements of the fingers; grasping and functional movements; and isometric, isotonic hand configurations and basic wrist movements. The data was collected using 10 surface EMG electrodes. For cohesion with first dataset, we have only utilized 9 hand movements from NinaPro DB1. The gestures utilized are: Thumb up, Hand close, Hand open, Pointer, Supination, Pronation, Flexion, Extension, and Wrist extension with closed hand. Fig. 1 (b) shows these gestures utilized from NinaPro DB1 for this study.

B. Preprocessing Technique

The preprocessing technique is applied to both datasets. EMG based pattern recognition techniques require smaller intervals of signals to extract useful classification features. However, when the processing window decreases, the performance drops significantly [22]; hence optimum duration is limited between $150ms$ to $250ms$ [23], [24]. We divide each raw EMG signal into smaller intervals (windows) with a length of $200ms$ and an overlapping increment of $29ms$. For first dataset, each EMG signal results in 4400×6 windowed signals with data in six channels for 11 hand gestures. This was done

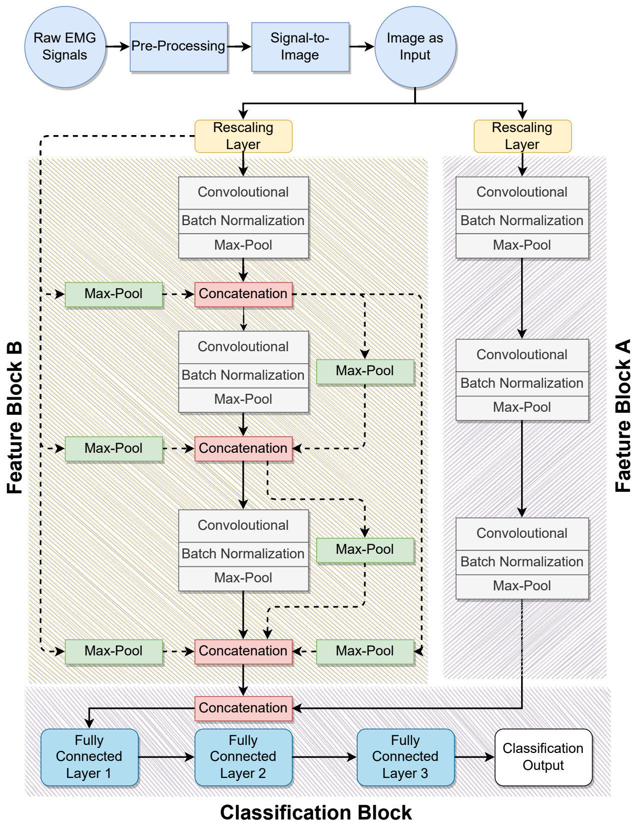


Fig. 4. Block Diagram of the proposed strategy implemented in this paper

for all seven days for each subject and resulted in 30800×6 windowed signals for each subject. For second dataset, each EMG signal results in 1791×10 windowed signals with data in ten channels for 9 hand gestures. To get better interpretation of one dimensional, stochastically distributed EMG signal, the windowed sEMG signals were converted into Log-Mel Spectrograms using librosa package in python.

C. Signals to Log-Mel Spectrograms

Let $s_w(n)$ be an windowed EMG signal, with the length L, sampling frequency f_{s_w} in Hertz as shown in Fig. 2(a). Then its Short-Term Fourier Transform (STFT) S_w will be:

$$S_w(x,y) = \sum_{n-0}^{N-1} s_w(n+xH) \cdot w(n) \cdot e^{-\iota 2\pi y \frac{n}{N}} \tag{1}$$

where, $H \in \mathbb{N}$ is hop length, $w : [0 : N-1] \in \mathbb{R}$ is Hann window $w = 0.5 - 0.5\cos(\frac{2\pi n}{N-1})$, $N \in \mathbb{N}$ is length of w, $x \in [0 : \frac{L-N}{H}]$ and $y \in [0 : \frac{N}{2}]$ indicates time and frequency indices, respectively.

