2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Decoding of Hand Gestures from Electrocorticography with LSTM Based Deep Neural Network

Jathurshan Pradeepkumar[1], Mithunjha Anandakumar[1], Vinith Kugathasan[1]
Thilina D. Lalitharatne[2], Anjula C. De Silva[1] and Simon L. Kappel[3]

Abstract— **Hand gesture decoding is a key component of controlling prosthesis in the area of Brain Computer Interface (BCI). This study is concerned with classification of hand gestures, based on Electrocorticography (ECoG) recordings. Recent studies have utilized the temporal information in ECoG signals for robust hand gesture decoding. In our preliminary analysis on ECoG recordings of hand gestures, we observed different power variations in six frequency bands ranging from 4 to 200 Hz. Therefore, the current trend of including temporal information in the classifier was extended to provide equal importance to power variations in each of these frequency bands. Statistical and Principal Component Analysis (PCA) based feature reduction was implemented for each frequency band separately, and classification was performed with a Long Short-Term Memory (LSTM) based neural network to utilize both temporal and spatial information of each frequency band. The proposed architecture along with each feature reduction method was tested on ECoG recordings of five finger flexions performed by seven subjects from the publicly available 'fingerflex' dataset. An average classification accuracy of 82.4% was achieved with the statistical based channel selection method which is an improvement compared to state-of-the-art methods.**

I. INTRODUCTION

A Brain-Computer Interface (BCI) is a communication bridge between the brain and external devices, and translates electrical activity in the brain to commands. BCIs have many applications in assistive technology to e.g. paralysed and amputees. Several signals can be used as an input to BCI systems, including Electroencephalography (EEG), Magenetoencephalography (MEG), and Electrocorticography (ECoG). Although EEG is a non-invasive method and enables easy data acquisition, the signals are generally bandwidth limited [1], because of the low-pass filtering caused by the skull. In contrast, ECoG is an invasive method, with the advantage of higher SNR and broader bandwidth of the signals.

When considering frequencies from 0.5 to 200 Hz, previous studies have found that high gamma frequencies (above 65 Hz) are the most descriptive, to distinguish individual hand gestures [2]–[5]. However, the classification accuracy is also influenced by other factors, including complexity of the experiment, electrode location and density, and processing methods and parameters [10]. In previous studies, classifiers such as Naive Bayes (NB) [4], template matching [5], Support Vector Machines (SVM) [3], [6], Linear Program Machines (LPM) [2], Linear Discriminant Analysis (LDA) [1], [7]–[9], Time Variant Linear Discriminant Analysis (TVLDA) [10] and Recurrent Neural Networks (RNN) [11] have been used for gesture classification. Among these, the studies using NB, LDA, SVM and LPM have not utilized the temporal information. Since a gesture comprises a series of motor actions, temporal information in ECoG signals contain vital information for a gesture classifier. Thus, recent studies have used classifiers such as template matching, RNN and TVLDA to incorporate temporal information. Although the classifiers in recent studies utilized the temporal information, they did not consider variations in different frequency bands separately. Variations in different frequency bands are important because each band has its own activation pattern during a gesture, such as Event Related Synchronization (ERS) and Event Related Desynchronization (ERD) [6].

Excessive features generally lead to overfitting of machine learning models, increased computational complexity and inclusion of irrelevant information. In the context of hand gesture classification, feature reduction methods such as feature selection, [5]–[9], [11], Common Spatial Patterns (CSP) [7], [12], and Principal Component Analysis (PCA) [10] have previously been studied.

In previous researches [1], [10], channels that exhibit significant power variations (ERD/ERS) in higher number of frequency bands were considered during feature reduction, while the channels that exhibit significant power variations in lower number of frequency bands were rejected. However, the rejected channels might exhibit significant power variations in certain frequency bands than the selected channels. Therefore, this approach might not always utilize power variations occurred in some frequency bands.

