

## 1. データ処理の工夫点と考察

本コンペでは 16 ch. の 1000 点の筋電位の時系列データを基に、XYZ 各軸の速度データ 30 点を推定するモデルを作成する。対象とする速度データの区間は 0.5 秒と短く、速度データのパターンも限られることから、ある時刻の速度ベクトルを推測する際に、時系列的には後の筋活動も速度パターンを捉える情報が含まれるため有用であると考えた。そこで 1000 点の筋電データ全体から速度データ 90 点を推定するモデルを構築した。

マルチチャンネルの時系列データの取り扱いにおいては、時系列データを画像化して解析した。通常、音声解析等で用いられるスペクトログラム画像への変換とは異なり、このコンペでは時系列データを直接画像化する手法を採用した[ref]。各筋電データを 1024 点(24 点をパディング)に変換し、これを折り返して画像データとした後(Figure 1)、16 ch.のデータを正方形の画像になるように配置した。筋電データを 1024 点にすることで 2 の指数乗の点数となり、 $32 \times 32$  や  $16 \times 64$  等の画像データに変換しやすく、またチャンネル数も  $16 (=2^4)$  であることから  $4 \times 4$  等に並べやすい点も画像モデルを採用した理由の 1 つである。

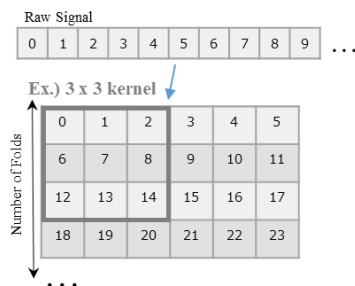


Figure. 1 Conversion from time-series data to image format. (Here, the data is folded every 6 points and analyzed with a 3x3 kernel)

時系列データを画像化する際の配置は、GradCAM を用いた画像特徴量の重要度評価と、スコアの評価により実施した。評価では、被験者 ID1-5 を学習用データに用いて、EfficientNetb0 を用いた画像モデルをバックボーンに設定し、さまざまな折り返し点数の影響を検証した。スコアを評価する際には、後述する公開スコアの他に、学習用データの CV のスコア、被験者 ID5 のテストデータのスコアを基に検討を進めた。

1024 点の時系列データの折り畳み方法として、128/64/32/16/8/4 点毎の折り畳みを検討した。Figure 2 は 16/8/4 点毎に折り畳んだ際の配置を示している。4 点で折り返す際は、各筋電データ毎に  $4 \times 256$  の列画像が得られる。この列画像を 4 列並べて  $16 \times 256$  の列画像を生成し、これを 16 ch.分並べた。提供された筋電データは、下肢から肩にかけての筋肉 8 か所について、左右の筋電データが与えられている。Figure 2 に示すように、(a)左データと右データを連続した配置、(b)左データと右データを交互に配置、の 2 つを検討したところ、(b)左右交互に配置した方が高いスコアが得られやすかった。左右の筋電データを交互に配置した方が左右の筋肉の力の違いを捉えやすく、高スコアに繋がると考えられる。折り畳みの点数は、短かく折り返すほどスコアが良くなる傾向が見られた(左右交互配置の

公開スコア : 1.288(128 点), 1.166(64 点), 1.110(32 点), 1.017(16 点), 1.007(8 点), 0.990(4 点))。短く折り返すと時系列的に近いデータが近くに配置されやすいため、本タスクのような短時間の変化を捉えるタスクでは高スコアが得やすいと考えられる。

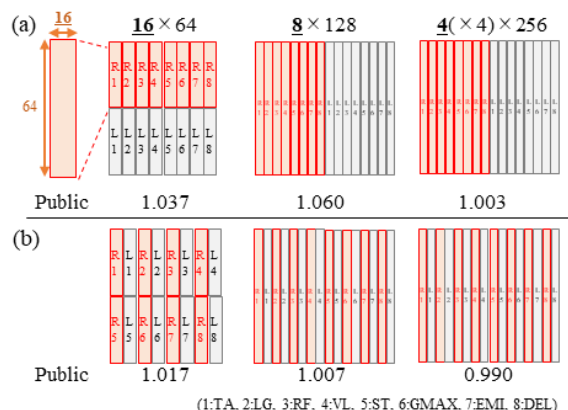


Figure 2 Arrangement when data is folded every 16/8/4 points. (a) Images showing a continuous sequence of L and R (b) Images showing an alternating sequence of L and R