The STFT spectrogram of S_w as illustrated in Fig. 2(b) can be achieved by 2:

$$S_{STFT}(x,y) = |S_w(x,y)|^2. \tag{2}$$

The relationship between the Mel spectrum and the frequency is $f_{mel} = 2959 \times \log_{10}(1 + \frac{f}{700})$. The Log-Mel spectrogram can be estimated using 3:

$$S_{LM}(x,y) = \sum_{f(y)=f_c(x-1)}^{f_c(x+1)} \log_{10}(M(x,y) \cdot S_{STFT}(x,y)) \tag{3}$$

where, $M(x,y)$ is Mel filter banks and can be computed from 4:

$$M(x,y) = \begin{cases} \frac{f(y)-f_c(x-1)}{f_c(x)-f_c(x-1)} & \text{for } f_c(x-1) \le f(y) < f_c(x) \\ \frac{f(y)-f_c(x+1)}{f_c(x)-f_c(x+1)} & \text{for } f_c(x) \le f(y) < f_c(x+1) \\ 0 & \text{others.} \end{cases} \tag{4}$$

where $f(y)$ is linear frequency and $f_c(x) = x \cdot \delta f_{mel}$ are centre frequencies on Mel-scale. Figure 2(c) illustrates an image of LM spectrogram for the EMG windowed signal s_w.

Each windowed signal is individually converted into a LM spectrogram, this process is repeated for each channel resulting in six LM spectrograms. These six LM spectrograms are then combined vertically and converted into an image as shown in Fig. 2(d). As a result, we get 30800 EMG images as LM spectrograms as our input dataset to CNN for each subject. The process for first dataset is illustrated in Fig. 2. Similar to first dataset, the same technique is applied to second dataset and we get 1791 EMG images as LM spectrograms for each subject with ten LM spectrograms combined vertically in each image. The process is depicted in Fig. 3.

D. Efficient Concatenated Convolutional Neural Network (E2CNN) Architecture

The experimental process in this study involves a deep CNN with max-pooling functions [25]. Let's name our model as Efficient-Concatenated CNN (E2CNN). The proposed E2CNN is implemented on surface EMG based Log-Mel (LM) spectrogram images with the input size of 128×128.

We have designed the proposed E2CNN to reduce the number of parameters while maintaining performance. This is achieved by employing convolutional layers with large kernel sizes and max-pool layers with large pooling sizes. Consequently, the number of parameters is reduced, but numerous features are lost. Concatenation layers are used to mitigate with feature loss. Concatenation layers append the original input to the output of subsequent convolutional and max-pool layers. This again combines the features that might be overlooked during heavy padding and strides.

The total number of epochs in the network in E2CNN is 100, and the batch size is 32. For optimization, Adam optimiser is used. Except for the final layer, which uses the Softmax activation function, all layers of the proposed E2CNN use the Rectified Linear Unit (ReLU) activation function. The model is divided into three sections: Feature Block A, Feature Block B, and the Classification Block. Using the rescaling layer (RL), the input images are resized to $(0,1)$. The following is more information about the proposed E2CNN:

Feature Block A:

- Layer 1a: The image from RL is fed to the model's first layer (L1a) having 16 filters, each with a 7×7

receptive field. It is followed by a 2×2 strided max-pooling function with batch normalisation in between. The activation function utilized in this layer is ReLU.

- Layer 2a: The 2nd layer (L2a) consists of double filters (32) in comparison to L1a with a reduced receptive field of $5{\times}5$ with a max-pooling with strides of $2{\times}2$ performed using ReLU as the activation function. Similar to L1a, batch normalisation is used.
- Layer 3a: The 3rd layer (L3a) is made up of double filters (64) with respect to L2a with a further reduced receptive field of 3×3. A max-pooling function with strides of 2×2 using ReLU as the activation function is applied following a batch normalisation layer.

Feature Block B:

- Layer 1b+Concatenation: The input image passed through RL is fed to this layer too. This layer (L1b) has the same dimension and parameters as L1a. However, the output from RL is passed through another max-pooling layer, and combined with the output of L1b using a concatenation layer (Concat1).
- Layer 2b+Concatenation: The output from Concat1 becomes the input to this layer (L2b). The layer has the same dimensions as L2a. Similar to L1b, the output from RL, Concat1, and output of L2b is combined using another concatenation layer (Concat2). Max-pooling layers are used to match the input requirement of Concat2.
- Layer 3b+Concatenation: The 3rd layer (L3b) has the same parameters and dimensions as L3a. The output of L3b, Concat1, Concat2, and RL are combined using another concatenation layer (Concat3). The output of Concat1, Concat2, and RL are passed through individual max-pooling layers to match the input size required by Concat3.

