Capstone Project Report AI/ML

CIMT College of Information Management and Technology

Post Graduate Artificial Intelligent and Machine Learning

Detecting Behavioral Patterns Using Wearable Sensor Data

Albert Tchaptchet Womga

Instructor : Priya Virdi

June 23, 2025

Abstract

This capstone project explores the detection of Body-Focused Repetitive Behaviors (BFRBs) using multi-modal sensor data from the Helios wrist-worn device. BFRBs, such as hair pulling and skin picking, can significantly impact mental health and well-being. The project aims to develop a machine learning model that can classify BFRB-like gestures from non-BFRB gestures using time-series data from inertial measurement units (IMUs), thermopile sensors, and Time-of-Flight (ToF) sensors.

Contents

Abstract 1

Introduction **3**

1 Problem Definition **4**

2 Methodology **5**

2.1 Data Collection 5

2.1.1 Helios Device data structure 5

2.1.2 Demographic data structure 8

2.2 Data Cleaning 9

2.2.1 Filtering Relevant Sequences 9

2.2.2 Handling Missing Values 9

2.2.3 Normalization 10

2.2.4 Summary 10

2.3 Algorithm: Classical Machine Learning Approach 10

2.3.1 Problem Formulation 10

2.3.2 Feature Extraction from Time-Series 11

2.3.3 Algorithms Used 11

2.3.4 Comparative Analysis of Algorithms 12

2.3.5 Evaluation Metric 15

3 Results and discussion **16**

3.1 Feature Exploration and Signal Processing 16

3.2 Model Training and Evaluation 17

3.3 Random Forest Model Results 17

3.3.1 Binary Classification (BFRB vs. Non-BFRB) 17

3.3.2 Multiclass Classification (Gesture Types) 18

3.4 XGBoost Model Results 18

3.4.1 Binary Classification (BFRB vs. Non-BFRB) 18

3.4.2 Multiclass Classification (Gesture Types) 19

3.5 Model Selection and Comparison 21

Conclusion **22**

Future Scope **23**

Bibliography **24**

Introduction

Body-Focused Repetitive Behaviors (BFRBs) are a group of compulsive actions, including hair pulling (*trichotillomania*), skin picking (*excoriation*), and nail biting. While these behaviors are relatively common, they can become chronic and significantly impair physical and mental well-being, especially when associated with anxiety disorders and obsessive-compulsive disorder (OCD). Despite their prevalence, BFRBs often go undetected in clinical settings, leading to delayed diagnosis and treatment.

Recent advances in wearable technology and artificial intelligence have opened promising avenues for the real-time monitoring and classification of human behavior. Among these innovations is the Helios wrist-worn device, developed by the Child Mind Institute. Unlike traditional wearables that rely solely on Inertial Measurement Units (IMUs), the Helios device incorporates a combination of:

- **Inertial Measurement Units (IMUs)** for motion tracking,
- **Thermopile sensors** for detecting infrared radiation (body heat),
- **Time-of-Flight (ToF) sensors** for measuring proximity to nearby surfaces.

This multi-modal sensing approach enables a more comprehensive capture of hand and wrist movements, temperature gradients, and proximity patterns—key indicators of BFRB-like activity.

This project investigates the feasibility of using sensor data from the Helios device to distinguish between BFRB-like gestures and everyday, non-BFRB hand movements. By leveraging time-series sensor data and machine learning models, we aim to build a robust classification system capable of not only detecting BFRB tendencies but also identifying the specific gesture type.

The ability to distinguish these behaviors with high accuracy has profound implications for mental health monitoring, behavioral therapy, and the broader domain of human activity recognition in Internet of Things (IoT) applications.

In this capstone project, we explore time-series analysis, feature engineering, and both classical and deep learning methods to develop a behavioral detection model. Our ultimate goal is to evaluate the effectiveness of multimodal sensor fusion in improving detection accuracy and to provide practical insights into the deployment of intelligent wearable systems for real-world behavioral health support.

Chapter 1

Problem Definition

The central challenge is to distinguish BFRB-like gestures from visually and kinesthetically similar everyday gestures using multi-sensor time-series data. The complexity arises due to high-dimensional signals, missing data points, and the similarity between gestures.

We aim to:

- Detect if a given gesture is BFRB-related (binary classification).
- Classify the specific type of BFRB gesture (multi-class classification).
- Handle partial sensor availability and noisy input.

Body-Focused Repetitive Behaviors (BFRBs) are often subtle and varied, presenting a major challenge for detection using conventional sensor modalities. The fine-grained nature of these gestures—such as scratching the neck or pulling hair above the ear—can closely resemble benign, everyday hand movements like adjusting glasses or drinking from a cup. This ambiguity makes it difficult for standard motion-sensing wearables to reliably differentiate between harmful and non-harmful behaviors.

The core problem addressed in this project is to develop a predictive model capable of classifying time-series sensor data collected from the Helios wrist-worn device into either BFRB-like or non-BFRB-like gestures. More specifically, the challenge involves two key classification tasks:

1. **Binary classification:** Determine whether a gesture belongs to the "target" class (i.e., BFRB-like) or the "non-target" class (i.e., non-BFRB everyday gestures).
2. **Multi-class classification:** Identify the specific type of BFRB-like gesture (e.g., "Pull hair above ear", "Scratch forehead") or categorize the action as a general non-BFRB gesture (i.e., "non-target").

The complexity of the problem is further amplified by the lack of labeled data in the test set, the high dimensionality of sensor inputs (including 320+ ToF channels), and the need to capture temporal dependencies over sequences of gesture execution. Furthermore, to evaluate the utility of multimodal sensor integration, the model must be robust enough to perform well under two different sensor availability scenarios: (1) IMU-only data, and (2) full sensor data (IMU, thermopile, and ToF).

