EEG-based Motion Intention Recognition via Multi-task RNNs

Weitong Chen* Sen Wang† Xiang Zhang‡ Lina Yao ‡

Lin Yue § Buyue Qian ¶ Xue Li *‖

Abstract

Recognition of human intention based on Electroencephalography (EEG) signals attracts strong research interest in pattern recognition because of its promising applications that enable non-muscular communications and controls. Over the past few years, most EEG-based recognition works make significant efforts to learn extracted features to explore specific patterns between a segment of EEG signals and the corresponding activities. Unfortunately, vectorization-based feature representations, either vector-like or matrix-like ones, suffer from massive signal noise and difficulties of exploiting signal correlations between adjacent sensors of EEG signals. Most importantly, EEG signals are represented by one unique frequency and then fed into the subsequent learning model. Neglecting different frequencies of EEG signals can be detrimental to activity recognition because a particular frequency of EEG signals is more helpful to recognize some activities. Inspired by this idea, we propose to extract EEG signals with different frequencies and introduce a novel Multi-task deep learning model to learn the human intentions. We have conducted extensive experiments on a publicly available EEG benchmark dataset and compared our method with many state-of-the-art algorithms. The experimental results demonstrate that the proposed Multi-task deep recurrent neural network outperforms all the compared methods in a multi-class scenario.

1 Introduction

The human brain is apparently a complex system that always drives researchers in different areas to explore its mystery. Fortunately, Brain-Computer Interfaces (BCI) can translate neuronal activities into signals, which thus permits completely new research possibilities of discovering the correlation between the brain activities and behaviors. Electroencephalography (EEG) signal analysis, which is a non-invasive technique to acquire the brain dynamics through BCI, can reflect brain activities when a subject is performing specific tasks. Therefore, EEG-based intention recognition has been widely studied in recent years and become one of the important research topics in the community of pattern recognition [12, 20, 18, 14]. Many promising applications that recognize human intentions via EEG signals, such as EEG signal controlled wheelchairs [6], brain typing [31], and brain wavelet-controlled exoskeleton [7], restore the ability to those individuals who are currently suffering from high degrees of motor disability or locked-in syndrome.

1.1 Motivation Most works on EEG-based intention recognition represent features of EEG signals at only one frequency before feeding the features into the subsequent learning model [18, 23, 30]. Learning EEG signals at only one frequency may be detrimental to the performance of intention recognition. In fact, the information within an EEG signal segment can be further divided into different frequency ranges, each of which has different relevance levels to specific brain activities. A certain frequency range was specifically associated with particular neuron activities. Specifically, most of the neuronal activities can be reflected by EEG data in the range of 0.5Hz to 28Hz, which primarily falls into six frequency bands: *Alpha*, *Beta1*, *Beta2*, *Beta3*, *Theta*, and *Delta*. These segmented EEG rhythms have unique biological significances. For instance, Alpha wavelets fluctuate with closing eyes and relaxation in mind, while the Beta waveforms are closely linked to motor behaviors and are attenuated during active movements [10]. Through signal decomposition, the transient features can be accurately captured and localized in a frequency context, which contributes to better intention recognition performance[33, 32, 28]. In [16], only *Alpha* rhythm has been used to recognize intentions achieving around 60% of accuracy. By considering the spatial correlations between the segregated rhythms, Korik et al. [13] and Kim et al. [12] precisely recognize different hand movements. However, little temporal information, such as correlations between different rhythms over time, has

*The University of Queensland, Australia.
†Griffith University, Australia.
‡University of New South Wales, Australia.
§Northeast Normal University, China.- Corresponding Author
¶Xi'an Jiaotong University, China.
‖Nanjing University of Aeronautics and Astronautics, China.

been taken into account, which may be beneficial to recognition performance.

1.2 Challenges Although EEG-based intention recognition has advanced so rapidly with promising achievements, there still exists some major challenges because of the technological limitations: **(1)** The quality of signals directly determines the distinctive power of the extracted features. Unlike other invasive techniques that have direct contacts with neurons, EEG-based BCI system acquires signals via sensors placed over the skull leading to strength variations and a mass of noises in signals; **(2)** Learning signals to recognize a particular intention can achieve high performance, because all ambiguous information will much less affect the supervised learning model that only targets at a binary classification. Unfortunately, the relevance between EEG wavelets and multiple different intentions is difficult to model, which results in low performance of multi-intention recognition. **(3)** It has witnessed that the same intention can yield different shapes of EEG signals leading to the large intra-class variation problem. This may blame to mental status as it can greatly influence the shape of EEG signals [25], even for the same intention. Such large intra-class variation may further worsen the performance of EEG-based multi-intention recognition.

1.3 Solution Given these challenges, we proposed a novel framework in which a multi-task recurrent neural network learns segregated EEG signals. Because noise may only be dominant at a particular frequency, we firstly decompose EEG signals into different frequency channels to reduce noise from the other frequencies ranges. In this way, a more robust feature learning can be achieved comparing to non-segregated signal representation methods. For each frequency channel, signals with ground-truth information, i.e. intention labels, will be learned by a Recurrent Neural Networks (RNNs) to obtain sophisticated representations at that frequency. Specifically, a special kind of RNNs, Long Short Term Memory (LSTM) networks, is applied to forge temporal features from signals with different frequencies. To overcome the large intra-class variation problem, a shared layer has been used to exploit temporal correlations between features of signals with different frequencies. The proposed model is not only capable of identifying a specific intention, but also accurately recognize multiple motion intentions.

