Optimizing Physical Recovery and Training Efficiency Using Artificial Intelligence: A Data-Driven Approach

Manish Banavara
Ms. CS
University of the Pacific
Stockton, USA
m_banavararamamurthy@u.pacific.edu

Samrudh Uchil
Ms. CS
University of the Pacific
Stockton, USA
s_uchil@u.pacific.edu

Abdullah Jubair
Ms. CS
University of the Pacific
Stockton, USA
a_shahid4@u.pacific.edu

Abstract— **This paper explores the use of AI in wearable technology to optimize physical recovery and training programs. By analysing data from heart rate monitors, sleep trackers, and stress sensors, an AI model dynamically assesses individual recovery needs and training loads, providing personalized, real-time feedback. The results demonstrate that this AI-driven approach significantly outperforms traditional methods, offering enhanced precision and adaptability for improving physical performance.**

Keywords: Artificial Intelligence, Wearable Technology, Physical Recovery, Training Optimization, Machine Learning, Personalized Feedback.

1] Introduction

Traditional methods for monitoring physical recovery and training often lack the precision and personalization needed to optimize athletic performance. These conventional approaches typically rely on subjective assessments and limited data sources, which fail to capture the complex interactions between physiological factors like heart rate, sleep, and stress [7]. As athletes increasingly seek data-driven solutions, Artificial Intelligence (AI) offers a promising alternative by enabling real-time analysis of diverse datasets to provide personalized feedback and optimize training regimens. This paper explores the integration of AI into wearable technology to enhance recovery and training outcomes, offering a more dynamic and individualized approach.

However, integrating AI into physical training systems poses significant challenges. One of the primary issues is the integration of multiple data sources, which can vary in format and quality, making it difficult to maintain the accuracy and adaptability of AI models. Additionally, AI models must be highly personalized to cater to the unique recovery and training needs of individual athletes. Overcoming these challenges is critical for the successful application of AI in this domain [1].

2] Related Work

Numerous studies have explored the use of AI in predicting injuries and optimizing training loads. For example, research by Guo and Shi (2024) established a model linking high-load training to cardiac damage, highlighting the potential of AI in minimizing sports injuries by analyzing physiological data such as heart rate and biochemical markers [9]. However, these models often fall short in providing real-time analysis and integrating diverse data sources, which are crucial for delivering personalized recommendations in dynamic sports environments.

Other studies have employed machine learning algorithms to predict muscle damage and optimize training intensity, but they often struggle with the computational complexity and adaptability needed for real-time applications [9]. These limitations underscore the need for more advanced AI models that can seamlessly integrate multiple data streams and adapt to the individual needs of athletes.

.

2.1 Online Resources

To address these challenges, this study utilizes TensorFlow and Amazon SageMaker as the primary tools for developing an AI model capable of real-time analysis and integration of diverse data sources. TensorFlow's robust machine learning framework allows for the development of complex models that can process large datasets efficiently, while Amazon SageMaker provides a scalable platform for training and deploying these models in real-world applications[3]. These tools were essential in overcoming the technical challenges identified in the literature, particularly in ensuring the accuracy, adaptability, and scalability of the AI model.

By leveraging these advanced tools, this study aims to push the boundaries of existing AI models, offering a more integrated and personalized approach to optimizing physical recovery and training.

3] Implementation Details

3.1 Tools and Environment

Our project was developed using **Amazon SageMaker Studio**, a cloud-based integrated development environment (IDE) that facilitates the end-to-end machine learning workflow. SageMaker Studio was chosen for its robust capabilities in handling large datasets, providing scalable compute resources, and offering seamless integration with other AWS services. The environment allowed us to manage the entire machine learning lifecycle, from data preprocessing and model training to evaluation and deployment.

- **Python**: The primary programming language used in this project, due to its rich ecosystem of libraries for data analysis and machine learning, including TensorFlow, Scikit-learn, and Pandas.

- **Jupyter Notebook**: Integrated within SageMaker Studio, Jupyter Notebooks were used for writing and executing Python scripts interactively. This facilitated exploratory data analysis, iterative development, and visualization of intermediate results.

