perplexity

. Dataset for Model Training

This dataset will be used to train your machine learning model to identify user body types from biometric data and images. It should include the following features for a large and diverse group of individuals:

Biometric Data: Age, weight, height, and activity level.
Body Type: A categorical label (e.g., endomorph, mesomorph, ectomorph) or a more granular classification that the model will learn to predict.
Body Image: A collection of images from different angles (front, side, back) and in various poses, as an optional input for the model.
This dataset will allow the machine learning model to learn the relationships between a user's biometrics and their body type. The model will analyze the data to identify the user's body type

are there such datasets available
find.

Yes, there are several datasets available for machine learning model training involving body types, biometrics, and images, though they vary in their inclusion of detailed features, images, and somatotype labels.

Datasets with Biometric Data and Images

- **Celeb-FBI Dataset** contains 7,211 full-body images of celebrities along with detailed biometric data: height, weight, age, and gender. Images are front-facing and labeled, but body type (endomorph, mesomorph, ectomorph) may not be directly labeled. Access requires contacting the authors for permission.[1]
- **D4FLY Multimodal Biometric Dataset** includes six biometric modalities (3D face, thermal face, iris, somatotype, activity level via smartphone sensor data) and somatotype categories (endomorph, mesomorph, ectomorph). It has images from various angles and scenarios. The dataset contains 31 subjects of diverse demographics. Access requires signing a license agreement.[2]
- **Body Size Dataset (Kaggle)** features body measurement datasets with photos and corresponding body metrics, but direct somatotype labeling is not confirmed.[3]
- **SOMASet** is a synthetic dataset of 100,000 images of 50 human prototypes (25 female, 25 male), mixing three somatotypical "seeds." This provides simulated somatotype data, potentially useful for initial model experimentation.[4]

Datasets Focused on Somatotype Classification

- **D4FLY Dataset** explicitly records somatotype (endomorph, mesomorph, ectomorph) with corresponding biometric and image data, making it particularly well-suited for your task.[2]
- Synthetic options such as **SOMASet** provide body shapes and classes in simulated images, helpful for initial prototyping or augmenting real datasets.[4]

Other Relevant Resources

- **Body Segmentation Datasets** on Kaggle offer large-scale labeled segmentation data (undergarments and full body), useful for training image-based models or for body shape estimation after additional processing.[5]
- Research such as "Image-Based Somatotype as a Biometric Trait for Non-Cooperative Person Identification" demonstrates the application of body type recognition using image data, and may cite datasets or provide methodological insights.[6]

Accessibility and Usage

- Most datasets require either institutional affiliation, a license agreement, or requesting permission from dataset creators.
- For direct somatotype (endomorph, mesomorph, ectomorph) labeling in real images, D4FLY is the closest, whereas Celeb-FBI provides biometric attributes with labeled images.

These datasets collectively cover biometric data, images, and, in some cases, body type classes suitable for advanced model training in body type prediction tasks.[5] [1] [3] [6] [2] [4]

⁂

which is best for my use of AI-Driven Workout Generation: The onboarding process captures user biometrics (age, weight, height, activity level) and an optional body image. A machine learning model analyzes this data to identify the user's body type and generate a tailored workout regimen.

The **D4FLY Multimodal Biometric Dataset** is the best suited for AI-driven workout generation using biometric onboarding and body type identification. This is because D4FLY contains comprehensive biometric data (age, weight, height, activity level via sensors), explicit body type (somatotype) categorization (endomorph, mesomorph, ectomorph), and high-quality body images from multiple angles and scenarios—all of which match your requirements precisely.[21]

Why D4FLY Dataset Fits Your Use Case

- **Explicit Body Type Labels**: D4FLY records the user's somatotype (endomorph, mesomorph, ectomorph), which aligns directly with body type-based workout tailoring.[21]
- **Rich Biometric Data**: Includes age, gender, weight, height, and smartphone-captured activity level metrics.[21]
- **Body Images**: Captures RGB images in varied scenarios (static, walking) from multiple camera angles, supporting image-based analysis.[21]
- **Granular Sensor Data**: Smartphone sensor logs add depth to user activity level measurements, potentially enhancing personalization.[21]
- **Diversity**: Features subjects from various age, gender, and ethnicity backgrounds, increasing model generalizability.[21]

Comparison with Celeb-FBI

Feature	D4FLY [21]	Celeb-FBI [22]
Somatotype	Directly labeled (endomorph, etc.)	Not labeled
Biometric Data	Age, weight, height, gender, activity level	Age, weight, height, gender
Images	Multi-angle, body, walking/static	Mostly front-facing, static
Activity Level	Sensor-based logs	Not included
Scale	31 subjects (diverse)	7,211 subjects (mostly celebrities)
Access	License agreement; research use	Request to authors; research use

If your machine learning model relies on both biometric features and body type labels for personalized workout plan generation, **D4FLY excels due to its explicit somatotype classification, sensor-based activity assessment, and comprehensive image data**. Celeb-FBI is ideal for biometric attribute estimation but does not provide body type classification, making it less optimal for your use case.[22] [21]

⁂

give the link to download the dataaset

The **D4FLY Multimodal Biometric Dataset** can be requested for download from the official project page. Researchers must sign a license agreement before receiving access.

Download Link and Access Process

- **D4FLY Dataset:**
  - Request the dataset and access licensing details at the official project page: https://d4fly.eu/dataset/[23]
  - After submitting a license agreement, the dataset will be provided for research use.

This is the authoritative source for the D4FLY biometric dataset required for AI-driven workout model development.[23]

⁂

Applications for the D4FLY Multimodal Dataset are temporarily disabled due to technical issues.