Classification Block:

- Concatenation Layer: The output of Concat3 and L3a is concatenated using concatenation layer (Concat4) and the output is provided to the next layers. Max-pooling is not required for Concat4.
- Layer 4 & Layer 5: Both of these layers are fully connected (fc) dense layers, and each consists of 12 hidden units. ReLU is again used as an activation function for both of these layers.
- Layer 6: The final layer is the another fc dense layer with Softmax activation function. The final layer is composed of hidden units equal to the number of classes. Therefore, it is composed of 11 hidden units for the first dataset. This layer is modified for second dataset and made up of 9 hidden units.

E. Statistical Tests and Performance Evaluation Metric

The two-way layout of Friedman's tests was used to evaluate the performance comparison. Results from the classification are shown as a mean accuracy and standard deviation. Kruskal-Willis test is used to evaluate the performance comparison on the second dataset due to different number of classes. All p-values below 0.05 were deemed statistically significant.

TABLE I
ACCURACY ACHIEVED BY E2CNN ON ABLE-BODIED AND AMPUTEE SUBJECTS FROM FIRST DATASET

Subject	Accuracy	Training Time (s)	Prediction Time (s)
Able-bodied 1	99.31 %	35.522	0.095
Able-bodied 2	97.57 %	39.967	0.084
Able-bodied 3	98.32 %	37.975	0.090
Able-bodied 4	98.41 %	38.772	0.097
Able-bodied 5	97.20 %	40.247	0.098
Able-bodied 6	97.50 %	38.246	0.100
Able-bodied 7	99.35 %	34.984	0.095
Able-bodied 8	98.78 %	38.784	0.089
Able-bodied 9	97.67 %	39.124	0.081
Able-bodied 10	99.23 %	40.285	0.088
Mean ± SD	**98.31 % ± 0.77 %**	**38.062 ± 1.781**	**0.0935 ± 0.0050**
Amputee 1	98.10%	39.732	0.097
Amputee 2	95.95%	40.45	0.083
Amputee 3	99.94%	41.278	0.102
Amputee 4	97.88%	37.158	0.113
Amputee 5	98.12%	39.901	0.100
Amputee 6	98.38%	36.814	0.116
Mean ± SD	**97.97 % ± 1.41**	**39.655 ± 1.541**	**0.0988 ± 0.0107**

Accuracy is used as the primary performance evaluation factor. Accuracy for multi-class classification is provided by 5:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

where, TP represents the True Positive and TN represents True Negative values. FN and FP represents False positive and False Negative values, respectively.

III. EXPERIMENTAL RESULTS

A. Performance Assessment of Proposed Method

1) First Dataset: The first dataset comprises of able-bodied and amputee people; thus, the dataset is categorised based on subject type. The data is then divided into an 80-20 structure, with 80% of the data being used for training and 20% being utilised for testing. For seven days, training and testing were carried out on every subject, after which the mean accuracy, training time, and prediction time were determined. Mean Accuracy, Training Time, and Prediction Time are shown in Table I for both able-bodied and amputee subjects, respectively. E2CNN was able to attain an accuracy of 98.31% ± 0.77% for able-bodied participants and 97.97% ± 4.41% for amputee subjects, respectively. Fig. 6 (a) illustrates the confusion matrix of the suggested model for able-bodied subjects, while Fig. 6 (b) illustrates the confusion matrix for amputee subjects. Here, the True and Predicted labels refer to the real hand motions, while the Predicted labels refer to the relevant labels that were predicted using E2CNN.