Thus, the problem extends beyond simple classification to include the engineering of time-aware features, careful handling of missing or noisy data, and the design of model architectures that can generalize across subjects and sensor modalities.

Chapter 2

Methodology

This section describes the methodology adopted to solve the problem of classifying Body-Focused Repetitive Behaviors (BFRBs) from non-BFRB gestures using time-series sensor data. The process follows a structured data science pipeline that includes data acquisition, cleaning, feature engineering, and classical machine learning modeling. Due to time constraints, this study focuses exclusively on traditional supervised learning techniques, excluding deep learning approaches.

2.1 Data Collection

The dataset analyzed in this research comes from the Kaggle competition *"CMI: Detect Behavior with Sensor Data"* [3]. Data was gathered using the Helios wrist-worn device, created by the Child Mind Institute, along with demographic information about participants—details that can serve as valuable features to enhance model performance [1]. The Helios device contains multiple sensors that record multi-modal time-series data, allowing for in-depth examination of hand and wrist movements related to BFRB-like behaviors. It incorporates three types of sensors:

2.1.1 Helios Device data structure

- **Inertial Measurement Unit (IMU):** providing acceleration $\vec{a}(t) = [a_x(t), a_y(t), a_z(t)]$ and orientation in quaternion form $\vec{q}(t) = [w(t), x(t), y(t), z(t)]$.

- **Thermopile Sensors:** measuring infrared-based temperature from five directions, $\vec{Thm}(t) = [thm_1(t), thm_2(t), thm_3(t), thm_4(t), thm_5(t)]$.

- **Time-of-Flight (ToF) Sensors:** recording proximity across 320 spatial bins, denoted as $\vec{Tof}(t) \in \mathbb{R}^{320}$.

Each gesture recording is structured as a sequence indexed by $t \in \{0, 1, \ldots, T\}$, where T is the maximum sequence length (typically 63 time steps). Each row in the data corresponds to a timestamp within a sequence and includes:

- Temporal context: `sequence_id`, `sequence_counter`

- Label annotations: `gesture` (18-class), `behavior` (binary target - BFRB-like behavior (positive class) /non-target - non-BFRB-like behavior (negative class))

- Sensor readings: IMU, thermopile, and ToF
- Participant metadata: subject ID, age, sex, handedness, etc.

A visual summary of the Helios data format is provided in Figure 2.1.

Figure 2.1: Sensor placement and signal flow in the Helios device. IMU, thermopile, and ToF sensors provide complementary spatial and temporal information.

A subset of the sequences was annotated with phase labels such as `Moves hand to target location` or `Performs gesture`. This study retains only the `Performs gesture` phase, which corresponds to the actual execution of the behavioral pattern.

For modeling purposes, the original gesture label $y \in \mathcal{G}$ was collapsed to:

$$y' = \begin{cases} y, & \text{if behavior is target (i.e., BFRB-like)} \\ \text{non\_target}, & \text{otherwise} \end{cases}$$

resulting in a simplified label set $\mathcal{G}'$ with 9 gesture classes (8 BFRB types + 1 non-BFRB class).

2.1.2 Demographic data structure

Demographic information was provided in separate files (train_demographics.csv and test_demographics.csv), and included the following variables:

- **participant_id**: Unique identifier for each participant
- **adult_child**: 0 = child (< 18 years), 1 = adult ($\geq$ 18 years)
- **age**: Participant's age in years
- **sex**: 0 = female, 1 = male
- **handedness**: 0 = left-handed, 1 = right-handed
- **height_cm**: Height in centimeters
- **shoulder_to_wrist_cm**: Distance from shoulder to wrist in centimeters
- **elbow_to_wrist_cm**: Distance from elbow to wrist in centimeters

Demographics will be added to sequence-level features for both training and test predictions. The merge is typically done after we extract our per-sequence features, but before model training or inference. Some BFRBs or gestures might be easier/harder to detect depending on age, sex, handedness, or body size. For example: Movement patterns can differ for children vs adults, males vs females, left-handed vs right-handed users.

Table 2.1: Summary Statistics Table for Demographic Features

Metric	**Mean**	**Std Dev**	**Min**	**Max**
Age	21.81	10.29	10.0	53.0
Height (cm)	167.99	10.61	135.0	190.5
Shoulder to Wrist (cm)	51.58	4.89	41.0	71.0
Elbow to Wrist (cm)	25.47	3.03	18.0	44.0
Female (%)	38.27	–	–	–
Left Handed (%)	12.35	–	–	–

Table 2.1 presents summary statistics for the study participants. The average age was approximately 21.8 years (SD = 10.3), with ages ranging from 10 $to 53\ years.$

The mean height was 168 cm ($SD = 10.6$), with participants' heights ranging from 135 $to 190.5\ cm$. Shoulder-to-wrist and elbow-to-wrist distances averaged 51.6 $cm(SD = 4.9)$ and 25.5 $cm(SD = 3.0)$, respectively. The cohort was 38.3% female and 12.4% left-handed. These demographic characteristics highlight the diversity of the sample in both age and physical attributes, which were incorporated as additional features in the modeling process to help account for inter-individual variation in sensor-derived gesture patterns.

2.2 Data Cleaning

Before building any predictive model, it is essential to preprocess the data to ensure consistency, integrity, and suitability for analysis. The raw Helios dataset includes over 300 sensor channels and several metadata columns. However, it contains some inconsistencies that must be addressed:

1. Missing values represented by the constant -1 (especially in ToF sensors),
2. Sensor readings with vastly different units and scales,
3. Non-relevant phases (e.g., transition sequences) that must be filtered.

