Too Little Data? Teach Machines to Detect Flow Anyway.

Bachelor/Master project "Information Systems"

Table of Contents

Prof. Dr. Mario Nadj
University of Duisburg-Essen
Faculty of Computer Science

Chair and Website
Information Systems and AI in Marketing
ai.ris.uni-due.de

Address Campus Essen
Universitätsstraße 9
D-45141 Essen

1: Project Overview

Background: Machine learning enables the classification of flow states using physiological data from wearable devices. However, research progress is limited by data scarcity. One possible solution approach is transfer learning.

Project Goal: The goal of this project is to identify relevant datasets for flow state classification from flow-related and flow-specific contexts. Based on the identified datasets, a machine learning pipeline including transfer learning methods is developed to classify flow states based on physiological data.

Highlights: Gaining practical insights into applied machine learning problems; close collaboration with the supervising chair, providing guidance and support throughout the project to achieve its objectives; and the development of a complete machine learning pipeline suitable for inclusion in a GitHub portfolio.

2: Project Background

In recent years, machine learning (ML) has been increasingly applied across a wide range of domains. This rapid adoption is also reflected in market projections. The ML market is expected to grow at an annual growth rate of 31.72 % between 2026 and 2031, reaching a projected volume of USD 445.27 billion by 2031 (Statista, 2026). These advances have opened up new possibilities in applications that rely on complex pattern detection, such as the recognition of the flow state through physiological data. However, a major challenge for data scientists when developing ML models, in particular more advanced deep learning models, remains data scarcity (Alzubaidi, 2023).

Researchers focusing on flow state classification are likewise affected by this challenge. Flow, defined as the feeling of being completely absorbed into an activity, is regarded as a desirable state due to its association with improved performance and well-being (Peifer, 2020). Consequently, recent research approaches focus on the development of tools capable of detecting the flow state with wearable devices, such as ECG chest belts, to support individuals in entering or maintaining this state. Researchers have already accomplished important initial steps toward implementing such tools (Irshad et al., 2023; Rissler et al, 2020). For example, Rissler et al. (2020) developed a classification model using a random forest algorithm that was able to distinguish between high and low flow states with an accuracy of 70.6 %, based on data from a field study in which software developers wore an ECG chest belt during their workday.

Despite these advances, the limited number of publicly available datasets hinders further progress in flow state classification. One possible solution is transfer learning, where knowledge learned from related physiological classification tasks is transferred to flow-specific classification. Initial results demonstrate the potential of transfer learning within the flow context by improving classification performance by 5 % (Irshad et al., 2023). Thus, the objective of this student project is to identify relevant datasets and to develop an ML pipeline including transfer learning to classify flow based on physiological states.

3: Project Objective

The objective of this specific Bachelor/Master project is to classify flow states based on physiological data. To achieve this, students will identify publicly available physiological datasets from (1) flow-related contexts (e.g., DEAP by Koelstra et al. (2006) or LAUREATE by Laporte et al. (2023)) and (2) flow-specific contexts (PhySF by Irshad et al., (2023)), which are then used to apply transfer learning methods. Students will get access to the results of an overview of flow-specific datasets identified by previous student groups and to the university's high-performance-cluster.

The **objectives of this specific project** are the following:

- Review of the relevant literature on flow physiology
- Review of current transfer learning and machine learning methods in physiological flow classification
- Identification of publicly available datasets from a flow-related and a flow-specific context
- Preprocessing of data, e.g.:
  - Filtering and processing of physiological signals
  - Feature extraction
- Training and testing of different ML algorithms, e.g.:
  - Implementation of selected algorithms
- Evaluation of model performance
  - Comparison of different ML pipelines
  - Ablation analysis to evaluate the impact of individual pipeline elements

Depending on group size, adjustments to the project scope can be made. Adjustments will be announced at the kickoff and/or in subsequent milestone meetings.

Each finished artifact has to be properly documented to facilitate understanding and further development by students in follow-up projects.

