...

Using Smart-textiles and Virtual Reality for Artificial Intelligence Enabled Monitoring and Management of Sleep, Fatigue and Mental Health for Deep Space Exploration

[Development of enabling space technologies, PT-7 Artificial Intelligence-Enabled Solutions for Crew Health and Wellness on Deep Space Missions]

D8 - Final Report

M5: Machine Learning Models

Date	Author	Revision	Description of Change
28-06-2025	Mohamamd Jambar	1.0	Creating final report document
17-06-2025	Mohamamd Jambar	1.1	Revisions based on CSA review comments
08-08-2025	Mohamamd Jambar	1.2	Revisions on dashboard visualizations and information
02-09-2025	Mohamamd Jambar	1.3	Changes made to “TRL 6 justification of final prototype” subheading to focus on current prototype use.

Table of Contents

1 Abbreviations and Definitions

Abbreviation	Full Form
AEKF	Adaptive Extended Kalman Filter
BACC	Balanced Accuracy
Bi-LSTM	Bidirectional Long Short-Term Memory
CNN	Convolutional Neural Network
CSA	Canadian Space Agency
CSV	Comma Separated Values
DL	Deep Learning
ECG	Electrocardiogram
EEG	Electroencephalography
GPS	Global Positioning System
GUI	Graphical User Interface
HR	Heart Rate
HRV	Heart Rate Variability
IMU	Inertial Measurement Unit
ISB	International Society of Biomechanics

JCS	Joint Coordinate System
LightGBM	Light Gradient Boosting Machine
LSTM	Long Short-Term Memory
MARS	Morning Assessment Restful Sleep
ML	Machine Learning
PPG	Photoplethysmography
PSG	Polysomnography
RQ	Research Question
SOTA	State-of-the-Art
STAR	Stress and Tiredness Assessment in Real World
SVM	Support Vector Machine
TCNN	Temporal Convolutional Neural Network
TRRA	Technology Readiness and Risk Assessment
VR	Virtual Reality

2 Introduction

The report covers the summary of each step of the development done as part of the project. Recommendations and future development avenues for the project are also provided. Finally, a Technology Readiness and Risk Assessment (TRRA) has been done for the project.

Our health monitoring dashboard represents a cutting-edge prototype developed through Canadian Space Agency (CSA) partnership funding, demonstrating advanced technical capabilities essential for both terrestrial healthcare and future space mission applications. Built on Apache Superset and containerized with Docker, this system showcases our organization's growing expertise in autonomous health monitoring systems while delivering immediate value for space medicine research and operational deployment.

This prototype project, supported by the Canadian Space Agency, serves dual purposes: advancing our technical capabilities in autonomous health monitoring systems while developing technology directly applicable to astronaut health monitoring during extended space missions. The CSA partnership validates our approach to building resilient, self-contained health monitoring systems capable of operating in isolated environments with limited connectivity—core requirements for both space applications and remote terrestrial deployments.

Through this project, our team has developed specialized expertise in edge computing architectures, autonomous database management, and real-time physiological data processing—capabilities that position us as leaders in space medicine technology development. The successful implementation demonstrates our ability to build mission-critical systems that operate reliably in resource-constrained environments.

Through this project, our team has developed specialized expertise in edge computing architectures, autonomous database management, and real-time physiological data processing—capabilities that position us as leaders in space medicine technology development. The successful implementation demonstrates our ability to build mission-critical systems that operate reliably in resource-constrained environments.

The solution eliminates ongoing operational expenses through its open-source foundation and self-contained architecture. Organizations avoid expensive licensing fees, subscription costs, and cloud hosting charges while maintaining enterprise-level functionality. The Docker-based deployment reduces IT overhead and ensures consistent performance across diverse infrastructure environments.

Complete data sovereignty ensures that all sensitive health information remains within the organization's controlled environment. This architecture addresses critical regulatory requirements such as HIPAA, GDPR, and other healthcare data protection standards by eliminating external data transmission and maintaining full audit trails within the client's infrastructure.

The containerized architecture enables rapid deployment across laptops, local servers, or enterprise infrastructure with a single command. This flexibility supports both pilot implementations and large-scale organizational rollouts, allowing for seamless scaling as user requirements grow.

3 Summary of Different Project Phases

The work done in the project is divided into five phases, i) Protocol Design and Data Collection, ii) Ml Modelling, iii) Model Evaluation, iv) Dashboard Development, and v) Final Prototype Development.

3.1 Protocol Design and Data Collection:

The literature was reviewed on modelling and detection of mental health related variables using physiological data collected through continuous monitoring using the smart-textile based solution. The main variables chosen included, stress, fatigue, and sleep quality and sleep staging. These variables play a key role in overall mental health and performance of astronauts. The focus on the literature review was to answer the following research questions (RQ):

- RQ1: What kind of protocols have been used in the literature for real world stress and fatigue detection using wearable sensors?
- RQ2: Has subjective sleep quality been measured in real world conditions before and what are sleep staging related variables that impact sleep quality?
- RQ3: What kind of pattern recognition approaches have been used in the literature to correlate physiological measurements to mental health variables?

The protocols developed used Myant Skiin textile-integration technology for collecting physiological data. The electrodes are integrated directly into garments using knitted silver yarns. This allows the final form factors to be flexible, washable, reusable and unobtrusive thus making them ideal for deploying in space where resource constraints, reusability and unobtrusive monitoring are important. Currently, the garments are available in the following form factors:

- Underwear
- Chestband
- Tanktop
- Bra

A small electronic pod is used to interface with the textile form factors and used for transmitting data to a recording device via bluetooth. For the current generation of Skiin garments, 3-lead ECG is recorded at 320 Hz along with a 3-axis IMU recorded at 25 Hz.

The following protocols were proposed:

i) Stress and Tiredness Assessment in Real World (STAR): The STAR study was designed to collect subjective stress and fatigue labels from participants throughout the day while simultaneously monitoring physiology using the textile-based wearable solution. The study consisted of an onboarding phase where habits (frequency of exercise, caffeine and alcohol) as well as demographic information was collected from the participants. Following this, the data collection phase consisted of participants wearing continuous physiological monitoring (using Skiin chestband and the Skiin App) along with answering survey questions about their stress and fatigue levels four times a day (9AM, 1PM, 5PM, 9PM). The questionnaire also collected information about sleep quality from the previous night as well as caffeine and alcohol intake. The goal was to capture daily variability in stress and fatigue levels. The surveys were developed on Google forms and sent as email or Calendar notifications.

The different components of the study are shown in the figure below:

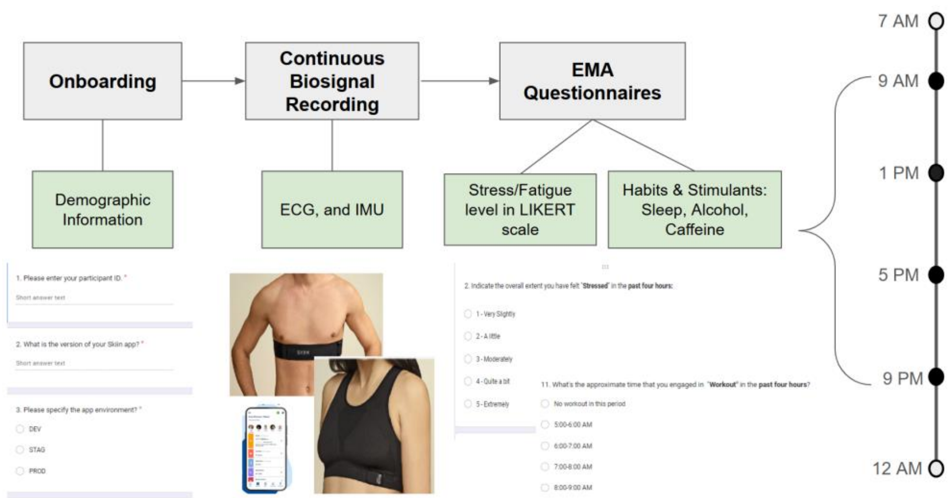


Figure 1: Three components of STAR study

Based on the study size in the literature, budget, and time constraints. We aimed to collect data from 25 to 30 participants for a 14 day period. In total 27 individuals participated in the study.

ii) Morning Assessment Restful Sleep (MARS): The goal of the MARS study was to collect overnight physiological variability along with subjective assessment of sleep next morning. The study had a similar onboarding process as STAR with subjects answering demographic and habit related questions. In the data collection period, participants were asked to wear the Skiin

Chestband while sleeping and answering questions about their sleep quality the next morning. The questionnaire was developed on Google form and sent as an email and calendar notification every morning at 7AM. The questionnaire inquired about different components of sleep including: sleep quality, mental energy, sleep sufficiency etc. Participants also reported their physical and mental stress and fatigue levels from the previous day. The summary of the protocol looks is shown below:

Figure 2: Summary of the MARS protocol

The aim of the study was to recruit a small sample of participants that we can monitor for a long period of time. This would allow for personalized and longitudinal modelling as well as could reveal insights into long-term physiological changes with stress, sleep, fatigue and overall health. As a result, the aim of the study was to recruit between 5-10 participants with 30 days of data collected from each one. We ended up recruiting 20 participants with 16 subjects enrolled in both STAR and MARS studies giving a unique dataset of near 24-hour physiological monitoring along with mental health related labels.

iii) Polysomnography (PSG) Study: The changes in overnight physiology correspond to changing sleep stages throughout the night. The time spent in different sleep stages and number of transitions to and from and to various stages have been linked to objective measures of sleep quality and recovery. As a result, a clinical PSG study was proposed with the purpose of collecting clinically assigned sleep staging labels along with overnight physiological data. The physiological data included both standard clinical signals recorded during a PSG study (SpO2, respiration, airflow, ECG, and EEG etc.) as well as a Skiin chestband. The Skiin chestband provided 3-lead ECG, as well as 3-axis IMU measurements. The protocol along with the clinician generated sleep staging labels are shown in the figure below:

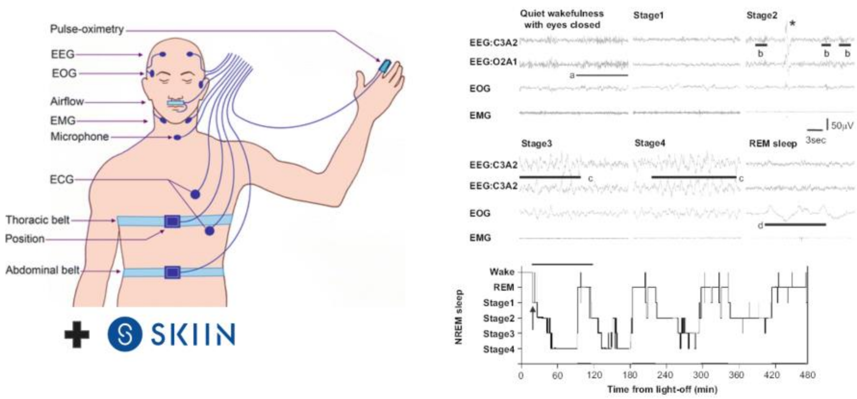


Figure 3: (Left) Signals recorded in standard PSG study, (Top Right) How different signal characteristics are used to determine sleep stages, (Bottom Right) Sleep stags changing through the entire sleep cycle

We aimed to record 80 participants for the study with 70 subjects recruited from patients visiting the sleep clinic and suffering from various sleep disorders along with 10 health participants. Due to massive amounts of data loss and noise issues, our final dataset only consisted of 40 participants.

iv) Physical Health Assessment Study: This study aimed to quantify kinematics during cardiovascular and strength exercises using textile-integrated IMU sensors while gravity information is removed. Traditional kinematic measurements typically rely on expensive, camera-based motion capture systems, which necessitate specialized laboratory environments and lack versatility for broader applications. The proposed protocol involved participants performing a range of physical activities, including those common in astronaut physical health training, while instrumented with eight IMUs strategically placed on the right side of the body. Concurrent data collection was performed using a gold-standard optical motion capture system. Ten participants were enrolled in the study.

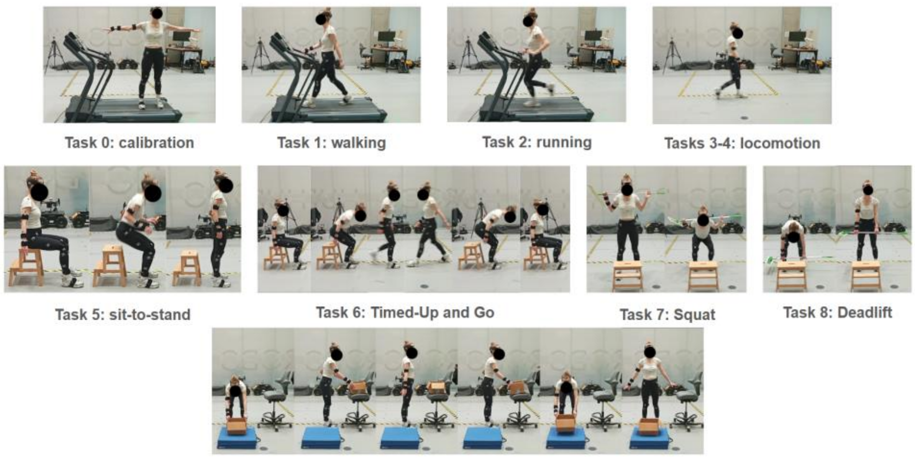


Figure 4: The different tasks in the physical health assessment protocol.

3.2 ML Modelling

3.2.1 Preprocessing Pipeline

Due to similar modelling techniques used for the MARS, STAR and PSG datasets, they were preprocessed to a standard format. The preprocessing pipeline is shown below:

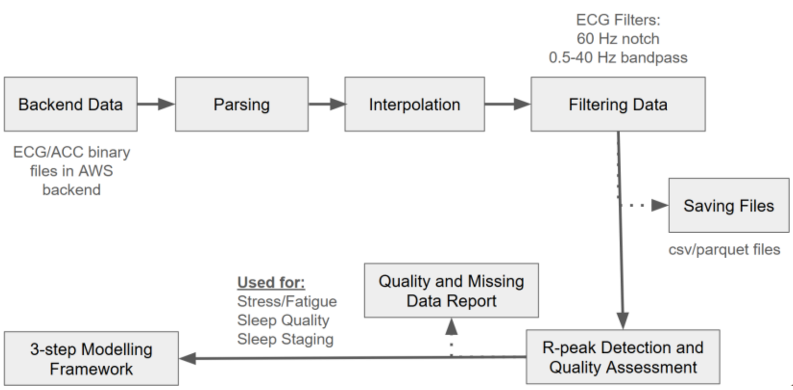


Figure 5: Preprocessing pipeline for data collected using Skiin chestbands for STAR, MARS and PSG study

The study data is originally in encrypted binary format on Myant's backend. For usability, this data is first parsed into dataframe format for both ECG and ACC. Following this, linear interpolation of missing segments is performed on the ECG data for ease of visualization. Next, the ECG signal is filtered using a bandpass filter (0.5-40 Hz) to preserve ECG signal content and with a 60 Hz notch filter to remove power line interference. The ECG and ACC files are then saved in parquet format. Using parquet instead of CSV helps allow for faster loading of data by trading off readability. Finally, Rpeak detection and quality assessment of RR intervals is performed using Myant's light weight quality algorithm. The RR intervals are stored as a CSV file along with a quality and missing data report. The raw and lightly processed RR and ACC files are then used for development of various pattern recognition models. The modelling stages as be described with a 3-step framework as follows:

i) Data Transformation: The raw RR interval and ACC data is first transformed to prepare for feature extraction using windowing, quality threshold and assessment, data augmentation, and label binarization. Data transformation is done differently based on how the data was collected and the frequency on the subjective labels. The transformation steps were also derived from the literature review and RQ3 done in the protocol design stage.

ii) Feature Extraction: Once the data has been transformed, features are extracted from this data. Heart Rate (HR), Heart Rate Variability (HRV), IMU-derived respiration, and activity based features were extracted for different algorithms.

iii) Pattern Recognition: Finally the feature-label pairs are used for pattern recognition algorithms either using classical machine learning or deep learning models.