With D4FLY applications disabled, the most accessible alternatives for AI-driven workout applications involving body type, biometrics, and images are:

Best Public Alternatives

- **BodyM Dataset (AWS Open Data)**
  - Contains 8,978 frontal/lateral images of 2,505 real subjects, with paired height, weight, and 14 body measurements.
  - Includes silhouette images for each individual, suitable for both biometric and image-based modeling.
  - Creative Commons Attribution-Non Commercial 4.0 license.
  - Download from: https://registry.opendata.aws/bodym/ [24]
- **SOMASet (Kaggle)**
  - Synthetic dataset: 100,000 images of 50 human prototypes (25 female, 25 male) created by mixing three somatotypical “seeds” (endomorph, mesomorph, ectomorph).
  - Useful for prototyping or data augmentation, less ideal for real population modeling.
  - Download from: https://www.kaggle.com/datasets/vicolab/somaset [25]
- **Body Size Dataset (Kaggle)**
  - Includes photos with corresponding body measurements for a diverse set of people.
  - Download from: https://www.kaggle.com/datasets/tapakah68/body-measurements-dataset [26]

Usage Notes

- BodyM (real silhouettes and biometrics) is best for developing models with real images and measurements.
- SOMASet is suitable for somatotype-based proof-of-concept or when real data is not mandatory.
- These datasets do not always explicitly label traditional somatotypes but provide biometric and image data sufficient for AI training.

These links provide immediate access to the best available public datasets for your project. [24] [25] [26]

⁂

BodyM (real silhouettes and biometrics) is best for developing models with real images and measurements. give link to download

The BodyM dataset—containing real silhouette images paired with biometric measurements—is available for download via the AWS Registry of Open Data.

Download link:
https://registry.opendata.aws/bodym/ [42]

This source provides official access and usage instructions for the BodyM dataset, suitable for model development using real images and biometric data. [42]

⁂

no download link

The BodyM dataset does not use a conventional website download button. Instead, it is distributed using Amazon S3, and can be accessed directly via AWS CLI (no AWS account needed):

Direct Access Method

- **AWS CLI Command (No account required):**

  ```
  aws s3 ls s3://amazon-bodym/ --no-sign-request
  ```

  This command lists files in the bucket. To download individual files or folders, use the CLI `cp` command:

  ```
  aws s3 cp s3://amazon-bodym/<file_or_folder> <local_path> --no-sign-request
  ```

License and Documentation

- **License:** Creative Commons Attribution-Non Commercial 4.0 International
- **Documentation:** https://adversarialbodysim.github.io/ [43]

This process allows for programmatic access and bulk downloading. Desktop download links are not provided; all access is managed through AWS S3 commands as described above. [43]

⁂

will it save in my pc locally?

Yes, using the AWS CLI commands, you can **download and save the BodyM dataset files locally on your PC**.

How it works:

- When you run the command

  ```
  aws s3 cp s3://amazon-bodym/<file_or_folder> <local_path> --no-sign-request
  ```

  the specified file or folder from the AWS S3 bucket will be copied and saved into the folder (`<local_path>`) on your computer.
- For example, to download a file named `data.zip` from the bucket to your current local directory, you would run:

  ```
  aws s3 cp s3://amazon-bodym/data.zip ./ --no-sign-request
  ```

  This saves `data.zip` locally.

Summary

- The dataset files are stored remotely on Amazon's S3 servers initially.
- You use the AWS CLI to copy these files down to your PC.
- After copying, the files reside on your local storage and are accessible like any other file.

So, the answer is **yes**, the data can be saved locally after download using the right AWS CLI commands. [45]

⁂

i ask you again , which dataset makes more sense for my use case?

For your use case—AI-driven workout generation based on user biometrics (age, weight, height, activity level) plus optional body images, with the goal to identify body type (endomorph, mesomorph, ectomorph) for personalized training—the **D4FLY Multimodal Biometric Dataset** still makes the most sense conceptually.

Why D4FLY is Ideal

- **Explicit somatotype labels** (endomorph, mesomorph, ectomorph) directly match your need for body type classification.
- **Comprehensive biometric features** including age, height, weight, gender, and smartphone activity sensor data align well with onboarding biometrics.
- **Multi-angle body images** enable the model to analyze visual features alongside numeric biometrics.
- **Designed for multimodal biometrics and somatotype classification**, supporting robust AI models for personalized health or fitness applications.

Caveat

- Availability is currently limited due to technical issues in application access, slowing immediate use.

Alternative: BodyM Dataset

- While showing real body images and many biometric measurements, BodyM **does not explicitly label somatotype categories** needed for workout tailoring by body type.
- This demands additional annotation/classification work and may limit out-of-the-box use for your application.

Summary Table

Dataset	Direct Somatotype Labels	Biometric Data	Images (Body Angles)	Activity Level Data	Access	Best for
D4FLY	Yes	Age, weight, height, activity via sensors	Multi-angle body images	Yes	Restricted	Body type identification & AI-driven personalized workouts
BodyM	No	Height, weight, 14 body measurements, gender	Frontal/lateral silhouettes	No	Open	General biometrics and silhouette modeling, no direct body type

Conclusion

If access to D4FLY becomes available, it is the **best fit** dataset to train your model effectively in predicting body types for workout generation. If immediate access is critical, BodyM can be useful but will require extra labeling or model adaptation to infer somatotypes.[46] [47]

⁂

are there any other strong alternatives to the D$FLY which has all the data

Besides the D4FLY dataset, currently, there are **no other comprehensive public datasets combining all the following for your use case**:

- Biometric data (age, weight, height, activity level)
- Explicit somatotype or body type labels (endomorph, mesomorph, ectomorph)
- Multi-angle body images for visual analysis

Closest Alternatives Found

Dataset	Somatotype Labels	Biometric Data	Images	Activity Level	Notes	Link
Celeb-FBI	No	Age, height, weight, gender	Full body front-facing images	No	Large dataset (7,211+ subjects) with biometrics and images but no somatotype labels. Requires request for access.	Celeb-FBI
Body Size Dataset	No	Various body measurements	Images	No	Photos with body measurements but no somatotype labels or activity data.	Kaggle Body Size Dataset
SOMASet	Synthetic	None or simulated	Synthetic images	No	Synthetic data with somatotype variations, useful for experiments but not real data.	SOMASet
BodyM	No	Height, weight, gender, 14 body measurements	Frontal and lateral silhouettes	No	Real biometric and silhouette data without somatotype labeling or activity data.	BodyM

Summary

- **D4FLY is unique** in offering real somatotype labels, biometric data, multi-angle images, and activity sensor data in one dataset for somatotype-based AI applications.
- Other datasets lack explicit **somatotype labels** or activity level data, meaning additional annotation or sensor data collection would be necessary.
- Synthetic datasets like SOMASet might supplement but cannot fully replace real data.