In the current study, we considered six frequency bands: 4-8 Hz, 8-12 Hz, 12-40 Hz, 40-70 Hz, 70-135 Hz and 135-200 Hz. In order to give equal importance to power variations in each frequency band separately, PCA and statistical based channel selection methods were performed for each frequency band individually. In addition, to capture the temporal power variations from each frequency band separately, a RNN based architecture with two sequential Long Short-Term Memory (LSTM) blocks was implemented. In the first block, features from each frequency band were given as input to separate LSTM layers. The LSTM layer in the second block combined the outputs from the first block.

[1]Jathurshan Pradeepkumar, Mithunjha Anandakumar, Vinith Kugathasan and Anjula C. De Silva are with Department of Electronic and Telecommunication Engineering, University of Moratuwa, Sri Lanka.

[2]Thilina D. Lalitharatne is with Dyson School of Design Engineering, Imperial College London, UK.

[3]Simon L. Kappel is with Department of Electrical and Computer Engineering, Aarhus University, Denmark.

II. Methodology

The proposed method comprises A) Power Spectral Density (PSD) based feature extraction, B) statistical and PCA based feature reduction, and C) LSTM based architecture for gesture classification. The publicly available 'fingerflex' dataset was used to validate the method.

A. Feature Extraction

The feature extraction was based on the PSD of the signal, as described in the following and shown in Fig. 1. Let $x_{l,m}$ be an ECoG signal recorded with channel l of L channels, while the subject was performing a gesture in trial m of M trials. $x_{l,m}$ contains ECoG samples recorded within the time interval of -2 s to $+2$ s, with time = 0 corresponding to the onset of the gesture. $x_{l,m}$ was segmented into T overlapping segments, with a length of 250 ms and an overlap of 50 ms. Let $x^{(t)}_{l,m}(n)$, $n = 0; \ldots; N-1$, be N samples of segment t extracted from channel l and trial m. The PSD, $S^{(t)}_{l,m}(k)$, was estimated for each segment as given in Eq. 1.

$$S^{(t)}_{l,m}(k) = \frac{1}{N}\left|\sum_{n=0}^{N-1} h(n)x^{(t)}_{l,m}(n)e^{-\left(\frac{j2\pi kn}{N}\right)}\right|^2 \quad (1)$$

where $h(n)$ is a hamming window and k is a frequency bin.

In the method proposed by Li et al., 2017 [6], five frequency bands were used for the analysis. Here we used the same frequency bands, with the addition of breaking the lower frequency band into Theta and Alpha. Thus, the frequency bands were: Theta (4-8 Hz), Alpha (8-12 Hz), Beta (12-40 Hz), low Gamma (40-70 Hz), high Gamma (70-135 Hz), and a high frequency band (135-200 Hz). The Theta and Alpha bands were considered separately, because we observed different power variations within these frequency bands before the onset of a gesture.

For frequency band f, the average PSD value $A^{(t)}_{l,m,f}$ was calculated for each channel and segment as follows:

$$A^{(t)}_{l,m,f} = \frac{1}{N_f}\sum_{i=1}^{N_f} S^{(t)}_{l,m}(k_{f,i}) \quad (2)$$

where $k_{f,i}$ is the i^{th} frequency bin for frequency band f and N_f is the total number of frequency bins in the f^{th} frequency band, where $f = 1, 2, ..., 6$.

The time interval from $-2s$ to $-1.5s$ of each segment was considered as the relaxation period, as indicated in Fig. 2. The average of this interval was used to normalize the average PSD as given in Eq. 3 and shown in Fig. 1. In order to improve signal stationarity and Gaussianity, the normalized PSD was log transformed [10].

$$A^{(t)}_{norm,l,m,f} = 10\log_{10}\left(\frac{A^{(t)}_{l,m,f}}{\bar{A}_{relax,l,m,f}}\right) \quad (3)$$

where $A^{(t)}_{norm,l,m,f}$ is the normalized PSD and $\bar{A}_{relax,l,m,f}$ is the average relaxation PSD.