3.2.2 Stress and Fatigue Modelling (STAR):

One stress and fatigue label was collected for every four hour window in a day. As a result, augmentation was performed on the data by dividing the 4 hour period into 1 hour chunks each corresponding to the same subjective label. This augmentation allowed for 4 data-label pairs to be generated out of a single one increasing the number of samples by 4, helping with model training

For each of these one hour windows, the data was epoched into 5 minute epochs with a 2.5 minute overlap and HR, HRV and IMU based activity features were extracted for each epoch. This resulted in a 24 sample series for every feature for a 1 hour window. After rejecting noisy epochs due to poor quality and feature sequences that are too short this sequence of features is either input into a sequential deep learning model or summarized using feature aggregation by calculated mean, standard deviation, range etc. for the full window. The whole process is summarized in Figure 6.

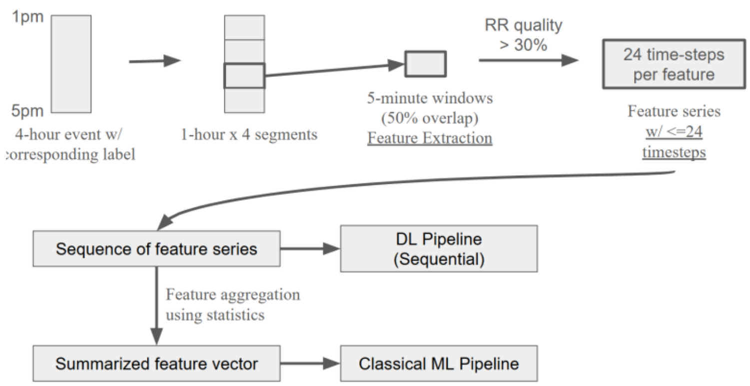


Figure 6: Data transformation and feature extraction for stress and fatigue detection

3.2.3 Sleep Quality and Mental Energy Modelling (MARS)

For modelling sleep quality and mental energy, the overnight RR and ACC data corresponding to a single sleep quality label is transformed by isolating the period where the participants were asleep by using an IMU based actigraphy algorithm. Following this, multi-resolution features (HR, HRV, ACC-derived Respiration and ACC) in 150, 90, and 30s windows are calculated for the asleep period in 30s steps. Once the features are calculated, features are summarized by either using statistical aggregates (similar to stress modelling) or by extracting domain specific features such as time to minimum HR, slope to minimum HR etc. This summarized vector along with the sleep quality and mental energy labels are used to train a classical machine learning model. No deep learning options were explored in this case, due to the small size of the dataset (575 samples). The data transformation and feature extraction are summarized below:

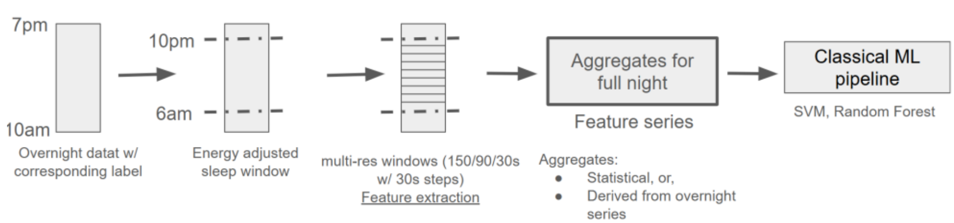


Figure 7: Data transformation and feature extraction for mental energy and sleep quality detection

3.2.4 Sleep Staging Model

In PSG studies, clinicians generate sleep staging labels every 30s. For each of this window, we extract features in different ways for the classical ML or Deep learning pipeline.

- For classical ML pipeline: Multi-resolution features (similar to mental energy/sleep quality modelling) are extracted in 150, 90, and 30s windows with the 150s and 90s windows being centered around the 30s window corresponding label. The feature-label pair is then used for model training.
- For DL pipeline: For RR intervals median, standard deviation and quality are extracted for 30s intervals. For the respiration signal (derived from ACC data), respiration rate, median and standard deviation are calculated. Median, norm of the 3-axis accelerometer and standard deviation are calculated as ACC features. Finally, sensor independent circadian features are extracted and added to the feature vector. This creates a feature vector of length 15 and the full overnight series of features and corresponding labels are used to train a sequential model.

3.2.5 Physical Health Assessment Modelling:

Motion Capture Data Preprocessing: Raw motion capture data, consisting of three-dimensional trajectories of retroreflective markers affixed to anatomical landmarks, underwent a comprehensive preprocessing pipeline. Initially, markers were labeled using the motion capture system's proprietary software. Subsequently, data gaps within the trajectories were interpolated utilizing the same software. To construct rigid body models representing individual body segments, markers associated with each segment were spatially grouped. The data were then subjected to a dual-pass Butterworth low-pass filter to mitigate high-frequency noise while maintaining signal integrity. Finally, the orientation of each body segment was computed from the trajectories of its anatomical landmarks and represented as quaternions.

IMU Data Preprocessing: Triaxial accelerometer and gyroscope data were acquired from eight Inertial Measurement Unit (IMU) pods, with each data packet containing twelve samples. The raw data were subsequently parsed into structured data frames, yielding three-dimensional acceleration and angular velocity vectors for each pod. A dual-pass Butterworth low-pass filter was applied to the IMU data to attenuate high-frequency noise while preserving signal integrity.

Orientation Estimation: An Adaptive Extended Kalman Filter (AEKF) was implemented to estimate the orientation, represented as a quaternion, and the gravity vector for each IMU pod. This estimation utilized both accelerometer and gyroscope data. The hyperparameters of the orientation filter were optimized by minimizing the discrepancy between the estimated orientation and a reference orientation obtained from the motion capture system. Gravitational acceleration was subsequently eliminated from the IMU readings to obtain gravity-free accelerometry for further analysis.

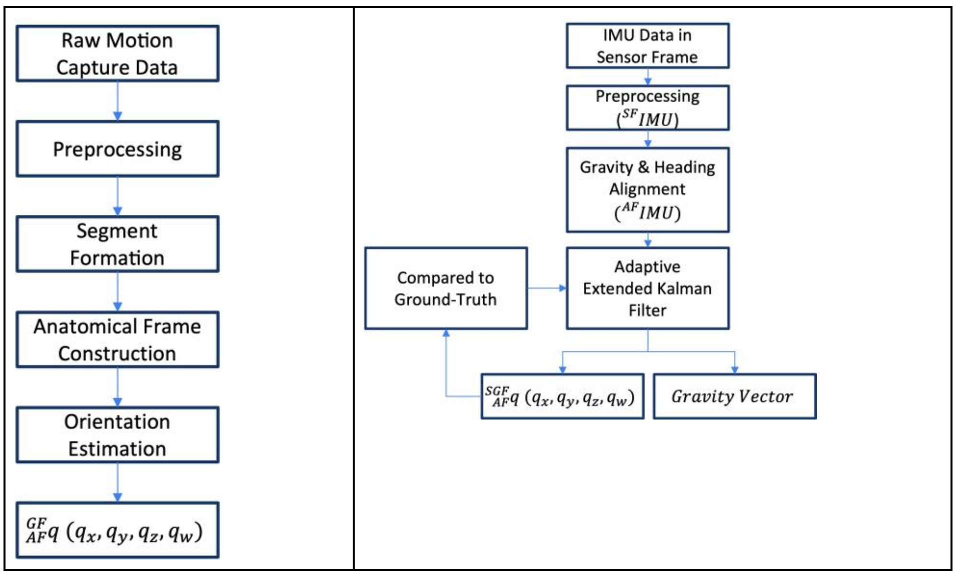


Figure 8: motion capture (left) and IMU (right) data processing and orientation estimation

Deep Learning Model: This study employs a temporal deep neural network to emulate the behavior of a tuned AEKF for orientation (quaternion) estimation in the absence of gravitational cues. The network's input consists of a 1-second window (100 samples) encompassing 3D gyroscope data, 3D gravity-free acceleration data, strap-down integration of 3D gyroscope data, and a one-hot encoded vector representing pod location. The network architecture comprises two primary components: a feature encoder block and a regression block. The feature encoder integrates a convolutional block for spatial feature extraction, an LSTM block for modeling temporal dependencies, and a Transformer encoder block to capture dynamic input behaviors. The regression block, composed of feedforward layers, is succeeded by a custom quaternion normalization layer designed to enforce the unique mathematical properties of quaternions. To accommodate the specific characteristics of quaternions—namely, that a quaternion and its negative represent the same orientation, and that a quaternion's norm must be unity—a custom loss function was developed and implemented for model training, consistent with established literature. A custom hyperparameter tuning using Bayesian Optimization was conducted to optimize the hyperparameters of the model.

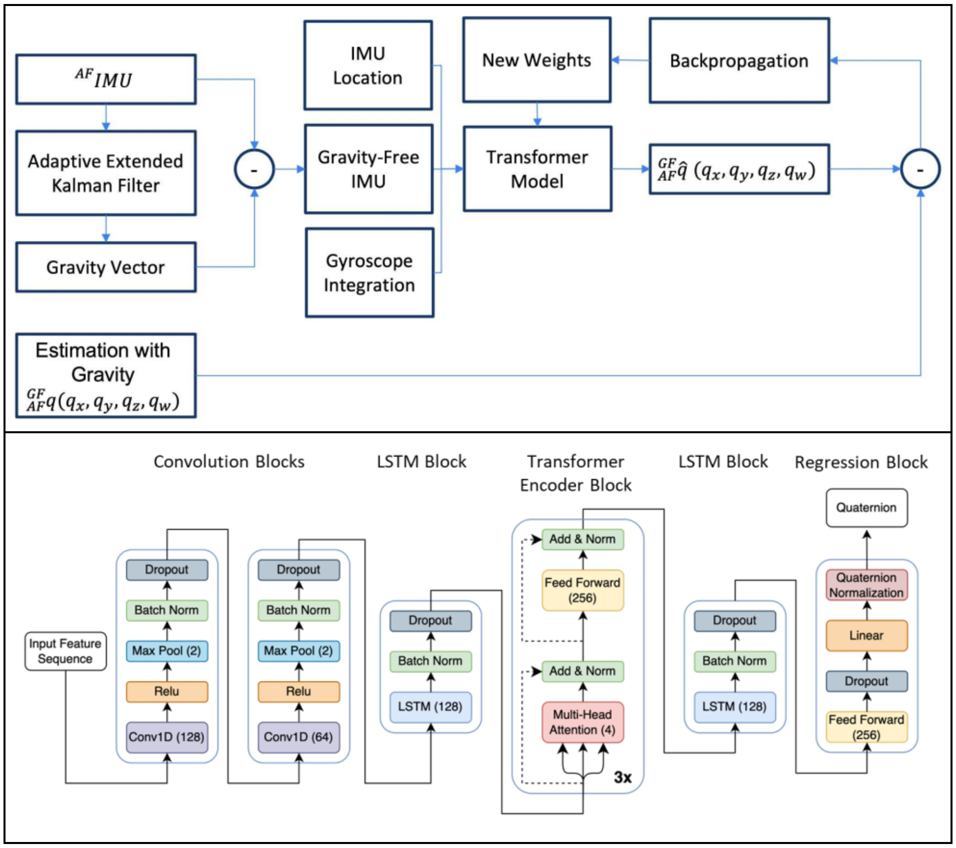


Figure 9: (top) training logic, (bottom) model architecture

Joint Angle Estimation: Joint angles were estimated by calculating the relative orientation between superior and inferior segments. These joint angles were subsequently expressed as Euler angles or in accordance with the Joint Coordinate System (JCS) Convention proposed by the International Society of Biomechanics (ISB).

Range of Motion: The range of motion for key joints (ankle, knee, hip, elbow, and shoulder) across various tasks will be determined by calculating the difference between the maximum and minimum joint angles. This information is critical for evaluating exercise efficacy and conducting ergonomic assessments.

Joint Coordination: Joint coordination, specifically for two-joint (ankle-hip) and three-joint (ankle-hip-L5/S1) kinematic chains, was quantified using the Cancellation-Index equation as defined in the literature, derived from the corresponding joint angles.

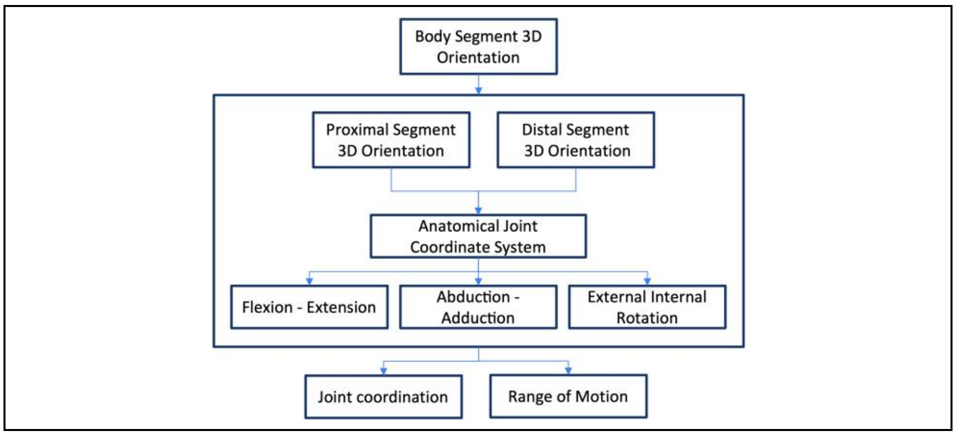


Figure 10: biomechanics including joint angles, range of motion, and joint coordination

3.2.6 Classical ML Pipeline

For datasets with smaller number of samples i.e. STAR and MARS studies, models were developed and evaluated with a strategy that ensures true generalizability of the final results. As a result, the datasets were first divided into three groups:

- D_Fs: This dataset is used for selection of optimal number of features. This step is important as having a very large feature set might reduce generalizability of the model developed due to overfitting.
- D_CV: After feature selection has been performed, the cross validation dataset is used for evaluating the best model and hyper-parameters setting. The performance of the model is defined as the average over all the different folds thus ensuring better generalization. For each of the folds, the data is normalized, followed by sampling and training of the model with a given hyper-parameter set. As a result, the optimal model hyper-parameters and sampling parameters can be determined using the evaluation process
- D_Te: Once the optimal model has been selected using D_CV, the results on the final test set D_Te are generated. This result should be representative of true model performance on any unseen data. To prevent optimistic bias in reporting results, D_Te has not been used for training or making any modelling decisions in any capacity.

The division of the entire dataset into the above mentioned three groups can be done using two different subject-data splitting strategies. For subject-independent splits, the training and test datasets both have unique subjects i.e. the model is trained and evaluated on completely different individuals. Typically, these models may not lead to the best classifier performance, however, they are advantageous as they can be used on completely new data without any changes to the classifier. In contrast, for the subject-dependent split where the data from the same subject may appear in train and test data yield higher performance. However, they require the model to be retrained with part of the new subjects' data. Typically, this kind of split is useful when the model needs to be implemented in a fixed closed team such as firefighters, police officers, or astronauts.