2.2.1 Filtering Relevant Sequences

Each row in the dataset contains a `sequence_type` column that annotates the current action phase. For classification purposes, only rows with:

$$\texttt{sequence\_type} = \texttt{"Performs gesture"}$$

were retained, as these correspond to the actual behavior (target or non-target) being performed.

2.2.2 Handling Missing Values

Time-of-Flight (ToF) sensor readings frequently include -1 to indicate the absence of a reflected signal. These values are replaced by the standard missing value placeholder `NaN`, allowing mathematical imputation to be applied consistently.

Let $x_i^{(j)}$ be the i-th reading of sensor j. Define:

$$x_i^{(j)} = \begin{cases} \texttt{NaN}, & \text{if } x_i^{(j)} = -1 \\ x_i^{(j)}, & \text{otherwise} \end{cases}$$

We apply column-wise imputation:

- For classical machine learning models, **median imputation** was used:

$$\text{median}\left(\left\{x_k^{(j)} \mid x_k^{(j)} \neq \texttt{NaN}\right\}\right)$$

- For advanced temporal modeling, optional **rolling window imputation** can be applied within each sequence to preserve local patterns.

2.2.3 Normalization

Sensor readings come from different modalities and have varying physical units (e.g., acceleration in m/s^2, temperature in $^\circ C$, proximity in mm). To ensure fair treatment during learning, all continuous sensor values were standardized using z-score normalization:

$$\hat{x}_i^{(j)} = \frac{x_i^{(j)} - \mu_j}{\sigma_j}$$

where μ_j and σ_j are the mean and standard deviation of sensor j computed from the training set.

2.2.4 Summary

After completing these steps, the cleaned dataset contained:

- No missing values or placeholders,
- Normalized sensor readings,
- Sequence rows filtered to only the gesture performance phase,
- A harmonized format ready for feature engineering and model input.

2.3 Algorithm: Classical Machine Learning Approach

Due to time constraints and the need for explainability, this study focuses on classical supervised learning algorithms instead of deep learning models. These methods are well-suited for structured tabular data, particularly when paired with carefully engineered time-series features.

2.3.1 Problem Formulation

Let $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_{i=1}^N$ denote the dataset, where:

- $\mathbf{x}^{(i)} \in \mathbb{R}^d$ is the feature vector derived from the i-th gesture sequence,
- $y^{(i)} \in \mathcal{Y}$ is the corresponding gesture label (9-class gesture_slim),
- $\mathcal{Y} = \{\text{non\_target}, g_1, g_2, \dots, g_8\}$, where g_k are BFRB gestures.

The task is to learn a function $f : \mathbb{R}^d \to \mathcal{Y}$ that minimizes the classification error:

$$f^* = \arg\min_f \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\left\{f(\mathbf{x}^{(i)}) \neq y^{(i)}\right\}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function.

2.3.2 Feature Extraction from Time-Series

Each sequence was transformed into a fixed-size feature vector by aggregating time-series sensor values using statistical and temporal functions:

- **Statistical summaries:** mean, standard deviation, min, max, range
- **Signal deltas:** difference between time steps
- **Rolling features:** e.g., mean/std over a window size of $w = 5$
  - **Purpose:** Captures local trends, smooths noise, and summarizes recent values in a time window.
  - **Interpretation:** Helps the model detect gradual changes, movement trends, or persistent shifts in sensor readings.
  - **Example:** For an accelerometer signal, rolling mean/std can highlight sustained acceleration patterns, such as a hand moving toward the face.
  - **Implementation:** For each time step t, compute:

$$\text{rolling\_mean}_t = \frac{1}{w}\sum_{k=0}^{w-1} x_{t-k}, \quad \text{rolling\_std}_t = \sqrt{\frac{1}{w}\sum_{k=0}^{w-1}(x_{t-k} - \text{rolling\_mean}_t)^2}$$

- **Energy:** root mean square (RMS) of the IMU signals
- **FFT (Fast Fourier Transform) components:** low-frequency energy using Discrete Fourier Transform. Fast Fourier Transform (FFT) is a powerful signal processing technique that converts a time-domain signal (e.g., accelerometer or gyroscope data) into its frequency-domain representation. Extracting FFT-based features is valuable for gesture recognition and motion analysis because it allows us to capture periodic patterns and frequency components that may not be apparent in the raw time-series data. The FFT decomposes the signal into its constituent frequencies, enabling us to identify dominant frequencies and their amplitudes.
- **Time-domain features:** such as zero-crossing rate, signal energy, and entropy.

Let $x_j(t)$ be a time-series for feature j. The following engineered features were computed:

$$\text{mean}(x_j), \quad \text{std}(x_j), \quad \max(x_j), \quad \min(x_j), \quad \text{RMS}(x_j) = \sqrt{\frac{1}{T}\sum_{t=1}^{T} x_j(t)^2}$$

2.3.3 Algorithms Used

Three classical models were selected:

- Random Forest (RF): A robust ensemble method based on decision trees with bootstrap aggregation. It provides variable importance scores and is resilient to overfitting.
- XGBoost (Extreme Gradient Boosting): A powerful boosting-based algorithm that iteratively builds trees to correct residual errors. Known for high performance in tabular competitions.

2.3.4 Comparative Analysis of Algorithms

To effectively detect and classify body-focused repetitive behaviors (BFRBs), three classical machine learning models were employed: Random Forest and XGBoost. Each algorithm brings unique strengths in handling time-series sensor data with engineered features.