1.4 Contribution We summarize contributions of our method as follows::

- We advance the correlation between different EEG frequencies and biological significance by learning segregated signals for human motion intention recognition;
- In this paper, we proposed a novel framework that applies Multi-task Recurrent Neural Networks to learn distinctive features from EEG signals. More importantly, the proposed method learns not only each frequency channel but also exploits the temporal correlation between different channels. In this way, improvements of both binary and multiple intention recognition are achieved.
- To evaluate our method, we have conducted extensive experiments on a publicly available EEG benchmark dataset and compared our method with many state-of-the-art algorithms. The experimental results demonstrate our approach outperforms all the compared methods and hit a new record with 97.8% of accuracy. Also, we have validated the robustness of our methods on another real-world dataset.

The remainder of the paper is organized as follows. Related work is discussed in Section 2. Section 3 presents the details of the proposed framework. The experimental results are reported in Section 4. Finally, we conclude this paper in Section 5.

2 Related Work

Over a decade, analyzing EEG signals has attracted extensive research interest in pattern recognition because there is an additional approach to conduct non-muscular communication and control. Recent research works have demonstrated that EEG-based intention recognition applications are being able to restore patients' social interactions and movement capabilities [12, 27]. Since EEG is a non-invasive manner to acquire the neuronal activities, it is widely used for capturing the human intended activities [8, 18, 14, 30]. In general, great efforts have been made to improve either feature representation or learning models[29, 4, 3].

Due to the complexity and high dimensionality of EEG signals, the accuracy of the intention recognition models is highly dependent on the rusticity of the feature representation. Vizard et al. [26] achieved 71.59% of accuracy on a binary alertness states prediction by employing the common spatial pattern (CSP) on the feature extraction. Inspired by [26], Meisheri et al. [15] and Shiratori et al. [21] exploited multi-class CSP scenarios for the EEG feature extraction and only achieved 54.63% and 56.7% of accuraciess, respectively. By utilizing features extracted from the Mu and Beta rhythms separately, Kim et al. [12] produced a high accuracy of

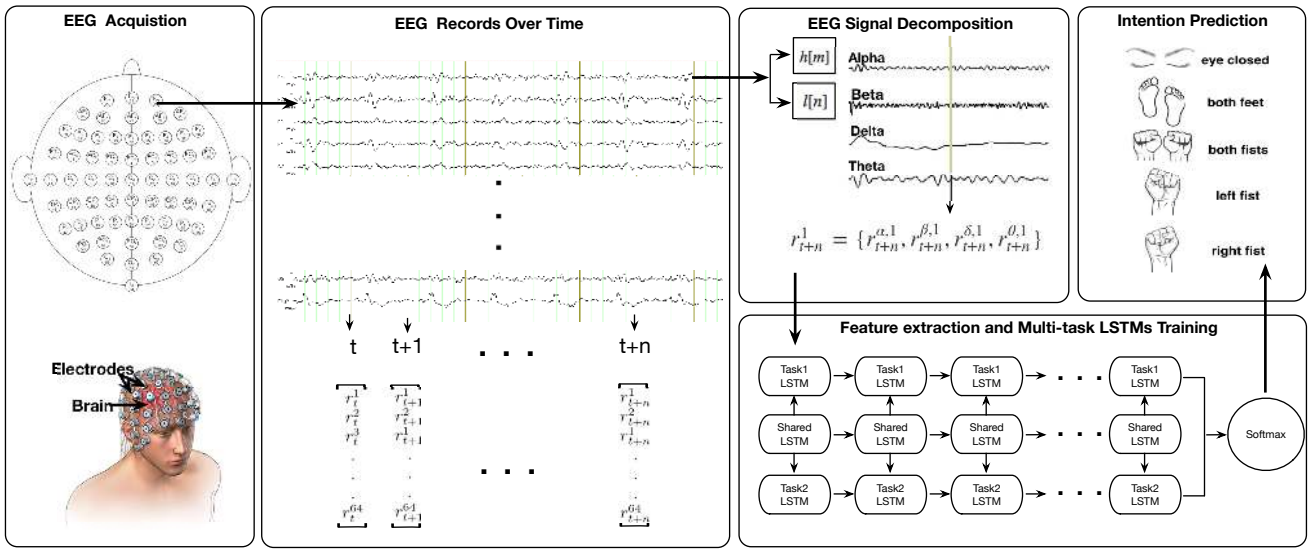


Figure 1: **The Workflow of the Proposed Approach.** EEG Signal will firstly be acquired and recorded by the BCI headset; it will then be decomposed to multi-resolution rhythms accordingly. Time series features embedded in the EEG signals will be decoded by the proposed Multi-task Model, and be used for recognizing intent of humans.

80.5% by adopting a random forest classifier.

The EEG signals from the brain continuously vary by adopting themselves. It is because the neurons react differently to the same brain activities [25]. It is requires the classifier able to capture the underlying consistency among EEG signals. Kang et al. [11] reached 70% of classification accuracy by using a Bayesian CSP model, along with an Indian Buffet process to capture a shared latent subspace across subjects. Inspired by the success of deep neural networks in different domains. Bashivan et al. [2] and Zhang et al. [30] took the advantage of Deep Neural Networks (RNNs) and related variations to capture the nature of time-series features over time, and achieved significantly high performance in intention recognition 85.05% and 95.5% of accuracy correspondingly.

3 Methods and Technical Solutions

Figure 1 illustrates the conceptual framework of the proposed approach. The goal of our approach is to precisely recognize human motion intentions based on raw EEG signals.