The classical machine learning pipeline has been shown below:

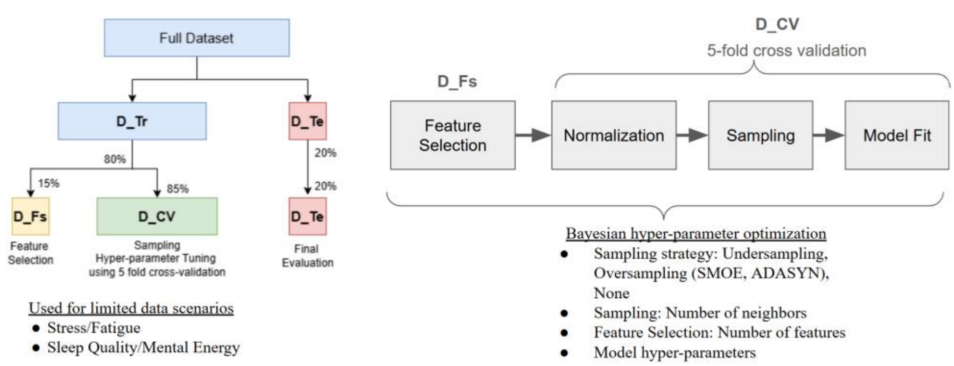


Figure 11: Classical machine learning pipeline with the dataset split (left) and different steps and hyper-parameters tuned (right)

The various hyper-parameters shown in Figure 11 were tuned using *Optuna,* a bayesian optimization python package. For most models, Balance Accuracy (BACC) was optimized as an evaluation metric

For a dataset with larger sample sizes (Physical Health Assessment and PSG), a standard training (60%), validation (20%) and test (20%) split was applied to the dataset. Similar hyper-parameter tuning and modelling steps were performed as described in Figure 11 (right).

The benchmark models were developed using the classical machine learning pipeline in order to assess the performance improvements we can get using a deep learning based solution. The choice of the classifier was based on literature review and the best ML models. As a result, Light Gradient Boosting Machine (LightGBM) model was used for PSG, STAR dataset based models while two alternative classical machine learning models were used for the MARS dataset. While both models are considered typical machine learning models, the benchmark SVM model is a much weaker model (both in terms of computational complexity and ability to fit complex datasets) compared to the proposed RandomForest model.

MYANT

...

3.2.7 Sequential Deep Learning Pipeline

Deep learning models can handle sequential data such as physiological signals really well compared to classical ML models as they are able to automatically learn temporal dependencies, hierarchical features, and complex patterns over time without the need for extensive manual feature engineering. As a result, a sequential deep learning package (using Keras and Tensorflow) was created for deep learning model development with a standard 60-20-20 train, validation and test split for the datasets.

The latest sequential models were explored including LSTM, Bi-LSTM, 1-d CNN + LSTM, TCNN, and Transformers. The package allows for ease of selection for different components of a deep learning pipeline including: type of optimizer, loss function, and regularization method. A bayesian optimizer was used to select hyper-parameters including:

- Learning rate
- Number of layers
- Size of layers
- Normalization method
- Dropout rate
- Regularization rate

The bayesian optimizer performance was tuned using Area Under Curve- Precision Recall Curve (AUC-PRC) scores that allow for better performance with imbalance datasets. The best model was picked based on the validation dataset performance and evaluated on the test set.

3.3 Model Evaluation and Results

3.3.1 Stress and Fatigue Modelling Results

The balanced accuracy performance for the benchmark and proposed stress models in a subject-dependent setting is shown in the table below:

Stress - Balanced Accuracy		
	Benchmark	**Proposed**
Model	**LightGBM**	**Conv + LSTM**

Validation	0.720	**0.755**
Test	**0.616**	0.507

The benchmark Light GBM model gives the best performance on the test set. The table below compares this performance to the state-of-the-art solutions that have been trained and tested on real world datasets.

Paper	**Performance**	**Notes**	**Model Improvement**
Tiwari et al. (2021)	BACC - 0.658	200 hospital nurses with 3 months of data collected	Only cross validation reported, no hold-out test subject - **Optimistic Bias**
Velmovitsky et al. (2022)	ACC- 0.55	33 subjects, 6 ratings a day with 30s ECG from apple watch	Performance optimized on test set itself - **Optimistic Bias**
Zhang et al. (2024)	ACC- 0.65	77 subjects, ECG + phone use + GPS + other sensing data, 12 ratings daily at 45 minute intervals	**Comprehensive analysis of models, epochs etc.** Uses extra data including sleep information, previous day's features,
STAR	ACC - 0.685 BACC - 0.61	25 subjects, 4 ratings per day	Hold out test set and other best ML practices used

Overall, our classical machine learning benchmark model performs similar to SOTA. With fewer participants while only using wearable signal data and no added modalities (screen time, GPS etc.) Additionally, we have a reduced optimistic bias in our model as a result of good ML development practices.

For fatigue, the results look like:

Fatigue - Balanced Accuracy

	Benchmark	Proposed
Model	**LightGBM**	**TCNN**
Validation	**0.670**	0.595
Test	**0.647**	0.573

Once again, the benchmark model outperforms the proposed deep learning model. Looking into the literature we found that fatigue detection studies have not looked into real world settings aside from a few driver fatigue detection cases [1]. As a result, there is no established benchmark of expected performance in the literature. But the lack of studies highlights the unique nature and importance of this dataset.

Overall, the proposed models have the following advantages:

- **Multimodal data**
  - They makes use of ECG and IMU data collected from textile form factor
  - Stress ratings are collected every 4 hours to capture daily fluctuations compared to a single stress rating in larger studies
- **Novel feature extraction methods**:
  - Literature focuses simply on time and frequency domain features
  - Our models used extensive non-linear and recently proposed features (Tiwari et al. 2021) that have shown better overall performance
- **Robust ML pipeline** that considers latest classical ML model developments:
  - Bayesian optimization was used for efficient hyper-parameter tuning
  - Light Gradient Boosting Tree Model has been recently proposed and shows promising results with imbalanced real world datasets.
  - Proposed model evaluation technique: The proposed evaluation strategy allows for model generalizability in limited data sample scenarios.
- **Metrics available for interpretation:**
  - Models are not completely black box and metrics such as HR and HRV can be visualized
- **Fatigue Modelling:**
  - No studies have developed fatigue models on **real world datasets**.

3.3.2 Mental Energy and Sleep Quality Model

The mental energy and sleep quality models were based on subject-independent data split strategy. Due to a small number of samples, both the benchmark and proposed models were based on classical machine learning pipeline. Support Vector Machine was used as a benchmark model, while a RandomForest was used as a proposed model.

Mental Energy - Balanced Accuracy		
	Benchmark	**Proposed**
Model	**SVM**	**Decision Tree**
Validation	**0.625**	0.620
Test	**0.603**	0.531

Sleep Quality - Balanced Accuracy		
	Benchmark	**Proposed**
Model	**SVM**	**Decision Tree**
Validation	**0.622**	0.634
Test	**0.578**	0.542

In both cases the benchmark model achieved better performance compared to the proposed ones. Similar to the fatigue modelling scenario, subjective sleep quality and mental energy using overnight data is a novel assessment. Commonly used sleep scores are weighted averages of objective measures and heuristics but they haven't been validated with subjective sleep quality results.

3.3.3 Sleep Staging Model

The sleep staging model performance was assessed using various metrics with a special focus on Cohen's Kappa as it has been widely used as a performance metric in sleep staging literature.

Sleep Staging				
Metrics	**Benchmark**		**Proposed**	
Model	**LightGBM**		**Bi-LSTM**	
Dataset	**Val**	**Test**	**Val**	**Test**
acc	0.468	0.49	0.41	0.37
f1	0.474	0.517	0.39	0.35
kappa	**0.233**	**0.217**	**0.28**	**0.24**
bacc	0.454	0.485	0.45	0.47
wake-acc	0.748	0.822	0.70	0.71
rem-acc	0.769	0.742	0.50	0.49
deep-acc	0.832	0.853	0.84	**0.94**
light-acc	0.587	0.563	0.40	0.41

The proposed Bi-LSTM was the best performing model based on Kappa performance. Additionally, the deep learning model was trained to optimize deep sleep detection as it is important for rest and recovery in individuals.

3.3.4 Physical Health Assessment Modelling

The core of the physical health assessment is the orientation estimation model, representing orientation as quaternions. Quaternions exhibit specific characteristics that must be considered when evaluating the model: (1) a quaternion and its negative represent the same orientation, and (2) a quaternion's norm must be unity. To accommodate these characteristics, a custom loss function and an angular error metric were developed and implemented for model training and evaluation consistent with established literature as shown below:

$$Loss = \frac{1}{N}\sum_{i=1}^{N} \left(1 - |q_{true}^{(i)} \cdot q_{pred}^{(i)}|^2\right)$$

$$angular\ error = \frac{1}{N}\sum_{i=1}^{N} \theta^{(i)} = \frac{1}{N}\sum_{i=1}^{N} 2 \, . \, cos^{-1}(|q_{true}^{(i)} . q_{pred}^{(i)}|)\frac{180}{\pi}$$

where $q_{true}^{(i)}$ and $q_{pred}^{(i)}$ represent true and prediction quaternions at sample (i); $\theta^{(i)}$ represents angular error at sample (i); and N is the number of samples. The loss and angular error were calculated for the train, validation, and test sets as demonstrated in the following table:

	Loss	**Angular Error (deg)**
Train	0.0055	5.1991
Validation	0.0379	11.8110
Test	0.0253	9.0208

The observed angular error aligns with findings reported in the literature. For example, Golroudbari et al. demonstrated a comparable angular error using a temporal deep neural network (see figure below). It is crucial to note that our proposed model utilizes gravity-free IMU data as input, in contrast to models in the literature that specifically incorporate gravitational information to enhance accuracy.

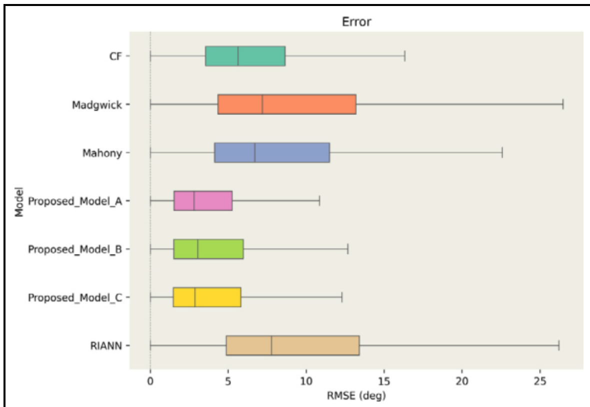


Figure 12 Golroudbari, Arman Asgharpoor, and Mohammad Hossein Sabour. "Generalizable end-to-end deep learning frameworks for real-time attitude estimation using 6DoF inertial measurement units." Measurement 217 (2023): 113105.

Segment-wise evaluation of the model demonstrated comparable loss and angular error across all body segments, as illustrated in the figure below.

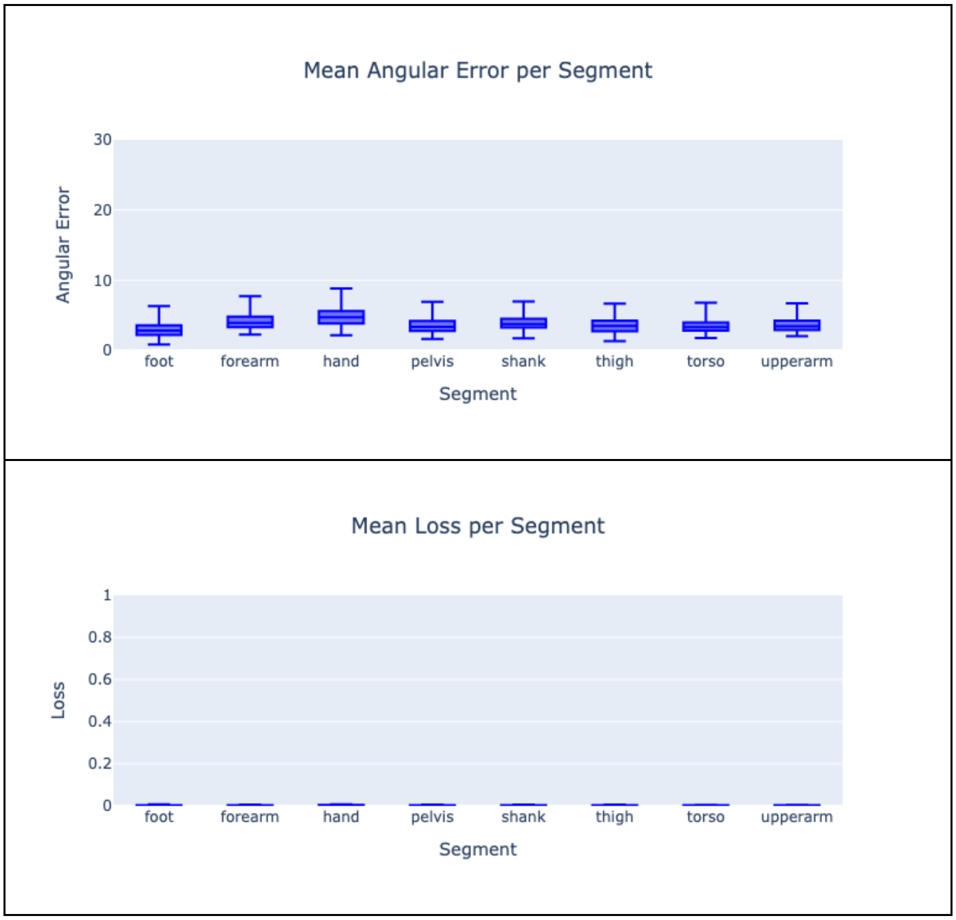


Figure 13. Quaternion loss per body segment (top) and quaternion angular error per body segment (bottom)

3.4 Dashboard Development

Commented [1]: @miladam@myant.ca @behzad.amini@myant.ca dashboard development information.

Commented [2]: @amirali.toossi@myant.ca

3.4.1 Integration of software packages

The pipelines described in Section 2.3 were integrated together and packaged in order to create an end to end pipeline that takes data recorded by Skiin chestbands and outputs relevant metrics. These metrics are then written into the database and can finally be visualized by using the

dashboard. The overall architecture and the packages the different parts belong to are shown in Figure 14. These include:
i) Preprocessing Package: Takes data collected using Skiin chestband and app and preprocesses it using filtering, R-peak detection and interpolation. The files are converted into uniform and easy to read dataframes stored in parquet or csv files. Additionally generates information about the collected data i.e. quality, amount missing, total duration of recording.
ii) Feature Extraction Package: Once the physical health assessment IMU data is separated from data collected using Skiin chestbands, the chestband based physiological data is divided into sleep and awake segments using IMU-energy expenditure information. Following this, feature extraction is performed on the data for stress, fatigue, sleep quality, and mental energy detection. Specialized HR, HRV, Activity and IMU-derived respiration features are extracted and saved as csv files. Overnight HR and HRV series and averages are also generated.
iii) Model Inference: These packages transform the extracted features using windowing, aggregation, normalizing, and feature selection. Following this, the best trained models are used to generate various predictions.
Finally, all the generated information is stored into the database which is read by the dashboard.