2) Second Dataset (NinaPro DB1): The second dataset consists of 27 able-bodied subjects; however, we have only utilized 18 subjects to validate the performance of E2CNN. Similar to first dataset, data is split into 80-20 format with 80% assigned for training and 20% used for testing. Training

TABLE II
PERFORMANCE OF E2CNN ON NINAPRO DB1

Subject	Accuracy	Training Time (s)	Prediction Time (s)
Able-bodied 1	87.73	49.202	0.056
Able-bodied 2	90.38	43.510	0.062
Able-bodied 3	92.71	46.587	0.063
Able-bodied 4	88.63	45.582	0.069
Able-bodied 5	94.22	47.113	0.070
Able-bodied 6	93.34	44.576	0.071
Able-bodied 7	94.71	43.351	0.068
Able-bodied 8	95.73	45.277	0.062
Able-bodied 9	90.74	47.960	0.052
Able-bodied 10	90.31	44.327	0.069
Able-bodied 11	89.97	41.102	0.063
Able-bodied 12	91.04	49.579	0.059
Able-bodied 13	92.14	43.684	0.058
Able-bodied 14	93.02	41.039	0.060
Able-bodied 15	88.25	42.856	0.072
Able-bodied 16	91.72	46.981	0.069
Able-bodied 17	92.71	41.025	0.065
Able-bodied 18	91.26	40.513	0.062
Mean ± SD	**91.2575 % ± 2.33 %**	**44.829 ± 2.738**	**0.06415 ± 0.005914**

and testing were performed for each subject based on the mean accuracy, training time, and prediction time. Table II show the Accuracy, Training Time, and Prediction Time for able-bodied subjects from NinaPro DB1. E2CNN achieved an accuracy of 91.2757% ± 2.33% for this dataset. Figure 6 (c) shows the confusion matrix of the proposed model on able-bodied subjects from NinaPro DB1. Here, the True and Predicted labels refer to the real hand motions, while the Predicted labels refer to the relevant labels that were predicted using E2CNN.

B. Comparison of Results with Previous and Baseline Methods

In this study, we have employed two datasets: (i) a longitudinal data set collected for seven consecutive days from able-bodied and amputee subjects, and (ii) Publicly available NinaPro DB1 dataset.

Since the first dataset was previously analysed with a different CNN [9] and Stacked Sparse Autoencoders (SSAE) [14], therefore we are comparing our results with these previous results based on the same dataset. E2CNN achieved a higher mean classification accuracy than SSAE for able-bodied and amputee subjects, with 98.31% ± 0.77% and 97.97% ± 1.41%, respectively. However, the number of parameters and processing time for SSAE were not, however, calculated. In addition, the same data set analysed by a different CNN [9] by the same authors achieved a higher level of accuracy than E2CNN. This CNN [9] performed worse in processing with a longer training and prediction time. Furthermore, CNN [9] has 207,275 parameters, while E2CNN has only 86,119 parameters. The table III compares the E2CNN model to the SSAE and CNN [9] models. It can be seen that E2CNN provides a better training prediction time with less computational cost for the SSAE and CNN models. The training and prediction time E2CNN took are $47.5ms$ and $116.25\mu s$ per sample fed to E2CNN.

We have tested and validated the performance of our proposed E2CNN on NinaPro DB1. Several studies have