3.1 Data Acquisition The neuronal activities in brain are translated into EEG signals by using a BCI system with wearable headsets. When an *imagery action* is performed by a subject in mind, fluctuations of the voltage from the scalp will be continuously captured by multiple electrode sensors attached on the wearable headset (as illustrated in Figure 1). Therefore, an EEG reading $\boldsymbol{r}$ at time step $\boldsymbol{t}$ can be represented with a single n-dimension vector $\boldsymbol{r}_t = \left[r_t^1, r_t^2, ..., r_t^n\right]$, where the r_t^i is the reading of ith electrode sensor at time step $\boldsymbol{t}$. The EEG record of a single subject in an experiment run is R_t^i, where $i \in \Re$ is the sequence electrode sensor attached on the wearable headset, and $t \in T$ is the duration of an experiment run T.

Table 1: Correlation between the Spectrum of EEG bandwidths and Brain Functions.

Bandwidth	Rhythm	Function
0.5-3.5 Hz	Delta	Continuous-attention tasks
3.5-8 Hz	Theta	Inhibition of elicited responses
8-12.5 Hz	Alpha	Relaxed and closing the eyes
12.5-16 Hz	Beta 1	Mental and physical stress
16.5-20 Hz	Beta 2	Sustained attentional processing
20.5-28 Hz	Beta 3	Mental alertness power

3.2 EEG Signal Decomposition The EEG record contains multiple time series corresponding to measurements across different frequency bands. The most salient features are correlated with different human behaviors resided in different frequencies. The spectrogram of signals can be used for studying the richness of signals, and the Fast Fourier Transform (FFT) can be applied to observe the power of the EEG signals. Inspired by the EEG signals used in clinical prac-

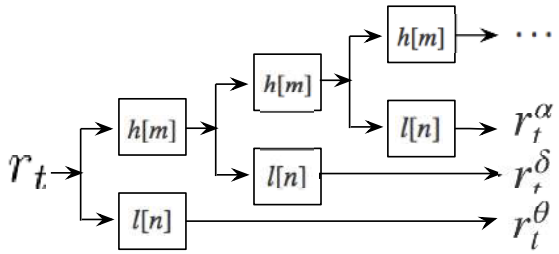


Figure 2: **EEG signal decomposition.** The $h\,[n]$ is high-band filter, and the $l\,[n]$ is the low-band filter.

tice, each frequency range has uniquely biological significance and signals are well correlated with specific brain functions [6, 24], as shown in Table 1. Therefore, an EEG signal can be quantized according to its pathological characteristics in the frequency range from 0.5Hz to 28Hz [24]. The raw EEG signal r can be segmented into different categories of bandwidth c, where $c = \{\delta, \theta, \alpha, \beta 1, \beta 2, \beta 3\}$. The decomposed EEG reading R of ith electrode sensor at time step t can be represented as a linear combination of a particular set of wavelets $\boldsymbol{R}_t^i = \left[r_t^{\alpha,i}, r_t^{\beta,i}, r_t^{\beta 2,i}, r_t^{\beta 3,i}, r_t^{\theta,i}, r_t^{\delta,i}\right]$. In this study, we focus on six frequency bands in the range from 0.5Hz to 28Hz.

A specified frequency range of EEG signals can be obtained by applying the band-pass filters. A band-pass filter that only passes required range frequency of a signal and rejects frequencies outside the range. In general, a band-pass filter consists of a high-pass filter and a low-pass filter. The high-pass filter, $h\,[n]$, passes signals with a frequency higher than a certain cut-off frequency n. The low-pass filter, $l\,[n]$, only attenuates signals with higher frequency than the cut-off frequency m. The procedure of a signal $\boldsymbol{s}_t$ multi-resolution is schematically illustrated in Figure 2, where $\{m, n\} = \{x \in \mathbb{Q} \mid 0 < x < 28\}$. The down-sampled outputs of the first filter provide the r_t^θ and the input of the next decomposition. By comparing filtered $r_t^{\beta 3,i}$ and the original signal r_t^i at same time in the same channel, where $i = 3$ and $t = 1.8$ million seconds, the first two graphs in Figure 2 show that the features embedded in the EEG signal become distinguished over time after band-pass filtering.

On the other hand, by converting the signal $r_t^{\beta 3,i}$ from its time domain to a representation in the frequency range, as illustrated in the third and fourth subgraphs in Figure 3, the connection between different frequency wavelets is lost. However, there are correlations between the EEG wavelets, and the corresponding intentions are ambiguous [25]. Therefore, we employ a shared layer in our model to capture the correlation

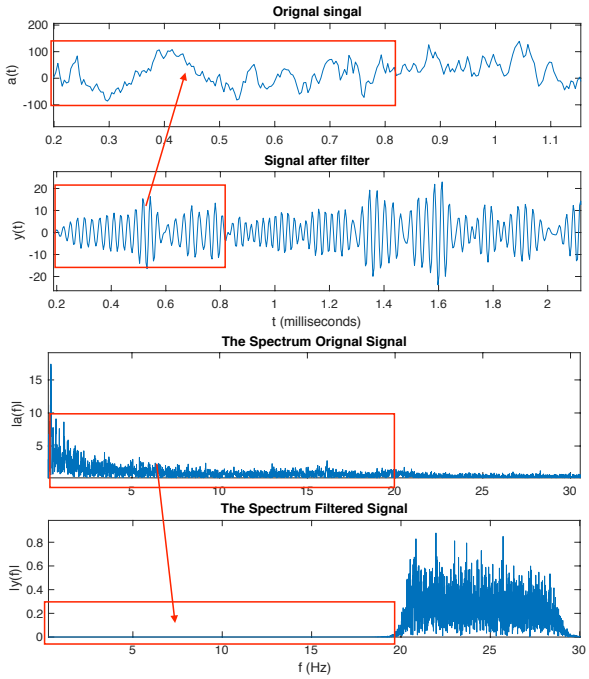


Figure 3: **Decomposed EEG signal samples.**

between different decomposed wavelets, which will be detailed in the next session.