The system delivers four specialized dashboard interfaces powered by Apache Superset's enterprise visualization engine with custom D3.js components::

- **Daytime Report**: Gives overall report about the recording characteristics (recording duration, quality and missing data information) as well as stress and fatigue variation information (hourly changes and overall scores).
- **Night Report**: Provides information about sleep using sleep staging analysis as well as overnight physiological changes. This includes overnight HR and HRV variation as well as changing sleep stages throughout the night.
- **Physical Health Assessment**: For physical health assessment protocol, insights about joint coordination and other kinematics metrics will be visualized.
- **Summary and Recommendation**: Visualization of trends for various metrics (stress/sleep/fatigue score, average HR and HV) over a 7 day period as well as a recommendation to use the bWell, system in case of consistently poor scores (scores < 50 for more than 3 days in a 7 day period)

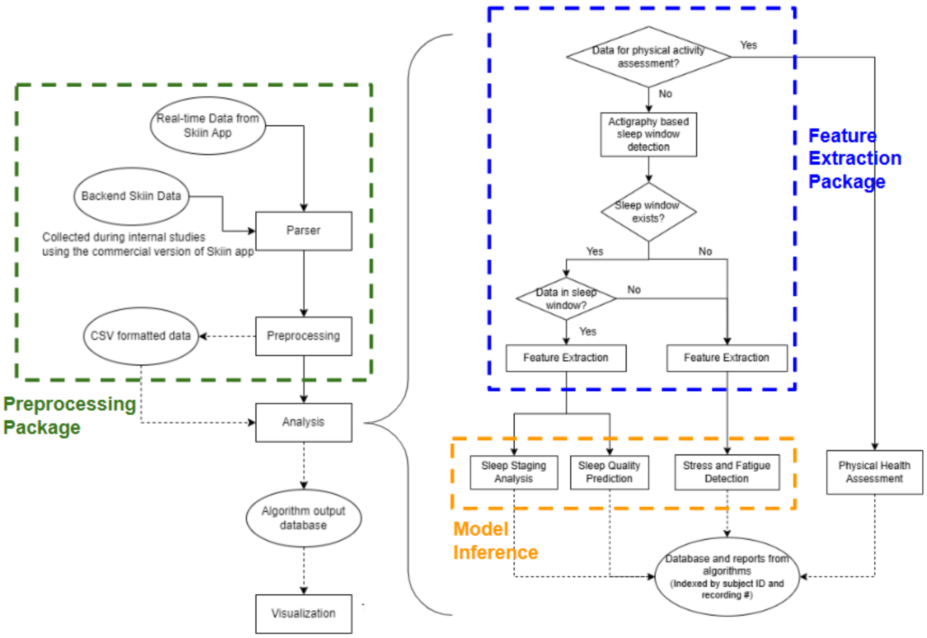


Figure 14: Integration of different software packages developed

3.4.2 Dashboard Features

The dashboard shows the following personalized health metrics, scores and outputs for the different participants

Construct	**Metrics per recording**	**For 7-day summary**
Signal Integrity	RR Signal Quality ECG Missing % ACC Missing %	
Stress and Fatigue	Changing stress and fatigue levels hourly per recording Stress and fatigue score	Stress and fatigue score
Sleep	Sleep stages changing overnight	Sleep score per night

	Sleep score per night Percent/Minutes of sleep stages (light, deep, REM, awake) Mental Energy Sleep Quality	Percent/Minutes of sleep stages (light, deep, REM, awake)
Physiological metrics	Time series of overnight HR Time series of overnight HRV Average overnight HR Average overnight HRV	Average overnight HR Average overnight HRV
Physical health assessment metrics	Metabolic rate (MET), VO2, calories, and physical intensity level Joint angles, joint coordination, and range of motion	

All of the above information will be stored in a standardized dictionary format (dictionary key details have been provided in Final Technical Report CDRL 11). This format allows for easy integration of these inputs into any other system.

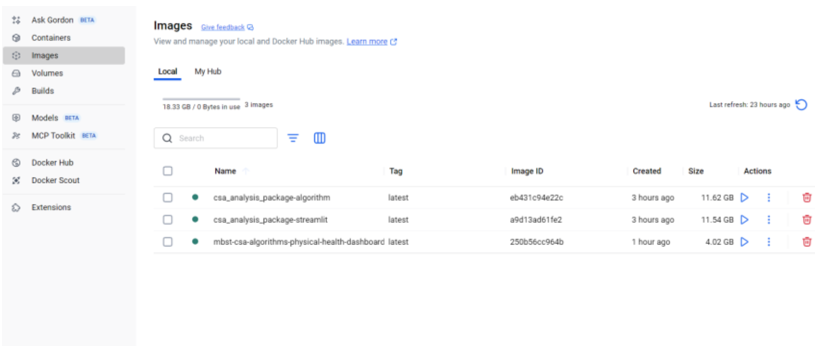
3.4.2 Dashboard Visualizations

The dashboards for visualization have been built using Streamlit. The Streamlit dashboard comes bundled in a Docker image. For simplicity, the Skiin Chestband Algorithms and Physical Health Assessment have been packaged separately, with their respective dashboards. Each of them are packaged in their individual docker containers and can be launched locally via Docker Desktop once the image has been built and loaded onto the target machine.

3.4.3 Skiin Algorithms Dashboard

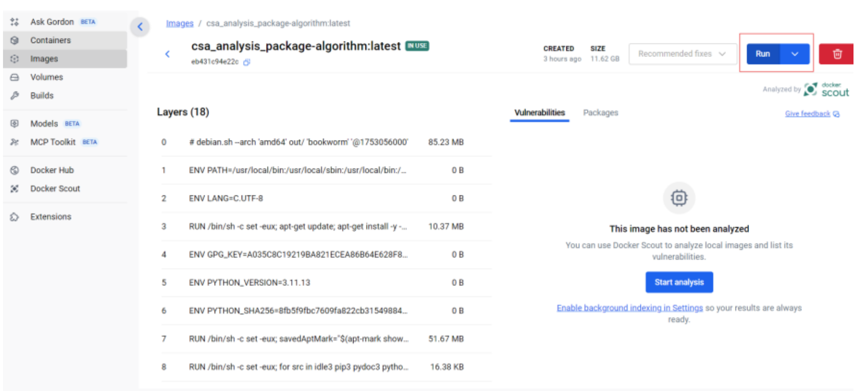
For starting up the service, the following steps are identified:

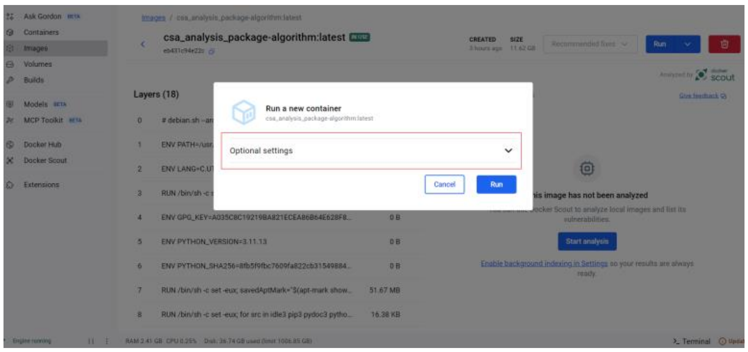
1. Docker Desktop is launched and the skiin-algorithms image is seen, as shown in the image below. There are two different images, one for running the analysis package and updating the database (csa-analysis-package-algorithm) and another for the dashboard (csa-analysis-package-streamlit)



Click on the analysis package container.

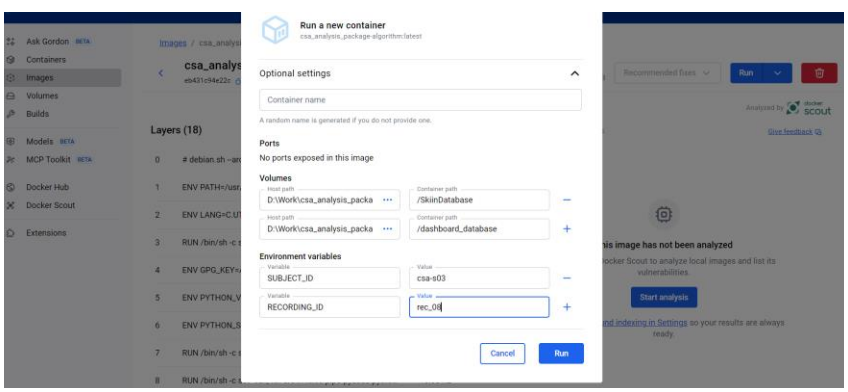
2. Next, press the run button as shown below, which will lead you to open new container window where you need to click optional settings.



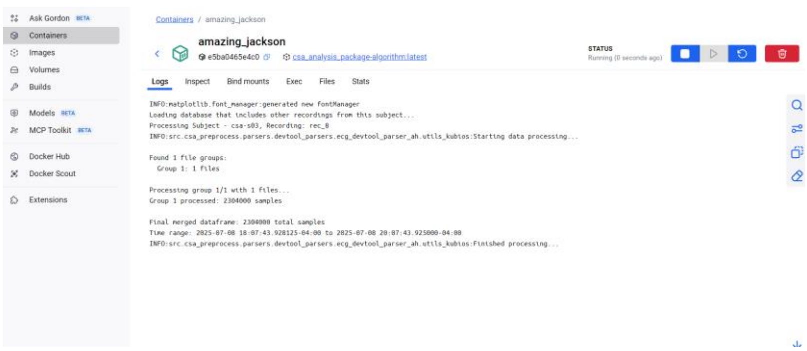


3. In the optional settings with will add two pieces of information. The Skiin data folder and dashboard database folder (in Volumes) and SUBJECT_ID and RECORDING_ID information (in environment variables). The following variables need to be added:
   - In Volumes:
     - First host and container path: (host) /path/to/SkiinDatabase/ where all the data is stored, and container (/SkiinDatabase)
     - Second host and container path: (host) /path/to/dashboard_database/ where all the data is stored, and container (/dashboard_database)
   - In Environment Variables:
     - Variable: SUBJECT_ID, Value: csa-s{XY} where XY is subject number
     - Variable: RECORDING_ID, Value: rec-{XY} where XY is recording number

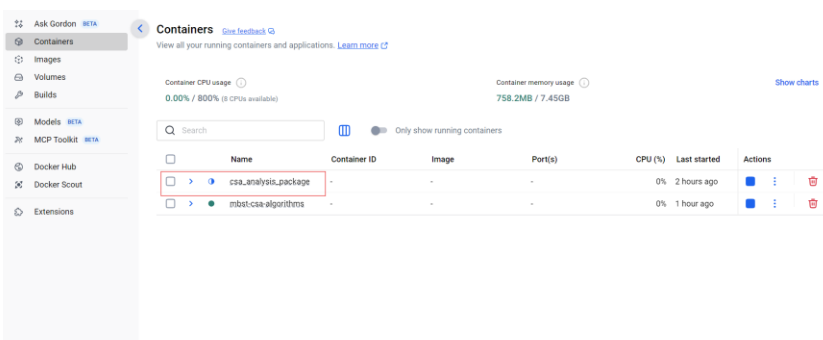
The completed information looks as follows:



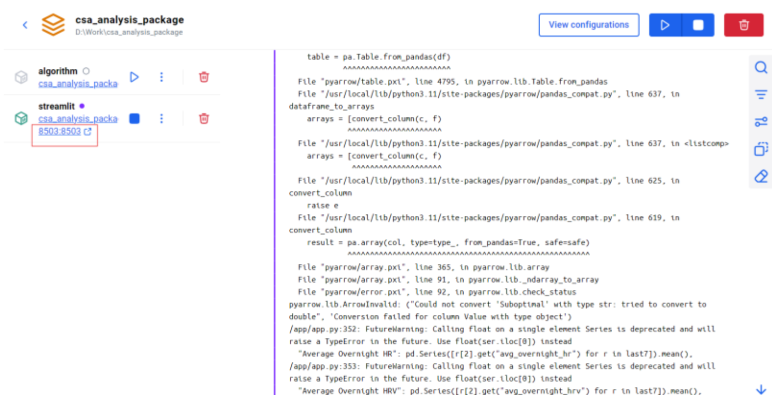
4. The analysis package will run while showing the processing information. Once the analysis is completed, the container will stop by itself. **Important:** Delete the container once its been run (delete button on top right corner)



5. Now you have to run the streamlist dashboard, go to the containers option and click on csa-analysis-package. Inside this container, you will see streamlit as active and click on the link as shown. This will open the dashboard for visualizing the data at its login page.



...



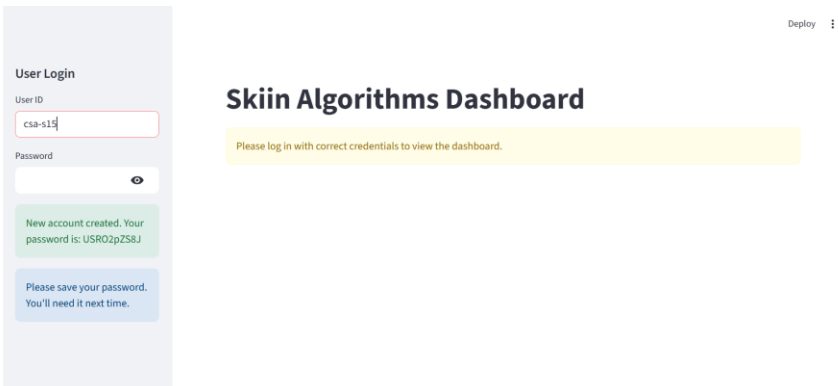
The workflow of the dashboard looks like this:

1. Log in Overview: The dashboard displays **user-specific information** based on the **User ID** entered at login. All analytics are loaded from a **local database**.

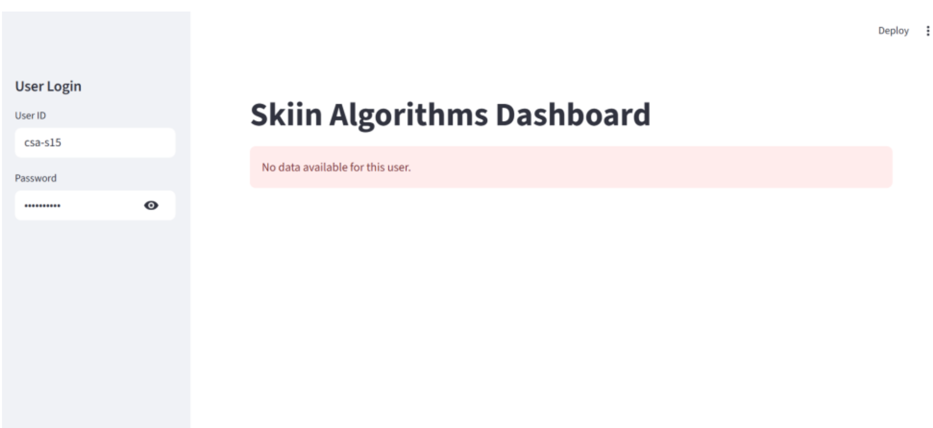
2. Naming Convention: For the dashboard to find your data, save it using the following naming convention for User IDs:
   csa-s01, csa-s02, csa-s03 and so on
   For each subject, we recommend saving recordings for each subject as: rec_01, rec_02 etc
   Use the exact pattern csa-s{**XY**} followed by a **two-digit** number. Keep letter case and hyphenation exactly as shown.

3. First-Time Login (Password Creation): When a new User ID logs in for the first time, the dashboard automatically assigns a password to that User ID and displays it once.
   - Open the dashboard.
   - Enter your User ID (e.g., csa-s01)
   - Select Log In.
   - On first use of a new User ID, a password will be generated and shown on screen (see Figure 1).
     Important: There is no "Forgot Password" or reset option. You must save this password securely to access the dashboard in the future.

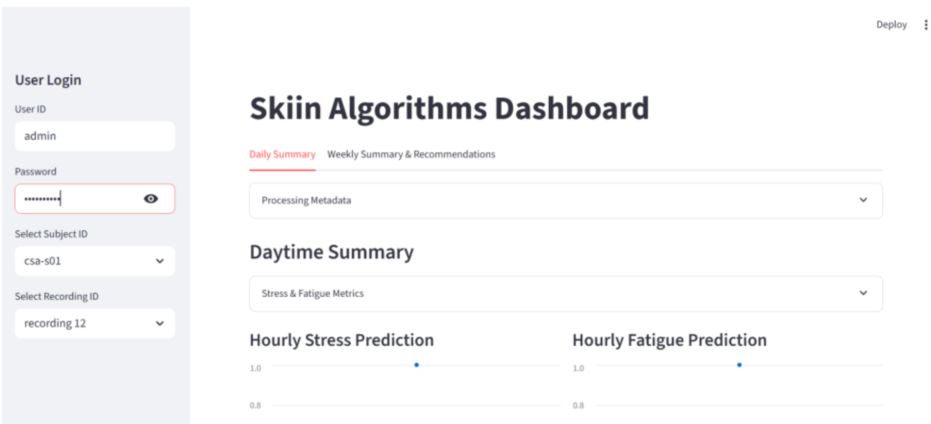
...