Recommendation

If D4FLY remains inaccessible, combining multiple datasets or creating a labeled internal dataset with biometric, image, and somatotype information might be necessary for accurate AI-driven workout generation.

No other single public dataset currently offers all data types as comprehensively as D4FLY for your specific use case.[48] [49] [50] [51] [52]

⁂

so which is the next close one

The next closest dataset to D4FLY for your use case is the **Body Size Dataset on Kaggle**.

Why Body Size Dataset?

- It contains **photos of individuals** paired with detailed body measurements (biometric data like height, weight, etc.).
- Though it lacks explicit somatotype labels (endomorph, mesomorph, ectomorph), you could potentially **derive or label body types based on the measurements**.
- It includes **real images**, supporting image-based feature extraction for body type classification.
- It is publicly accessible and easier to acquire compared to D4FLY.

Important Note

- You may need to add custom annotation or create somatotype labels using biometric thresholds or expert input.
- Activity level data is absent, so consider capturing separate activity data or integrating sensor inputs.

Summary

Dataset	Somatotype Labels	Biometric Data	Images	Activity Level	Access
Body Size Dataset	No (can be derived)	Body measurements, height, weight	Real photos	No	Public

Link: Body Size Dataset [67]

This dataset is the best available alternative to D4FLY for developing a model using combined biometric and image data tailored to body type prediction. [67]

⁂

what about the bodyM dataset?

The BodyM dataset is a strong option but has some limitations compared to D4FLY for your use case.

Strengths of BodyM Dataset

- Contains **8,978 frontal and lateral silhouette images** of 2,505 real subjects.
- Provides **height, weight, gender, and 14 detailed body measurements** such as ankle girth, arm length, calf girth, chest girth, thigh girth, waist girth, and more.
- Images show full body in frontal and side poses, suitable for visual analysis or silhouette-based modeling.
- Publicly accessible via AWS Open Data with a Creative Commons license.
- Useful for biometric and image-based modeling of body shapes.

Limitations for Your Use Case

- **No explicit somatotype or body type labels** (endomorph, mesomorph, ectomorph), so body type classification needs to be inferred or annotated separately.
- **No activity level data**, so additional sensor or user input data will be required.
- Silhouettes are black-and-white, not full RGB images, which may limit image-based deep learning approaches that benefit from color and texture.

Summary Table vs D4FLY

Feature	BodyM	D4FLY
Somatotype Labels	No	Yes
Biometric Data	Height, weight, 14 measurements	Age, weight, height, activity
Images	Silhouettes, frontal & lateral	Multi-angle body RGB images
Activity Level	No	Yes (smartphone sensor data)
Access	Public, AWS S3 bucket	Restricted (license, application)

Conclusion

BodyM is excellent for biometric and silhouette image modeling but requires **extra labeling and data augmentation** to fully support body type classification and activity-based workout personalization. It is a practical alternative when D4FLY is not available but involves more preprocessing effort.[68] [69]

⁂

any other strong alternative

A strong alternative dataset relevant to your use case is described in this research work:

Image-Based Somatotype Biometric Dataset

- The dataset investigates full-body image-based somatotype as a biometric trait for person recognition at a distance and on-the-move.
- It uses publicly available datasets by merging them into a unified large dataset containing many identities with somatotype features visible in full-body images.
- The somatotype trait (endomorph, mesomorph, ectomorph) can be captured easily through standard 2D cameras without special equipment.
- Deep learning models (Inception V3, Siamese networks) were used successfully for identification and verification tasks based on somatotype.
- The dataset is suitable for biometric fusion schemes, providing a valuable soft biometric feature.

Key Reference and Access

- Article: "Image-Based Somatotype as a Biometric Trait for Non-Collaborative Person Recognition at a Distance and On-The-Move" (2020)
  PMC Article with Dataset info

Relevance to Your Use Case

- Includes somatotype classification derived from publicly available full-body image datasets.
- Enables development of models using images and somatotype labels.
- While it may require assembling or access requests, this dataset aligns well with multimodal biometric, body type classification, and image data needs for AI-driven workout personalization.

This is currently an important academic dataset focused on image-based somatotype recognition and presents a strong alternative to D4FLY when direct biometrics combined with somatotype labels and images are needed.[70] [71]

⁂

i have decided to modify my prject in such a way that only requires user to upload their pic , the model model will compute the metrics and i will write a code with if else to recommend workouts

That is a practical and streamlined approach.

Benefits of Image-Only Body Type Detection

- **User Convenience:** No need to input detailed biometric data manually—just upload a photo.
- **Simplified Data Collection:** Focuses on computer vision models trained to extract body measurements and body type features automatically from images.
- **Scalable:** Easier to deploy on mobile/web apps as image upload is straightforward.
- **Automation:** The model computes necessary metrics (e.g., body shape, estimated girths) from the image, then your rule-based system makes workout recommendations.