3.3 Multi-Task Learning The recurrent neural networks (RNNs) [5] has sufficient ability to process arbitrary sequential inputs by recursively applying a transaction function to its *hidden vector* $\mathbf{h}_t$. The activation function f of the current hidden state $\mathbf{h}_t$ at t time step can be computed as followings:

$$\mathbf{h}_t = \begin{cases} 0 & t = 0 \\ \int (\mathbf{h}_{t-1}, x_t) & otherwise \end{cases} \tag{3.1}$$

where x_t is the current state input, and h_{t-1} is the previous hidden state. However, RNNs with transition function of this form has difficulties to learn long-range dependencies. The components of the gradient vector can vanish or explode exponentially over a long sequence. LSTM (Long Short-Term Memory) network [9] was proposed to address the vanishing gradient problem by incorporating gating functions. At each time step, an LSTM maintains a hidden vector h and a memory vector m for controlling state updates and outputs [8]. The LSTM unit at each time step t is defined as a collection of vectors in R^d. Each unit includes i, f, o, c and h, which are the input gate, forget gate, output gate, memory cell, and hidden state respectively. The forget gate controls the amount of memory in each unit to be "forgotten", the input gate

rules the update of each unit, and the output gate checks the exposure of each internal memory state. The LSTM transition equations are defined as follows:

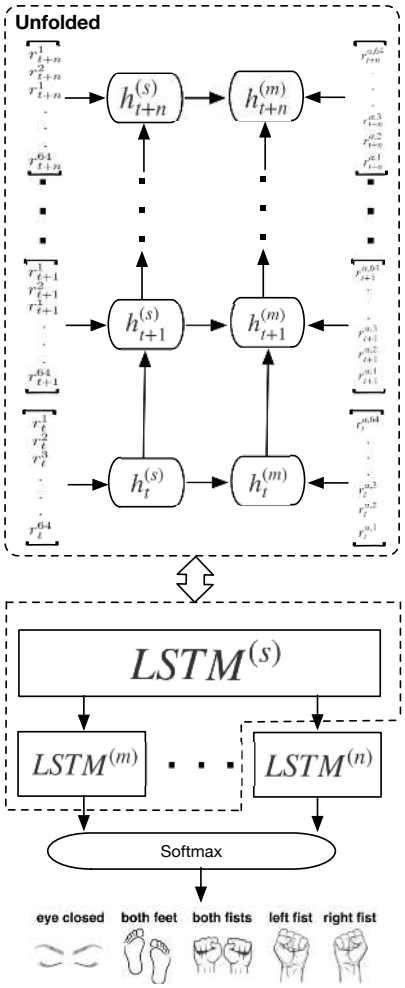


Figure 4: **Multi-task LSTM Architecture.** Unfolded structure illustrated the iteration between a shared-task S and one of the task m.

$$
\begin{aligned}
i_t &= \sigma(W_{ri} r_t + W_{hi} h_{t-1} + W_{ci} c_{t-1} + b_i) \\
f_t &= \sigma(W_{rf} r_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f) \\
o_t &= \sigma(W_{ro} r_t + W_{ho} h_{t-1} + W_{co} c_{t-1} + b_o) \\
c_t &= f_t c_{t-1} + i_t \tanh(W_{rc} r_t + W_{hc} h_{t-1} + b_c) \\
h_t &= o_t \tanh(c_t),
\end{aligned} \tag{3.2}
$$

where x_t is the input at the time step t, W are weights, bs are bias terms, and σ denotes the logistic sigmoid function. Therefore, we employed multi LSTM tasks to extract distinguished sequential features and embedded them in different frequency domains. As shown in Figure 4, $LSTM^{(m)}$ and $LSTM^{(n)}$ are different tasks, and they are assigned to capture the transitory features and embed decomposed EEG signals overtime, where $m, n = \{x \in \mathbf{Z} \mid 1 \leqslant x \leqslant 6\}$. In order to capture the correlation between decomposed signals, we introduced a share hidden layer $LSTM^{(s)}$, as shown in Figure 4. The hidden layer of LSTM cells are fully connected to the task-specific LSTM cell $LSTM^{(m)}$, and the activation function f of the current hidden state for the shared LSTM $\mathbf{h}_t^{(s)}$ remains the same as an ordinary LSTM. The activation function for the task-specific LSTM $\mathbf{h}_t^{(m)}$ at time step t will be updated as below:

$$
\mathbf{h}_t = \begin{cases} 0 & t = 0 \\ \int \left(\mathbf{h}_{t-1}^{(m)} \odot \mathbf{h}_{t-1}^{(s)}, r_t^{c,i)} \right) & otherwise \end{cases} \tag{3.3}
$$

A new gating mechanism is proposed in the task neuron $h_{t-1}^{(m)}$ in order to process the output $h_{t-1}^{(s)}$ passed by the neuron in the shared layer $h_{t-1}^{(s)}$. The new state $c_t^{(m)}$ for the task-specific $LSTM^{(m)}$ at time step t can be computed as follows:

$$
\begin{aligned}
c_t^{(m)} = f_t c_{t-1}^{(m)} + i_t^{(m)} \tanh(W_{rc} r_t^{c,i} \\
+ W_{hc} h_{t-1}^{(m)} \odot h_{t-1}^{(s)} + b_c^{(m)}),
\end{aligned} \tag{3.4}
$$

where $r_t^{c,i}$ is the input at the time step t, $h_{t-1}^{(m)}$ is the output from $h_{t-1}^{(m)}$ at $t-1$, $h_{t-1}^{(s)}$ is the output from the shared task $h_{t-1}^{(s)}$ at $t-1$ time step, and $\odot$ denotes as the concatenate operation.