Random Forest (RF)

Random Forest is an ensemble model that aggregates the predictions of multiple decision trees trained on bootstrapped datasets. Its prediction is defined as:

$$f_{\mathrm{RF}}(\mathbf{x}) = \frac{1}{T}\sum_{t=1}^{T} h_t(\mathbf{x})$$

where h_t is the output of the t^{th} decision tree. It uses Gini impurity or entropy to select splits:

$$\mathrm{Gini}(D) = 1 - \sum_{i=1}^{C} p_i^2$$

This model is robust to overfitting, captures non-linear dependencies, and is suitable for high-dimensional sensor-derived features.

Random Forest Model Architecture

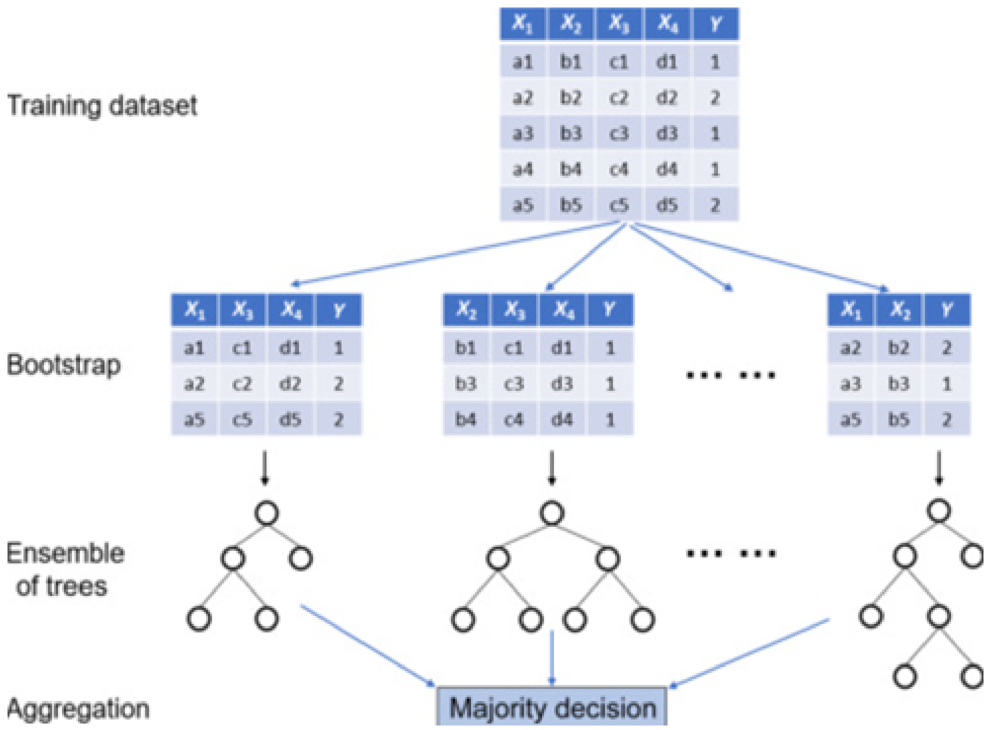

Figure 2.2: Random Forest Modeling Approach 1

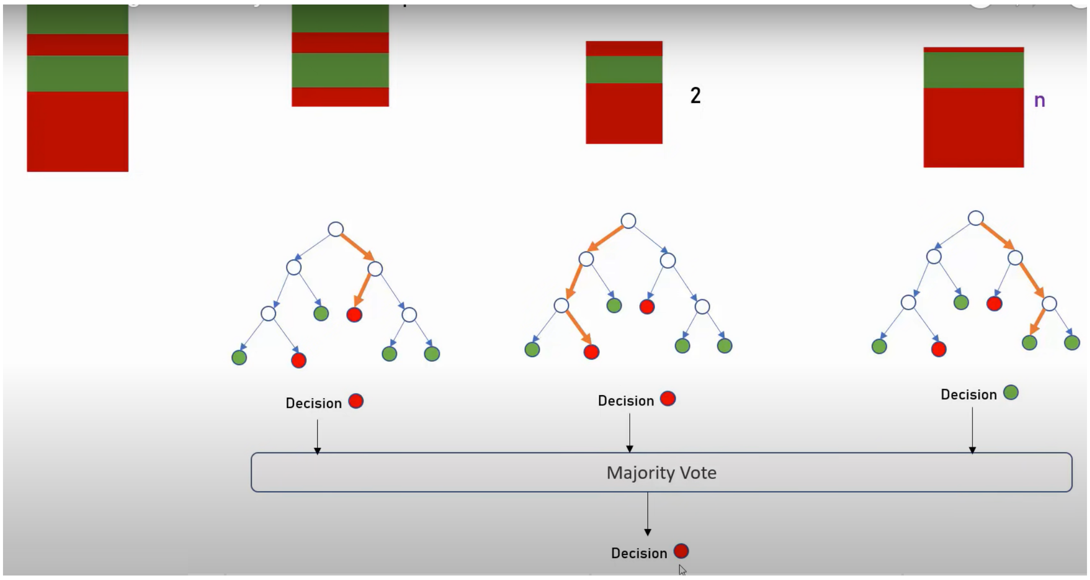

Figure 2.3: Random Forest Modeling Approach 2

XGBoost (Extreme Gradient Boosting)

XGBoost is a boosting-based algorithm that sequentially adds decision trees to correct previous errors. The optimization objective at stage t is:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t)$$

where l is the differentiable loss function (e.g., log-loss), and Ω is a regularization penalty for controlling model complexity.