Finally, the normalized PSD values for a gesture trial were arranged into feature matrices, one for each frequency band. The rows of a features matrix represented the channels, and the columns represented the segments, corresponding to different time windows, as described above. Thus, six feature matrices were extracted for each trial, as shown in Fig. 2. Before feeding the feature matrices to the classifier, they were truncated to the time interval of -0.5 s to $+2$ s from onset.

B. Feature Reduction

In the current study, the features were calculated for each channel separately, and thus reducing the dimension of the channel space resulted in a proportional reduction of the feature space. Two different methods were used for feature reduction, as described below.

*1) **Statistical Based Channel Selection**:* This method removes irrelevant channels and retains informative channels with features that show significant difference between the gestures. Statistical channel selection was carried out for each frequency band separately. In order to preserve the temporal information, three time intervals were considered : before onset (-0.5 s to 0 s), during (0 s to $+0.5$ s) and after onset ($+0.5$ s to $+1.5$ s). These time intervals were considered because significant power variations were observed in those intervals during visual inspection of the data. For each gesture trial and time interval, the average PSD value was calculated. Then, the average PSD values for a gesture was collected in three separate sample populations, corresponding to the three time intervals. Channels with

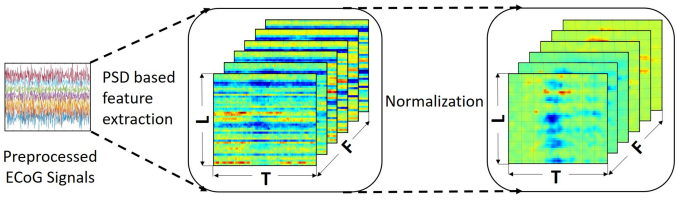

Fig. 1. Illustration of the feature extraction process based on PSD values.. The normalized data is then rearranged into a 3-D matrix, where L, T and F represent number of channels, number of segments and number of frequency bands, respectively.

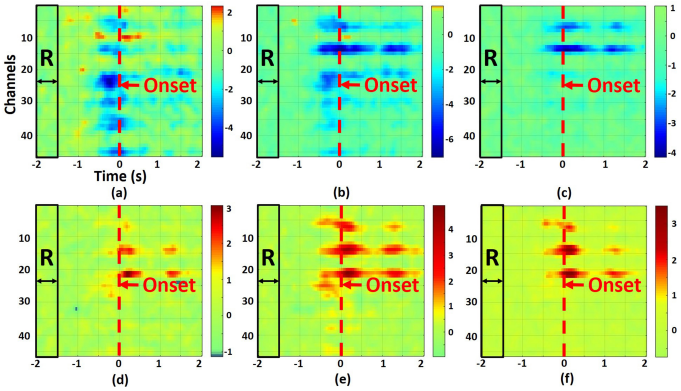

Fig. 2. Illustration of the extracted PSD based feature maps for each frequency band after normalization for subject *bp* in the fingerflex dataset. Figures (a), (b), (c), (d), (e), and (f) represent feature maps for 4-8 Hz, 8-12 Hz, 12-40 Hz, 40-70 Hz, 70-135 Hz and 135-200 Hz frequency bands, respectively. R represents the relaxation period (-2 to -1.5 s) before performing the gesture.

power variations relevant for gesture classification were determined by comparing the sample populations for all possible pairs of gestures in each time interval separately. Channels were included in the feature space if the p-value for at least three pairs were below 0.01 in a two-tailed paired t-test.

*2) **Principal Component Analysis**:* A disadvantage of the channel selection method described above is that it entirely discards the information within unselected channels and it is incapable of combining the joint information from the selected channels, which leads to redundancy. PCA can be used to perform dimensionality reduction and overcome these shortcomings. PCA based dimensionality reduction was carried out for each frequency band separately. The feature matrices for all trials were combined to a single matrix for each frequency band, resulting in matrices with the following dimensions $(N_{segments} * N_{trials}) \times N_{channels}$. Dimensionality reduction was performed in the channel dimension, and therefore the temporal information of the data was not affected.