4 Experiment

To evaluate the performance of the proposed model, we performed three types of experiments. First, we benchmarked our model on a publicly available dataset *eegmmidb*[1] in terms of accuracy. Then we conducted extensive experiments to examine and analyze the influence of the multi-resolution wavelets. Lastly, we investigated the robustness of the model on a real-world case. The source code and sample data were made publicly available on github[2].

4.1 Dataset A publicly and freely available benchmark EEG dataset, also known as *eegmmidb*, from PhysioNet[19], was used as a benchmark dataset in this study. EEG signals were acquired using a BCI 2000 system[3] at a 160Hz sample rate and referenced to a wearable headset with 64 electrodes sensors. The dataset contains more than 1,500 one-to-two-minute experiment

[1] https://physionet.org/pn4/eegmmidb/
[2] https://github.com/AnthonyTsun/Multitask_RNN/
[3] http://www.schalklab.org/research/bci2000

runs of 109 subjects. Each of subject performed 14 experiment runs in five different imagery tasks: *open and close left or right fist*, *open and close both fists*, *open or close both feet*, *open and close left fist*, and *open and close right fist*. The *eye closed* records are considering a baseline.

For the subsequent evaluations, we sampled ten subjects from the dataset. Each subject included $28,000$ EEG signal readings. It is denoted as $R_t^c = \left\{r_t^{c,i}\right\}_{i=1}^{64}$, where i is the number of electrodes sensors. In addition, $Alpha$ and $Beta2$ rhythms were extracted from the raw EEG records because of their high correlation to the experimental tasks. R_t^{α}, $R_t^{\beta 2}$, and R_t^c were fed into $LSTM^{(m)}$, $LSTM^{(n)}$ and $LSTM^{(s)}$ tasks respectively. During the training phase, the model uses 70% of the samples for training, and 30% for the validation and testing.

Table 2: Class annotation in two different datasets.

Dataset	Class 1	Class 2	Class 3	Class 4	Class 5
eegmmidb	eye closed	left hand	right hand	both hands	both feet
emotiv	confirm	up	down	left	right

4.2 Evaluation In this session, we showed the overall performance of our model and evaluated the effect of the multi-resolution signals. We named our Method **MTLEEG**.

4.2.1 Compared models To evaluate the proposed method, we have compared it with many of state-of-the-art methods, as well as the baseline methods on the dataset *eegmmidb*. A brief introduction of compare model is listed below:

1. Almoari et al. [1] employed the Support Vector Machine, along with features extracted from multi-resolution EEG signals for the binary classification.
2. Shenoy et al. [20] used regularized signal filtering to capture EEG features and adopted Fishers Linear Discriminant (FDA) for the binary classifications.
3. Rashid et al. [17] used an Neural Network (NN) algorithm to classify the decomposed EEG signals.
4. Kim et al. [12] extracted features from *Alpha* and *beta* rhythms, and then predicted EEG signals by using a random forest classifier.
5. Sita et al. [22] achieved decent classification performance by adopting an LDA classier, along with selected EEG signals.
6. Zhang et al. [30] had the best recognition result by employing the Recurrent Neural Networks (RNNs) with raw EEG signals.

Table 3: Comparison between the proposed methods and all the compared methods.

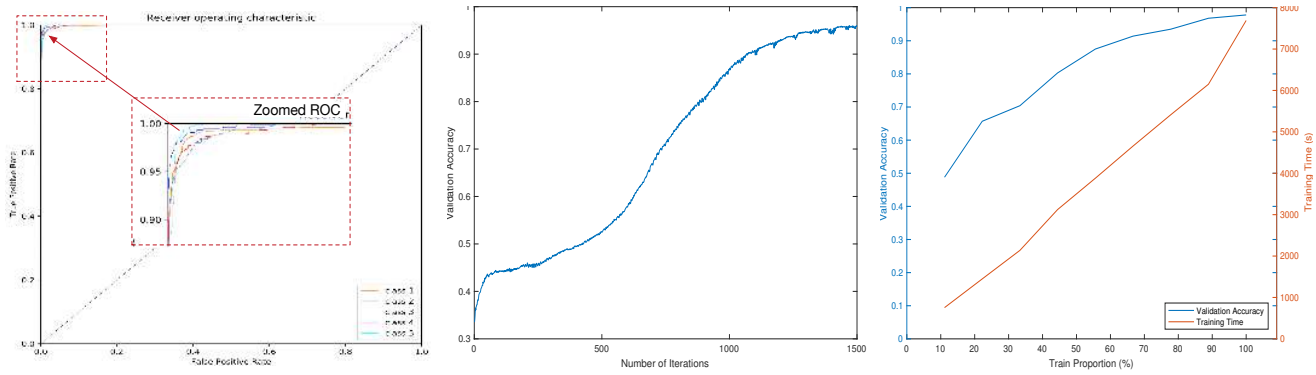
Index	Method	Class	Accuracy
1	Almoari[1]	Binary	0.749
2	Shenoy[20]	Binary	0.8206
3	Rashid[17]	Binary	0.92
4	Kim [12]	Multi (3)	0.805
5	Sita[22]	Multi (3)	0.8724
6	Zhang[30]	Multi (5)	0.9553
7	RNN	Multi (5)	0.9327
8	CNN	Multi (5)	0.8410
9	RNN-r^{α}	Multi (5)	0.8834
10	RNN-r^{β}	Multi (5)	0.9084
11	RNN-r	Multi (5)	0.9374
12	RNN-$r^{\alpha,\beta,}, r$	Multi (5)	0.9380
13	**MTLEEG**	Multi (5)	**0.9786**