4. Returning User: For subsequent logins:
   (a) Enter the same User ID (e.g., csa-s01)
   (b) Enter the password that was assigned at first login
   (c) Select Log In
   (d) If the password is lost, you will need to register a new User ID and re-ingest/save data under that new ID following the naming convention.



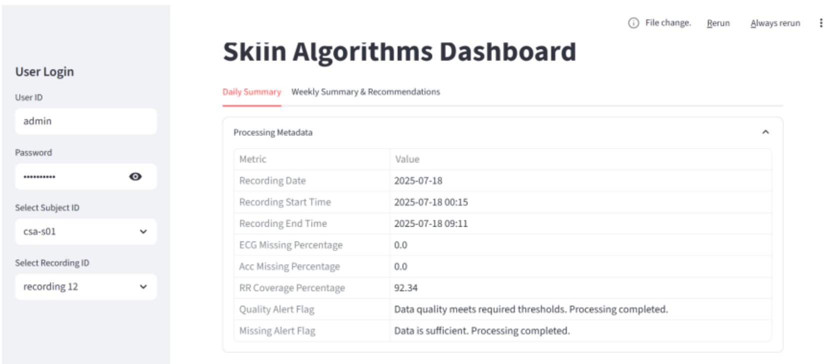
5. Admin Account:
   (a) The system includes an **administrator account** with elevated privileges.
   Username: admin
   Password: Provided **separately** by the system owner.



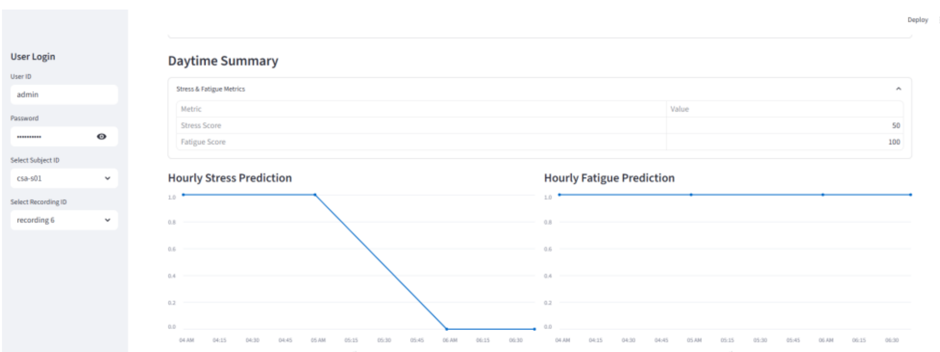
   (b) An administrator can **view all users and all associated data** stored in the local database.

6. Daytime Summary: The Daytime Summary tab opens with a quick-scan Metadata panel that audits the raw file before any analysis begins. It checks that both ECG and accelerometer streams are present, then runs two flags, Data Missing (enough minutes recorded) and **Data Quality** (signal integrity above threshold). If either flag fails, the panel marks the stream as "Insufficient," and processing stops. When both flags pass, the panel displays **"**Data sufficient, processing complete**"** and unlocks the metrics below.

Once the file is cleared, the tab reports two metrics: Overall Stress Score and Overall Fatigue Score, calculated from the full recording. Beneath these, an hourly trend chart traces stress and fatigue from morning through evening, allowing the user to visualize the variation during the day.



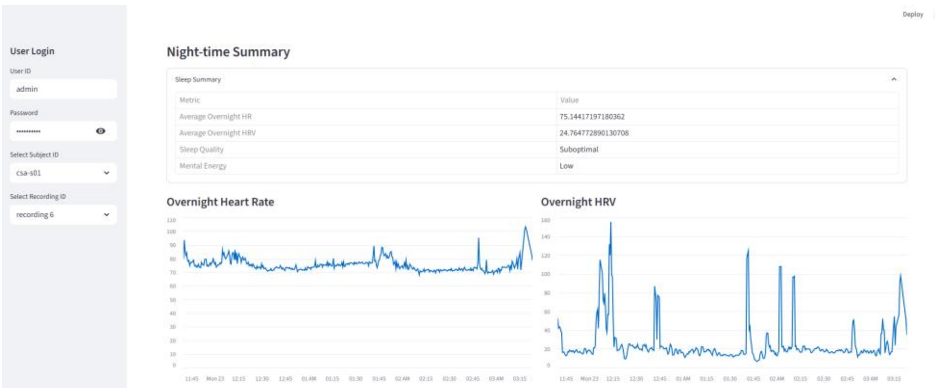
7. Night-time Summary: The Night-time Summary tab summarizes the events while the user slept. It

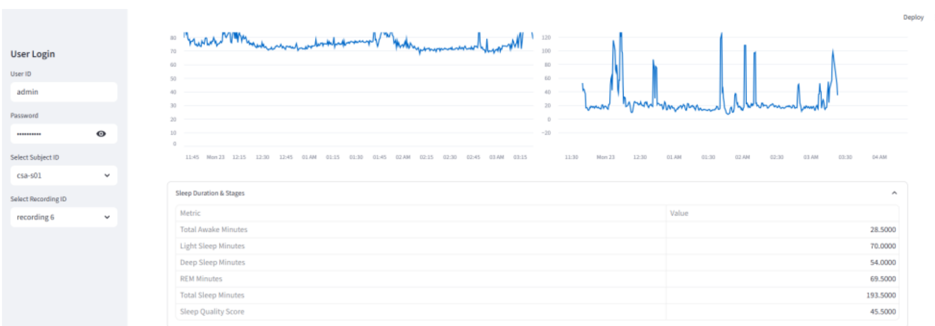


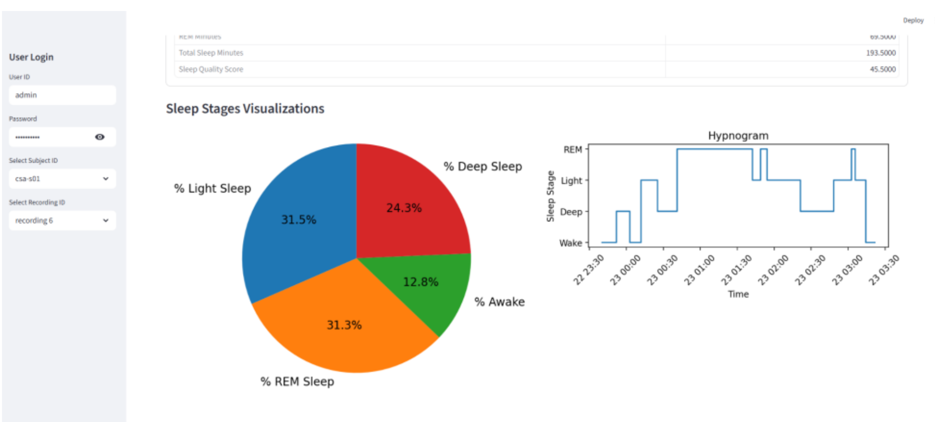
opens with a compact table listing the average overnight heart-rate (HR) and heart-rate–variability (HRV) values, then interprets them for the user: it determines whether Mental Energy and Sleep Quality are *high, low, optimal, or sub-optimal*. Directly beneath, a line chart plots HR and HRV hour-by-hour so any late-night spikes or steady improvements stand out.

A second table focuses on sleep architecture. It breaks the session into total minutes spent awake, in light sleep, deep sleep, and REM, and assigns an overall Sleep Quality Score. These same stage proportions are visualised in a pie chart just below. Finally, a timeline graph traces the flow of stages across the night, giving the user a clear picture of how often (and when) the user cycled between awake, light, deep, and REM periods.

...


Deploy
User Login
User ID
admin
Password
Select Subject ID
csa-s01
Select Recording ID
recording 6
Night-time Summary
Sleep Summary
Metric
Value
Average Overnight HR
75.14417197180362
Average Overnight HRV
24.764772890130708
Sleep Quality
Suboptimal
Mental Energy
Low
Overnight Heart Rate
Overnight HRV


Deploy
User Login
User ID
admin
Password
Select Subject ID
csa-s01
Select Recording ID
recording 6
Sleep Duration & Stages
Metric
Value
Total Awake Minutes
28.5000
Light Sleep Minutes
70.0000
Deep Sleep Minutes
54.0000
REM Minutes
69.5000
Total Sleep Minutes
193.5000
Sleep Quality Score
45.5000

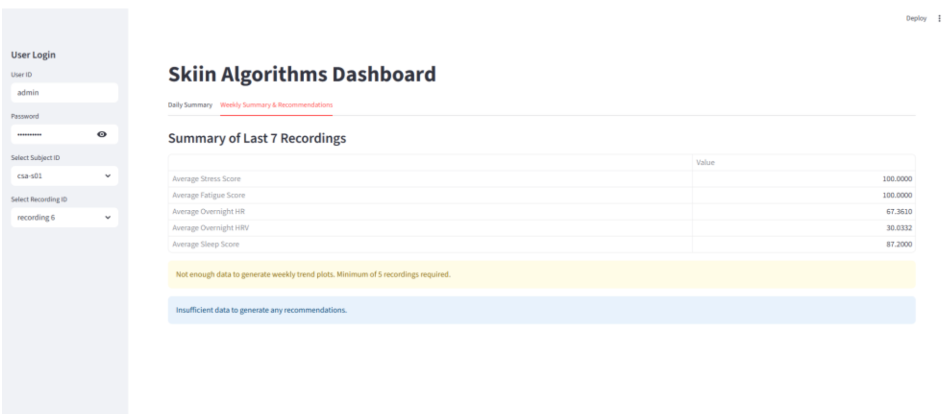

Deploy
User Login
User ID
admin
Password
Select Subject ID
csa-s01
Select Recording ID
recording 6
Total Sleep Minutes
193.5000
Sleep Quality Score
45.5000
Sleep Stages Visualizations
% Light Sleep
31.5%
% Deep Sleep
24.3%
12.8%
% Awake
31.3%
% REM Sleep
Hypnogram
REM
Light
Deep
Wake
Sleep Stage
Time

8. Weekly Summary and Recommendation

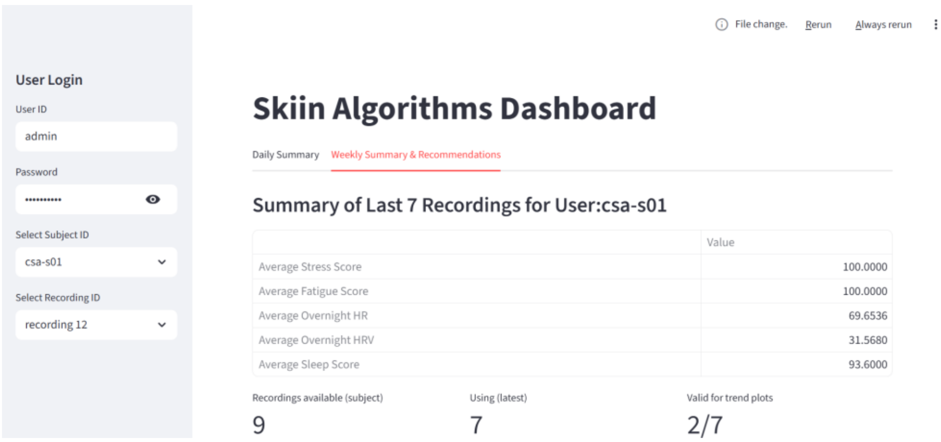
Commented [3]: @sarah.gulzar@myant.ca is it possible to show what the trend would look like with some dummy data of scores etc??

The Weekly Summary tab gives users a quick, seven-day overview of their recent activity and health metrics. When you open this tab, the dashboard automatically looks back at the last seven recordings stored for the selected user and counts how many are available. If at least five of those seven recordings are present, the page displays full trend plots, line or bar charts that track each metric day-by-day, alongside a table of weekly averages and personalized recommendations.

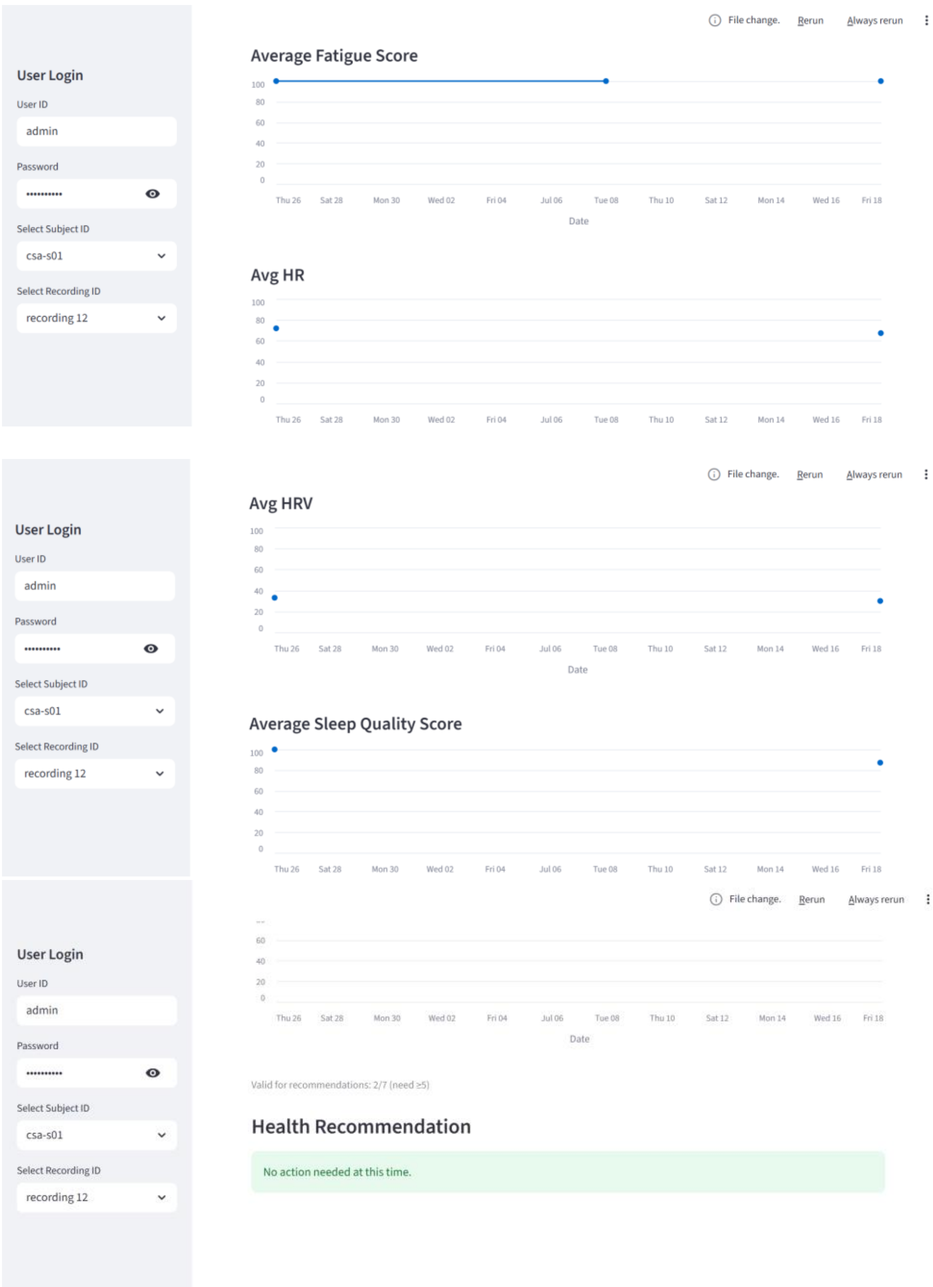
If fewer than five recordings are found, the dashboard switches to a data-light mode. In that case, the trend plots disappear and you will see only the summary table with averages across whatever recordings do exist. The banner at the top of the tab always tells you how many recordings were detected, for example, “6 / 7 recordings detected” enables plots, while “3 / 7 recordings detected” hides them, which helps determine whether missing data is the reason charts aren’t showing. When plots or metrics seem inaccurate, the first thing to verify is that the required recordings are named correctly (csa-s01, csa-s02, etc.) and have been processed without errors in the Metadata panel.