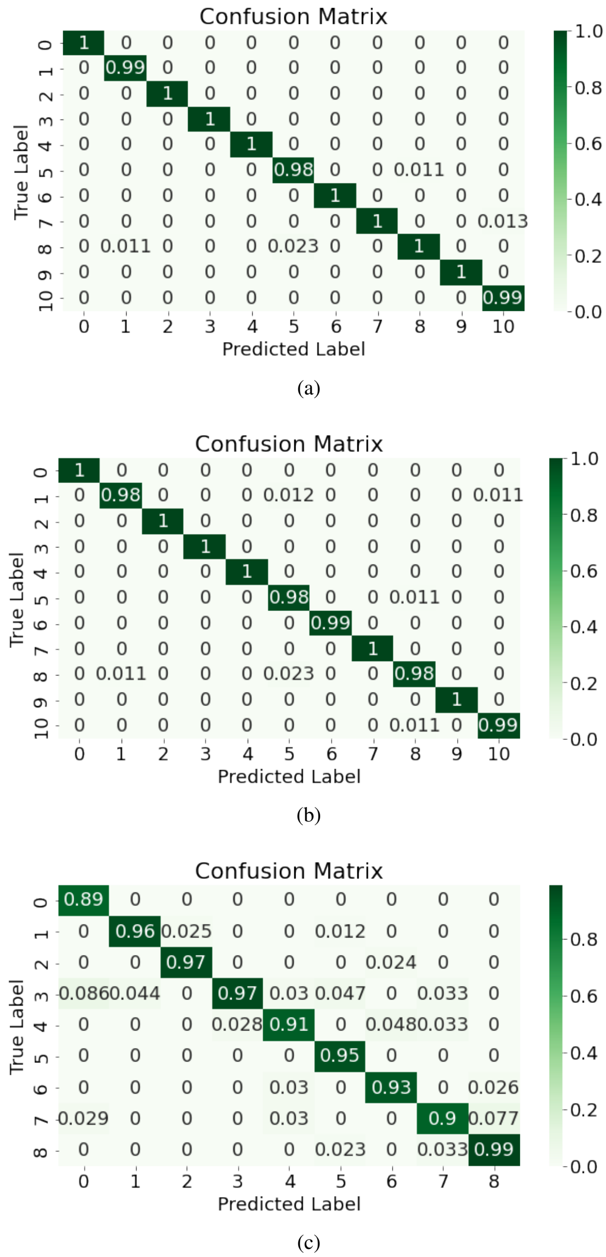


Fig. 6. Confusion matrices for (a) Able-bodied and (b) Amputee subjects from first dataset; (c) NinaPro DB1

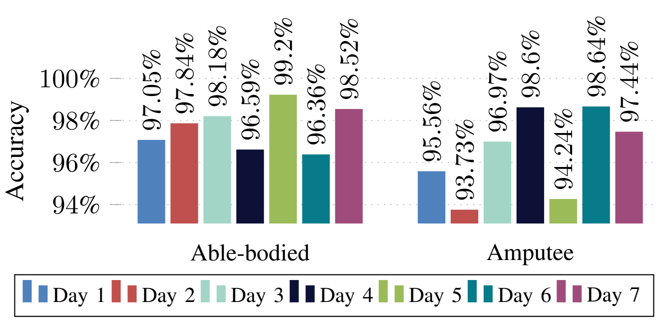


Fig. 5. Performance of E2CNN on Each Day for Able-bodied and Amputee Subjects from first dataset

TABLE III
COMPARISON OF E2CNN (BOLD FONT) WITH PREVIOUS STUDIES IMPLEMENTED ON THE FIRST DATASET

Classifier	Mean Accuracy ± SD Able-bodied	Mean Accuracy ± SD Amputee	Parameters	Training Time (s) for 3600 samples	Prediction Time (s) for 800 samples
SSAE [14]	96.78 % ± 6.40 %	86.89 % ± 11.35 %	-	-	-
CNN [9]	99.42 % ± 0.42 %	98.00 % ± 0.58 %	207,275	1042.355 ± 2.820	0.741 ± 0.050
E2CNN	**98.31 % ± 0.77 %**	**97.97 % ± 1.41 %**	**86,119**	**38.062 ± 1.781**	**0.093 ± 0.005**

TABLE IV
COMPARISON OF E2CNN (BOLD FONT) WITH PREVIOUS STUDIES IMPLEMENTED ON THE SECOND DATASET (NINAPRO DB1)

Reference	Year	Classes	Classifier	Accuracy
[11]	2016	52	CNN	66.60%
[26]	2016	8	CNN	78.90%
[27]	2016	6	CNN	60.00%
[28]	2017	10	CNN	88.42%
[29]	2019	52	CNN	85.00%
[30]	2020	5	Compact CNN	66.90%
[31]	2021	17	Dual-View CNN	65.43%
[19]	2022	53	DLPR	92.18%
[32]	2022	29	Deformable CNN	81.80%
[33]	2022	12	Dual-View CNN	86.29%
This Work	**2023**	**9**	**E2CNN**	**91.27%**

used this dataset to validate their works. Table IV shows the comparison of E2CNN with other well-known works. It can be seen that E2CNN has achieved a mean accuracy of 91.27% when averaged over 18 subjects. The proposed E2CNN has increased the accuracy by 24.67% with respect to baseline CNN model [11]. Furthermore, E2CNN and DLPR [19] has an accuracy difference of less than 1% which suggests that both models perform comparably well. DLPR requires extraction of features from EMG signals, while E2CNN utilizes LMS images. The work [19] also explored its performance on STFT images and achieved an accuracy of 70.14%. LMS representation enables E2CNN to capture both temporal and spectral information from the input signal and achieved an accuracy of 91.27%.