Recommended Workflow

1. **Train or use a pre-trained body segmentation and measurement model** on datasets like BodyM or image-based somatotype datasets to extract relevant body metrics from a single user photo.
2. **Classify body type** (endomorph, mesomorph, ectomorph) automatically based on these measurements or visual features.
3. **Apply conditional logic (if-else rules)** to recommend personalized workouts corresponding to each body type.
4. Optionally, use data augmentation or pose estimation to improve measurement accuracy from varied photo angles.

Suitable Datasets for Model Training

- BodyM (silhouettes + measurements) for metric extraction from images.
- Image-based somatotype datasets (from academic papers) for training classification from images.

This approach leverages computer vision to eliminate manual biometric input, simplifying onboarding while still enabling personalized recommendations.Since the project now involves only image input, the best course is to train or fine-tune a computer vision model that extracts body metrics and classifies body types from user photos automatically. Then your custom if-else logic can recommend workouts based on the inferred body type.[89] [90]

This image-only streamline:

- Improves user convenience (no manual data entry).
- Uses datasets like BodyM (silhouette images with biometric measures) or academic image-based somatotype datasets.
- Leverages body segmentation, pose estimation, and measurement extraction models.
- Enables personalized workout recommendations purely from images.

This approach remains well aligned with your goals and relies on widely accessible visual datasets and techniques for automated biometric analysis and body type classification.[90] [89]

⁂

datasets like BodyM (silhouette images with biometric measures) is it perfect for my project now?

Yes, the BodyM dataset is **very well suited** for your updated project where users upload photos, and the model computes body metrics to classify body type for workout recommendations.

Why BodyM Fits Your Image-Only Project

- It contains **frontal and lateral silhouette images** of real subjects, which are ideal inputs for computer vision models focusing on shape and measurement extraction.
- Paired with **detailed biometric measurements** (e.g., girths, heights), BodyM enables supervised learning to predict these metrics directly from images.
- Although it lacks explicit somatotype labels, you can **derive body type classifications** using the biometric data or silhouette features.
- The dataset supports building automated pipelines for **body parameter estimation from images**, aligning perfectly with your goal of inferring body type from just user photos.
- It is publicly available and comes with a clear usage license, making it practical for development and deployment.

Summary

Feature	BodyM Dataset	Suitability for Image-Only Project
Image Type	Frontal & lateral silhouettes	Perfect for image-based body metric extraction
Biometric Data	Extensive body measurements	Enables training models to predict metrics from images
Somatotype Labels	No	Can derive based on biometric data or silhouettes
Activity Data	No	Not needed in image-only model
Availability	Public, AWS S3 bucket	Ready for use

Thus, BodyM provides a robust foundation for training AI models to process uploaded photos, estimate body metrics, and support your if-else workout recommendation code.Yes, the BodyM dataset is well suited for your updated image-only project. It provides frontal and lateral silhouette images paired with detailed biometric measurements, enabling you to train models that extract body metrics directly from photos. While it lacks explicit somatotype labels, you can derive body type classification from the biometric data or silhouette features. This makes BodyM a strong, practical base for your AI to compute metrics from user images and support your rule-based workout recommendations effectively.[91] [92]

⁂

help me download it

To download the BodyM dataset, follow these steps to use the AWS CLI, as the dataset is hosted on an Amazon S3 bucket:

Step 1: Install AWS CLI

- If not already installed, download and install AWS CLI from:
  https://aws.amazon.com/cli/

Step 2: List Files in the BodyM Bucket

Open your terminal or command prompt and run the command below to list the files available in the BodyM S3 bucket (no AWS account or credentials needed):

```
aws s3 ls s3://amazon-bodym/ --no-sign-request
```

Step 3: Download Files

To download a file or entire folder from the bucket to your local machine, use this command:

```
aws s3 cp s3://amazon-bodym/<file_or_folder> <local_folder_path> --recursive --no-sign-request
```

Replace <file_or_folder> with the specific file or folder name listed from Step 2, and <local_folder_path> with the path on your PC where you want to save the data.

Example:

```
aws s3 cp s3://amazon-bodym/ ./bodym_data --recursive --no-sign-request
```

This will download the entire dataset into a local folder named `bodym_data`.

Additional Resources

- BodyM dataset info and documentation:
  https://registry.opendata.aws/bodym/
- Dataset license: Creative Commons Attribution-Non Commercial 4.0 International (CC BY-NC 4.0)

This method allows you to securely and fully download the BodyM dataset for your project.[93]

⁂

C:\Users\bened>aws s3 ls s3://amazon-bodym/ --no-sign-request

```
                           PRE testA/
                           PRE testB/
                           PRE train/
```

The output indicates there are three main folders in the BodyM dataset bucket:

- `testA/`
- `testB/`
- `train/`

To download the entire BodyM dataset locally, run this command:

```
aws s3 cp s3://amazon-bodym/ C:\Users\bened\bodym_data --recursive --no-sign-request
```

This will recursively copy all files from the bucket into a local folder named `bodym_data` on your C: drive.

To download a specific folder only, for example `train/`, use:

```
aws s3 cp s3://amazon-bodym/train/ C:\Users\bened\bodym_data\train --recursive --no-sign-request
```

Summary

- Use `--recursive` to download folders and subfolders.
- Specify the local path after the bucket path to save the data.
- No AWS credentials or account is required with `--no-sign-request`.

Let me know if help is needed with specific commands or next steps!

C:\Users\bened>aws s3 cp s3://amazon-bodym/ C:\Users\bened\Documents\Alyster Coding\PROJECTS\AI_GYM\bodym_data --recursive --no-sign-request

Unknown options: Coding\PROJECTS\AI_GYM\bodym_data

The error occurs because the file path contains spaces, and the command line interprets them as separate arguments.

Solution: Enclose the local path in double quotes

Try this command instead:

```
aws s3 cp s3://amazon-bodym/ "C:\Users\bened\Documents\Alyster Coding\PROJECTS\AI_GYM\bodym_data" --recursiv
```

This tells the CLI to treat the entire path with spaces as a single argument.