Its key advantage lies in optimizing for both **bias and variance**, handling class imbalance well, and delivering state-of-the-art performance on tabular and time-series data.

Gradient boosting of shallow trees Model Architecture

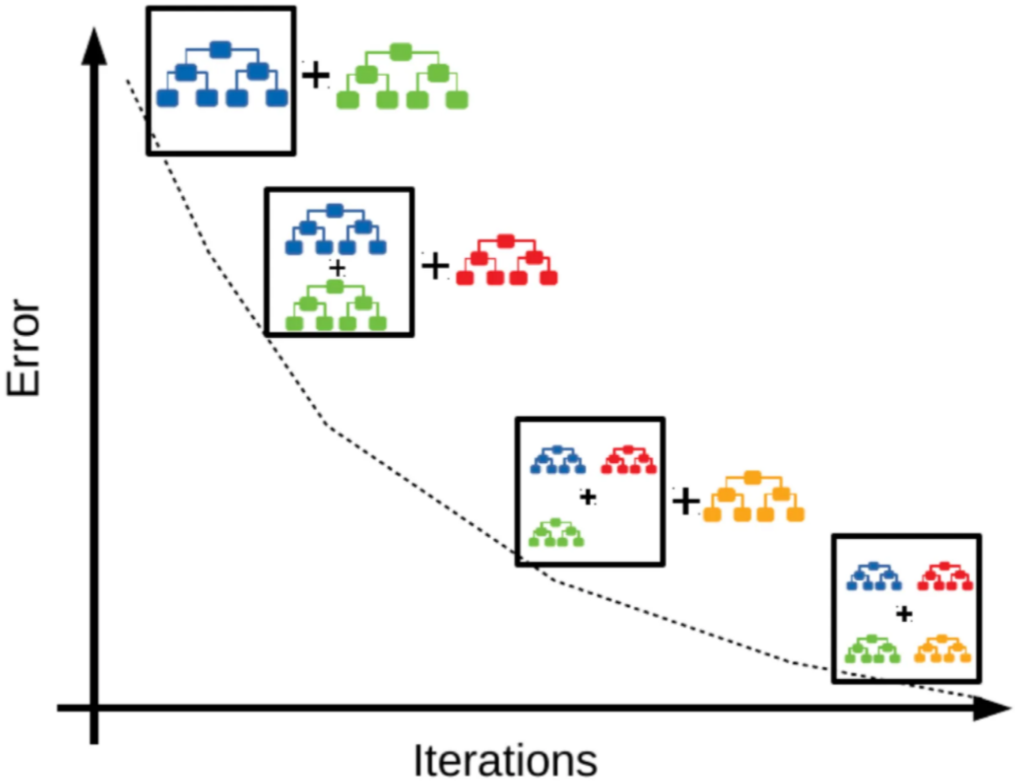

Figure 2.4: Gradient boosting of shallow trees Modeling Approach 1

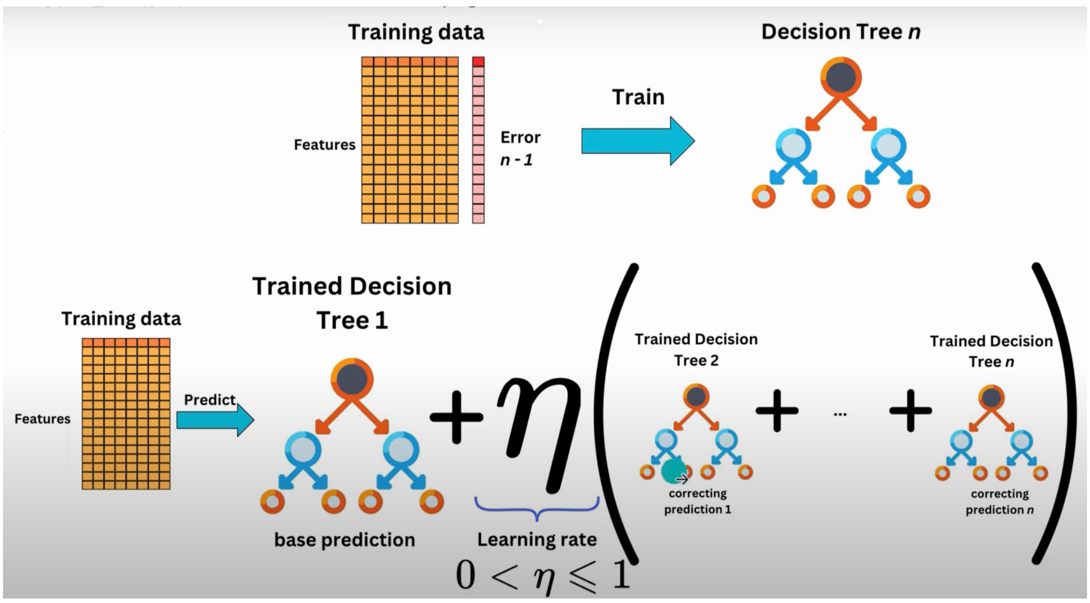

Figure 2.5: Gradient boosting of shallow trees Modeling Approach 2

Impact on Classification Metrics

Each algorithm contributes differently to performance:

- **Accuracy**: XGBoost generally achieves higher accuracy through boosting.
- **Loss Minimization**: XGBoost minimizes log-loss.
- **Confusion Matrix Balance**: XGBoost and RF handle false positives/negatives better via ensemble averaging.

Each model was trained using k-fold cross-validation ($k = 5$) for robustness, and hyperparameters were optimized using grid search.

