

コンペ概要

「CMI - Detect Behavior with Sensor Data」は、Child Mind Institute（米国の児童精神衛生研究機関）が主催するコンペで、手首装着型センサデバイス（Helios）から取得した時系列データを用いて**体中心型反復行動（BFRB）**を識別するモデルを構築することが目的です。具体的には、髪を引っ張る・皮膚を引っかくなどのBFRB的ジェスチャーを、食事や眼鏡の調整といった一般的動作（非BFRB）と区別できるかを競います¹²。賞金総額は5万ドル、参加登録締切は2025年8月26日（提案期限）とされています²¹。この課題は、Heliosデバイスなどのウェアラブルによる**BFRBの検出技術**を進展させ、治療やモニタリングへつなげる研究に貢献します。実際にChild Mind InstituteはこのHelios手首デバイスで熱センサーを追加した場合のBFRB検出精度向上を示す研究成果も発表しています³。BFRBは全人口の約5%以上に見られる比較的高頻度な症状であり⁴、患者自身も気づきにくいいため、ウェアラブルデバイスで検出しリアルタイムにフィードバックを与える技術の実用化が期待されています。

データ構成

提供データはHeliosデバイスで収集されたセンサ時系列情報です。Heliosには**複数の種類のセンサー**が搭載されています：三軸加速度・角速度・磁気計（IMUセンサー、BNO080等）、5個の赤外線（サーモパイル）センサー、そして5個の距離（Time-of-Flight）センサーです³²。これにより、手首の動き（動作の動き・姿勢変化）だけでなく、皮膚や周囲との温度変化や距離情報も同時に取得できます。

データセットには、**8種類のBFRBジェスチャー**と**10種類の非BFRBジェスチャー**（合計18種）が記録されています。参加者は立位・座位など複数の体勢で各ジェスチャーを行っており、各シークエンス（例：髪引っ張り動作）に対して3つの区間「移行動作（Transition）→本動作（Gesture）→停止（Pause）」が含まれます。各シークエンスには固有の `sequence_id` が振られ、トレーニングデータではその**ジェスチャー名（ラベル）**と**タイプ（target: BFRB系 or non-target: 非BFRB系）**が付与されています。

提供ファイル例（推測）を挙げると、トレーニング用のセンサーデータ（例：`train_features.csv`）には行毎に `sequence_id` と各時刻のセンサー値列（例：

`acc_x, acc_y, acc_z, gyr_x,..., thm_1,..., thm_5, tof_1,..., tof_5`）が並び、`train_labels.csv` で `sequence_id` と対応するジェスチャー（文字列）やターゲットフラグが与えられている構成と考えられます。テスト用データ（例：`test_features.csv`）にはラベル無しのシークエンスが含まれ、さらに `test_demographics.csv` など参加者属性（身長・体重など）の補足情報が付与されている可能性があります（Kaggleノート上で言及あり）。データは生のセンサ信号なので、ノイズ除去やセンサー間の基準合わせ（キャリブレーション）、正規化など**前処理**が重要となります。例えば、熱・距離センサーは被験者の利き手（右手・左手）によって向きが変わるため、利き手情報でセンサーチャンネルを反転する処理などがNotebook上で試されています（“handedness”に応じたチャンネル正規化）⁵。

評価指標

公式ルールには明記されていませんでしたが、Kaggle上の参加者ノートブックでは**二値F1スコア**（BFRB系か否か）と**マルチクラスF1スコア**（非BFRB系を1クラスにまとめたジェスチャー分類）の平均を最終評価値とする設定が使用されているようです（最終スコア = (binary_F1 + macro_F1)/2）⁶。そのため、BFRB検出性能とジェ

スチャー識別精度の両方が重視されます。不正なジェスチャー名（訓練セットにないラベル）で予測するとスコアリングができない仕様になっています。

提出形式

提出ファイルはCSV形式で、`sequence_id`と`gesture`の2列を持ちます。`sequence_id`は各シークエンスのID、`gesture`には予測したジェスチャー名を記入します。BFRB系ジェスチャーであれば具体名（例：`hair_pull`、`skin_pick`など）を、非BFRB系であれば“non_target”という共通ラベルを使います（下表はサンプルの一例です）。

sequence_id	gesture
SEQ_000001	hair_pull
SEQ_000002	non_target
SEQ_000003	non_target
SEQ_000004	skin_pick