C. Gesture Classification

The aim was to develop a classification model which utilizes the spectral information while preserving the temporal variations. To achieve this, we used a LSTM model, which is a type of RNN designed to model the temporal relationships within sequential data. LSTMs are specialized in learning long-term dependencies of a sequence, since it includes memory cells which can preserve information for a long time interval. This decides which information from preceding time windows contributes to the classification output.

As shown in Fig. 3, the proposed architecture contains two sequential LSTM blocks. The first block consists of a separate LSTM layer for each frequency band. The outputs from these LSTM layers are then concatenated and fed into a another LSTM layer. The proposed architecture was implemented using Keras with a TensorFlow backend. The feature matrices after feature reduction were fed into the 6-layer LSTM networks with 128 units, tanh activation function and dropout of 0.1. Outputs from the 6-layer LSTM networks remained as temporal sequences. The outputs were concatenated together and fed into a single LSTM layer with 32 units, LeakyReLU (alpha = 0.015) activation function and 0.1 dropout. A Dropout layer with a probability of 0.3 and a classification layer with softmax activation function was incorporated after the LSTM layers. The number of units for each LSTM layer was selected based on results from experiments conducted on the data using different combinations. RMSprop optimizer was used to update the weights during model training with a learning rate of 0.001. Categorical cross entropy function was used to calculate the loss of the model, and the performance metric was accuracy, expressed as number of correctly classified trials. Stratified 10-fold cross validation was used to evaluate the performance of the model, and also to tune the hyper-parameters and model architecture.

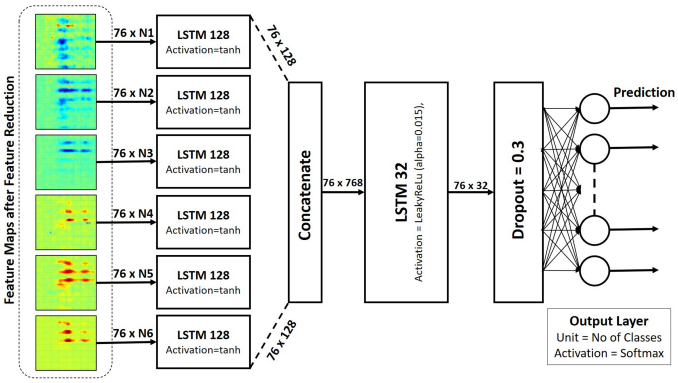

Fig. 3. Illustration of proposed architecture. $N1$, $N2$, $N3$, $N4$, $N5$ and $N6$ represent number of selected channels in each frequency band after feature reduction.

D. Dataset

The proposed method was evaluated with the publicly available 'fingerflex' dataset from Kai Miller (Miller K.J., 2012 [13]). The study was approved by the Institutional Review Board at the University of Washington and all patients participated voluntarily*. As shown in Table I, the dataset contains ECoG data acquired from nine subjects. However, we did not consider data from the subjects *mv* and *wm* for evaluation, because the recordings from *mv* were corrupted while *wm* data did not have a sufficient number of trials per gesture. The data was acquired using a Synamps 2 biosignal amplifier (Compumedics Neuroscan, North Carolina, USA). The data was sampled at 1 kHz and was bandpass filtered from 0.3 to 200 Hz [10]. Visual cues indicating which finger to flex out of the five fingers, were provided on a bedside monitor. The trials were 2 s long and 2 to 5 flexions were carried out during each trial. Each trial was succeeded by a resting period of 2 s. Altogether 150 trials were conducted for each patient, 30 trials corresponding to each finger. To supplement the ECoG data, a glove sensor (Fifth Dimension Technologies, Irvine, CA) with 5 degrees of freedom was used to record the finger positions.

