

Lactate Threshold Prediction

AI-Based Performance Data Analysis for Training Optimization

Students



Marcel Schubert



Musa Rochi

Introduction: Blood lactate responses are collected during graded exercise testing and are used to assess physical fitness. In particular, they help estimate lactate thresholds. Lactate thresholds indicate endurance capacity and are used to define training zones. These training zones are important for avoiding under- or overtraining. At our project partner, Davos Sports and Health, these thresholds are currently adjusted manually by experts, which limits standardization and requires additional recalculation effort. To address these limitations, this project aims to develop an AI model that predicts the aerobic threshold based on data characteristics and to implement a demonstrator that enables practical application.

Approach: We analyze graded exercise test data from young athletes to model aerobic threshold definitions using physiological and machine learning approaches. The dataset comprises 510 tests from 328 participants (median age 24 years, 62 percent male), collected using treadmill and cycle ergometer protocols and including approximately 430 anthropometric, physiological, and performance related variables. After manual data cleaning and application of a standardized preprocessing pipeline with automated validation, 500 tests are retained for analysis.

Exploratory data analysis is conducted to assess data quality, distributions, and feature correlations. Physiologically informed feature engineering reduces the feature space to 32 dimensions, including parameters derived from exponential lactate intensity curves, physiologically meaningful points along these curves, and two independent approximations of the aerobic threshold. To mitigate modality specific effects, exercise intensity is normalized to individual maximum intensity.

We train and evaluate multiple statistical and machine learning models such as linear regression, ElasticNet, principal component regression, and partial least squares regression. More complex models include k nearest neighbors, support vector regression, probabilistic Gaussian process regression, multilayer perceptron, and XGBoost. The models are used to predict heart rate, lactate, and intensity based aerobic threshold definitions.

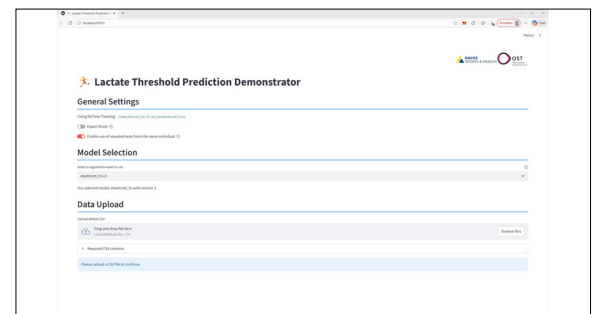
The models are trained using a stratified 70 / 30 train test split with ten fold cross validation. Performance is assessed using mean squared error, root mean squared error, and coefficient of determination, and feature contributions are analyzed using SHAP values.

Result: We find that simple models achieve strong predictive performance, while more complex approaches provide no substantial improvement under our data conditions. Heart-rate-based targets

are considerably easier to predict than lactate- and intensity-based targets. The best performing ElasticNet model achieves an RMSE of 5.102 bpm and an R^2 of 0.869 and predicts across both modalities. We implemented an application deployable via Docker, supporting automated predictions, interactive visualizations, clinician-driven adjustments, PDF report generation, and an expert analysis mode.

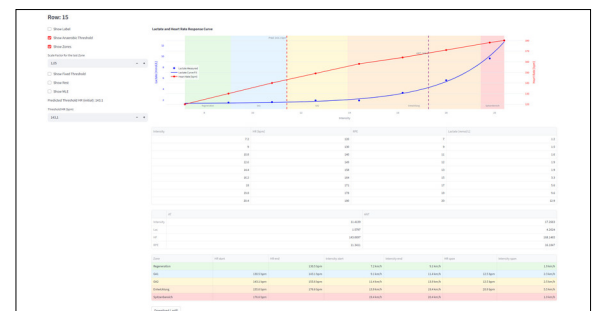
Lactate Threshold Demonstrator

Own presentation



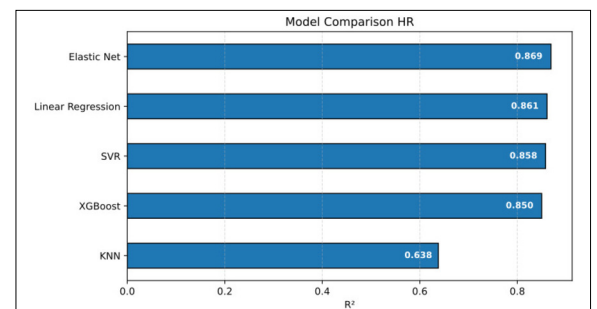
Example of an Aerobic Threshold Prediction

Own presentation



R^2 Model Comparison of the HR Label

Own presentation



Advisors

Prof. Dr. Christoph Gebhardt, Robert Schöch

Subject Area

Artificial Intelligence, Data Science

Project Partner

Davos Sports & Health, 7270 Davos Platz, Graubünden