Let me know if that works or if further help is needed!

whats the size of the dataset?

The BodyM dataset size is not explicitly stated as a total file size on the AWS page, but based on the number of images and measurements it includes:

- **8,978 frontal and lateral silhouette images** for 2,505 subjects.
- Accompanied with height, weight, gender, and 14 detailed body measurements.

Considering the file types are mainly black-and-white silhouette images (which have small file sizes compared to RGB photos) and accompanying textual measurement data, the dataset is estimated to be **in the range of several gigabytes (GBs), likely under 10 GB**.

The lack of detailed size info means you should ensure sufficient storage (10-15 GB recommended) and a stable internet connection for download.

For exact sizes of files in the bucket after listing them with AWS CLI, you can run:

```
aws s3 ls s3://amazon-bodym/ --recursive --no-sign-request
```

This command will display individual file sizes, so you can sum them or check the largest components before downloading.[94]

⁂

BodyM Dataset Details

- **Contents:**
  - 8,978 black-and-white silhouette images (frontal and lateral views) of 2,505 real subjects.
  - Detailed biometric data including height, weight, gender, and 14 precise body measurements such as ankle girth, arm length, bicep girth, calf girth, chest girth, forearm girth, hip girth, leg length, leg girth, shoulder breadth, thigh girth, waist girth, wrist girth, and shoulder-to-crotch length.
- **Data Splits:** Training set and two test sets with diverse environmental conditions to test robustness.
- **License:** Creative Commons Attribution-Non Commercial 4.0 International.
- **Access:** Publicly available on AWS S3 without credentials using AWS CLI.
- **Use Case:** Designed for body shape modeling, biometric research, and automated metric extraction from silhouette images.

Why BodyM is Used for Your Project

- **Image-Based Metric Extraction:** Your project involves users uploading photos to automatically compute body metrics. BodyM contains silhouette images paired with extensive ground-truth biometric data, enabling training of computer vision models that estimate body measurements directly from images.
- **Supports Body Type Classification:** While it lacks explicit somatotype labels, the comprehensive biometric measurements can be used to derive or classify body types (endomorph, mesomorph, ectomorph) through data-driven methods or rules, supporting your if-else workout recommendation logic.

- **Simplifies User Input:** The dataset aligns with your updated project goal of only requiring a user image as input, as models trained on BodyM can extract key body features without manual biometric entry.
- **Accessibility and Practicality:** Public availability with a clear license and no access restrictions facilitates swift development and deployment of your AI system.
- **Robustness:** Multiple viewing angles and varying test conditions in BodyM ensure model generalization to real-world images uploaded by users in diverse environments.

Summary

Aspect	Explanation
Images	Frontal and lateral silhouette images for metric learning
Biometric Data	Extensive body measurements to supervise learning
User Input Alignment	Supports image-only input pipeline
Body Type Classification	Enables derivation of body types from numeric metrics
Accessibility	Open dataset, easy to download and use
Practical for AI Model	Enables automatic metric extraction from simple images

BodyM provides the ideal balance of image data and biometric ground truth needed to build automated, image-only body type classifiers that power personalized workout recommendations in your project.[104] [105]

⁂

give the directory structure of the dataset including my path

Based on the BodyM dataset structure visible from AWS S3 listing and your local download path, here is the expected directory structure including your local path:

```
C:\Users\bened\Documents\Alyster Coding\PROJECTS\AI_GYM\bodym_data\
│
├── train\
│   ├── [image files: silhouette images of training subjects]
│   ├── [measurement files or metadata for training subjects]
│
├── testA\
│   ├── [image files: silhouette images of test set A subjects]
│   ├── [measurement files or metadata for test set A subjects]
│
├── testB\
│   ├── [image files: silhouette images of test set B subjects]
│   ├── [measurement files or metadata for test set B subjects]
│
├── README or documentation files (if available)
├── metadata.csv or similar file containing biometric data (height, weight, girths, gender)
```

Explanation

- The dataset is divided into **three main partitions**: `train`, `testA`, and `testB`, reflecting typical training and testing splits.
- Each partition contains silhouettes (images) and corresponding measurement data files.
- Your AWS CLI command downloads these folders into the specified local path, preserving the same folder structure.

This structure helps organize labeled biometric and image data clearly for efficient training and evaluation of your computer vision model.Based on the BodyM dataset AWS S3 bucket and your given local path, the expected directory structure after downloading is:[106]

```
C:\Users\bened\Documents\Alyster Coding\PROJECTS\AI_GYM\bodym_data\
│
├── train\
│   ├── [silhouette images and measurement data for training subjects]
│
├── testA\
│   ├── [silhouette images and measurement data for test set A]
│
├── testB\
│   ├── [silhouette images and measurement data for test set B]
│
├── README or metadata files (if any)
```

- The dataset consists of three main folders: `train`, `testA`, and `testB`, each holding images and biometric measurements.
- The local path after `bodym_data` matches the bucket folder structure exactly as downloaded using the AWS CLI with `--recursive`.
- Inside these folders are the silhouette images (frontal and lateral) and corresponding body measurement data to be used for model training and evaluation.

This structure facilitates organized access to distinct training and test splits for your AI body metric extraction and classification project.[106]

⁂

this is the actual directory

modify ur response

Based on your attached screenshots, here is the **actual directory structure** of your downloaded BodyM dataset:

```
C:\Users\bened\Documents\Alyster Coding\PROJECTS\AI_GYM\bodym_data\
│
├── testA\
│
├── testB\
│
├── train\
│   ├── mask\
│   ├── mask_left\
│   ├── hwg_metadata.xlsx
│   ├── measurements.xlsx
│   ├── subject_to_photo_map.xlsx
```

Folder & File Descriptions

- **testA, testB, train** — The three main partitions for model training and evaluation.
- **mask, mask_left** — Contain segmentation or silhouette images used in metric extraction.
- **hwg_metadata.xlsx** — Contains subject metadata such as height, weight, gender.
- **measurements.xlsx** — Contains detailed biometric measurements for subjects.
- **subject_to_photo_map.xlsx** — Provides mapping between subject IDs and photo filenames for data alignment.