TABLE I
DESCRIPTION OF THE DATASET.

Patient Code	Age	Gender	Handedness	Hemisphere**	No.of electrodes
bp	18	F	Right	Left	46
cc	21	M	Right	Right	63
zt	27	F	Right	Left	61
jp	35	F	Right	Left	58
ht	26	M	Right	Left	64
mv	45	F	Right	Left	43
wc	32	M	Right	Left	64
wm	19	F	Right	Right	38
jc	18	F	Right	Left	47

*__Ethics Statement__ : All patients participated in a purely voluntary manner, after providing informed written consent, under experimental protocols approved by the Institutional Review Board of the University of Washington (No. 12193). All patient data was anonymized according to IRB protocol, in accordance with HIPAA mandate. These data originally appeared in the manuscript "Human Motor Cortical Activity Is Selectively Phase- Entrained on Underlying Rhythms" published in PLoS Computational Biology in 2012 (Miller K.J., 2012 [13]).

**Hemisphere indicates the brain region covered by the electrodes (left/right).

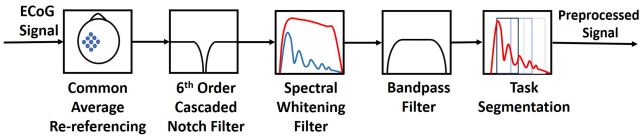

Fig. 4. Data preprocessing pipeline

E. Data Preprocessing

The preprocessing pipeline we used is summarized in Fig. 4. This predominantly follows the procedure described by Grunewald et al., 2019 [10]. First, the ECoG data was re-referenced to the average of all channels. Then, to reduce the first three harmonics of power-line interference in the data, sixth order cascaded Butterworth notch filters were applied at frequencies 60 Hz, 120 Hz and 180 Hz. Then, the data was whitened by applying the spectral whitening filter proposed by Grunewald et al., 2019 [10] to equalize the spectral contributions. Finally, the data was bandpass filtered (Chebyshev type II, 3rd order) between 0.5 Hz and 200 Hz.

For onset detection, derivatives of the glove signals were acquired and their peaks were detected. The detected peaks were compared with the expected pattern for the instructed gesture (paradigm signal) and the first peak corresponding to the gesture was considered as the onset. Wrong execution of gesture was detected by comparing the paradigm signal with the peaks detected from derivatives of glove signals and such trials were removed from the data.

III. Results and Discussion

The performance of the proposed gesture classification method was evaluated for each feature reduction technique by performing stratified 10-fold cross validation. The obtained accuracies were compared with the TVLDA based classification method proposed by Gruenwald et al. [10]. The average accuracy for the proposed architecture with statistical channel selection was higher for all the subjects except *wc*, as shown in Table II. However, the standard deviation values were higher with our proposed architecture. This might be because Gruenwald et al., [10] evaluated using 20 repetitions of 10-fold cross validation, whereas we evaluated using a single 10-fold cross validation. Overall, Gruenwald et al., achieved an average classification accuracy of 79.6% across the seven subjects as opposed to 82.4% with statistical channel selection and 77.0% with PCA based feature reduction.

Using the first 3 subjects in the 'fingerflex' dataset, Onaran et al., 2011 [12] achieved an average classification accuracy of 86.3% with a redundant spatial projection framework based on common spatial patterns. However, we were able to surpass this accuracy for these three subjects achieving an average classification accuracy of 87.3% with the statistical channel selection method. However,the PCA based method produced a lower accuracy of 83.7%.