For the baseline models, we kept the same structures and settings and fed baselines with different kinds of features extracted from the same dataset to evaluate the influence of the multi-resolution signals.

4.2.2 Experiment result It is clear that the proposed model outperforms state-of-the-art methods and the baseline methods. The best accuracy was reported in Table 3, with an improvement of 2.33% over the second best method [30]. Although method #1-3 focused on relative scenarios, our model surpasses their method significantly. Furthermore, the single RNN with $r^{\alpha,\beta,}, r$ multi-resolution signal achieved a relatively competitive result. This implies that the importance of signal segregation.

The Receiver Operating Characteristic (ROC) curve can demonstrate the discrimination capability of a classifier by plotting the *True Positive Rate* against the *False Positive Rate* in a range of threshold values. The area under the ROC curve (AUC) measures the accuracy. However, it is typically used for evaluating the binary classification problem. To evaluate multi-class classification performance, we binarized the output by considering each task separately (one task vs. the others). Figure 5a along with the Table 4 shows that ROC curves in all categories are above 0.999, where shows very high performance of our classifier. Figure 5b illustrates the accuracy improves with the increases in training iterations during the feature learning phase. It is clear that the proposed method can achieve stable performance ($\geq 90\%$) in 1,000 iterations.

Data size is can affect the performance of deep neural networks. Therefore, we tested our model with different data proportion and evaluated the corresponding



(a) The ROC Curves (b) Accuracy vs Iterations (c) Accuracy vs Training Proportion

Figure 5: **Performance evaluation of the proposed method**. (a) Zoomed ROC curves for 5-class classification, (b) The effectiveness of normalization method, (c) The relationship between accuracy and training data proportion.

accuracy (left y-axis) and running time (right y-axis) in Figure 5c. As illustrated in Figure 5c, it is observed that the performance positively correlates to the data size. 70% of the original data can contribute an 90.76% of accuracy, which shows that our approach is weakly dependent on the size of data. For computational complexity, training time varies linearly with respect to the training data size.

Table 4: Evaluation on the influence of decomposed EEG signals.

	Method	**Class1**	**Class2**	**Class3**	**Class4**	**Class5**	**Average**
Percision	RNN-r^{α}	**0.9824**	0.8295	0.8756	0.7681	0.8869	0.8872
	RNN-$r^{\beta 2}$	0.9059	0.9364	0.9219	0.8882	0.8953	0.9092
	RNN-r	0.9517	0.9280	0.9463	0.9214	0.9517	0.9376
	RNN-$r^{\alpha,\beta 2,r}$	0.9660	0.9315	0.8942	0.9304	0.9392	0.9383
	7 layer RNN	0.9618	0.9618	0.9574	0.9732	0.9396	0.9545
	MTLEEG	**0.9824**	**0.9721**	**0.9687**	**0.9796**	**0.9735**	**0.9763**
Recall	RNN-r^{α}	0.9549	0.8779	0.8176	0.8754	0.8275	0.8834
	RNN-$r^{\beta 2}$	0.9701	0.8348	0.8735	0.9182	0.8887	0.9084
	RNN-r	0.9532	0.9301	0.9239	0.9378	0.9304	0.9380
	RNN-$r^{\alpha,\beta 2,r}$	0.9512	0.9273	0.8949	0.9532	0.9457	0.9374
	7 layer RNN	0.9380	0.9084	0.9257	0.9028	0.9392	0.9228
	MTLEEG	**0.9919**	**0.9573**	**0.9687**	**0.9805**	**0.9869**	**0.9770**
F1-Score	RNN-r^{α}	0.9685	0.8530	0.8456	0.8183	0.8562	0.8842
	RNN-$r^{\beta 2}$	0.9369	0.8827	0.8971	0.9029	0.8920	0.9078
	RNN-$r^{\alpha,\beta 2,r}$	0.9446	0.9276	0.9199	0.9370	0.9487	0.9373
	RNN-r	0.9596	0.9308	0.9088	0.9341	0.9348	0.9381
	7 layer RNN	0.9497	0.9241	0.9413	0.9413	0.9394	0.9382
	MTLEEG	**0.9822**	**0.9643**	**0.9687**	**0.9800**	**0.9801**	**0.9751**
AUC	RNN-r^{α}	0.9685	0.8530	0.8456	0.8183	0.8562	0.8680
	RNN-$r^{\beta 2}$	0.9925	0.9874	0.9915	0.9918	0.9916	0.9904
	RNN-r	0.9596	0.9308	0.9088	0.9341	0.9347	0.9336
	RNN-$r^{\alpha,\beta 2,r}$	0.9935	0.99466	0.9932	0.9964	0.9973	0.9950
	7 layer RNN	0.9982	0.9977	**0.9990**	0.9990	0.9987	0.9985
	MTLEEG	**0.9989**	**0.9987**	**0.9990**	**0.9996**	**0.9996**	**0.9991**

