

# Retention 機構に基づく頭蓋内脳波の分類と BMI の構築

## ECoG-Based Motor Imagery Classification Using Retentive Network

長嶋 隼矢 \*1  
Shunya Nagashima

兼田 寛大 \*1  
Kanta Kaneda

飯田 紡 \*1  
Tsumugi Iida

田口 美紗 \*2  
Misa Taguchi

平田 雅之 \*2  
Masayuki Hirata

杉浦 孔明 \*1  
Komei Sugiura

\*1 慶應義塾大学  
Keio University

\*2 大阪大学  
Osaka University

Speech impairments from conditions like Amyotrophic Lateral Sclerosis and muscular dystrophy severely restrict patient communication, affecting daily and social life. Decoding technology based on Electroencephalography (ECoG) is essential for supporting these patients' communication. In this study, we propose a novel architecture combining a specialized convolutional layer for electrode feature extraction and a retentive network for ECoG signal classification of motor imagery, outperforming all baselines in accuracy.

### 1. はじめに

筋萎縮性側索硬化症 (ALS) や筋ジストロフィーなどの神経変性疾患に伴う発話障害は、患者のコミュニケーションを大幅に制限し、社会生活や日常の活動に深刻な影響をもたらす。この問題への解決策として、頭蓋内脳波を利用したデコーディング技術が注目されている。この技術は、発話障害を持つ患者が非言語的手段 (例: 脳波によるキーボード操作) で意思を伝達できるようにすることを目指し、社会参加の機会拡大および患者の生活の質を向上させることが期待されている。本研究では、上肢の運動想起時に記録した頭蓋内脳波から想起した行動を分類する、頭蓋内脳波の多クラス分類タスクを扱う。本タスクでは、被験者が M 種類の行動の中の任意の行動を想起し、そのときにおける頭蓋内脳波を正しく分類できることが望ましい。

図 1 に本タスクの例を示す。この例では、手の掌握を想起した際の頭蓋内脳波が与えられる。この頭蓋内脳波をモデルの入力とし、各行動に対応する確率の予測値を得る。この例の場合、手の掌握クラスに対応する確率の予測値が最も大きいことが望ましい。

既存手法である EEG Conformer [Song 23] に代表されるように、脳波分類タスクにおけるモデルの性能は向上している。しかし、複数の電極を持つ複雑な信号である頭蓋内脳波の分類性能は、以前として不十分である。これは既存の手法が、複雑な頭蓋内脳波における長期の依存関係を捕捉することが困難であることが考えられる。

本研究では、運動想起時における頭蓋内脳波の分類において、電極間の特徴量抽出に特化した畳み込み層および retentive 機構 [Sun 23] を組み込んだブランチ構造を組み合わせた新しいアーキテクチャを提案する。これにより、電極間特徴量の効果的な抽出および長期の依存関係を捕捉することが可能である。

### 2. 問題設定

本タスクの入力は、運動野近辺に埋め込まれた脳表電極から計測された頭蓋内脳波である。出力は運動に対応する確率の予測値である。本論文では、電極は運動野近辺に配置された硬膜下電極と定義する。また、本研究では、特定の個人に合わせた BMI システムの開発に焦点を置いている。これは、被験者が ALS という希少疾患を患っており、開頭手術を受けている

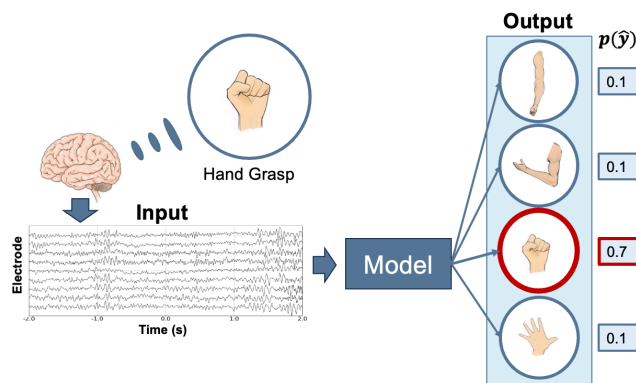


図 1: 頭蓋内脳波の多クラス分類タスクの例。

ためである。仮に複数の被験者が存在したとしても、脳の形状の違いにより、各電極が被験者間で同じ機能領域の信号を表すわけではない。したがって、本研究では、多数の被験者から得た大規模データセットに基づくモデルの構築ではなく、各被験者に特化した BMI システムの開発を前提とする。また、本タスクでは評価尺度として、分類精度を使用する。