IV. DISCUSSION

The proposed model for sEMG gesture classification is based on the Log-Mel Spectrogram (LMS) images generated from the raw EMG signals. The LMS approach has been widely used in speech and sound classification as it better represents the human perception of sound. We have converted the raw sEMG signals to LMS images. The LMS images provide a more compact and informative representation of the raw EMG signal. However, the utilization of images for gesture classification also results in an increase in the number of parameters of the model. This increase in the number of parameters does lead to longer training and prediction times, which can limit the practical application of the model. To address this issue, the number of filters in the model was reduced while the filter sizes were increased, which resulted in a lower number of parameters and improved training and prediction times. Despite the improvement in the training and prediction times, the reduction in the number of filters resulted in a decrease in accuracy. To mitigate this, the concatenation layers were introduced to the model. These concatenation layers combined the input images from all preceding layers to each succeeding layer, resulting in a more comprehensive representation of the data. The introduction of the concatenation layers resulted in improved accuracy and reduced the trade-off between training and prediction times, making the model a suitable choice for real-time EMG gesture classification.

The results of the statistical analysis on the first dataset showed that E2CNN outperformed previously published work with faster training and prediction times, and comparable accuracy with a p-value of 0.014 from the Kruskal-Willis test and an H-statistic of 14.19. This demonstrates the effectiveness and efficiency of the E2CNN model in performing classification tasks. In order to further validate the performance of the model, we used the NinaPro DB1 dataset, which contains 53 movements, but we only used 9 movements that were similar to the movements available in the first dataset. We compared our results with earlier published works that used DB1 and took into consideration that some studies used all available movements while others used a lower number. The statistical analysis on the NinaPro DB1 dataset was conducted using the Kruskal-Willis test due to different pre-processing and different numbers of movements. The results showed that E2CNN outperformed other classifiers with an H-statistic of 25.92 and a p-value of 0.000002355 ($p < 0.05$), demonstrating its superior performance. The results of the statistical analysis on the first dataset and the NinaPro DB1 dataset show that the proposed E2CNN model has improved performance compared to previous work with faster training and prediction times and high accuracy.

It is essential to consider that an increase in the number of classes in a dataset has a direct impact on the accuracy of the model. This issue will be the focus of future studies, as the goal is to enhance the accuracy of the E2CNN model even with an increased number of classes. These future studies will provide a deeper understanding of the capabilities of the E2CNN model in EMG gesture classification tasks and provide valuable insights for future developments in the field.

V. CONCLUSION

In this study, the performance of a proposed E2CNN was investigated for its application in sEMG classification based on Log-Mel spectrogram images. We conducted tests on the proposed E2CNN using two different datasets: a longitudinal dataset that included ten able-bodied (healthy) subjects as well as six transradial amputee subjects and spanned the data collected for a period of seven days; and the NinaPro DB1 dataset. In this study, we were able to show that Log-Mel spectrogram EMG images provide much more information and illustrate better results than a raw EMG signal does. The achieved results are comparable to those obtained using SSAE and other CNN models. Furthermore, E2CNN is associated with reduced training and prediction time, making it a potential candidate for real-time classification of sEMG based on LM spectrogram images.

REFERENCES

[1] H. M. Al-Angari, G. Kanitz, S. Tarantino, and C. Cipriani, "Distance and mutual information methods for EMG feature and channel subset selection for classification of hand movements," *Biomedical Signal Processing and Control*, vol. 27, pp. 24–31, May 2016.

[2] Y. Geng, Y. Ouyang, O. W. Samuel, S. Chen, X. Lu, C. Lin, and G. Li, "A Robust Sparse Representation Based Pattern Recognition Approach for Myoelectric Control," *IEEE Access*, vol. 6, pp. 38326–38335, 2018.

[3] M. Awais, C. Chen, X. Long, B. Yin, A. Nawaz, S. F. Abbasi, S. Akbarzadeh, L. Tao, C. Lu, L. Wang, R. M. Aarts, and W. Chen, "Novel framework: Face feature selection algorithm for neonatal facial and related attributes recognition," *IEEE Access*, vol. 8, pp. 59100–59113, 2020.