4: Project Registration and further Information

All Bachelor- and Master students interested in this IS-project may apply via our online application form. The link to the application will be **available here** (Link hinterlegt) as of the start of the application period on **Monday, 28.09.2026.** The deadline for the application is **Sunday, 11.10.2026, 23:59.**

For successful completion of the project, Master students receive 12 ECTS (corresponding to 360 working hours) and Bachelor students receive 6 ECTS (corresponding to 180 working hours)

- Master students will invest 18 hours per week into the project.
- Bachelor students will invest 9 hours per week into the project.

In addition to project-specific deliverables during the process, we also require some fixed deliverables from each group. On the 04.12.2026, the project team needs to submit a one-page management summary on the current state of the project. A template will be provided by the Chair. As final deliverables, the project team has to submit four items by 28.02.2027. These are (1) a 5-minute video presenting the results (video for orientation), (2) a scientific poster summarizing the results (whether the poster will be considered for presentation will be determined by our Chair, and a final centralized vote among the Information Systems Chairs), and (3) a project report along with the (4) project code. The project report will contain the final results on 20 to 50 pages (Bachelor) or 30 to 60 pages (Master). The exact number of pages depends on group size and the number of bachelor and master students and will be communicated at kickoff. A template is provided on the WIPRO page. The (4) code needs to be accessible as Open Source via GitHub with appropriate documentation. We want to put up the best posters of each semester (2) at the Chair's premises to highlight your achievements. For outstanding projects, the final video (1) and a link to your code (4) will be shown on the Chair's page.

5: Project Schedule (Dates are to be announced)

Kickoff	Kickoff with accepted participants Clarify project scope and responsibilities
Milestone 1	Identification of publicly available datasets in a flow-related and flow-specific context
Milestone 2	Development of a machine learning pipeline including transfer learning methods
Milestone 3	Execution of the machine learning pipeline
Milestone 4	Evaluation of flow classification model
Milestone 5	Submission of one-page management summary, project poster, report, code & 5-minute video illustrating the project

6: Grading

The final grade will be calculated based on the sum of the following milestones:

- Milestone 1: 15%
- Milestone 2: 15%
- Milestone 3: 15%
- Milestone 4: 15%
- Milestone 5: 40%

7: Sources

- Alzubaidi, L., Bai, J., Al-Sabaawi, A. *et al.* A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *J Big Data* 10, 46 (2023). https://doi.org/10.1186/s40537-023-00727-2
- Irshad, M. T., Li, F., Nisar, M. A., Huang, X., Buss, M., Kloep, L., Peifer, C., Kotzsnik, B., Pollak, A., Pyszka, A., Flak O. & Grzegorzek, M. (2023). Wearable-based human flow experience recognition enhanced by transfer learning methods using emotion data. *Computers in Biology and Medicine*, *166*, 107489.
- Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., & Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, *3*(1), 18-31.
- Kutt, K., Drążyk, D., Żuchowska, L., Szelążek, M., Bobek, S., & Nalepa, G. J. (2022). BIRAFFE2, a multimodal dataset for emotion-based personalization in rich affective game environments. *Scientific Data*, *9*(1), 274.
- Laporte, M., Gjoreski, M., & Langheinrich, M. (2023). LAUREATE: A dataset for supporting research in affective computing and human memory augmentation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *7*(3), 1-41.
- Peifer, C., Syrek, C., Ostwald, V., Schuh, E., & Antoni, C. H. (2020). Thieves of flow: how unfinished tasks at work are related to flow experience and wellbeing. *Journal of Happiness Studies*, *21*(5).
- Rissler, R., Nadj, M., Li, M. X., Loewe, N., Knierim, M. T., & Maedche, A. (2020). To be or not to be in flow at work: physiological classification of flow using machine learning. *IEEE transactions on affective computing*, *14*(1), 463-474.
- Statista. (2024). *Machine learning market size worldwide from 2017 to 2031*. Statista. https://www.statista.com/outlook/tmo/artificial-intelligence/machine-learning/worldwide?srsltid=AfmBOooAAsiihUXLuEqEjzji0XIevLk2YE267ZMvhUtAsk5tp3qx7fym

Recommendation:

- Knierim, M. T., Rissler, R., Dorner, V., Maedche, A., & Weinhardt, C. (2018). The psychophysiology of flow: A systematic review of peripheral nervous system features. *Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2017*, 109-120.