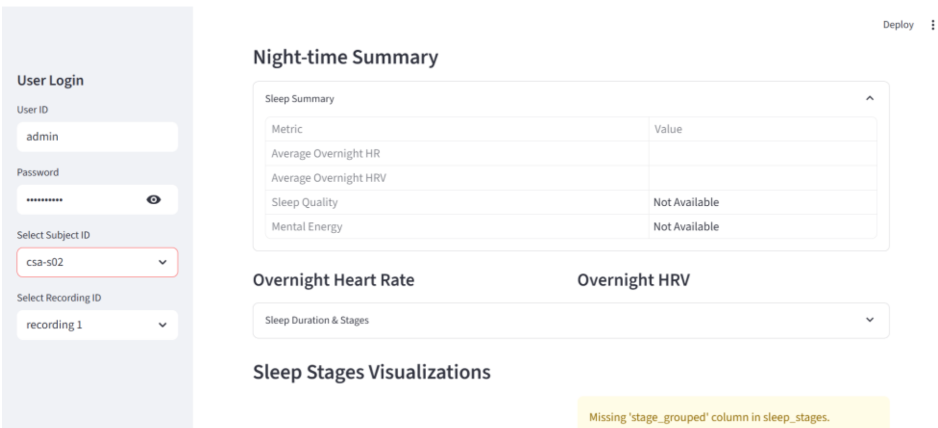
Weekly Trends shown with 2 recordings - will always default to 5:


File change.
Rerun
Always rerun
User Login
User ID
admin
Password
Select Subject ID
csa-s01
Select Recording ID
recording 12
Skiin Algorithms Dashboard
Daily Summary
Weekly Summary & Recommendations
Summary of Last 7 Recordings for User:csa-s01
Value
Average Stress Score 100.0000
Average Fatigue Score 100.0000
Average Overnight HR 69.6536
Average Overnight HRV 31.5680
Average Sleep Score 93.6000
Recordings available (subject)
9
Using (latest)
7
Valid for trend plots
2/7


File change.
Rerun
Always rerun
User Login
User ID
admin
Password
Select Subject ID
csa-s01
Select Recording ID
recording 12
Recordings available (subject)
9
Using (latest)
7
Valid for trend plots
2/7
Metric Trends Over Last 7 Recordings
Average Stress Score
100
80
60
40
20
0
Thu 26
Sat 28
Mon 30
Wed 02
Fri 04
Jul 06
Tue 08
Thu 10
Sat 12
Mon 14
Wed 16
Fri 18
Date


User Login
User ID
admin
Password
Select Subject ID
csa-s01
Select Recording ID
recording 12
File change. Rerun Always rerun
Average Fatigue Score
Thu 26 Sat 28 Mon 30 Wed 02 Fri 04 Jul 06 Tue 08 Thu 10 Sat 12 Mon 14 Wed 16 Fri 18
Date
Avg HR
Avg HRV
Average Sleep Quality Score
Valid for recommendations: 2/7 (need ≥5)
Health Recommendation
No action needed at this time.

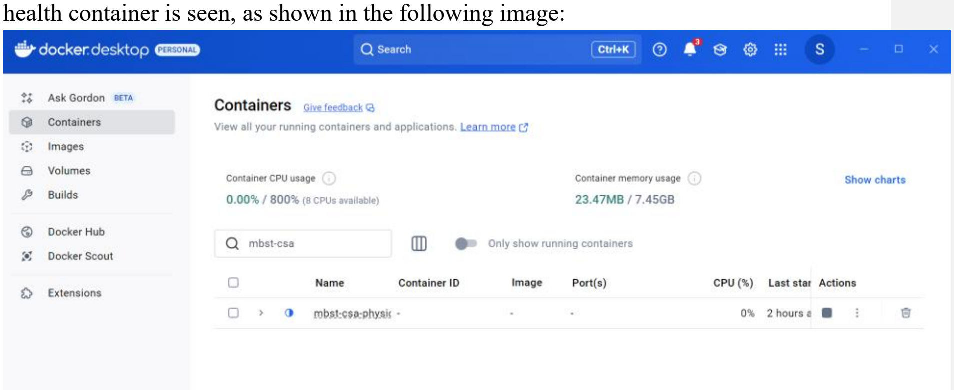
...



3.4.4 Physical Health Assessment Dashboard

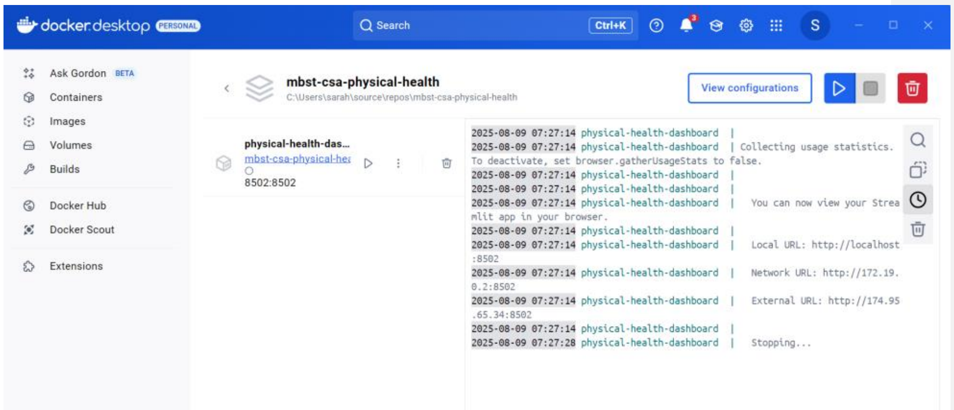
For starting up the service, the following steps are identified:

1. Docker Desktop is launched and inside "Containers" on the left-panel, mbst-csa-physical-health container is seen, as shown in the following image:

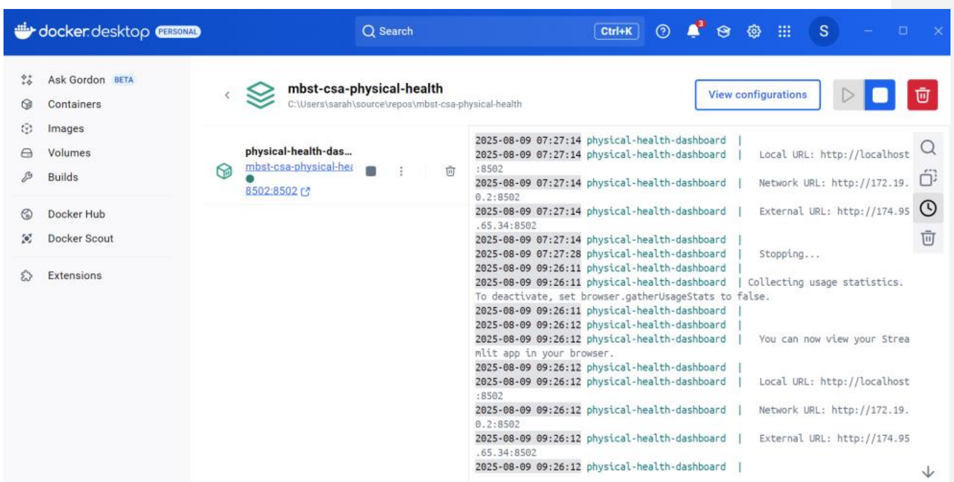


2. Once the container is clicked, the following window will appear:

...



3. The start/play button can be pressed which starts the algorithm and streamlit service. The following image shows the outputs seen on the terminal on the right-side once the service has been started:

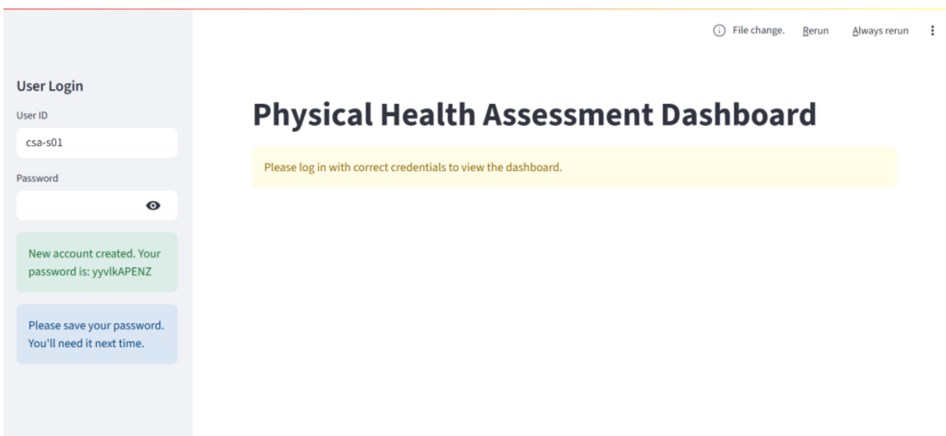


4. The local URL shown should be clicked which will launch the dashboard. The algorithm is run via the dashboard as shown below.

The workflow of the dashboard looks like this:

1. Log in Overview: The dashboard requires a User ID and password to be able to access it. Each user should pick a username after which the dashboard will assign a password. This password should be saved in a secure location as it will not be re-assigned for the same user ID. The user should always use the same credentials for logging in and saving the data.
   Important: There is no "Forgot Password" or reset option. You must save this password securely to access the dashboard in the future.

   The following image shows the dashboard view once opened:



2. After the User ID and password is entered, the left panel changes to show the user inputs required to run the algorithm service.

   The user is required to enter the following user inputs:
   1. User ID: This could be the same or different as the one entered for log in. This is required to save the users data to their local computer.
   2. Body weight in kilograms
   3. Body height in metres
   4. Gender

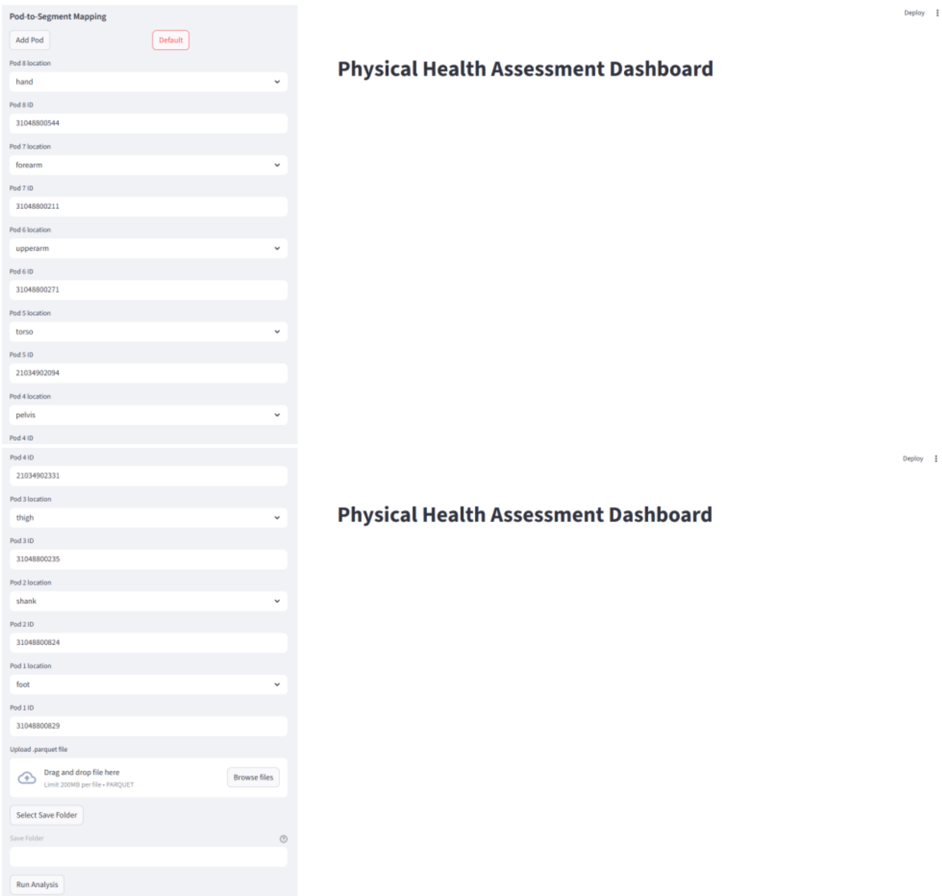
After the inputs have been entered, the user must select the pod type. For this algorithm service, the pod types supported are Skiin and SFLP. Skiin pods were used primarily for data collection for training data. For the data collection software built for collecting data from participants using Raspberry Pi, SFLP pods have been utilized. Thus, the user must always select SFLP for pod type for any new data collected. Once the pod type has been selected, there is a default mapping of pods that have been handed over for data collection. The "Default" button can be pressed to load this mapping onto the dashboard. The default mapping should be checked with the mapping provided in the following document: User Instructions for Physical Health Assessment Data Collection Using IMU Bands and Raspberry Pi

Commented [4]: updated @abhishek.tiwari@myant.ca

If the mapping is correct according to the document, the user can proceed to the next step of loading the data.

...


Body height (m)
1.75
Gender
male
Pod Type
SKIIN
SKIIN
SFLP
Add Pod
Default
Please add at least two pods before you can run the analysis.
Deploy
Physical Health Assessment Dashboard

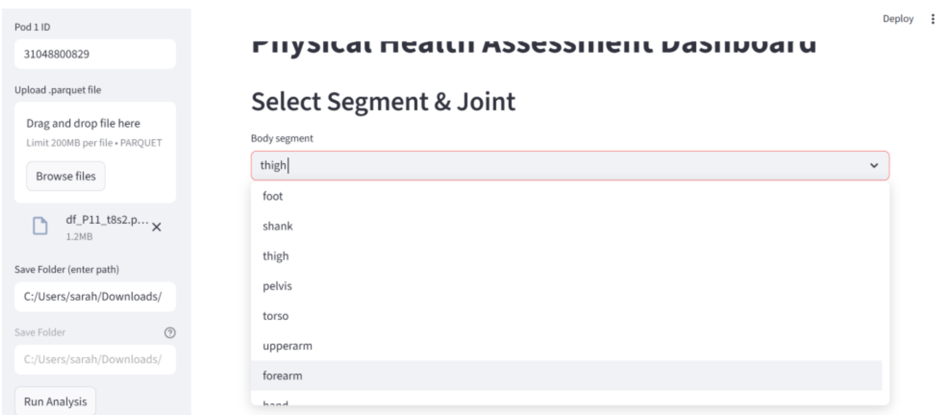


3. Data Processing: The user must type in the path for the location where the recording folder is saved on the user's laptop. The recording folder must contain all the CSV files from the data collection session (8 CSV files total, 2 per board). The CSV files must be placed in a single folder.