Table 5: Accuracy comparison on Emotive Dataset

Index	Method	Class	Accuracy
1	RNN-r	Multi(5)	0.6751
2	7 layer RNN	Multi(5)	0.7980
3	RNN-$r^{\alpha,\beta 2,r}$	Multi(5)	0.8103
4	**MTLEEG**	Multi(5)	0.8847

4.2.3 Effects of the Multi-resolution Signals

To evaluate the influence of the multi-resolution EEG signals, we transformed our model to a single task RNN by evaluating it with the same data but in different resolutions. We compared our model with four baselines: RNN-r^{α} and RNN-$r^{\beta 2}$ models only learn α and $\beta 2$, respectively; RNN-r learns raw non-segregated EEG signals; RNN-$r^{\alpha,\beta 2,r}$ separately learns each rhythm ($\alpha, \beta 2$, and r) and concatenates all three features into a longer feature vector. Also, the second best counterpart in Table 3, which proposed to use a 7-layer RNN model [30], was adopted in the comparison. The evaluation was measured by Precision, Recall, F1 score, and AUC (area under ROC Curve). The detailed results in Table 4 illustrate that our model that considers the correlation between rhythms over time consistently outperforms all the baselines in the table. Such results have proved that the shared layer in the proposed methods contributes to further performance improvement.

4.3 Use Case To test the robustness of our proposed model, we tested our model on a real-world application. The application was developed by Zhang et al. [31] for a scenario of **Brain Typing**[4]. The experiment dataset was collected at a sampling rate of 128Hz from 7 participants for five different different executive commands, *up*, *down*, *left*, *right*, and *confirm (eye closed)*. The *alpha* rhythm and the *beta*2 rhythm were extracted from the signals as because of high relevance to the typing gestures. 70% data samples were used in the training phase, while the remaining records were used for testing and validation. The parameters followed the previous settings. For the use case, we compared our model with the second-best method (7 layer RNN) [30], RNN-r, and RNN-$r^{\alpha,\beta 2,r}$. The results were reported in Table 5.

It is observed that our model was surpassing all the comparing methods and achieved 88% of accuracy. Table 6 illustrate the accuracy of our model in each class. Decreased performance (from 97.8% to 88.4%) is blamed to much fewer signal sensors (14 sensors in EMOTIVE EPOC+ V.S. 64 sensors in BCI2000) were used to collect data. However, our model still robustly yielded satisfied performance in recognizing gesture commands, as illustrated in Table 7.

[4]http://www.cse.unsw.edu.au/ z5119405/nav/Demo.html

Table 6: Confusion Matrix.

Class	1	2	3	4	5
1	**0.9586**	0.0165	0.0018	0.0174	0.0058
2	0.0207	**0.8833**	0.0284	0.0399	0.0276
3	0.0048	0.0637	**0.8470**	0.0751	0.0570
4	0.0219	0.0627	0.0285	**0.8489**	0.0580
5	0.0062	0.0477	0.0354	0.0585	**0.8522**

Table 7: Model Evaluation.

Class	Precision	Recall	F1-Score	AUC
1	0.9763	0.9585	0.9673	0.9673
2	0.8383	0.8833	0.8602	0.8602
3	0.8839	0.8330	0.8445	0.8445
4	0.7840	0.8488	0.8151	0.8151
5	0.8668	0.8521	0.8594	0.8594
Avg	0.8869	0.8521	0.8852	0.8693

5 Conclusions

In this paper, we propose to extract EEG signals with different frequencies and introduce a novel Multi-task deep learning model to learn the human intentions. We advance the correlations between EEG frequencies and biological significances by learning different rhymes jointly, and maintain the correlations between different learning tasks via a shared layer embedded in the multi-task deep learning model. Experiments have shown that our method has achieved recorded high performance in human intention recognition.

6 Future Work

In our future work, for each frequency channel, we intend to incorporate a Convolutional Neural Network to encode the spatial information from the placement of the electrode sensors to improve the performance. Also, we plan to investigate most sophisticated sharing mechanisms in the RNNs based multi-task architecture to enhance the arbitrary signal into a robust spatio-temporal representation .

7 Acknowladgement

This research has been partially supported by the Fundamental Research Funds for the Central Universities (Grant NO. 2412017QD028), China Postdoctoral Science Foundation (Grant No. 2017M621192), the Scientific and Technological Development Program of Jilin Province (Grant No. 20180520022JH).

References

[1] M. H. Alomari, A. Abubaker, A. Turani, A. M. Baniyounes, and A. Manasreh, *Eeg mouse: A machine learning-based brain computer interface*, Int. J. Adv. Comput. Sci. Appl, 5 (2014), pp. 193–198.

[2] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, *Learning representations from eeg with deep recurrent-convolutional neural networks*, arXiv, (2015).

[3] X. Chang, Z. Ma, Y. Yang, Z. Zeng, and A. G. Hauptmann, *Bi-level semantic representation analysis for multimedia event detection*, IEEE transactions on cybernetics, 47 (2017), pp. 1180–1197.

[4] X. Chang and Y. Yang, *Semisupervised feature analysis by mining correlations among multiple tasks*, IEEE transactions on neural networks and learning systems, (2017).

[5] J. L. Elman, *Finding structure in time*, Cognitive science, 14 (1990), pp. 179–211.