### 3. 提案手法

本研究では、頭蓋内脳波の分類のための新しい並列機構を導入し、大域的な特徴抽出のための retention 機構および局所的な特徴抽出のための畳み込み層を効果的に組み合わせさせた ECoG Dual Context Network (ECoG DCNet) を提案する。本研究では、頭蓋内脳波の多クラス分類タスクにおいて提案手法を検証するが、本手法における長期依存を捉える構造は、脳波から特定の刺激をデコーディングする手法や時系列予測の手法にも広く適用可能であると考えられる。モデル全体は 2 つの主要モジュールから構成され、それぞれ Electrode-Aware Convolution Unit および Retentive Dual Context Module である。図 2 にモデルの構造を示す。モデルの入力  $X$  は  $c$  個の電極における系列長  $T$  の頭蓋内脳波である。

#### 3.1 Electrode-Aware Convolution Unit

Electrode-Aware Convolution Unit (EACU) では、入力  $X$  から時空間特徴  $H_{sp}$  を抽出する。本モジュールでは、時間および空間次元にそれぞれ沿った畳み込みにより、時間的特徴と電極間の空間的關係を抽出することで局所の特徴を捉える。脳波特徴抽出で一般的に使用されている従来の 1 次元畳み込み

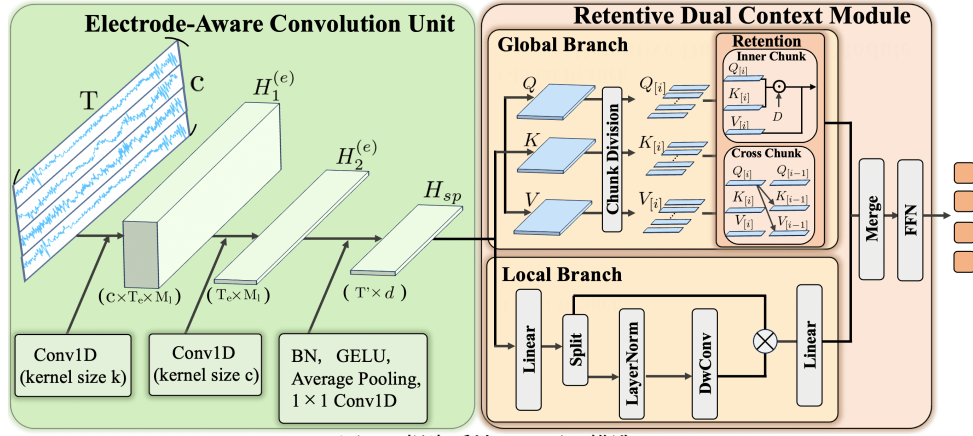


図 2: 提案手法のモデル構造.

(1D-Conv) では、一般的に比較的小さなカーネルサイズを採用する。この手法では空間的な文脈を十分に捉えることができず、電極間の関係を効果的にモデル化できない可能性がある。

一方で、EACU においては、電極全体を含む大きさのカーネルサイズを持つ 1 次元畳み込みアプローチを採用する。これにより、電極間の関係を直接モデル化することができ、脳活動を示す複雑な空間情報を効果的に捉えることができる。また、全ての電極を含むようにカーネルを拡張することで、空間的コンテキストをより包括的に統合し、カーネルサイズが小さい従来の手法の限界を克服する。具体的な手順は以下の通りである。まず、以下のように、時系列方向に沿ったカーネルサイズ  $k$  の 1D-Conv を用いて時間的特徴  $H_1^{(e)} \in \mathbb{R}^{c \times T \times M_1}$  を捉える。ここで、 $M_1$  は 1D-Conv におけるチャンネル数である。次に、全電極を包含するカーネルサイズ  $c$  の 1D-Conv を用い、電極間情報を埋め込んだ表現  $H_2^{(e)} \in \mathbb{R}^{T \times M_1}$  を得る。さらに、以下の処理により空間特徴  $H_{sp} \in \mathbb{R}^{T \times d}$  を得る。

$$H_{sp} = \text{Conv}_{1 \times 1} \text{ AvgPool GELU BN } H_2^{(e)} \in \mathbb{R}^{T \times d}$$