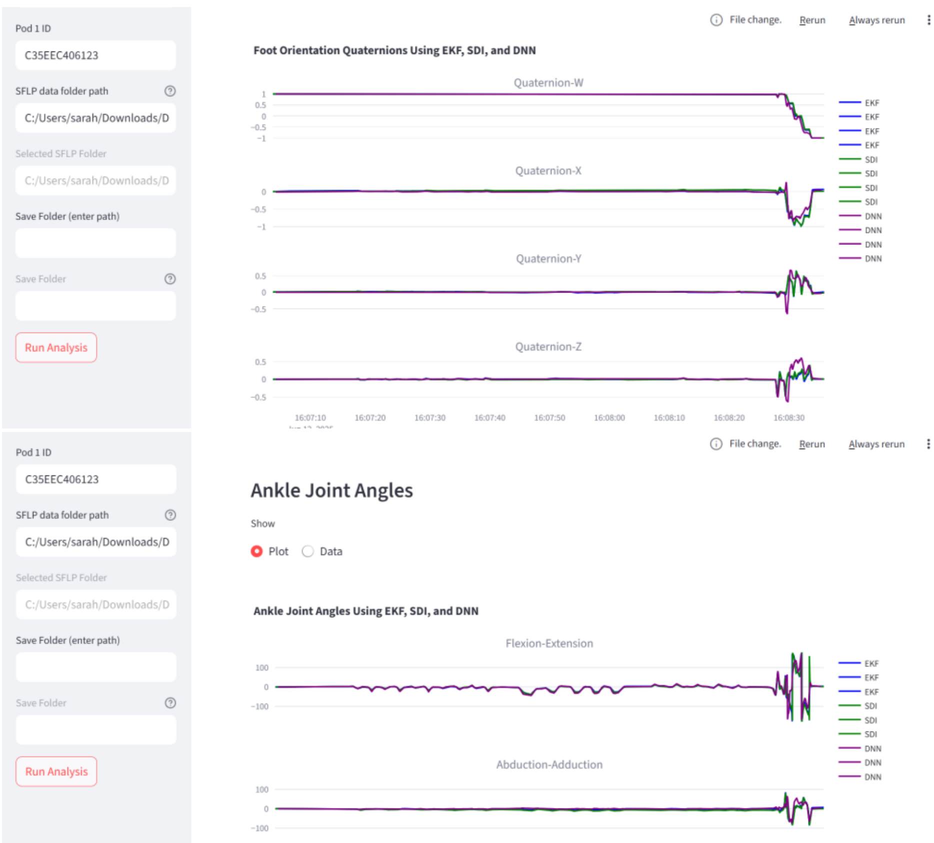
Commented [5]: updated this too: @abhishek.tiwari@myant.ca

Once the path has been typed in, the “Run Analysis” button can be pressed which runs the algorithm in the backend for the recorded data and displays the results.

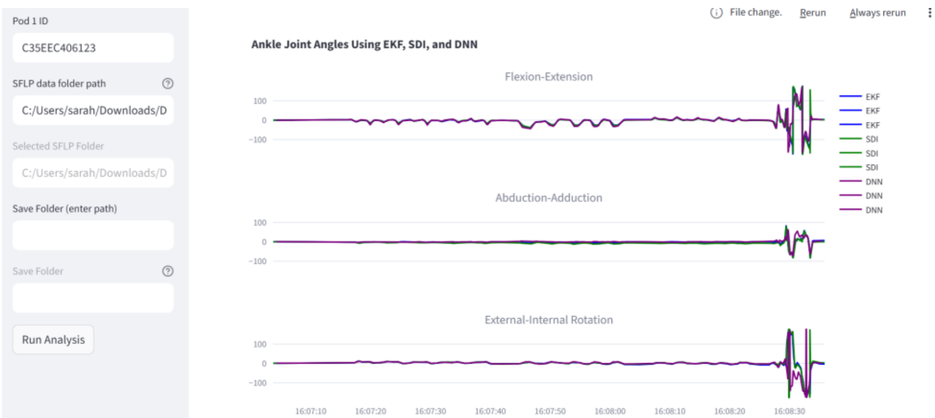
Pod 1 ID
31048800829
Upload .parquet file
Drag and drop file here
Limit 200MB per file • PARQUET
Browse files
df_P11_t8s2.p...
1.2MB
Save Folder (enter path)
C:/Users/sarah/Downloads/
Save Folder
C:/Users/sarah/Downloads/
Run Analysis
Deploy
Physical Health Assessment Dashboard
Select Segment & Joint
Body segment
thigh
Joint
hip
Thigh Orientation
Show
Plot
Data

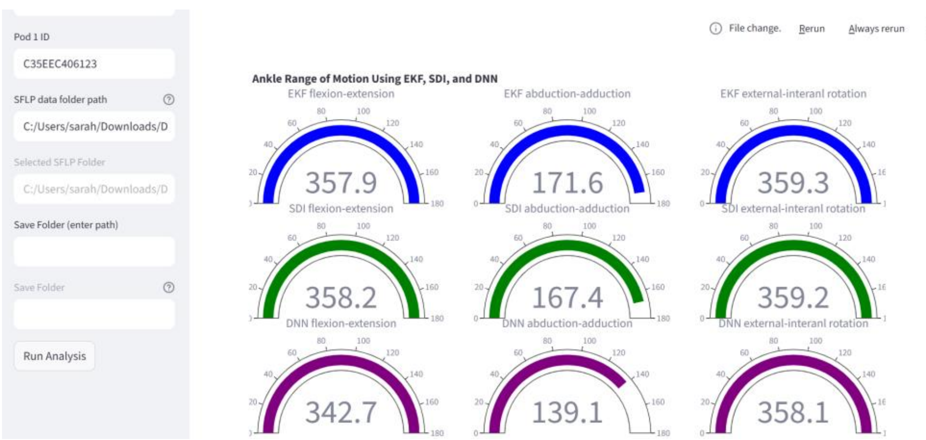


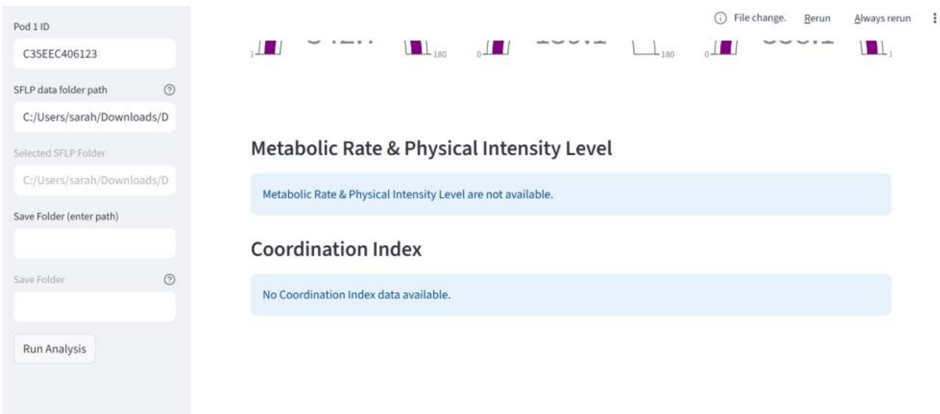
Pod 1 ID
31048800829
Upload .parquet file
Drag and drop file here
Limit 200MB per file • PARQUET
Browse files
df_P11_t8s2.p...
1.2MB
Save Folder (enter path)
C:/Users/sarah/Downloads/
Save Folder
C:/Users/sarah/Downloads/
Run Analysis
Deploy
Select Segment & Joint
Body segment
thigh
foot
shank
thigh
pelvis
torso
upperarm
forearm

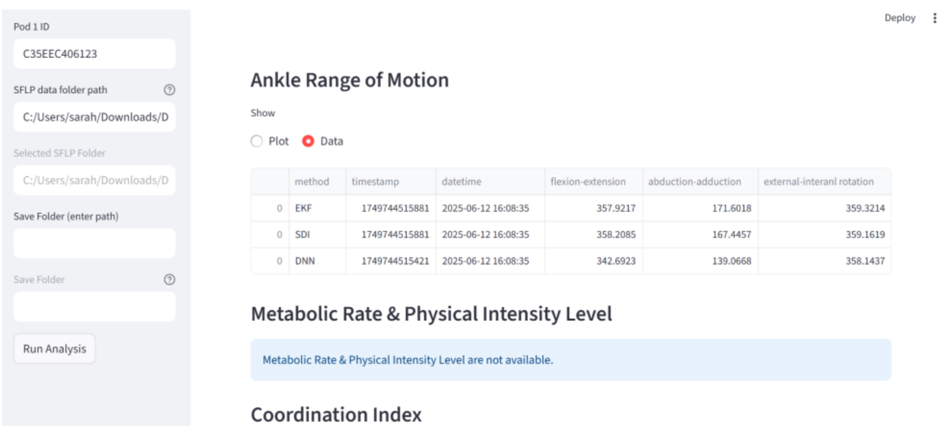

Pod 1 ID
C35EEC406123
SFLP data folder path
C:/Users/sarah/Downloads/D
Selected SFLP Folder
C:/Users/sarah/Downloads/D
Save Folder (enter path)
Save Folder
Run Analysis
File change.
Rerun
Always rerun
Foot Orientation Quaternions Using EKF, SDI, and DNN
Quaternion-W
Quaternion-X
Quaternion-Y
Quaternion-Z
EKF
SDI
DNN
16:07:10
16:07:20
16:07:30
16:07:40
16:07:50
16:08:00
16:08:10
16:08:20
16:08:30
Pod 1 ID
C35EEC406123
SFLP data folder path
C:/Users/sarah/Downloads/D
Selected SFLP Folder
C:/Users/sarah/Downloads/D
Save Folder (enter path)
Save Folder
Run Analysis
File change.
Rerun
Always rerun
Ankle Joint Angles
Show
Plot
Data
Ankle Joint Angles Using EKF, SDI, and DNN
Flexion-Extension
Abduction-Adduction
EKF
SDI
DNN

...


Pod 1 ID
C35EEC406123
SFLP data folder path
C:/Users/sarah/Downloads/D
Selected SFLP Folder
C:/Users/sarah/Downloads/D
Save Folder (enter path)
Save Folder
Run Analysis
File change.
Rerun
Always rerun
Ankle Joint Angles Using EKF, SDI, and DNN
Flexion-Extension
Abduction-Adduction
External-Internal Rotation
EKF
SDI
DNN
100
0
−100
16:07:10
16:07:20
16:07:30
16:07:40
16:07:50
16:08:00
16:08:10
16:08:20
16:08:30


Pod 1 ID
C35EEC406123
SFLP data folder path
C:/Users/sarah/Downloads/D
Selected SFLP Folder
C:/Users/sarah/Downloads/D
Save Folder (enter path)
Save Folder
Run Analysis
File change.
Rerun
Always rerun
Ankle Range of Motion Using EKF, SDI, and DNN
EKF flexion-extension
357.9
EKF abduction-adduction
171.6
EKF external-interanl rotation
359.3
SDI flexion-extension
358.2
SDI abduction-adduction
167.4
SDI external-interanl rotation
359.2
DNN flexion-extension
342.7
DNN abduction-adduction
139.1
DNN external-interanl rotation
358.1





3.5 Prototype Workflow

The overall system is a combination of three components:

- Data Collection Tools: Chestbands, Textile-integrated IMU bands, App and GUI
- Analysis Packages: End-to-end package for processing data and generating prediction outputs
- Dashboard: That visualizes the outputs generated by the analysis packages

The overall workflow of this system looks as follows:

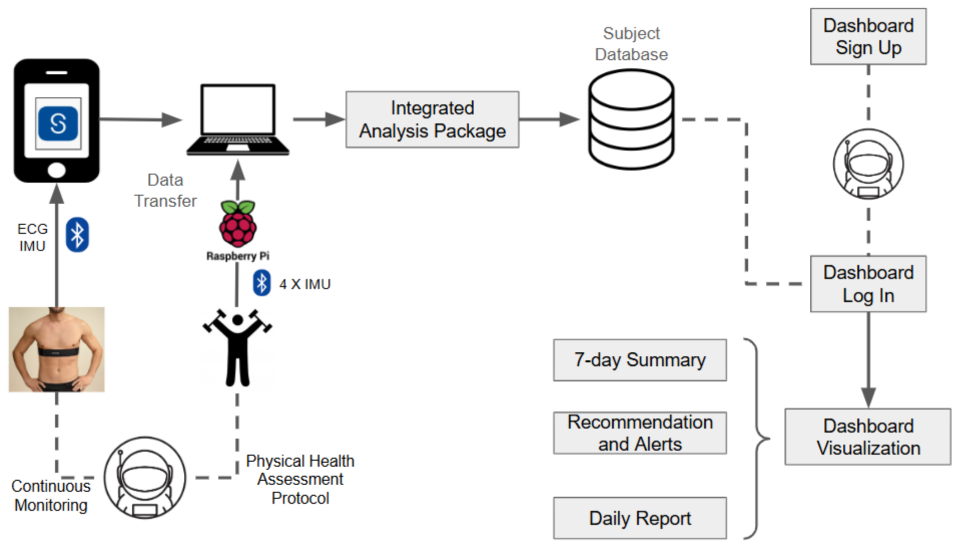


Figure 15: Diagram showing how the prototype components work together

First, the data is recorded using either the chestband with Skiin devtool app for ECG and IMU continuous monitoring or using textile integrated IMU bands and a Raspberry Pi for physical health assessment. Next, the data is manually transferred to a laptop in the appropriate subject and recording folder. A Docker container is used to run the analysis packages on this data that updates the data on the subject database. To access the dashboard, users first need to sign up and create an account to allow them access to their insights. After the account creation, they will be allowed to log-in and access the dashboard visualization.

Instructions for using the Devtool app:

1. **Connecting**: To begin recording data using the Devtool app, the first step is to connect the pod to the app. Upon opening the app, the screen shown below will appear. Tap **"Scan for Pods"**, then select the pod you wish to pair with.

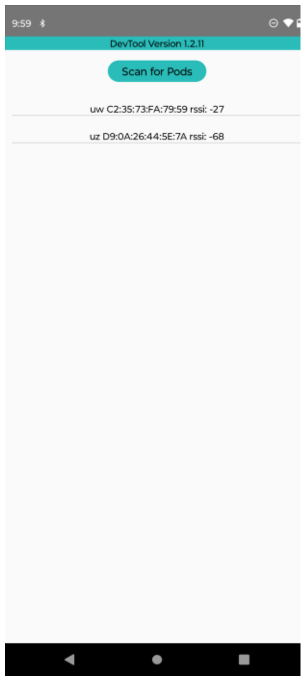


Figure 16: Opening screen of the Devtool app.

2. **Collecting**: When collecting the data using Devtool, it allows you to enter a name. This name will be included in the saved file's name, along with the pod's Bluetooth address and a timestamp. For example, in the Figure 17 below, "SHDT" is the default if the user doesn't enter any name, resulting in a file name that starts with **SHDT**, followed by the Bluetooth address and the timestamp. The files are saved in **CSV format**.
    a. For the CSA study, the recommended naming convention is to enter the **Subject ID** and the **Recording number** (e.g., CSA_ID01_rec1) to ensure clear and consistent file management.

...

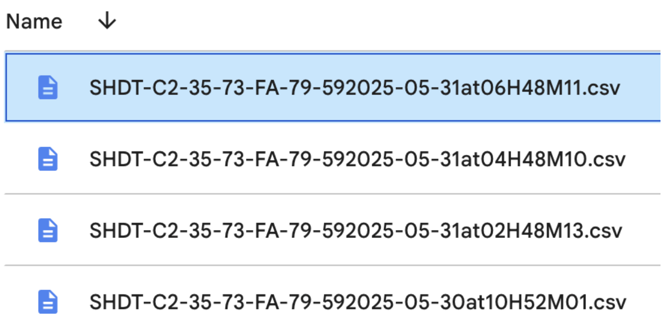


Figure 17: Name and format of recorded and saved files using the Devtool app.

b. In the Devtool app, as shown in the image below, there are two additional input fields: **Interval** and **Duration**.
   - **Duration** refers to the total length of the data collection session, measured in seconds. For example, if you plan to collect data for 10 hours, you should enter **36,000 seconds**.

   - **Interval** determines how often a new data file is created. It defines the duration of each individual file. For instance, if you set the interval to **3,600 seconds** (1 hour), the app will generate **10 separate files**, each covering one hour of data.

...

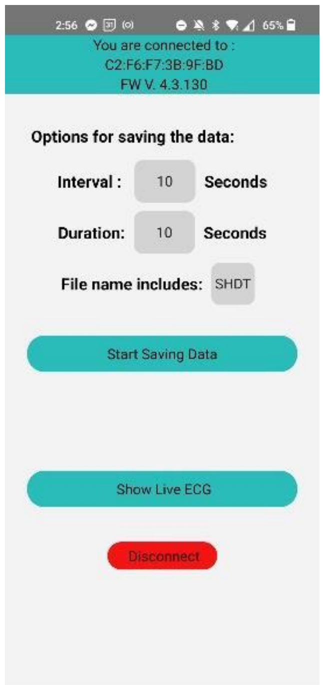


Figure 18: The devtool app set up prior to begin the data collection.