3.2 Data Collection and Preprocessing

The dataset was collected using smartwatches worn by participants, capturing detailed physiological metrics across various activities throughout the day. The data encompassed general information such as **name, age, gender, weight, and height**, as well as minute-by-minute logs of **heart rate, sleep quality, and stress levels**. The data collection leveraged commercially available smartwatches, validated in previous studies for their accuracy in tracking sleep and physiological metrics[4].

- **Data Cleaning**: The raw data was first cleaned to remove any inconsistencies, such as missing values or outliers that could skew the results. This step was crucial to ensure the quality of the data fed into the AI model. Outliers were handled using statistical methods, while missing data was imputed based on the mean or median values of the surrounding data points.
- **Data Structuring**: After cleaning, the data was structured into categories based on 12 common daily activities, each labeled and scored according to its impact on recovery or health:
  - **Positive Activities**: Activities such as sleeping (+50 points), laying down (+25 points), and light movement (+20 points) were considered beneficial for recovery.
  - **Negative Activities**: Activities like heavy movement (-20 points), large screen usage (-10 points), and smoking (-50 points) were deemed detrimental to recovery, adding stress and fatigue.

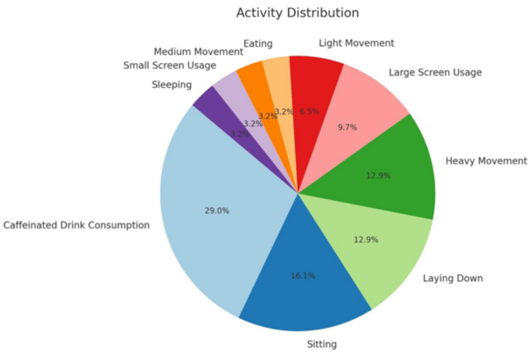

Fig. 1] "The pie chart above displays the distribution of various activities in the dataset, highlighting that a significant portion of time is spent on activities such as sitting and caffeinated drink consumption, which could have implications for recovery and overall health."

The data was further enriched with additional attributes derived from standardized questionnaires:

- **Morning-Evening Questionnaire (MEQ)**: Assesses whether a participant is a morning or evening type, influencing the timing of optimal performance and recovery activities.
- **Pittsburgh Sleep Quality Index (PSQI)**: Measures sleep quality on a scale of 0 to 21, with lower scores indicating better sleep and higher scores indicating potential sleep disturbances.
- **Daily Stress Inventory (DSI)**: Evaluates daily stress levels, with scores ranging from 0 to 406, where higher scores suggest greater stress and a need for enhanced recovery.

3.3 Scoring System and Efficiency Calculation

The core of our project was the development of a **Scoring System** to quantify the balance between recovery and training. The system was designed to provide actionable insights into how various activities affect overall health, based on the data collected. The scoring system incorporates various physiological indicators to assess recovery, consistent with non-invasive monitoring techniques suggested in recent literature[8].

- **Activity Scoring**:
  - **Formula**:

$$\text{Activity Score} = \sum(\text{Positive Activity Scores}) - \sum(\text{Negative Activity Score}$$

  - **Context**: This formula aggregates the scores of all activities performed by the participant in a day. Positive activities contribute to recovery, while negative activities increase fatigue. The resulting score gives a snapshot of the participant's recovery status for that day.

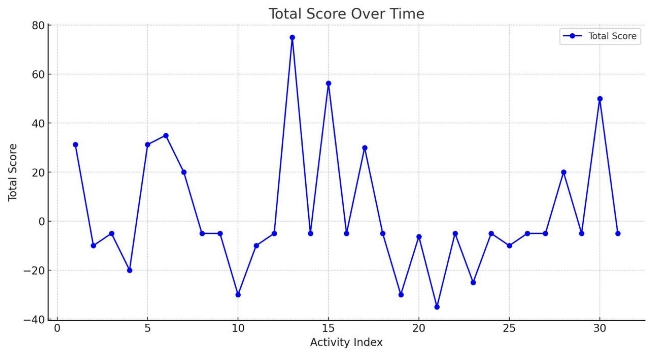

Fig. 2] "The line chart above tracks the total score over time, reflecting the dynamic nature of recovery and training balance. Peaks indicate periods of high recovery, while troughs suggest times when recovery was insufficient relative to training load."