[4] X. Li, O. W. Samuel, X. Zhang, H. Wang, P. Fang, and G. Li, "A motion-classification strategy based on sEMG-EEG signal combination for upper-limb amputees," *Journal of NeuroEngineering and Rehabilitation*, vol. 14, p. 2, Jan. 2017.

[5] A. Waris, I. K. Niazi, M. Jamil, K. Englehart, W. Jensen, and E. N. Kamavuako, "Multiday evaluation of techniques for EMG-based classification of hand motions," *IEEE journal of biomedical and health informatics*, vol. 23, no. 4, pp. 1526–1534, 2018.

[6] M. Akmal, M. F. Qureshi, F. Amin, M. Z. Ur Rehman, and I. K. Niazi, "SVM-based real-time classification of prosthetic fingers using myo armband-acquired electromyography data," in *2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE)*, pp. 1–5, 2021.

[7] J. J. Bird, J. Kobylarz, D. R. Faria, A. Ekárt, and E. P. Ribeiro, "Cross-domain MLP and CNN transfer learning for biological signal processing: EEG and EMG," *IEEE Access*, vol. 8, pp. 54789–54801, 2020.

[8] M. Zia ur Rehman, A. Waris, S. O. Gilani, M. Jochumsen, I. K. Niazi, M. Jamil, D. Farina, and E. N. Kamavuako, "Multiday EMG-Based Classification of Hand Motions with Deep Learning Techniques," *Sensors*, vol. 18, p. 2497, Aug. 2018.

[9] M. F. Qureshi, Z. Mushtaq, M. Z. ur Rehman, and E. N. Kamavuako, "Spectral Image-Based Multiday Surface Electromyography Classification of Hand Motions Using CNN for Human–Computer Interaction," *IEEE Sensors Journal*, vol. 22, pp. 20676–20683, Nov. 2022.

[10] Y. Cheng, G. Li, M. Yu, D. Jiang, J. Yun, Y. Liu, Y. Liu, and D. Chen, "Gesture recognition based on surface electromyography-feature image," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 6, p. e6051, 2021.

[11] M. Atzori, M. Cognolato, and H. Müller, "Deep Learning with Convolutional Neural Networks Applied to Electromyography Data: A Resource for the Classification of Movements for Prosthetic Hands," *Frontiers in Neurorobotics*, vol. 10, Sept. 2016.

[12] W. Wei, Q. Dai, Y. Wong, Y. Hu, M. Kankanhalli, and W. Geng, "Surface-Electromyography-Based Gesture Recognition by Multi-View Deep Learning," *IEEE Transactions on Biomedical Engineering*, vol. 66, pp. 2964–2973, Oct. 2019.

[13] L. Liu, N. Zhang, and K. Li, "Research on Gesture Recognition of Surface EMG Based on Machine Learning," in *Biometric Recognition* (W. Deng, J. Feng, D. Huang, M. Kan, Z. Sun, F. Zheng, W. Wang, and Z. He, eds.), Lecture Notes in Computer Science, (Cham), pp. 276–283, Springer Nature Switzerland, 2022.

[14] M. Zia ur Rehman, S. O. Gilani, A. Waris, I. K. Niazi, G. Slabaugh, D. Farina, and E. N. Kamavuako, "Stacked sparse autoencoders for EMG-based classification of hand motions: A comparative multi day analyses between surface and intramuscular EMG," *Applied Sciences*, vol. 8, no. 7, p. 1126, 2018.

[15] M. Jochumsen, I. K. Niazi, M. Zia ur Rehman, I. Amjad, M. Shafique, S. O. Gilani, and A. Waris, "Decoding Attempted Hand Movements in Stroke Patients Using Surface Electromyography," *Sensors*, vol. 20, p. 6763, Jan. 2020.

[16] D. Huang and B. Chen, "Surface EMG Decoding for Hand Gestures Based on Spectrogram and CNN-LSTM," in *2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI)*, pp. 123–126, Sept. 2019.