（実際のシークエンスIDやジェスチャー名は上記例と異なる可能性があります、フォーマットは同様です。）

コンペルール

本コンペはKaggleプラットフォーム上で実施される一般的なルールに従います。参加にはKaggleアカウント登録と公式ルールへの同意が必要です。チームは最大5名まで組成可能で、個人参加も可です。データおよびコードの取り扱いでは、主催者が提供したデータ以外の外部データ使用は原則不可（公開データやユーザ作成データも使用制限有）です。コンペ期間中のプライベートなコード共有は禁止され、公開Kaggleノートブックを通じた情報交換が奨励されます。提出回数や終了後のコード公開可否など詳細は公式Rulesタブを参照してください。

ベースラインコード

公式ページやKaggleチームから提供された**ベースラインコード**には、サンプルとして単純なモデルの例（例えばCatBoostや1D-CNNなど）でデータを読み込み、学習→推論する流れが示されています。ある投稿では単一のCatBoostモデルでCV約0.69の性能を得たと報告されており、これを出発点としてさらなる特徴量エンジニアリングやモデル改善を行うイメージです [7](#)（※参照のみ）。ベースラインの主な意図は「データフォーマットと評価方法を示し、初心者でも参入できる雛形を提供すること」です。ユーザはこれをもとに特徴抽出やモデル選択の実装を拡張していきます。

その他の技術的ポイント

本課題はマルチセンサー・マルチクラスの時系列分類問題です。一般に**時系列モデル**（1D-CNN、LSTM、Transformerなど）や**勾配ブースティング系モデル**（CatBoost/LightGBM等＋時系列統計特徴量）を組み合わせるアプローチが有効です。まずは各シークエンスの統計量（平均・分散・最大値など）やFFTスペクトル、相関特徴量などを特徴量として生成し、ツリー系モデルで学習する手法が良いスタートポイントとなります。

Hot tip: 先行研究でも指摘されているように、IMU（動き）のみならず**熱・距離センサー情報を統合**することで

BFRB検出精度が向上します³¹。例えば皮膚が近づく方向の熱チャンネルに特徴が出たり、特定のToFセンサーが近接を捉えたりするので、全チャンネルをモデルに投入するのが望ましいです。

さらに、**データ増強**（例えばノイズ追加、時間軸反転、センサー軸ごとのドロップアウトなど）や**アンサンブル**により性能改善が期待できます。Kaggleノートブックでは、IMUだけを用いたCNNモデルや、IMU+熱+距離センサーを統合したTabNet・LightGBMモデルなど様々なアプローチが試されています。

公開ディスカッション／ノートブック

Kaggleフォーラムや公開ノートブックには、EDAやモデル実装例が多く投稿されています。例えば「Sensor Pulse: Viz & EDA for BFRB Detection」では熱センサーデータの可視化と利き手補正について議論され、「Deep Dive EDA + CatBoost」では初期探索結果と簡易モデルのコードが共有されています。これらを参考にして「どのセンサーがどのジェスチャーで特徴的か」「特徴量生成のアイデア」「モデル構築手法」などの知見を得ることができます。ノートブック投稿例では、最終的に公開リーダーボードスコアが0.70~0.72程度を達成しており、与えられたベースライン（約0.69）からさらに改善されていることが報告されています。各ノートブックにはコードが含まれているので、実装の参考にすることが推奨されます。

参考: 提供情報によれば本コンペは手首デバイスHeliosの複数センサー（動き、熱、距離）を用いる点が特徴であるとされています²¹。また、Child Mind Instituteの研究では熱センサー追加がBFRB検出精度を大幅に改善することが示されており³、本コンペでもマルチモーダルセンサー統合が鍵となりそうです。

参考文献: Kaggleコンペページおよびリンク先資料²¹³。

¹ Excited to share this challenge from the Child Mind Institute and Kaggle! | Yuki Kotani, MBA
https://www.linkedin.com/posts/yukikotani_cmi-detect-behavior-with-sensor-data-activity-7337092026187804672-iv31

² Competition Launch Alert! | Kaggle
https://www.linkedin.com/posts/kaggle_cmi-detect-behavior-with-sensor-data-activity-7334205627998646272-b_Q7

³⁴ Gesture recognition wearable
<https://matter.childmind.org/gesture-recognition-device.html>

⁵ Sensor Pulse | Viz & EDA for BFRB Detection - Kaggle
<https://www.kaggle.com/code/tarundirector/sensor-pulse-viz-eda-for-bfrb-detection>

⁶ CMI 2025 | Only LightGBM Training - Kaggle
<https://www.kaggle.com/code/jiaoyouzhang/cmi-2025-only-lightgbm-training>

⁷ CMI2025 | Public | Baseline | V3 - Kaggle
<https://www.kaggle.com/code/ravi20076/cmi2025-public-baseline-v3>