3. **Observing**: The data that's being collected can be observed in "Show Live ECG", only one channel at a time is shown however, there is an option to choose between the three channels of the ECG signal.
4. **Storing Location:** Data files are saved on Android's phone local storage, in files/internal storage/downloads or file/downloads.

4 Recommendations for Future Work

4.1 Mental Health and Wellness Modelling

For the dataset shared, there are still many modelling improvements that can be done:

- Using Longitudinal Modelling Strategies: Sleep issues have been linked to both increased stress and fatigue during day time. During our data collection, we have a large number of participants who were involved in both MARS and STAR data collection. As a result, we have both day time stress and fatigue as well as overnight sleep quality labels for these participants. Using either the subjective sleep labels for overnight physiology changes for stress/fatigue modelling the next day could further improve model performance.

- Demographic information for modelling: Both age and biological sex have a well established impact on human physiology. Therefore using demographic information as input features may further help improve stress and fatigue prediction.
- Combining sleep staging model with subjective sleep quality prediction: Current solution assess sleep in two different ways, i) by correlating subjective sleep quality with overnight physiological, ii) looking at sleep stages generated by using clinically labelled (and objective) data as ground truth. To further improve modelling, we can try integrating insights of objective sleep stage assessment for prediction of subjective sleep quality and mental energy.
- Increasing dataset size: Given the protocol information shared, additional data can be collected for modelling. This can allow for improved performance even with deep learning models that usually require larger dataset for improved performance. Personalized modelling approaches can also be used with longitudinal data on subjects.
- Using semi-supervised pre-training: This method was tested as a potential strategy for improving performance with little success. However, further exploration of this strategy is highly recommended. This is in part due to large improvements seen in different domains achieved by Transformer based models. Additionally, pre-trained foundation models (specially those based on wearable device data [2]) have opened the possibility to transfer knowledge from such architectures easily. A potential example of such a pre-trained model is to train 1 hour of HR series to predict the next 30 minutes.

4.2 Physical Health Assessment Protocol

Building upon the current framework for physical health assessment modeling, future work can focus on several key directions to enhance robustness, generalizability, and clinical utility. First, expansion of the dataset to include a broader demographic spectrum—encompassing variations in age, sex, body morphology, and functional capability—can be essential to improve the model's generalization across diverse populations. This can also enable stratified analyses that could inform personalized intervention strategies.

Second, the integration of video-based, markerless motion capture approaches can be investigated to enable hybrid sensor fusion between inertial and visual modalities. This fusion aims to leverage the complementary strengths of each modality: the drift-free global pose estimation capabilities of vision-based methods and the high-frequency local motion tracking provided by IMUs. Advanced fusion strategies, including Kalman filtering, optimization-based approaches, and deep learning-based sensor fusion architectures, can be explored for robust and scalable multi-modal kinematic estimation.

Third, joint angle estimation can be refined using optimization-based inverse kinematics techniques that incorporate anatomical constraints and musculoskeletal priors. This approach is expected to improve accuracy, particularly in tasks involving complex joint articulation or partial camera occlusion in case of video-IMU fusion.

Fourth, trend analysis of kinematic outputs over time can be implemented to detect clinically meaningful deviations or improvements in motor performance. These trends can form the basis for automated feedback generation. Generative AI models can be used to provide personalized recommendations and coaching cues. These may include corrective strategies, performance goals, and alerts for potential deterioration, thereby supporting longitudinal monitoring and self-managed rehabilitation.

These directions aim to bridge the translational gap between biomechanical modeling and practical, real-world applications in rehabilitation, sports performance, and occupational health.

4.3 Expanding Multimodal Monitoring: Cardiovascular, Neuromuscular, and Sleep Health in Deep Space Missions

Astronauts on long-duration missions face complex, interrelated physiological challenges, including cardiovascular deconditioning, muscle atrophy, disrupted sleep, and autonomic dysregulation, arising from fluid shifts, unloading of the musculoskeletal system, and altered circadian dynamics. Myant's technologies that combine continuous sensing, targeted intervention, and ease of deployment offer a promising solution to enable autonomous health maintenance and risk mitigation in space.

4.3.1 Enhancing Cardiovascular and Autonomic Monitoring

Building on the capabilities of the Skiin Chestband, which is already deployed in this project and approved by Health Canada as a Class II 3-lead Holter monitor, future iterations currently being prepared for regulatory licensure (i.e., Skiin Gen 2) incorporate cuffless blood pressure monitoring via photoplethysmography (PPG). This integration would enable continuous, non-invasive measurement of cardiovascular parameters, critical during fluid redistribution in microgravity, where autonomic imbalance and preload reduction increase the risk of arrhythmias and orthostatic intolerance.

The existing arrhythmia detection algorithm developed by Myant allows for real-time recognition of cardiac abnormalities, and its integration with continuous blood pressure and respiration sensing could create a rich physiological dataset for assessing cardiovascular stability, sleep recovery, and mission readiness.

4.3.2 Integrating Level 3 Sleep and Respiratory Monitoring

To complement these capabilities, BresoTec's Level 3 Home Sleep Apnea Testing (HSAT) platform (FDA and Health Canada licensed), recently acquired by Myant, offers a path to expand the current cardiorespiratory monitoring framework. BresoTec's system includes non-invasive, non-canula, airflow, respiratory effort, and pulse oximetry sensors, enabling detection of sleep-disordered breathing and ventilation abnormalities without the need for full polysomnography.

When integrated with Skiin's ECG and respiratory rate tracking, this platform could provide:

- Objective detection of apneic events and oxygen desaturation during sleep,
- Ventilatory profiling to capture breathing dynamics impacted by microgravity,
- Enhanced sleep staging and continuity analysis,
- Predictive models for fatigue and recovery using synchronized cardiopulmonary data.

Given the well-established links between sleep quality, autonomic balance, and mental resilience, this multi-sensor platform could play a key role in long-duration flight operations and post-mission rehabilitation.

4.3.3 Musculoskeletal Preservation through Textile Neurostimulation

The Skiin Neurostimulation System consists of textile-integrated garments equipped with dry-contact electrodes and embedded inertial sensors for neuromuscular electrical stimulation (NMES) and motion tracking. Current form factors include:

- Calf Sleeves for gastrocnemius and soleus stimulation,
- Upper and Lower Limb Wraps for quadriceps, hamstrings, and upper arm muscles,
- Back Bands for trunk and postural muscle activation.

Muscle atrophy and vascular deconditioning are well-established consequences of prolonged microgravity exposure. NMES is a validated countermeasure shown to preserve muscle mass, slow fiber-type transformation, and maintain functional capacity in both spaceflight analogs and clinical populations.

Beyond muscle preservation, targeted calf stimulation has been shown to enhance peripheral blood flow and venous return, effectively mimicking the muscle pump mechanism. This is especially relevant in microgravity, where the elimination of hydrostatic gradients leads to lower-limb underfilling and contributes to orthostatic intolerance upon re-entry. Several studies demonstrate that NMES in the lower limbs can mitigate venous pooling, support baroreceptor function, and improve cardiovascular resilience during re-acclimatization to gravity.

Each Skiin garment includes integrated inertial sensors for real-time activity monitoring, compliance tracking, and closed-loop stimulation control, allowing the system to adapt dynamically to user movement and physiological status.

Importantly, the Skiin Neurostimulation System is currently undergoing preparation for regulatory license application with Health Canada, signaling a commitment to safety, efficacy, and clinical deployment. The platform is also being evaluated in two ongoing clinical trials:

- One onsite trial conducted at Myant Inc. for safety and usability assessment,
- And a second trial underway at the University of Alberta Hospital, focusing on feasibility and therapeutic outcomes in a clinical care setting.

These trials will inform device performance across varied patient populations and use cases, supporting further refinement of protocols and regulatory submissions.

Collectively, the Skiin Neurostimulation System can offer a scalable, wearable countermeasure to maintain neuromuscular and vascular health in space, and provides a strong translational bridge to terrestrial applications in rehabilitation, aging, and chronic disease management.

5 The Technology Readiness and Risk Assessment (TRRA)

5.1 Overview

The project titled “*Using Smart-textiles and Virtual Reality for Artificial Intelligence Enabled Monitoring and Management of Sleep, Fatigue and Mental Health for Deep Space Exploration*” targeted the development and validation of a multimodal mental health monitoring and management system. The technology encompasses wearable smart textiles, pattern recognition algorithms for physiological data interpretation, machine learning pipelines, and an interactive dashboard for decision support. The system’s goal is to enable comprehensive and continuous astronaut health monitoring and mental wellness assessment during deep space missions. All key components of the system—wearable sensors, signal preprocessing, ML pipelines, dashboard visualization, and integrated prototype—advanced from Technology Readiness Level (TRL) 4 to TRL 6, per the CSA's definitions in Appendix A-1 of the SOW.

5.2 TRL Assessment by Subsystem

...

Subsystem/Component	Initial TRL	Final TRL	Evidence of Advancement
Smart-textile Chestband (Skiin)	TRL 4	TRL 6	Validated in user studies (STAR, MARS, PSG). Demonstrated data reliability and wearability in daily and overnight protocols under realistic use-case scenarios.
IMU integrated textile bands	TRL 4	TRL 6	Validated in physical health assessment study. Demonstrated data reliability and wearability in under realistic use-case scenarios of physical activities.
Data Preprocessing Pipeline	TRL 4	TRL 6	Modular software pipeline developed and used in multiple studies. Implemented automated artifact correction, filtering, data formatting, and data analysis. Integrated into an end-to-end prototype.
Pattern Recognition Algorithms	TRL 4	TRL 6	Machine learning models (classical and deep learning) were developed, validated using (cross-) validation and test sets. Models include stress, fatigue, sleep staging, and physical health estimation. Subject-independent and dependent models demonstrated effectiveness.
Dashboard & Visualization Interface	TRL 4	TRL 6	Fully functional dashboard displaying analysis results. Includes login system, role based access control, daily and 7-day reports, and recommendations. Integrated with backend pipeline and database.
Kaptics VR Integrated headset	TRL 5	TRL 6	The kaptics integrated headset was tested with the bWell algorithms to validate it for the intended use. Before this, the system has already been used by various psychologists, research labs and institutions for data collection and analysis.

...

bWell Neurofeedback training	TRL 4	TRL 6	The neurofeedback aspect was integrated with the existing bWell platform using measures developed by Kaptics and a usability study was conducted showing that indeed the device can be used for wellness exercises. In addition to this, the benefits of the exercises were evaluated using sleep HR and HRV metrics from the Skiin during
Integrated Prototype System	TRL 4	TRL 6	System integration of wearable hardware, preprocessing software, ML pipelines, and dashboard. Validated through multiple workflows across studies, ensuring system-level functionality in relevant environments.

<u>TRL 6 justification of final prototype:</u>
The developed system has been tested by an independent team (not involved in any of the model or software development aspects) at Myant. These system level tests allowed us to assess system functionality in non-simulated environments. We also provide a use case for the system in a hospital for preventing burnout in nurses below.

Deploying system in a hospital setting for managing wellness in nurses:
Nurses report a very high rate of burnout due to high workload and job demands. The proposed system can be deployed in such an environment by using the following steps:

1. Training for system use: While the overall system (Skiin textile integrated bands, the dashboard and bWell VR platform) are very intuitive to use. Training the nurses for correctly using the system can optimize quality and reduce errors. We aim to provide the users with a manual for system use to ease the training process. This training will also emphasize the importance of wearing Skiin garments continuously in order to allow access to all dashboard features such as sleep, stress and fatigue analysis.
2. Credential and Log In: All the nurses will be asked to create a unique Login ID and password that will protect their data and information from other users. However, a system admin can access this information.
3. Ready to Use: Once the training and login phase are complete, the nurses are free to make use of the system. They can use and visualize the dashboard outputs at any time. In order to prevent burnout, they will be recommended the bWell exercises based on their stress, fatigue and sleep quality outputs. Additionally, the system administrator will also be able to access their data and check if a given nurse has been recommended the use of the bWell device and look at their overall trends.

Integration with other hospital outputs: The ability of the system outputs to be integrated into other databases allow for:

- Smart shift planning system: Nurses usually have to do long shifts, however, if an individual is showing signs of high stress, and fatigue along with declining sleep quality, they can be given some time off or moved to shorter shifts to give time for recovery.
- Patient safety improvement: Individuals reporting high stress levels, can also be matched with shifts with lower risk of such situations. This may help with patient safety as higher stress and fatigue has often been correlated with more medical errors by hospital staff.
- Return to work and recovery monitoring: Nurses who are coming back after a mental health leave can be monitored using the system to ensure they are easily able to integrate into the workflow at the hospital.
- Dynamic team assignment: Nurse teams can be assigned to establish a balance between high and low stress individuals to optimize team performance and well being.

5.3 Risk Assessment

5.3.1 Technical Risks

Risk	Description	Mitigation Strategy	Residual Risk
Data Loss or Noise	Some datasets (e.g., PSG study) suffered from data quality issues due to noise or hardware failures.	Robust preprocessing and filtering pipeline was developed; data quality monitoring integrated.	Low
Model Overfitting in Small Datasets	MARS and STAR had small sample sizes, posing risk of overfitting in ML models.	Implemented three-stage data splitting (D_Fs, D_CV, D_Te) and Bayesian optimization for generalizable models.	Low
Sensor Integration Failures	Integration between sensors and the app backend may encounter firmware or compatibility issues.	Standardized data format (parquet/CSV), automated parsing pipeline, and clear sensor usage protocol.	Low

Dashboard Reliability	Risk of bugs or lag in dashboard affecting user experience.	Modular architecture, backend testing, and real-time monitoring of input data quality.	Low

5.3.2 Programmatic Risks

Risk	**Description**	**Mitigation Strategy**	**Residual Risk**
Recruitment Delays	STAR/PSG studies faced recruitment and retention challenges.	Expanded cohort in MARS and relaxed protocol strictness to increase retention.	Moderate
Data Privacy & Security	Handling sensitive physiological and psychological data requires robust protection.	Data de-identified before analysis, stored securely with limited access. Consent procedures enforced.	Low
Technology Transfer Risk	Integration into space health systems requires TRL >6 and operational validation.	Delivered a system capable of further integration with real-time astronaut health monitoring platforms.	Moderate

5.4 Validation Activities Supporting TRL 6

- **User Trials in Simulated Environments**: Four human-subject studies (STAR, MARS, PSG, Physical Assessment) conducted with wearable systems and monitoring protocols.
- **Prototyping and Field Testing**: Complete end-to-end pipeline was packaged in a Docker container and deployed to analyze new incoming data, enabling real-world functional testing of software and hardware.
- **Integration Verification**: Multiple components (sensor data acquisition, signal processing, ML inference, dashboard display) were validated in a connected workflow.

MYANT

...

- **Performance Evaluation**: Models assessed using standard ML metrics (e.g., Balanced Accuracy, Cohen's Kappa, Loss, etc.) with separate test sets to ensure reliability and generalizability. Deep learning models showed promising performance.

5.5 Summary

The developed health monitoring and decision-support system demonstrated strong progress in technical maturity. From initial laboratory-based component demonstrations (TRL 4), the system was evolved and tested in conditions simulating operational environments (TRL 6). The project addressed all technical and integration challenges expected at this stage of development and delivered a credible prototype ready for future validation (TRL 7+). Continued R&D is recommended for space qualification and incorporation into crew health operations.

6 References

[1] Adão Martins, Neusa R., et al. "Fatigue monitoring through wearables: a state-of-the-art review." *Frontiers in physiology* 12 (2021): 790292.

[2] Abbaspourazad, Salar, et al. "Large-scale training of foundation models for wearable biosignals." arXiv preprint arXiv:2312.05409 (2023).