GradCAM による画像内の重要度の可視化結果を Figure 3(a)(b)に示す。明るい箇所は重要度が高いことを意味し、多数画像の重要度を平均化した結果を示している。(a)の縦方向(時間方向)の依存性をみると、上部(初期段階)が特に重要であり、次に終盤データが重要と分かった。一般的に、画像の端部は畳み込みプロセス中に学習が困難であるため、(c)(d)で示されるように、画像全体を 150 ピクセル下方にシフトする処理を適用した。この処理により、重要度が画像の中央付近にシフトし、スコアも向上した(公開スコア : 0.826(シフト前)→0.797(シフト後))。

以上の基礎評価を踏まえて、次の手順で画像配置した：①時系列データ 1024 点を 4 点ごとに折り返し、 $4 \times 256$  の列画像を 4 列並べて  $16 \times 256$  の列画像を作成、②その列画像を左右交互に 16 ch.並べて  $256 \times 256$  の画像に配置、③作成した画像全体を 150 ピクセル下方にシフトし、はみ出た部分をラップアラウンド。生成された画像の一例を(e)に示す。学習用と評価用の画像データ一式は.pkl ファイルで保存し、モデルから容易に呼び出せるように構成した。

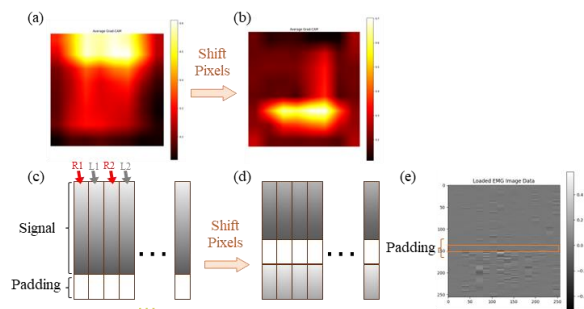


Figure 3 (a) Image showing heatmap of importance using GradCAM before and after shifting, (b) schematic diagrams of images before (c) and after (d) shifting, (e) example of data after shifting.

## 2. モデリングの工夫点と考察

各走行データには、LED サインによる「右キックターン」「左キックターン」「パンピング」のいずれかの指示が与えられており、これらは速度ベクトルに密接に関連する重要なメタ情報である。しかしこのコンペでは、走行データと走行指示の関係が公開されていないため、速度ベクトルのデータから LED サインを教師なしで3つに分類するモデルを開発した。

基礎評価として、提供された被験者 ID1-5 の学習用速度データを用いて、PCA で主成分二軸を抽出し、その後 K-means で3クラスに分類した(Figure 4(a))。主成分分析の結果は N 字型に分布しており、クラスの境界は明瞭でないものの、概ね3つのグループに分かれる傾向にあった。これに基づき、学習用データから PCA を用いて 10 次元の特徴量を抽出し、GMM を用いて3クラス分類を行った。そのクラス分類結果を「推定 LED サイン」として、(b)に示すように画像データのパディング領域の一部にメタ情報として埋め込んだ。

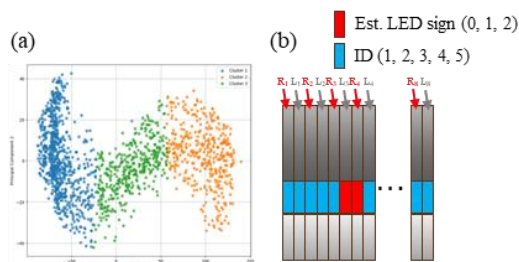


Figure 4 (a) Results of classifying into three classes using K-means after extracting the first two principal components from the training speed data using PCA. (b) Method of embedding estimated LED signs and participant IDs in the padding area.