[6] P. Fiala, M. Hanzelka, and M. Čáp, *Electromagnetic waves and mental synchronization of humans in a large crowd*, in Measurement, IEEE, 2017, pp. 241–244.

[7] A. A. Frolov, D. Húsek, E. V. Biryukova, P. D. Bobrov, O. A. Mokienko, and A. Alexandrov, *Principles of motor recovery in post-stroke patients using hand exoskeleton controlled by the brain-computer interface based on motor imagery*, Neural Network World, 27 (2017), p. 107.

[8] A. Graves, *Generating sequences with recurrent neural networks*, arXiv, (2013).

[9] S. Hochreiter and J. Schmidhuber, *Long short-term memory*, Neural computation, 9 (1997).

[10] A. K. Kaiser, M. Doppelmayr, and B. Iglseder, *Eeg beta 2 power as surrogate marker for memory impairment: a pilot study*, International Psychogeriatrics, (2017), pp. 1–9.

[11] H. Kang and S. Choi, *Bayesian common spatial patterns for multi-subject eeg classification*, Neural Networks, 57 (2014), pp. 39–50.

[12] Y. Kim, J. Ryu, K. K. Kim, C. C. Took, D. P. Mandic, and C. Park, *Motor imagery classification using mu and beta rhythms of eeg with strong uncorrelating transform based complex common spatial patterns*, Comp. intelligence and neuroscience, 2016.

[13] A. Korik, R. Sosnik, N. Siddique, and D. Coyle, *3d hand motion trajectory prediction from eeg mu and beta bandpower*, Progress in brain research, 228 (2016).

[14] T. C. Major and J. M. Conrad, *The effects of pre-filtering and individualizing components for electroencephalography neural network classification*, in SoutheastCon, IEEE, 2017, pp. 1–6.

[15] H. Meisheri, N. Ramrao, and S. K. Mitra, *Multiclass common spatial pattern with artifacts removal methodology for eeg signals*, in ISCBI, IEEE, 2016.

[16] M. R. Moore and E. A. Franz, *Mu rhythm suppression is associated with the classification of emotion in faces*, Cognitive, Affective, & Behavioral Neuroscience, 17 (2017), pp. 224–234.

[17] M. M. or Rashid and M. Ahmad, *Classification of motor imagery hands movement using levenberg-marquardt algorithm based on statistical features of eeg signal*, in ICEEICT, IEEE, 2016, pp. 1–6.

[18] O. R. Pinheiro, L. R. Alves, M. Romero, and J. R. de Souza, *Wheelchair simulator game for training people with severe disabilities*, in TISHW, IEEE, 2016.

[19] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, *Bci2000: a general-purpose bci system*, IEEE TBE, 51 (2004).

[20] H. V. Shenoy, A. P. Vinod, and C. Guan, *Shrinkage estimator based regularization for eeg motor imagery classification*, in ICICS, IEEE, 2015, pp. 1–5.

[21] T. Shiratori, H. Tsubakida, A. Ishiyama, and Y. Ono, *Three-class classification of motor imagery eeg data including rest state using filter-bank multiclass common spatial pattern*, in BCI, IEEE, 2015.

[22] J. Sita and G. Nair, *Feature extraction and classification of eeg signals for mapping motor area of the brain*, in ICCC, IEEE, 2013, pp. 463–468.

[23] Y. R. Tabar and U. Halici, *A novel deep learning approach for classification of eeg motor imagery signals*, JNE, 14 (2016).

[24] W. O. Tatum, *Ellen r. grass lecture: Extraordinary eeg*, The Neurodiagnostic Journal, 54 (2014), pp. 3–21.

[25] E. Vaadia and N. Birbaumer, *Grand challenges of brain computer interfaces in the years to come*, Frontiers in neuroscience, 3 (2009).

[26] L. Vézard, P. Legrand, M. Chavent, F. Faïta-Aïnseba, and L. Trujillo, *Eeg classification for the detection of mental states*, Applied Soft Computing, (2015).

[27] M. Wairagkar, I. Zoulias, V. Oguntosin, Y. Hayashi, and S. Nasuto, *Movement intention based brain computer interface for virtual reality and soft robotics rehabilitation using novel autocorrelation analysis of eeg*, in BioRob, IEEE, 2016.

[28] S. Wang, X. Chang, X. Li, G. Long, L. Yao, and Q. Z. Sheng, *Diagnosis code assignment using sparsity-based disease correlation embedding*, IEEE TKDE, 28 (2016), pp. 3191–3202.

[29] S. Wang, X. Li, L. Yao, Q. Z. Sheng, G. Long, et al., *Learning multiple diagnosis codes for icu patients with local disease correlation mining*, ACM TKDD, 11 (2017), p. 31.

[30] X. Zhang, L. Yao, C. Huang, Q. Z. Sheng, and X. Wang, *Intent recognition in smart living through deep recurrent neural networks*, in NIP, Springer.

[31] X. Zhang, L. Yao, Q. Z. Sheng, S. S. Kanhere, T. Gu, and D. Zhang, *Converting your thoughts to texts: Enabling brain typing via deep feature learning of eeg signals*, arXiv, (2017).

[32] X. Zhu, Z. Huang, Y. Yang, H. T. Shen, C. Xu, and J. Luo, *Self-taught dimensionality reduction on the high-dimensional small-sized data*, PR, 46 (2013).

[33] X. Zhu, L. Zhang, and Z. Huang, *A sparse embedding and least variance encoding approach to hashing*, TIP, 23 (2014).