ここで、BN, GELU, AvgPool, および  $\text{Conv}_{1 \times 1}$  はそれぞれバッチ正規化、活性化関数 [Hendrycks 16], 平均プーリング層、およびカーネルサイズが  $(1, 1)$  の 2 次元畳み込み層である。ここで、 $T^0$  は  $H_{sp}$  の縮小された時間次元であり、元の  $T$  に比べて時間的側面がよりコンパクトに表現されていることを示している。本モジュールは EEG Conformer の畳み込み層と構造的な類似性があるが、重要な違いは  $T^0$  のサイズにある。我々のモジュールは、プーリング層を通して系列方向の圧縮を抑えることにより、EEG Conformer に比べてより大きな  $T^0$  を維持する。これは、後続のモジュールで頭蓋内脳波の長期の依存関係を効果的に捕捉するために極めて重要である。

### 3.2 Retentive Dual Context Module

Retentive Dual Context Module は local ブランチと global ブランチを組み合わせることで、脳波の局所および大域的特徴を効果的に捉える。既存手法である EEG Conformer は畳み込み層と自己注意機構を直列に接続し、これらの特徴を抽出する。しかし、直列構造であるため、畳み込み層で捉えられた脳波の細かな変化が自己注意機構によって処理される際、データ全体の関連性に焦点を当てる自己注意機構の特性上、局所的な特徴が相対的に過小評価される可能性がある。Retentive Dual Context Module はこの問題を解決するために、局所および大域的特徴を並列に処理し、それぞれの特徴を効果的に捉えて統合する。本モジュールにおける local ブランチでは、短期間の脳波の変動を捉える一方で、global ブランチでは、脳波データ全体の長期的な依存関係を捉える。この局所的な特徴抽出および大域的な依存関係の捕捉を統合することにより、

脳波データの特性をより深く、包括的に捉えることが可能となり、既存手法よりも効果的に脳波データから有用な特徴を抽出することができると考えられる。local ブランチでは  $H_{sp}$  から局所的特徴  $Y_L$  を抽出する。

$$H^{(l)} = \text{GELU LN}(H_{sp})W_1^{(l)}$$

$$H^{(l)} = H_1^{(l)} \text{ DwConv LN}(H_2^{(l)})$$

$$Y_L = H^{(l)}W_2^{(l)}$$

ここで、 $H^{(l)} \in \mathbb{R}^{T^0 \times d_{\text{inter}}}$  は  $H_1^{(l)} \in \mathbb{R}^{T^0 \times d_{\text{inter}}}$  と  $H_2^{(l)} \in \mathbb{R}^{T^0 \times d_{\text{inter}}}$  に分割される。また、 $W_1^{(l)} \in \mathbb{R}^{d \times d_{\text{inter}}}$  および  $W_2^{(l)} \in \mathbb{R}^{d_{\text{inter}} \times 2 \times d}$  は学習可能な重み行列、 $\otimes$  はアダマール積である。また、LN, および DwConv はレイヤー正規化層、および depth-wise 畳み込み層である。

global ブランチでは  $H_{sp}$  から大域的特徴  $Y_G$  を得る。また、retention 機構において、並列および再帰表現を組み合わせた chunkwise recurrent 表現を用いた。これにより、長期的な依存関係を効率的に捉える。まず、入力  $X$  から線型変換により、以下のように  $Q \in \mathbb{R}^{T^0 \times d}$ ,  $K \in \mathbb{R}^{T^0 \times d}$ ,  $V \in \mathbb{R}^{T^0 \times d}$  を得る。

$$Q = XW_Q, K = XW_K, V = XW_V$$

ここで、 $W_Q \in \mathbb{R}^{d \times d}$ ,  $W_K \in \mathbb{R}^{d \times d}$ , および  $W_V \in \mathbb{R}^{d \times d}$  は学習可能な重み行列である。次に、 $Q$ ,  $K$ , および  $V$  を系列長が  $B$  のチャンクに分割する。

$$Q_{[i]} = Q_{B(i-1):B(i)}, K_{[i]} = K_{B(i-1):B(i)}, V_{[i]} = V_{B(i-1):B(i)}$$

ここで、 $[i]$  は  $i$  番目のチャンクを示す。つまり、 $x_{[i]} = [x_{(i-1)B+1}, \dots, x_{iB}]$  である。