How This Structure Supports Your Project

- The **'train' folder** is organized for supervised model training, combining images and corresponding biometric measurements.
- **Excel files** enable easy programmatic access to ground-truth biometrics for ML regression/classification.
- **Mask folders** store the input images (silhouettes) for computer vision-based metric prediction.
- The **mapping file** ensures accurate annotation of images with biometric data.

This layout is ideal for a pipeline where you feed user-uploaded images, have your model infer metrics using the same structure, and then make logic-based workout recommendations.[107]

⁂

├── mask\

│ ├── mask_left
│ ├── hwg_metadata.xlsx
│ ├── measurements.xlsx
│ ├── subject_to_photo_map.xlsx

is in train test A and Test B

Thank you for the clarification. Based on your update, the **corrected directory structure with these folders and files inside each partition** (`train`, `testA`, `testB`) is:

```
C:\Users\bened\Documents\Alyster Coding\PROJECTS\AI_GYM\bodym_data\
│
├── train\
│   ├── mask\
│   ├── mask_left\
│   ├── hwg_metadata.xlsx
│   ├── measurements.xlsx
│   ├── subject_to_photo_map.xlsx
```

```
│
├── testA\
│   ├── mask\
│   ├── mask_left\
│   ├── hwg_metadata.xlsx
│   ├── measurements.xlsx
│   ├── subject_to_photo_map.xlsx
│
├── testB\
│   ├── mask\
│   ├── mask_left\
│   ├── hwg_metadata.xlsx
│   ├── measurements.xlsx
│   ├── subject_to_photo_map.xlsx
```

Explanation

- Each of the three data splits (`train`, `testA`, `testB`) contains:
  - Two folders with silhouette images: `mask` (likely frontal or full body) and `mask_left` (possibly side or left view).
  - Three Excel files holding metadata, measurements, and mapping between subject IDs and photos relevant to that split.

This consistent structure ensures isolation of training and testing data along with their respective images and biometric measurements for proper model development and evaluation.

It is organized to support automated loading of images and corresponding measurements within each data partition. [111]

⁂

those are csv not xlsx

Thank you for the correction. Here is the updated directory structure reflecting that the metadata and measurement files are **CSV files**, not Excel XLSX files:

```
C:\Users\bened\Documents\Alyster Coding\PROJECTS\AI_GYM\bodym_data\
│
├── train\
│   ├── mask\
│   ├── mask_left\
│   ├── hwg_metadata.csv
│   ├── measurements.csv
│   ├── subject_to_photo_map.csv
│
├── testA\
│   ├── mask\
│   ├── mask_left\
│   ├── hwg_metadata.csv
│   ├── measurements.csv
│   ├── subject_to_photo_map.csv
│
├── testB\
│   ├── mask\
│   ├── mask_left\
│   ├── hwg_metadata.csv
│   ├── measurements.csv
│   ├── subject_to_photo_map.csv
```

Notes

- The `.csv` files provide tabular data for easy parsing in Python or any data processing tool.
- These contain the biometric metadata, body measurements, and mapping between photos and subjects within each data split.
- The structure supports straightforward data ingestion pipelines for your AI model training and evaluation.

This setup is optimal for programmatically linking silhouette images to biometric data in a clean, consistent format throughout all dataset partitions. [112]

⁂

extract 5 samples from each csv with table and names of columns

Here are 5 sample rows extracted from each of the three CSV files provided, along with the column names.

1. `hwg_metadata.csv`

subject_id	gender	height_cm	weight_kg
5a1e03cb9f17b800040cc77c	male	172.5	104.1
6JfuBS-duD_BLzV6gxsQjcPc7gv6mgQBTuNzdnp6gfA	female	159.0	87.2
6Xz1aGZP_g084xrBEy0UYJjzozZhj3G1wx_cFS9q-UM	female	163.0	81.45
hjG7-UJ-SuVAgiJ3WEgKdS2nc7fAwQ7iXJCP5oS9omg	female	169.0	84.2
Mrq4MhWptluvgMB-LCjGQl950hLNxVxQhBbzo--Il5E	male	180.0	94.6

Columns:

- subject_id: Unique identifier of subject
- gender: Gender of the subject
- height_cm: Height in centimeters
- weight_kg: Weight in kilograms

2. `measurements.csv`

subject_id	ankle	arm-length	bicep	calf	chest	forearm	height	hip
-JtWIHEvrLPWagPaAhm0eQ_UKMSg8By1yoe-SkdrKSs	22.73997	46.19904	26.39884	35.11439	97.74417	23.39306	161.3835	98.0668
1T9SoRhbCpQ0zXf4Le6WjwTOPdfnmEDebvWaECZg_HM	25.47771	48.23134	31.91501	38.68342	107.2672	28.92529	172.5558	103.8481
2LAYE-qYtPZMSeH-gUgP-H4o-WRPzC0Cg9Nf5_AkLBE	25.25466	50.60219	29.62661	37.51582	100.6735	27.38669	178.7748	99.3423
3KCeTD5hLgY5qDa-Dhg0IcrnsRM_EndCLV4DMSOQfQM	23.99533	51.21338	25.2306	35.23132	92.43688	24.25768	175.3322	95.4055
3beF_SfHubarfkfoyCeA6EDR7UqE8j-bgj4bzBFzGFM	24.03846	53.71963	28.88954	35.93877	107.579	26.35227	179.7179	97.62849

