Software Test Plan (STP)

Introduction

This test plan outlines the strategy and procedures for testing the **ThermoML** system - a machine learning-based tool developed to detect inflammation and pathologies in human hands using thermal imaging. The software combines automatic segmentation, anatomical landmark detection, and classification models within a web interface, and is based on clinical thermal images of surgeons and control subjects as part of the TH-SRG01 study.

Purpose of this test:

The purpose of this test is to verify the correctness, stability, and usability of both the web application and the detection pipeline. This includes validating the accuracy of inflammation detection, ensuring the reliability of joint landmark detection, confirming the proper functioning of the registration and masking process, and evaluating the user interface for ease of use and responsiveness.

Test Items

The following components of the system will be tested:

1. Image Preprocessing
2. Landmark Detection
3. Filtering & Validation Logic
4. Registration & Alignment
5. Classification Preparation
6. Classification Testing
7. Web Application Validation

Features to be Tested

1. Image Preprocessing

1.1 Optical and thermal images are successfully loaded from the dataset.

1.2 Optical images are enhanced using contrast adjustment (CLAHE).

1.3 Hand regions are segmented from images using the SAM model.

2. Landmark Detection

2.1 Automatic detection of hand landmarks on optical images using MediaPipe.

2.2 Landmark detection on thermal images when hands are clearly visible.

2.3 Visual rendering of detected landmarks on both optical and thermal images.

2.4 Automatic palm center calculation only on palmar-facing hands.

3. Filtering & Validation Logic

3.1 Scaling and padding the thermal image to match the actual physical comparable size of the optical image.

3.2 Each hand in an optical image must have 21 detected landmarks. Hands with fewer than 21 points are automatically discarded.

3.3 Matching between thermal and optical images based on consistent filename structure.

4. Registration & Alignment

4.1 Registration between optical and thermal images is performed using geometric transformations, based on masks or keypoint matching.

4.2 Visual verification of registration success is performed by overlaying the optical hand mask or landmarks on the thermal image.

4.3 Quantitative validation of registration is performed by computing the Euclidean distance between corresponding optical and thermal landmarks. Registration is considered valid if the average distance remains below a defined threshold (e.g., 10 pixels).

4.4 If the alignment error exceeds the defined threshold, the registration is considered failed and the image is discarded from further processing.

5. Classification Preparation

5.1 Extraction of 4-channel image patches centered around each joint, used as input for model training. Each patch includes:

(1) The thermal image after registration correction.
(2) A binary mask of the segmented hand region.
(3) The thermal image masked by the hand region.
(4) A distance map representing the Euclidean distance from each pixel to the center of the hand.

5.2 Training of 32 separate classification models (16 per hand)

5.3 Validation of file structures, naming conventions, and image input formatting.

6. Classification testing

6.1 Given a thermal patch for a specific joint, the corresponding model returns a valid classification result (e.g., inflamed / not inflamed).

6.2 Classification results are consistent for identical inputs (deterministic behavior).

6.3 The prediction latency per joint is within acceptable range for user-facing feedback.

6.4 Model outputs are correctly displayed and mapped back to the original joint on the hand image.

7. Web Application Validation

7.1 Users can view their previously uploaded hand images.

7.2 Thermal images are uploaded and processed through Firebase and HuggingFace pipeline.

7.3 Full preprocessing and classification are triggered through the frontend interface.

7.4 Results of inflammation classification are clearly displayed on the frontend with joint-level indicators.

7.5 Authentication, user access control, and top navigation functionality are verified.

Features Not to be Tested

The following features are outside the scope of this test plan and will not be explicitly tested:

- **Performance or internal logic of third-party libraries**

  The system relies on external models such as MediaPipe (for landmark detection) and SAM (for segmentation). Their internal algorithms and performance characteristics are assumed to function as intended and will not be validated by our test cases.

- **User authentication system**

  Although basic login/logout functionality exists, the authentication mechanism (e.g., Firebase Authentication) will not undergo security or penetration testing as part of this plan.

- **Scalability and concurrency**

  The system is designed for individual or small-scale clinical usage. We are not conducting stress testing or performance validation under concurrent access or high-throughput conditions (e.g., hospital-scale use).

  Firebase services (Authentication, Realtime Database, and Functions) handle scaling internally, and we rely on their infrastructure to ensure basic robustness.

  While Hugging Face inference endpoints do not offer strong scalability guarantees at our current budget tier, they were chosen as a cost-effective solution within the project's constraints.

Testing Strategy

Testing will be performed iteratively and in parallel with development milestones, as part of the final project workflow. The approach includes:

- **Unit Testing**

  Individual components such as segmentation (SAM), landmark detection (MediaPipe), palm center calculation, and joint distance logic will be tested using controlled inputs.

- **Integration Testing**

  Modules will be connected and validated together to ensure seamless flow from preprocessing to classification. This includes checking image matching, registration, and mask transfer.

- **System Testing**

  End-to-end flows will be tested through the actual frontend to confirm that the user can upload an image, process it, and receive a visual classification result.

- **Manual & Visual Verification**

  Visual checks will be used to validate overlay alignment between optical and thermal hand images after registration. This includes verifying hand contour overlap and landmark correspondence accuracy.

- **Web Application Testing**

  UI functionality will be tested for smooth navigation, upload interactions, result presentation, and authentication status management.

- **Classification Testing**

  The classification pipeline will be tested to ensure each model (15 joint models) receives the correct thermal patch, returns a valid prediction, and maps the result correctly back to the hand image. We will verify prediction accuracy, consistency, latency, and error handling for missing or invalid patches.

Test Environment

The ThermoML system will be tested in the following environment to ensure compatibility and stability:

- Backend: Hugging Face, Firebase Functions, Node.js server on Firebase Hosting.
- Database: Firebase Realtime Database, Firebase Storage.
- Frontend: React web application tested on Chrome, Safari and Firefox browsers.

Responsibilities

All testing activities are conducted jointly by both project team members, Shachar Zeharia and Maxim Feldman.

Schedule

Testing is aligned with final project delivery. Intermediate checks and logs will be completed before model training.

Risks and Contingencies

- Risk: Possible delays in integrating hugging face python and docker files.

  Mitigation: plan to use mock run on our hardware if delays occur.

- Risk: Data inconsistency (e.g., missing thermal-optical image pairs or incorrect file names).

  Mitigation: Automate validation of filename structure and directory scanning during preprocessing.