モデルの全体構成を Figure 5 を用いて説明する。このモデルは3段階構成で、第1段階では EfficientNetb0 を用いてパディング領域をゼロ埋めした画像データで学習した。このモデルで推論されたテスト速度データから「推定 LED サイン」を取得し、第2段階のテスト画像データのパディング領域に埋め込んだ。第2段階の学習用画像データのパディング領域には、提供された学習用速度データから取得した推定 LED サインを埋め込んだ。外側広筋・大腿四頭筋直腿筋・大腿四頭筋外側広筋・半腱様筋の筋電データは重要度が高い傾向が見られたことから、推定 LED サインは、重要度が高い下肢特徴量領域の付近の Padding 領域に配置し、その他の Padding 領域には、被験者 ID を識別するメタ情報を埋め込んだ(Figure 4(b))。

第2段階では、画像データを CNN 系、Transformer 系、Conv 系の複数モデルで学習し、推論を行った。学習データの用意には多様なアプローチを採用し、①被験者 ID1-5 の学習用データのみを用いたモデル、②①に被験者 ID5 のテストデータを追加したモデル、③被験者ごとに個別にチューニングしたモデルをそれぞれ準備した。第1段階と第2段階では 10-fold のクロスバリデーションを使用し、層化 K 分割による交差検証を行った。第3段階のモデルでは、第2段階で CV とスコアが優れていた 16 種類の予測モデルを、被験者 ID (1, 2, 3, 4, 5) と速度方

向 (X, Y, Z) ごとに Lasso 回帰でスタッキングし、最終的な提出ファイルを作成した。スタッキングによりスコアが改善する要因として、被験者 ID と速度方向によってモデルの得意不得意が分かれることが挙げられる。ハイパーパラメータ調整には、被験者 ID5 のテストデータのスコアを参考にした。

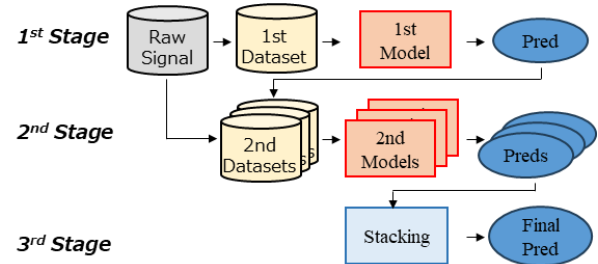


Figure 5 Overall structure of the model.

## 3. 分析結果から得られたインサイトと考察

本コンペでは、短時間(0.5 秒)の速度データを推論するための画像モデルを検討した。試行した複数の単体モデルの中で、EfficientNetb0 との相性が良く、第1段階のモデルで公開スコアは 0.804、第2段階では 0.747 を記録した。このタスクでは、短時間の筋電位の変化を学習しやすいよう、信号を短周期で折り返した画像を用意し、CNN 系モデルの適用が効果的と分かった。一方で、もし長期的な人の動き(例えば1時間や1日)を捉える場合は、大域的な傾向を捉えやすい高周期で折り返した画像や Transformer 系のモデルで高スコアを得やすいと予測される。画像モデルを用いたアプローチは、学習済のネットワークを持つ優れたモデルを簡単に試しやすく、最適なアーキテクチャを探索しやすい利点がある。

GradCAM を使用した可視化により、計測開始時刻付近の下肢筋電データの重要度が高いことが示唆された。この結果をふまえると、実験デザインを工夫してさらに精度向上させる方法として、①今回の計測開始時刻よりも前の筋電データを追加する、②下肢の筋電データを追加する、③下肢の筋電以外のセンサーデータ(3 軸加速度や足底圧力など)を追加することが有効と考えられる。画像化モデルは、時間区間を変えたり、サンプリングレートが異なる新たなセンサーデータを追加しても、画像化、学習、判定の一連の流れを変えることなく実施可能であり、実験サイクルを回しやすい。一般に実験デザインは実験者の知識や経験に基づいて設計されるが、可視化により、判定に有用な時間領域、センサー種別、センサー貼付け位置を効率的に探索できる。

時系列データを画像化するアプローチは、同時開催された「スケートボードトリック分類チャレンジ」にも適用したが、データ①が 22 位/44 位、データ②が 24 位/36 位であった。「重心位置予測チャレンジ」に注力したため、トリック分類コンペでは画像配置の工夫が十分でなかったと思われるが、上述したような配置の工夫により、精度向上の余地があると考えられる。

[ref] 5<sup>th</sup> solution, BirdCLEF 2024 in Kaggle (2024).

Available from <https://www.kaggle.com/c/birdclef-2024/discussion/511535>