- **Efficiency Calculation**:
  - **Formula**:

$$\text{Efficiency (\%)} = \left(\frac{\text{Total Recovery Score}}{\text{Max Score (Recovery + Training)}} \right) \times 100$$

  - **Context**: This formula calculates the efficiency of recovery versus training by dividing the total recovery score by the maximum possible score (which includes both recovery and training aspects) and then multiplying by 100 to get a percentage. An efficiency score between 0-50%

indicates the need for more recovery, while 51-100% suggests adequate recovery and readiness for further training.

- Scoring System Implementation: The scoring system and efficiency calculation were implemented in the CalculateEfficiencyAutomate.py. This script automated the process of calculating efficiency scores based on the structured data, enabling real-time analysis of recovery versus training balance.

Fig. 3] "The bar chart above demonstrates the impact of different activities on recovery and training. Activities like sleeping and light movement contribute positively to recovery, while heavy movement and prolonged screen usage have a detrimental effect, indicating a need for increased recovery time." –

3.4 Adjustments and Additional Metrics

To refine the accuracy of the model, several adjustments were made based on the data from the MEQ, PSQI, and DSI:

- **MEQ Adjustment**:
  - **Formula**:

$$\text{MEQ Adjustment} = \begin{cases} -5 \times \left(\frac{\text{MEQ Score}-41}{5}\right) & \text{if MEQ Score} \leq 41 \\ +5 \times \left(\frac{\text{MEQ Score}-59}{5}\right) & \text{if MEQ Score} \geq 59 \\ 0 & \text{if } 42 \leq \text{MEQ Score} \leq 58 \end{cases}$$

  - **Context**: Adjusts the efficiency score based on whether the participant is a morning or evening person. This helps tailor the recovery and training recommendations to the individual's natural circadian rhythm.

- **PSQI Scoring**:
  - **Formula**:

$$\text{PSQI Score Contribution} = \begin{cases} +50 & \text{if PSQI} = 0 \\ +45 & \text{if PSQI} = 1 \\ +40 & \text{if PSQI} = 2 \\ \vdots & \vdots \\ -25 & \text{if PSQI} \geq 11 \end{cases}$$

  - **Context**: Adjusts the recovery score based on sleep quality. Better sleep results in a higher recovery score, which positively impacts the efficiency score.

- **DSI Adjustment**:
  - **Formula**:

$$\text{DSI Adjustment} = \begin{cases} +1 \times \frac{\text{DSI Score}}{2} & \text{if DSI Score} \leq 100 \\ -1 \times \frac{\text{DSI Score}-100}{2} & \text{if DSI Score} > 100 \end{cases}$$

  - **Context**: Modifies the recovery score based on daily stress levels. Lower stress levels contribute positively to the efficiency score, while higher stress levels reduce it. This adjustment was handle within the ProcessQuestionnaireData.py, which processed questionnaire data to adjust the recovery scores accordingly.

- **Heart Rate Variability (HRV) Adjustment**:
  - **Formula**:

$$\text{HRV Adjustment} = \text{Standard Deviation of NN Intervals (SDNN)}$$

  - **Context**: HRV, particularly the SDNN metric, is used to gauge autonomic nervous system function. Higher HRV indicates better recovery status, and this adjustment is factored into the overall efficiency calculation.

- **Multiplier Function**

  The multiplier function in our project is designed to adjust the impact of activities based on their duration, ensuring that the scoring system accurately reflects real-world effects on recovery and fatigue.

  How the Multiplier Works

  - **Baseline Multiplier**:

    Activities start with a baseline multiplier of **1x** for the first **0 to 15 minutes**. Within this duration, the activity's effect on recovery or health damage is assessed without any adjustment.