$i$  番目のチャンクおよび入力  $X$  に対する retention は以下のように計算される。

$$R_i = K_{[i]}^T (V_{[i]} \zeta) + \gamma^B R_{i-1}, \quad \zeta_{ij} = \gamma^{B-i-1},$$

$$\text{Ret}(X_{[i]}) = (Q_{[i]} K_{[i]}^T - D) V_{[i]} + (Q_{[i]} R_{i-1}) \xi,$$

$$D_{nm} = \gamma^n \gamma^m, \text{ if } n > m; 0, \text{ otherwise; } \xi_{ij} = \gamma^{i+1},$$

$$\text{Retention}(X) = [\text{Ret}(X_{[0]}); \dots; \text{Ret}(X_{[n-1]})]$$

ここで、 $\gamma$  はハイパーパラメータであり、 $\zeta$  および  $\xi$  は  $i$  番目のチャンクに依存する係数である。 $\gamma$  は時間経過に伴う情報の重要度を制御する。 $\zeta$  および  $\xi$  は、それぞれチャンク内およびチャンク間での要素間の相互作用を調整する。Transformer と異なり、行列  $D$  の導入により相対距離に応じた指数関数的な減衰と特定の要素間の依存関係を制限する機能が実現され、これにより各時点でそれより前の時点との関係のみを考慮しつつ、空間計算量を抑えた効率的な学習を可能にする。本ブランチ

表 1: 定量的結果の比較. 太字は最良の値を示す.

Model	w/ retention	w/ EACU	Accuracy [%]
(a)EEGNet [Lawhern 18]			45.3 ± 11.6
(b)EEG Conformer [Song 23]			68.9 ± 4.7
(c)E-Branchformer [Kim 23]			68.8 ± 4.6
(d)Ours	✓		53.3 ± 3.0
(e)Ours		✓	72.8 ± 2.2
(f)Ours	✓	✓	<b>75.3 ± 4.0</b>

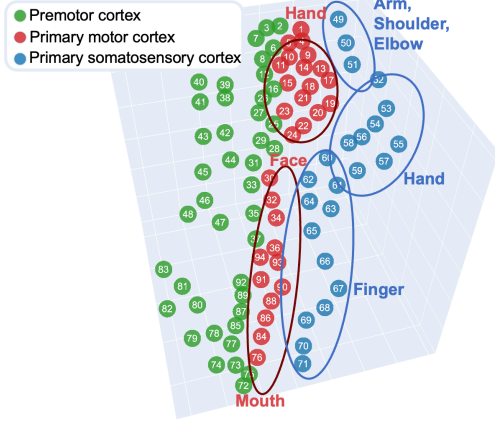


図 3: 電極の位置および皮質領域.

では, chunkwise recurrent 表現を用いた Gated Multi-Scale Retention (gMSR) を使用することで, モデルの表現力を向上させる. gMSR では, 各ヘッドごとに異なる  $\gamma$  を割り当てることで, モデルが脳波の複雑な時間的特性をより正確に捉えることができることを期待する. 本ブランチの入力である  $H_{sp}$  に対して gMSR は以下のように計算される.

$$\text{head}_j = \text{Retention}(H_{sp}, \gamma_j), \quad \gamma_j = 1 - 2^{-5} j,$$

$$Y_h^{(g)} = \text{GN}_h([\text{head}_0; \dots; \text{head}_{h-1}]),$$

$$Y_G = \text{gMSR}(H_{sp}) = \text{swish}(H_{sp} W_1^{(g)}) \quad Y_h^{(g)} W_2^{(g)}$$

ここで,  $h$  はハイパーパラメータ,  $W_1^{(g)} \in \mathbb{R}^{d \times d}$  および  $W_2^{(g)} \in \mathbb{R}^{d \times d}$  は学習可能な重み行列, swish は活性化関数 [Ramachandran 17] である. また GN [Wu 18] では, 各 head 毎の出力を正規化する. 次に2つのブランチの出力である  $Y_L$  および  $Y_G$  を以下のように統合する.

$$Y_C = [Y_L; Y_G], Y_D = \text{DwConv}(Y_C), Y_M = (Y_C + Y_D) W_O$$

ここで,  $W_O \in \mathbb{R}^{2d \times d}$  は学習可能な重み行列である. この  $Y_M$  に対して順伝播型ネットワーク層を適用することで, Retentive Dual Context Module の出力  $Y_R \in \mathbb{R}^{T \times d}$  を得る. 最後に以下のように本モデルの出力  $Y = \text{FFN}(\text{flatten}(Y_R)) \in \mathbb{R}^n$  を得る. ここで, FFN は順伝播型ネットワーク,  $n$  は運動想起の種類を示す. 本モデルの出力は, 上肢の行動確率の予測値  $p(\hat{y})$  である.