The performance of the models was evaluated using a composite metric that combines binary and multi-class classification scores. This approach ensures that both the detection of BFRB-like behaviors and the differentiation among gesture types are effectively captured.

2.3.5 Evaluation Metric

Models were evaluated using a composite metric as required by the competition:

1. **Binary F1-score** for detecting whether a gesture is target or non_target:

$$\mathrm{F1}_{\mathrm{binary}} = 2 \cdot \frac{\mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

   - ⋆Precision: Of all gestures predicted as "target," how many were actually "target"?
   - ⋆Recall: Of all actual "target" gestures, how many did the model correctly identify?
   - ⋆F1-score: Harmonic mean of precision and recall, balancing false positives and negatives.

2. **Macro F1-score** across all 9 gesture classes:

$$\mathrm{F1}_{\mathrm{macro}} = \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} \mathrm{F1}_y$$

3. **Final Score:**

$$\mathrm{Score} = \frac{1}{2}(\mathrm{F1}_{\mathrm{binary}} + \mathrm{F1}_{\mathrm{macro}})$$

This composite metric balances the need for both binary classification accuracy and multi-class performance, ensuring that models are effective at detecting BFRB-like behaviors while also distinguishing among different gesture types.

Model Selection Justification

Given the moderate dataset size and the structured nature of the features, these classical algorithms offered a balance of interpretability, training speed, and predictive accuracy. Deep learning models such as LSTM and CNN were excluded due to time constraints, but could be considered in future extensions of this work.

Chapter 3

Results and discussion

3.1 Feature Exploration and Signal Processing

Feature engineering was performed to create a rich set of inputs for classification. Key features included:

- **Statistical Features:** Mean, standard deviation, min, max, range for each sensor channel.
- **Temporal Features:** Rolling means, deltas, and FFT coefficients for IMU channels.
- **Demographic Features:** Participant metadata

To extract informative features from the time-series sensor data, we computed rolling mean features (window size 5) for IMU acceleration and rotation channels. This approach smoothed short-term fluctuations while preserving gesture dynamics (Figure 3.1). Additionally, Fast Fourier Transform (FFT) analysis was applied to normalized `acc_x` signals to characterize dominant frequency components associated with repetitive behaviors (Figure 3.2).

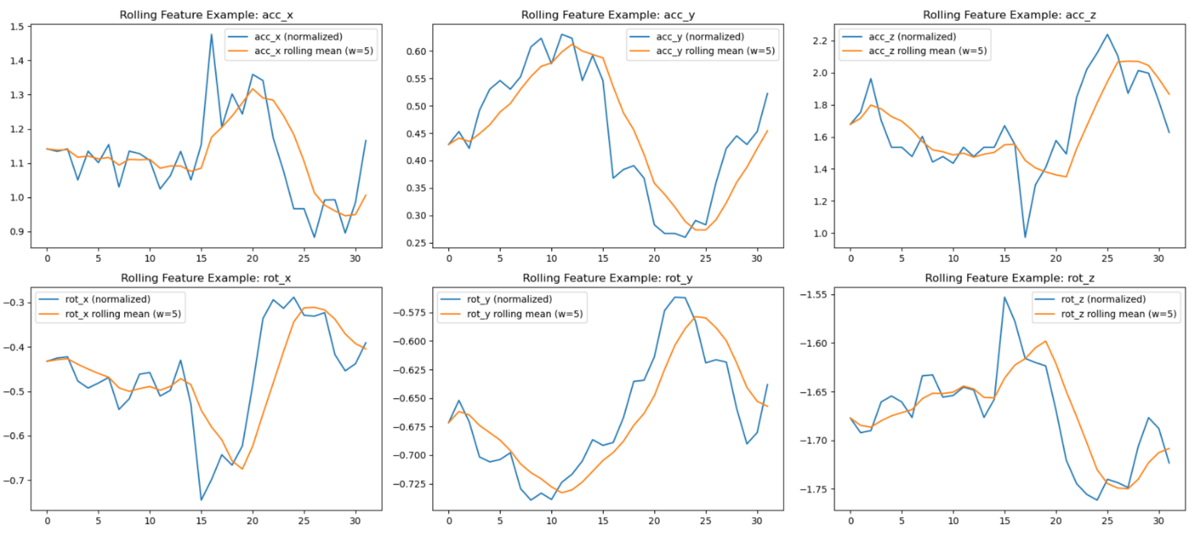

Figure 3.1: Rolling mean (orange) and raw (blue) values for representative IMU channels, highlighting temporal smoothing of gesture signals.

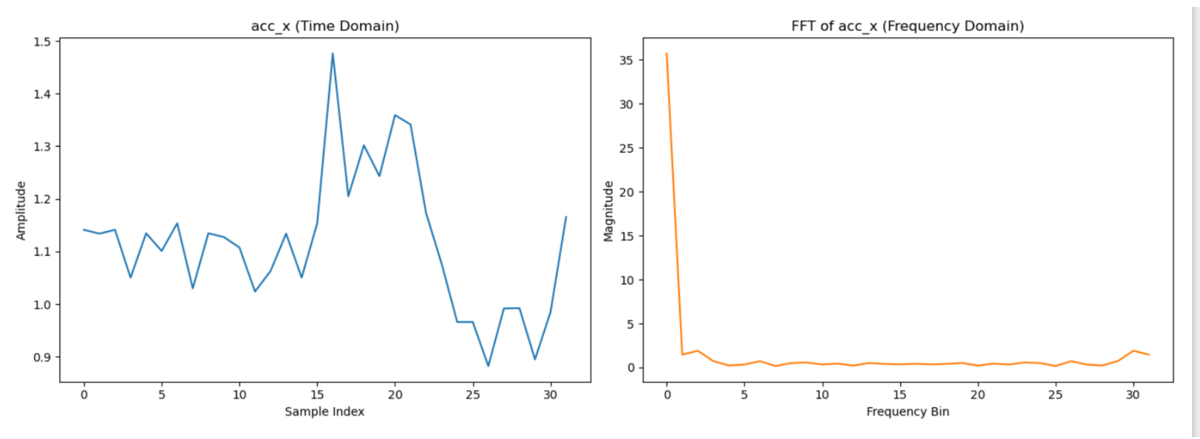

Figure 3.2: Time domain (left) and frequency domain (right, via FFT) for normalized `acc_x` channel.