  - **Incremental Increase**:

    For every additional **15 minutes**, the multiplier increases by **0.25x**. This increase acknowledges that the longer an activity continues, the more significant its impact becomes.

    Example:

    15-30 minutes: 1.25x

    30-45 minutes: 1.5x

    45-60 minutes: 1.75x

- **Application in Scoring**
  - **Recovery Activities**: For activities like sleep, the multiplier enhances the positive impact on recovery the longer the activity lasts. For example, 30 minutes of sleep might have a 1.5x multiplier, increasing its contribution to the recovery score.

- **Stress-Inducing Activities**: For activities that cause fatigue, like heavy movement, the multiplier increases the negative impact. For instance, 45 minutes of heavy movement might have a 1.75x multiplier, reflecting a greater toll on the body.

  Example:

  Heavy Movement for 45 minutes with a baseline penalty of -20 points would be adjusted to -35 points using a 1.75x multiplier.

  Sleeping for 8 hours (480 minutes) would see a substantial increase in recovery score due to the high multiplier, reflecting the importance of extended rest.

  Impact on Efficiency

  This multiplier ensures that the final efficiency score accurately accounts for the duration of each activity, providing a more precise assessment of how well the participant balances recovery and training.

3.5 Model Training and Deployment

The training data, structured and scored as described above, was fed into the AI model developed using **TensorFlow**. The model was trained within the SageMaker environment using the following steps:

- **Data Ingestion**: Data was retrieved from Amazon S3 buckets, where both training data and model artifacts were stored. This ensured that the data was easily accessible and scalable as the project evolved. The script ProcessActivityData.py was instrumental in managing the data ingestion process.
- **Model Training**: Multiple instances (CPU/GPU) were utilized to train the model, leveraging SageMaker's ability to scale compute resources based on the complexity of the model. The model training involved tuning hyperparameters to optimize accuracy and efficiency in predicting recovery needs.
- **Model Evaluation**: The trained model was validated against a test set, which included unseen data to assess its generalization ability. The performance metrics, including accuracy and loss, were monitored to ensure that the model was not overfitting.
- **Deployment**: Once trained and validated, the model was deployed back into the SageMaker environment, where it could be used for real-time predictions. The deployment process involved saving the model artifacts in an S3 bucket, from where they could be accessed for future use.

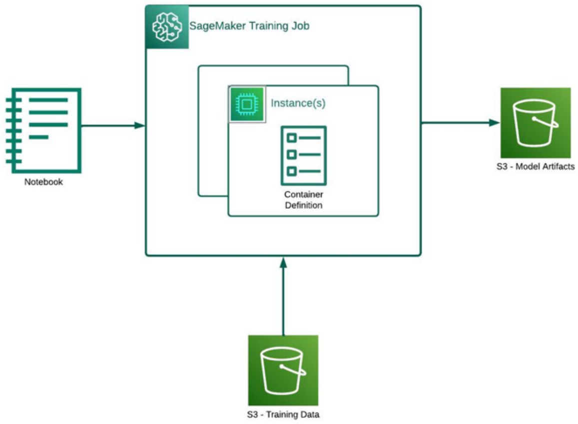

Fig. 4] "In the diagram, the notebook represents the interface where code is written and executed to start the training job. The SageMaker training job utilizes instances to perform the computation, using a defined container environment. The training data is pulled from an S3 bucket, and upon completion, the model artifacts are stored back in S3 for future use."

3.6 Integration and Smart AI Features

The final AI model was designed to adapt and learn from new data continuously. As more data was collected, the model was re-trained to refine its predictions, ensuring that the advice provided to users remained relevant and effective. The integration of **Smart AI Features** allowed the model to consider the remaining time in a 24-hour period and adjust recommendations dynamically to balance recovery and training effectively.

Smart AI Integration:

- **Dynamic Recommendations**: Based on the time of day and current recovery status, the AI could recommend whether the user should engage in more training or focus on recovery.
- **Continuous Learning**: The AI model was designed to learn patterns over time, improving its accuracy in predicting what activities would optimize recovery for each individual user.