## 4. 実験

### 4.1 データセット

本研究では, ALS を患った被験者 1 名から得られたデータセットを利用し, 運動の想起中に記録された頭蓋内脳波に焦点を当てた. 本タスクでは, 肘の伸展および屈曲, 手の伸展および握力の 4 種類の運動想起を含む. 各試行は 4 秒間続き, 運動想起開始 2 秒前から運動想起指示後 2 秒までが記録された. 図 3 は運動野の近くに配置された, 頭蓋内脳波の計測に使用された 94 個の硬膜下電極を示す. 電極の 10 番から 30 番は一次運動野付近に位置していると考えられ, 49 番から 59 番は一次体性感覚野付近に位置していると考えられる. ここで, 一次運

動野および一次体性感覚野は, 運動想起時の頭蓋内脳波の分類において神経生理学的に関連していると考えられる. 被験者は 8 日間にわたり, 4 種類の運動想起タスクを実験日ごとに 40 回, 合計 1280 回行なった. 被験者には, 運動想起開始の正確なタイミングを把握するために, 視覚および聴覚の手がかりが提供される. 具体的には, 被験者の前に置かれたディスプレイ上の矢印が音声の手がかりと同期することで運動想起開始のタイミングを知らせる. 想起すべき運動の種類はディスプレイ上の文章を通じて被験者に伝えられる.

我々の知る限り, 希少疾患である ALS を患った被験者からの多クラス分類用の頭蓋内脳波を含む公開標準データセットは存在しない. そのため, 臨床データセットを使用した. また, 事前処理として 60Hz のハムノイズをノッチフィルタにより除去し, 各電極ごとに標準化を行った. さらに, ダウンサンプリングを行い 250Hz のサンプリングレートとなるようにした. 本研究では, k-fold 交差検証 ( $k=4$ ) を採用し, 合計 8 日間のうち 6 日間のサンプルを訓練集合に, 1 日を検証集合, 残る 1 日をテスト集合に割り当てた. 1 日あたりのサンプル数は 160 サンプルであるため, 訓練, 検証, およびテスト集合は, それぞれ 960, 160, および 160 サンプルを含む. この分割法を採用することで, 未知の日付の頭蓋内脳波を使用する実際の被験者に対するシステムと同じ条件でモデルの性能を評価することができる. 訓練集合, 検証集合, およびテスト集合はそれぞれモデルの学習, ハイパーパラメータの調整, およびモデル性能の評価に使用した.

### 4.2 パラメータ設定

本研究では,  $d, d_{\text{inter}}, h$ , および  $B$  はそれぞれ 64, 256, 4, および 125 とした. ここで,  $d, d_{\text{inter}}, h$ , および  $B$  は Retentive Dual Context Module における隠れ層の次元数, local ブランチにおける隠れ層の次元数, Multi-Scale Retention における head 数, および retention の chunkwise recurrent 表現におけるチャンクサイズである. 最適化には Adam を使用し, 学習率は  $7.0 \times 10^{-6}$ , バッチサイズは 2 とした. 訓練および 1 サンプルあたりの推論時間は, それぞれ約 9.5 分および約 3.13ms であった. 各エポックごとに検証集合で精度を計算し, 最も高い精度を得たモデルを用いて, テスト集合における評価を行った.

### 4.3 定量的結果

表 1 に, ベースライン手法, 提案手法, および ablation study の定量的結果を示す. なお, ベースライン手法として, EEGNet [Lawhern 18], EEG Conformer [Song 23], および E-Branchformer [Kim 23] を用いた. Ablation study の詳細については, 4.5 節で説明する. 分類精度によりモデルを評価した. 表中の値は, k-分割交差検証 ( $k=4$ ) での 4 回の試行における平均値と標準偏差を示す. EEGNet および EEG Conformer は頭皮脳波の多クラス分類タスクにおいて良好な結果が得られている. 本研究では頭蓋内脳波の多クラス分類タスクを扱うが, 頭蓋内脳波および頭皮脳波は共に脳活動の計測値という点において共通しており, これらの手法は本タスクにも適用可能である. そのため, これらをベースライン手法とした. また提案手法と類似した, より簡易的な手法である E-Branchformer に