3.2 Model Training and Evaluation

All models were evaluated using the official composite metric, which averages:

- **Binary F1-score:** Measures detection of BFRB-like gestures versus non-target gestures.
- **Macro F1-score:** Averages F1-score across all 9 gesture classes (8 BFRBs + `non_target`).
- **Final Composite Score:**

$$\text{Final Score} = 0.5 \times (\text{F1}_{\text{binary}} + \text{F1}_{\text{macro}})$$

3.3 Random Forest Model Results

3.3.1 Binary Classification (BFRB vs. Non-BFRB)

Classification Report:

```
                precision   recall   f1-score   support
non-BFRB           0.97      0.92      0.94       607
BFRB               0.96      0.98      0.97      1023

accuracy                               0.96      1630
macro avg          0.96      0.95      0.96      1630
weighted avg       0.96      0.96      0.96      1630

Binary F1-score: 0.9681
Binary Accuracy: 0.9595
```

Confusion Matrix (Figure 3.3):

$$\begin{bmatrix} 561 & 46 \\ 20 & 1003 \end{bmatrix}$$

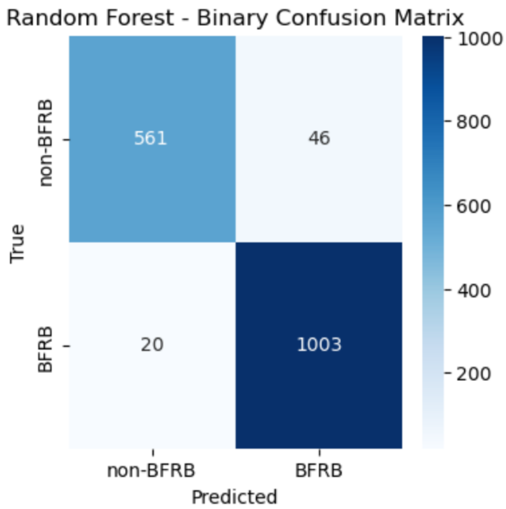

Figure 3.3: Random Forest Binary Confusion Matrix

3.3.2 Multiclass Classification (Gesture Types)

Classification Report:

	precision	recall	f1-score	support
Above ear - pull hair	0.72	0.62	0.67	128
Cheek - pinch skin	0.55	0.56	0.56	127
Eyebrow - pull hair	0.60	0.48	0.53	128
Eyelash - pull hair	0.64	0.48	0.55	128
Forehead - pull hairline	0.71	0.67	0.69	128
Forehead - scratch	0.69	0.74	0.71	128
Neck - pinch skin	0.57	0.55	0.56	128
Neck - scratch	0.59	0.57	0.58	128
non_target	0.86	0.98	0.91	607
accuracy			0.73	1630
macro avg	0.66	0.63	0.64	1630
weighted avg	0.72	0.73	0.72	1630

```
Macro F1-score: 0.6407
Multiclass Accuracy: 0.7294
```

Final Composite Score:

F1_binary: 0.9681
F1_macro: 0.6407
Final Score: 0.8044

3.4 XGBoost Model Results

3.4.1 Binary Classification (BFRB vs. Non-BFRB)

Classification Report:

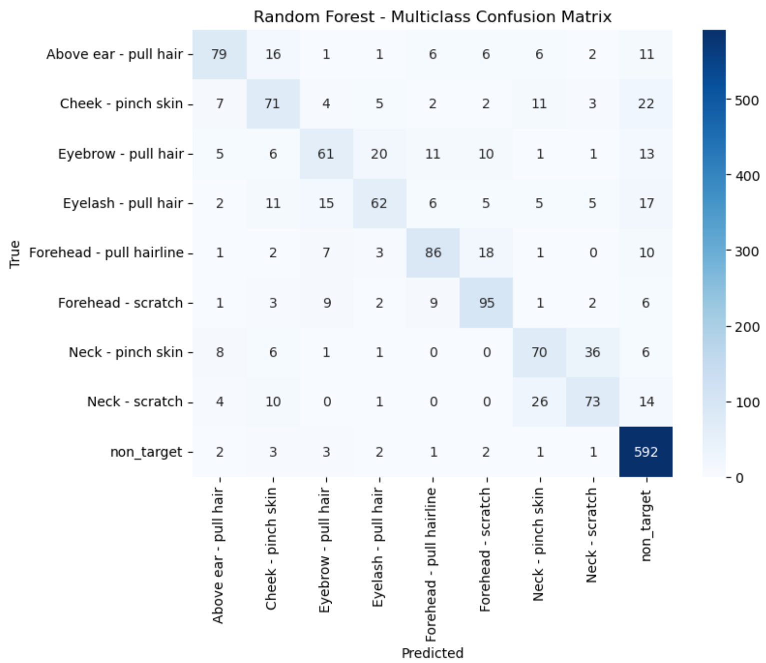

Figure 3.4: Random Forest Multiclass Confusion Matrix

	precision	recall	f1-score	support
non-BFRB	0.97	0.96	0.97	607
BFRB	0.97	0.99	0.98	1023
accuracy			0.97	1630
macro avg	0.97	0.97	0.97	1630
weighted avg	0.97	0.97	0.97	1630