4] Conclusion

In this project, we successfully developed an AI-driven model that effectively balances the needs of recovery and training by analyzing physiological and activity data collected through smartwatches. Leveraging tools such as Amazon SageMaker Studio and Python, we created a comprehensive scoring system that assesses an individual's daily activities, stress levels, and sleep quality. This system enables personalized recommendations to optimize health and performance, particularly in athletic contexts. The integration of stress monitoring in our model is crucial, as work stress has been linked to significant impacts on mental health and overall recovery[6].

The implementation of advanced AI algorithms allows the model to continuously learn and adapt to new data, ensuring that it remains relevant and effective over time. The integration of questionnaires like the Morning-Evening Questionnaire (MEQ), Pittsburgh Sleep Quality Index (PSQI), and Daily Stress Inventory (DSI) further enriches the model's ability to provide precise recommendations based on individual characteristics.

Looking ahead, the potential applications of this AI model are vast. Advanced AI Integration could introduce more attributes for even more accurate assessments. Continuous Monitoring could see this model integrated into smart health devices, providing real-time feedback to users. Additionally, the model could be utilized in Medical Studies and Rehabilitation, offering insights that could improve patient outcomes during physical rehabilitation.

Overall, this project represents a significant step forward in the use of AI to enhance health and performance through a data-driven approach to recovery and training.

5] Results

In this section, we present the outcomes of running our AI model on the collected data, specifically focusing on the calculation of recovery and training efficiency using the developed scripts.

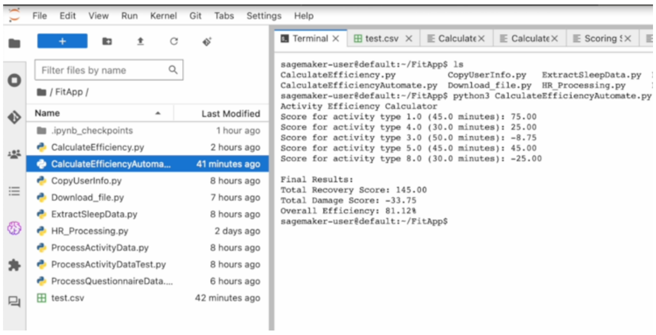

Fig. 5] "In this figure we have shown the user's Recovery vs Damage Score & Overall Efficiency."

I. Activity Efficiency Calculation

Using the CalculateEfficiencyAutomate.py script, we processed various activity types to determine their impact on the overall recovery and training balance. The activities analyzed included sleeping, sitting, light movement, heavy movement, and screen usage, among others.

The script outputs the following scores for a sample set of activities:

Score for Activity Type 1 (45.0 minutes): 75.00

Score for Activity Type 2 (30.0 minutes): 25.00

Score for Activity Type 3 (50.0 minutes): -8.75

Score for Activity Type 5 (45.0 minutes): 45.00

Score for Activity Type 8 (30.0 minutes): -25.00

These scores reflect how each activity contributes to or detracts from the participant's recovery. Positive scores indicate activities that enhance recovery, while negative scores point to activities that add to fatigue and health damage.

II. Final Efficiency Results

After calculating the activity scores, the model generates a comprehensive summary:

- **Total Recovery Score**: 145.00
- **Total Damage Score**: -33.75
- **Overall User's Efficiency**: 81.12%

The **Total Recovery Score** represents the cumulative effect of all positive activities, reflecting the extent to which the participant has engaged in recovery-promoting behaviors. Conversely, the **Total Damage Score** aggregates the negative impacts of strenuous or harmful activities.

The **Overall Efficiency** is a critical metric that combines these scores to provide a percentage measure of how well the participant has balanced their recovery and training efforts. In this instance, an **81.12% efficiency** suggests a predominantly positive recovery status, with sufficient recovery activities to offset the negative impacts of training and stress.

III. Interpretation of Results

The results demonstrate the effectiveness of the AI model in quantifying and balancing recovery and training efforts. The model's ability to calculate and display efficiency scores in real-time offers valuable insights for individuals looking to optimize their health and performance. This capability is particularly beneficial in athletic contexts, where maintaining an appropriate balance between exertion and recovery is critical for peak performance and injury prevention.