[17] A. Sengur, M. Gedikpinar, Y. Akbulut, E. Deniz, V. Bajaj, and Y. Guo, "DeepEMGNet: An Application for Efficient Discrimination of ALS and Normal EMG Signals," in *Mechatronics 2017* (T. Březina and R. Jabłoński, eds.), Advances in Intelligent Systems and Computing, (Cham), pp. 619–625, Springer International Publishing, 2018.

[18] G. Hajian and E. Morin, "Deep Multi-Scale Fusion of Convolutional Neural Networks for EMG-Based Movement Estimation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 486–495, 2022.

[19] S. Pancholi, A. M. Joshi, and D. Joshi, "DLPR: Deep Learning-Based Enhanced Pattern Recognition Frame-Work for Improved Myoelectric Prosthesis Control," *IEEE Transactions on Medical Robotics and Bionics*, vol. 4, pp. 991–999, Nov. 2022.

[20] S. A. Tuncer and A. Alkan, "Classification of EMG signals taken from arm with hybrid CNN-SVM architecture," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 5, p. e6746, 2022.

[21] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. M. Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Scientific Data*, vol. 1, p. 140053, Dec. 2014.

[22] N. Parajuli, N. Sreenivasan, P. Bifulco, M. Cesarelli, S. Savino, V. Niola, D. Esposito, T. J. Hamilton, G. R. Naik, and U. Gunawardana, "Real-time EMG based pattern recognition control for hand prostheses: A review on existing methods, challenges and future implementation," *Sensors*, vol. 19, no. 20, p. 4596, 2019.

[23] X. Zhang, X. Li, O. W. Samuel, Z. Huang, P. Fang, and G. Li, "Improving the Robustness of Electromyogram-Pattern Recognition for Prosthetic Control by a Postprocessing Strategy," *Frontiers in Neurorobotics*, vol. 11, p. 51, Sept. 2017.

[24] F. Riillo, L. Quitadamo, F. Cavrini, E. Gruppioni, C. Pinto, N. C. Pastò, L. Sbernini, L. Albero, and G. Saggio, "Optimization of EMG-based hand gesture recognition: Supervised vs. unsupervised data preprocessing on healthy subjects and transradial amputees," *Biomedical Signal Processing and Control*, vol. 14, pp. 117–125, Nov. 2014.

[25] Z. Mushtaq, S.-F. Su, and Q.-V. Tran, "Spectral images based environmental sound classification using CNN with meaningful data augmentation," *Applied Acoustics*, vol. 172, p. 107581, Jan. 2021.

[26] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Scientific Reports*, vol. 6, p. 36571, Nov. 2016.

[27] K.-H. Park and S.-W. Lee, "Movement intention decoding based on deep learning for multiuser myoelectric interfaces," in *2016 4th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–2, Feb. 2016.

[28] X. Zhai, B. Jelfs, R. H. M. Chan, and C. Tin, "Self-Recalibrating Surface EMG Pattern Recognition for Neuroprosthesis Control Based on Convolutional Neural Network," *Frontiers in Neuroscience*, vol. 11, p. 379, July 2017.

[29] W. Wei, Y. Wong, Y. Du, Y. Hu, M. Kankanhalli, and W. Geng, "A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface," *Pattern Recognition Letters*, vol. 119, pp. 131–138, Mar. 2019.

[30] L. Chen, J. Fu, Y. Wu, H. Li, and B. Zheng, "Hand Gesture Recognition Using Compact CNN via Surface Electromyography Signals," *Sensors*, vol. 20, p. 672, Jan. 2020.

[31] Z. Yu, J. Zhao, Y. Wang, L. He, and S. Wang, "Surface EMG-based instantaneous hand gesture recognition using convolutional neural network with the transfer learning method," *Sensors*, vol. 21, no. 7, p. 2540, 2021.

[32] H. Wang, Y. Zhang, C. Liu, and H. Liu, "sEMG based hand gesture recognition with deformable convolutional network," *International Journal of Machine Learning and Cybernetics*, vol. 13, pp. 1729–1738, June 2022.

[33] Y. Zhang, F. Yang, Q. Fan, A. Yang, and X. Li, "Research on sEMG-Based Gesture Recognition by Dual-View Deep Learning," *IEEE Access*, vol. 10, pp. 32928–32937, 2022.