TABLE II

Classification Accuracy (%)

Subject name	Gruenwald et al., 2019 [10]	Proposed arch. with PCA	Proposed arch. with stat. channel selection
bp	89.4±1.3	82.6±10.3	89.8± 6.7
cc	82.8±1.2	83.7± 7.2	85.4± 6.7
zt	85.7±1.2	84.9± 7.2	86.6± 9.2
jp	77.3±2.0	70.4±11.3	79.2±12.1
ht	64.5±3.2	66.1± 8.6	69.7± 6.5
wc	80.1±1.7	71.4±11.8	79.7± 6.0
jc	77.5±1.7	80.2± 7.5	86.7± 4.2
Average	79.6	77.0	**82.4**

IV. Conclusion

In order to provide equal importance to power variations in six frequency bands ranging from 4 to 200 Hz, we have implemented a RNN model with a separate LSTM layer for each frequency band along with PCA feature reduction and statistical based channel selection. The method was tested on publicly available ECoG data, and an average accuracy of 82.4% was obtained with the statistical based channel selection and 77.0% with PCA. The results are on par with the state-of-the-art methods, and justify that the method is suitable for gesture classification of ECoG data. Looking forward, the proposed method could be further improved to allow real-time hand gesture recognition. More optimal ECoG electrode placements might improve the classification, and in that context it would be relevant to test the proposed method on EEG data, as the non-invasive nature of EEG allows more flexibility in the electrode placements.

References

[1] T. Jiang et al., "Characterization and Decoding the Spatial Patterns of Hand Extension/Flexion using High-Density ECoG," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 25, no. 4, pp. 370-379, 2017.

[2] P. Shenoy, K. Miller, J. Ojemann, and R. Rao, "Finger movement classification for an electrocorticographic BCI,", IEEE Neur Eng. Conf., 2007.

[3] T. Yanagisawa et al., "Real-time control of a prosthetic hand using human electrocorticography signals: Technical note," J. Neurosurg., vol. 114, no. 6, pp. 1715-1722, 2011.

[4] C. Chestek et al., "Hand posture classification using electrocorticography signals in the gamma band over human sensorimotor brain areas," J. Neural Eng., vol. 10, no. 2, 2013.

[5] M. Bleichner, Z. Freudenburg, J. Jansma, E. Aarnoutse, M. Vansteensel, and N. Ramsey, "Give me a sign: decoding four complex hand gestures based on high-density ECoG," Brain Struct. Funct., vol. 221, no. 1, pp. 203-216, 2014.

[6] Y. Li et al., "Gesture Decoding Using ECoG Signals from Human Sensorimotor Cortex: A Pilot Study," Behav. Neurol., vol. 2017, pp. 1-12, 2017.

[7] C. Kapeller et al., "Single trial detection of hand poses in human ECoG using CSP based feature extraction," 36th Ann. Int. Conf. IEEE Eng. Med. Biol. Soc., 2014.

[8] G. Hotson et al., "Individual finger control of a modular prosthetic limb using high-density electrocorticography in a human subject," J. Neural Eng., vol. 13, no. 2, 2016.

[9] T. Xie, D. Zhang, Z. Wu, L. Chen, and X. Zhu, "Classifying multiple types of hand motions using electrocorticography during intraoperative awake craniotomy and seizure monitoring processes-case studies," Front. Neurosci., vol. 9, no. OCT, 2015.

[10] J. Gruenwald, A. Znobishchev, C. Kapeller, K. Kamada, J. Scharinger, and C. Guger, "Time-variant linear discriminant analysis improves hand gesture and finger movement decoding for invasive brain-computer interfaces," Front. Neurosci., vol. 13, no. SEP, 2019.

[11] G. Pan et al., "Rapid decoding of hand gestures in electrocorticography using recurrent neural networks," Front. Neurosci., vol. 12, no. Aug, 2018.

[12] I. Onaran, N. Ince, and A. Cetin, "Classification of multichannel ECoG related to individual finger movements with redundant spatial projections," 2011 Ann. Int. Conf. IEEE Eng. Med. Biol. Soc., 2011.

[13] K. Miller et al., "Human Motor Cortical Activity Is Selectively Phase-Entrained on Underlying Rhythms," PLoS Comput. Biol., vol. 8, no. 9, 2012.