Columns:

- subject_id: Unique identifier of subject
- ankle, arm-length, bicep, calf, chest, forearm, height, hip, leg-length, shoulder-breadth, shoulder-to-crotch, thigh, waist, wrist: Body measurements in relevant units (cm)

3. `subject_to_photo_map.csv`

subject_id	photo_id
-JtWIHEvrLPWagPaAhm0eQ_UKMSg8By1yoe-SkdrKSs	1982951758ea65ae715d9d1a95a8ecd1
-JtWIHEvrLPWagPaAhm0eQ_UKMSg8By1yoe-SkdrKSs	dc567997daef596835cce05a0c6f770f
-JtWIHEvrLPWagPaAhm0eQ_UKMSg8By1yoe-SkdrKSs	f72236f1ab2ddb4a7b2863019c1c943c
-JtWIHEvrLPWagPaAhm0eQ_UKMSg8By1yoe-SkdrKSs	f6aff3e542d4101fe9ac7cfa6ee4439a
-JtWIHEvrLPWagPaAhm0eQ_UKMSg8By1yoe-SkdrKSs	00e11c860f7b081a01e9c7f04405c40f

Columns:

- subject_id: Unique identifier of subject
- photo_id: Unique photo/image identifier matching those in image folders

If more samples or any other details are needed, please ask!

conda install pytorch torchvision torchaudio pytorch-cuda=12.1 -c pytorch -c nvidia

conda create --name metrics python=3.10 -y

these (versions)are compatible with each other ? check

Your commands:

```
conda create --name metrics python=3.10 -y
conda install pytorch torchvision torchaudio pytorch-cuda=12.1 -c pytorch -c nvidia
```

Compatibility Check

- Python 3.10 is **compatible** with recent PyTorch versions.
- PyTorch versions supporting CUDA 12.1 and compatible with Python 3.10 are available.
- The official PyTorch instructions for CUDA 12.1 recommend installing with:

```
conda install pytorch torchvision torchaudio pytorch-cuda=12.1 -c pytorch -c nvidia
```

- This installation correctly pairs PyTorch, torchvision, and torchaudio with CUDA 12.1 support for Python 3.10 environments.

Summary

Package	Compatibility
Python 3.10	Supported by PyTorch
PyTorch + torchvision + torchaudio with CUDA 12.1	Supported & officially available via conda `pytorch` and `nvidia` channels

Recommendation

Your commands are **compatible and recommended** for installing PyTorch with CUDA 12.1 in a Python 3.10 Conda environment.

If you want to verify after installation:

```
import torch
print(torch.__version__)
print(torch.cuda.is_available())
print(torch.version.cuda)
```

This will confirm the correct CUDA-enabled PyTorch version is installed and working.[116] [117] [118]

⁂

1. https://arxiv.org/html/2407.03486v1
2. https://d4fly.eu/dataset/
3. https://www.kaggle.com/datasets/tapakah68/body-measurements-dataset
4. https://www.kaggle.com/datasets/vicolab/somaset
5. https://www.kaggle.com/datasets/trainingdatapro/human-segmentation-dataset
6. https://pmc.ncbi.nlm.nih.gov/articles/PMC7348910/
7. https://bidalab.eps.uam.es/listdatabases?id=BioGIGA
8. https://pmc.ncbi.nlm.nih.gov/articles/PMC11860359/
9. https://arxiv.org/html/2501.14070v1
10. https://www.sciencedirect.com/science/article/pii/S2352340924007996
11. https://repository.fit.edu/cgi/viewcontent.cgi?article=1833&context=etd
12. https://study.com/academy/lesson/body-types-mesomorph-ectomorph-endomorph.html
13. https://openaccess.thecvf.com/content/WACV2024/papers/Liu_FarSight_A_Physics-Driven_Whole-Body_Biometric_System_at_Large_Distance_and_WACV_2024_paper.pdf
14. https://www.vedantu.com/biology/somatotype
15. https://www.sciencedirect.com/science/article/pii/S1319157820305577
16. https://www.medicinenet.com/body_type_ectomorph_endomorph_or_mesomorph/article.htm
17. https://pmc.ncbi.nlm.nih.gov/articles/PMC4121156/
18. https://cvrl.nd.edu/projects/data/
19. https://www.kaggle.com/datasets/ninadmehendale/multimodal-iris-fingerprint-biometric-data