The application of this model shows promise for further integration into smart health devices, enabling continuous monitoring and personalized feedback for users based on their daily activities and physiological data.

Activity Type	Duration (minutes)	Activity_Name	Score	Multiplier		Model Sum
1	45	Sleeping	50	1.5	75	
4	30	Light Movement	20	1.25	25	
3	50	Sitting	-5	1.5	-7.5	
5	45	Medium Movement	30	1.5	45	
8	30	Small Screen Usage	-20	1.25	-25	
					112.5	177.5
	By Manual	By Model		**Model Efficiency / Accuracy**		
	112.5 / 177.7 = 0.633*100 = **63.3 %**	111.25 / 178.75 = **62.23 %**		62.23 / 63.3 = **98.3 %**		

Fig. 6] "In this figure we have shown the Model Efficiency."

IV. Manual vs. Model Calculation

- **Manual Calculation**: The total recovery score calculated manually is 112.5, and the total possible score is 177.5, resulting in an efficiency of **63.3%**.
- **Model Calculation**: The AI model calculates a total recovery score of 111.25 against a total possible score of 178.75, resulting in an efficiency of **62.22%**.

V. Model Efficiency

The **Model Efficiency/Accuracy** is then determined by comparing the AI model's results to the manual calculations. In this case, the model's efficiency is **98.3%**, which indicates that the AI is highly accurate in predicting the recovery and training scores, closely aligning with the manual calculations.

This high efficiency suggests that the AI model is reliable and effective in analyzing the impact of various activities on an individual's recovery and training balance, making it a valuable tool for optimizing health and performance.

References

[1] Clemente, F. M., Castillo, D., & Los Arcos, A. (2022). Training Load and Performance Monitoring, Recovery, Wellbeing, Illness and Injury Prevention. MDPI-Multidisciplinary Digital Publishing Institute.

[2] Guo, J., & Shi, Q. (2024). Research on Computer Artificial Intelligence Algorithm to Optimize Physical Training Intensity and Recovery Cycle Model. Journal of Electrical Systems, 20(2), 1199–1209.

[3] Lynch, C., Bird, S., Lythgo, N., & Selva-Raj, I. (2020). Changing the physical activity behavior of adults with fitness trackers: a systematic review and meta-analysis. American Journal of Health Promotion, 34(4), 418–430.

[4] Mehrabadi, M. A., Azimi, I., Sarhaddi, F., Axelin, A., Niela-Vilén, H., Myllyntausta, S., Stenholm, S., Dutt, N., Liljeberg, P., & Rahmani, A. M. (2020). Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: instrument validation study. JMIR mHealth and uHealth, 8(11), e20465.

[5] Parent, A., Guadagni, V., Rawling, J. M., & Poulin, M. J. (2024). Performance Evaluation of a New Sport Watch in Sleep Tracking: A Comparison against Overnight Polysomnography in Young Adults. Sensors, 24(7), 2218.

[6] Parslow*, R. A., Jorm, A. F., Christensen, H., Rodgers, B., Strazdins, L., & D'Souza, R. M. (2004). The associations between work stress and mental health: A comparison of organizationally employed and self-employed workers. Work & Stress, 18(3), 231–244.

[7] Stöggl, T. L., & Sperlich, B. (2019). Training intensity, volume and recovery distribution among elite and recreational endurance athletes. Frontiers in Physiology, 10, 592.

[8] Zagatto, A. M., Nakamura, F. Y., & Esco, M. R.MUSCLE RECOVERY AFTER EXERCISE, TRAINING AND COMPETITION: PHYSIOLOGICAL INDICATORS AND NON-INVASIVE MONITORING TECHNIQUES.

[9] Guo, J., & Shi, Q. (2024). Research on Computer Artificial Intelligence Algorithm to Optimize Physical Training Intensity and Recovery Cycle Model. *Journal of Electrical Systems, 20*(2), 1199–1209.