Binary F1-score: 0.9801
Binary Accuracy: 0.9748

Confusion Matrix (Figure 3.5):

$$\begin{bmatrix} 581 & 26 \\ 15 & 1008 \end{bmatrix}$$

3.4.2 Multiclass Classification (Gesture Types)

Classification Report:

	precision	recall	f1-score	support
Above ear - pull hair	0.84	0.76	0.80	128
Cheek - pinch skin	0.66	0.70	0.68	127
Eyebrow - pull hair	0.56	0.55	0.56	128
Eyelash - pull hair	0.63	0.56	0.60	128
Forehead - pull hairline	0.77	0.72	0.74	128
Forehead - scratch	0.74	0.77	0.76	128

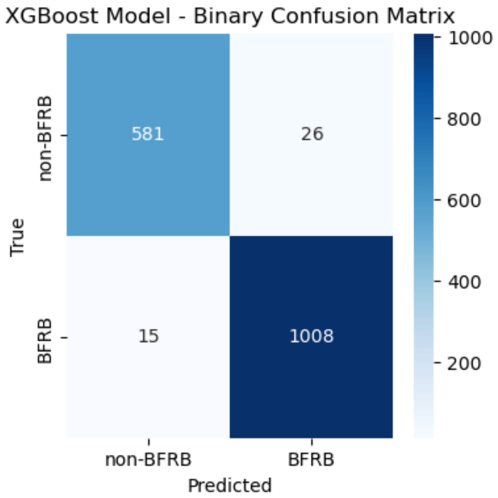

Figure 3.5: XGBoost Binary Confusion Matrix

```
Neck - pinch skin        0.54      0.50      0.52      128
Neck - scratch           0.67      0.62      0.64      128
non_target               0.95      0.99      0.97      607

accuracy                                      0.77     1630
macro avg              0.70      0.69      0.69     1630
weighted avg           0.77      0.77      0.77     1630

Macro F1-score: 0.6898
Multiclass Accuracy: 0.7736
```

Figure 3.6: XGBoost Multiclass Confusion Matrix

Final Composite Score:

F1_binary: 0.9801
F1_macro: 0.6898
Final Score: 0.8349

3.5 Model Selection and Comparison

Both Random Forest and XGBoost demonstrated strong binary classification performance. However, XGBoost outperformed Random Forest in multiclass gesture discrimination, yielding a higher composite score.

Summary Table:

Model	F1_binary	F1_macro	Final Score
Random Forest	0.9681	0.6407	0.8044
XGBoost	0.9801	0.6898	0.8349

In this study, we systematically evaluated state-of-the-art machine learning models for BFRB detection using sensor data. While both Random Forest and XGBoost achieved excellent binary detection, XGBoost provided superior multiclass performance and the highest composite score. This demonstrates its enhanced ability to capture subtle distinctions between gesture types. Therefore, XGBoost is recommended as the optimal model for deployment in BFRB detection systems.

Future work may focus on deep learning architectures, advanced feature engineering, and sensor fusion to further improve gesture subtype classification.

Conclusion

This project demonstrated the feasibility of detecting Body-Focused Repetitive Behaviors (BFRBs) using time-series data from the Helios wearable device. By leveraging classical machine learning models—Random Forest and XGBoost—we achieved strong classification accuracy in distinguishing BFRB-like gestures from everyday hand movements.

Both models excelled in binary detection, with F1-scores above 0.96. However, the challenge of differentiating among similar BFRB gesture subtypes was reflected in lower macro F1-scores. XGBoost notably outperformed Random Forest, achieving a composite score of 0.8349 compared to 0.8044, attributed to its superior ability to capture nuanced sensor patterns.

Based on these results, XGBoost is recommended as the optimal model for BFRB detection. Its robust binary accuracy ensures reliable identification of clinically relevant events, while its improved multiclass performance facilitates more accurate gesture differentiation. These strengths are especially valuable for real-world deployment and can directly inform future hardware and software development for wearable BFRB detection systems.

Overall, our results establish a strong baseline and demonstrate that data-driven, multimodal sensor approaches can meaningfully advance the state of automated BFRB behavior analysis. The ability to accurately classify BFRB-like gestures holds significant promise for improving behavioral health monitoring and outcomes.

Future Scope

Building on this foundation, several avenues exist for future research and development:

- Extend modeling to deep learning architectures (e.g., LSTM, CNN) for advanced sequence learning.
- Integrate contextual data (such as time of day and user activity) with sensor signals.
- Collaborate with clinical teams to maximize real-world behavioral health impact.

Bibliography

[1] Child Mind Institute. (2024). Helios: A Wearable Device for Behavioral Health Monitoring. Retrieved from `https://www.childmind.org/helios`

[2] Garey, J. (2025). What Is Excoriation, or Skin-Picking? Child Mind Institute. `https://childmind.org/article/excoriation-or-skin-picking/`

[3] Child Mind Institute. (2024). CMI: Detect Behavior with Sensor Data. Kaggle Competition. Retrieved from `https://www.kaggle.com/c/cmi-detect-behavior-with-sensor-data`

[4] Martinelli, K. (2025). What is Trichotillomania? Child Mind Institute. `https://childmind.org/article/what-is-trichotillomania/`