20. https://ebooks.inflibnet.ac.in/antp06/chapter/sheldon-and-parnell-classification-of-somatotype/
21. https://d4fly.eu/dataset/
22. https://arxiv.org/html/2407.03486v1
23. https://d4fly.eu/dataset/
24. https://registry.opendata.aws/bodym/
25. https://www.kaggle.com/datasets/vicolab/somaset
26. https://www.kaggle.com/datasets/tapakah68/body-measurements-dataset
27. https://arxiv.org/html/2407.03486v1
28. https://www.kaggle.com/datasets/trainingdatapro/human-segmentation-dataset
29. https://blog.nasm.org/body-types-mesomorph-ectomorphs-endomorphs-explained
30. https://www.shutterstock.com/search/endomorph-mesomorph-ectomorph
31. https://www.nature.com/articles/s41597-024-04189-w
32. https://www.nasm.org/resource-center/blog/body-types-how-to-train-diet-for-your-body-type
33. https://pmc.ncbi.nlm.nih.gov/articles/PMC7348910/
34. https://bidalab.eps.uam.es/listdatabases?id=BioGIGA
35. https://trainingdata.pro/datasets/body-measurements
36. https://d4fly.eu/dataset/
37. https://centaur.reading.ac.uk/101891/1/AVSS2021_Chen_etal_cameraready.pdf
38. https://d4fly.eu/data-capture-event-in-uk/
39. https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5db9c91b8&appId=PPGMS
40. https://centaur.reading.ac.uk/101891/
41. https://arxiv.org/html/2312.03227v1
42. https://registry.opendata.aws/bodym/
43. https://registry.opendata.aws/bodym/
44. image.jpg
45. https://registry.opendata.aws/bodym/
46. https://d4fly.eu/dataset/
47. https://registry.opendata.aws/bodym/
48. https://www.kaggle.com/datasets/tapakah68/body-measurements-dataset
49. https://arxiv.org/html/2407.03486v1
50. https://d4fly.eu/dataset/
51. https://www.kaggle.com/datasets/vicolab/somaset
52. https://registry.opendata.aws/bodym/
53. https://stock.adobe.com/search?k=ectomorph
54. https://data.ncl.ac.uk/articles/dataset/Human_Body_Shape_Classification_Dataset/19307300
55. http://www.ndl.iitkgp.ac.in/re_document/ieee_xplore/1234567_ieeeconf_4/206613
56. https://eprints.soton.ac.uk/493707/1/Facial_Profile_Recognition_Using_Comparative_Soft_Biometrics.pdf
57. https://www.shutterstock.com/search/body-type-ectomorph-endomorph-mesomorph?image_type=illustration&page=4
58. https://scpe.org/index.php/scpe/article/view/2524/954
59. https://www.medicinenet.com/body_type_ectomorph_endomorph_or_mesomorph/article.htm
60. https://arxiv.org/html/2503.22912v1
61. https://www.istockphoto.com/photos/endomorph
62. https://www.gettyimages.com/photos/endomorph
63. https://pmc.ncbi.nlm.nih.gov/articles/PMC7348910/
64. https://arxiv.org/html/2501.14070v1
65. https://www.sciencedirect.com/science/article/pii/S2665917422002276
66. https://pubmed.ncbi.nlm.nih.gov/32560464/
67. https://www.kaggle.com/datasets/tapakah68/body-measurements-dataset
68. https://registry.opendata.aws/bodym/
69. https://d4fly.eu/dataset/
70. https://pmc.ncbi.nlm.nih.gov/articles/PMC7348910/
71. https://pubmed.ncbi.nlm.nih.gov/32560464/
72. https://trainingdata.pro/datasets/body-measurements
73. https://cvrl.nd.edu/projects/data/
74. https://wjarr.com/sites/default/files/WJARR-2024-2741.pdf
75. https://pmc.ncbi.nlm.nih.gov/articles/PMC4121156/
76. https://www.sciencedirect.com/topics/computer-science/multimodal-biometric

77. https://www.nature.com/articles/s41597-024-04189-w
78. https://www.ijfmr.com/papers/2024/3/20409.pdf
79. https://arxiv.org/html/2501.14070v1
80. https://bidalab.eps.uam.es/listdatabases?id=BioGIGA
81. https://www.sciencedirect.com/science/article/pii/S1319157820305577
82. https://www.kaggle.com/datasets/trainingdatapro/human-segmentation-dataset
83. https://d4fly.eu/dataset/
84. https://arxiv.org/html/2407.03486v1
85. https://pmc.ncbi.nlm.nih.gov/articles/PMC11178413/
86. https://gts.ai/dataset-download/body-segmentation-dataset/
87. https://www.kaggle.com/datasets/tapakah68/body-measurements-dataset
88. https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0291084
89. https://registry.opendata.aws/bodym/
90. https://pmc.ncbi.nlm.nih.gov/articles/PMC7348910/
91. https://registry.opendata.aws/bodym/
92. https://www.kaggle.com/datasets/tapakah68/body-measurements-dataset
93. https://registry.opendata.aws/bodym/
94. https://registry.opendata.aws/bodym/
95. https://www.kaggle.com/datasets/tapakah68/body-measurements-dataset
96. https://www.kaggle.com/datasets/saurabhshahane/body-measurements-dataset
97. https://aws.amazon.com/marketplace/pp/prodview-w3762gcflmhzk
98. https://dagshub.com/datasets/bodym-dataset/
99. https://unidata.pro/datasets/body-measurements-image-dataset/
100. https://gts.ai/dataset-download/body-size-recognition-dataset/
101. https://data.mendeley.com/datasets/bjv6c9pmp4
102. https://universe.roboflow.com/urmi/body-measurement-zmbgv
103. https://huggingface.co/datasets/UniqueData/body-measurements-dataset/blob/36bcafd86b0396517c5d6345c157ba7f25e59de1/Body Measurements Dataset.pdf
104. https://registry.opendata.aws/bodym/
105. https://www.kaggle.com/datasets/tapakah68/body-measurements-dataset
106. https://registry.opendata.aws/bodym/
107. https://registry.opendata.aws/bodym/
108. image.jpg
109. image.jpg
110. image.jpg
111. https://registry.opendata.aws/bodym/
112. https://registry.opendata.aws/bodym/
113. hwg_metadata.csv
114. measurements.csv
115. subject_to_photo_map.csv
116. https://pytorch.org/get-started/previous-versions/
117. https://raidenworks.com/2024/01/20/torch-cuda-is_available-false-□/
118. https://discuss.pytorch.org/t/cant-to-install-pytorch-for-cuda-12-4/200672
119. https://github.com/unslothai/unsloth/issues/3308
120. https://discuss.pytorch.org/t/install-pytorch-with-cuda-12-1/174294
121. https://discuss.pytorch.org/t/cannot-for-the-life-of-me-get-pytorch-and-cuda-to-install-work/197088
122. https://stackoverflow.com/questions/76678846/pytorch-version-for-cuda-12-2
123. https://www.gcptutorials.com/post/how-to-install-pytorch-with-conda-step-by-step-guide-for-python-3.11-&-machine-learning
124. https://www.reddit.com/r/pytorch/comments/11z9vkf/pytorch_installation_with_cuda_121/
125. https://pytorch.org/get-started/locally/