MotiVAtor: Analyzing Physical Activity Study Data in Lab and Life through Visual Analytics

Julius Rauscher*
University of Konstanz
Manuel Schmidt*
University of Konstanz
Udo Schlegel†
LMU Munich & Munich Center for Machine Learning (MCML)
Maik Bieleke*
University of Konstanz
Julia Schüler*
University of Konstanz
Daniel A. Keim*
University of Konstanz

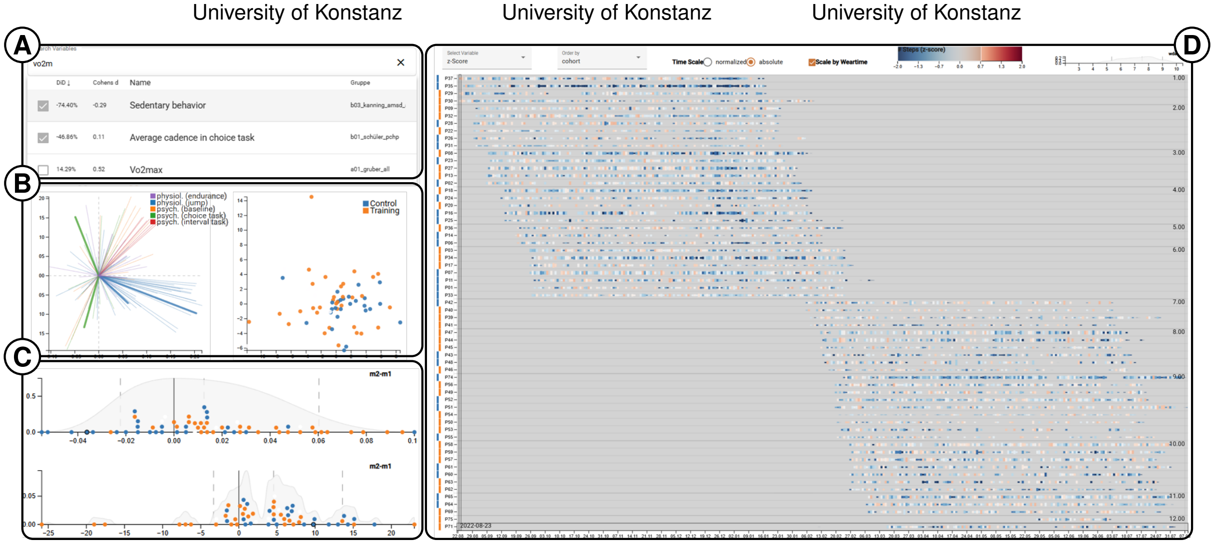

Figure 1: Overview of our proposed Visual Analysis application. (A) Variable Selection Search bar to select lab measurements. (B) PCA BiPlot that supports similarity search in participant and measurement space. (C) Variable Distribution Plots for every selected variable show the distribution of all participants. (D) PixelPlot displays the activity of participants in everyday life.

Abstract

The complexity of intervention studies to assess physical activity (PA) is increasing, resulting in vast amounts of data being recorded in laboratory settings. Recent studies extend datasets with measurements outside the lab using wearable devices, allowing for a bridge to be built between the lab and real-life applications. Such heterogeneous, multigranular datasets impose various challenges for data analysis and visualization, and require tailored approaches to support domain experts. Contrarily, it enables data-driven hypothesis generation, which is particularly valuable in interdisciplinary contexts where theory-driven approaches fall short due to a lack of well-established theoretical foundations. While lab conditions are extensively handled in existing visual interfaces, measurements in everyday life situations are often neglected, yet have an essential impact on the capturing of PA. To facilitate exploration of this data, we propose a visual analytics application consisting of multiple linked views comprising a BiPlot, Variable Distribution plots, and a dense-pixel visualization allowing experts to generate novel interdisciplinary hypotheses based on laboratory and everyday life measurements. We evaluate our application by conducting an expert study with an end user of the application, showcasing the application's benefits to support experts in solving tasks regarding exploration, pattern identification, association, and comparison.

Index Terms: Visual Analytics, Physical Activity, Intervention Studies, Multigranular Visualizations.

*e-mail: {first.last}@uni.kn
†e-mail: udo.schlegel@lmu.de

1 Introduction

Promoting physical activity (PA) through behavioral interventions is an essential aspect in health research, as the WHO identified physical inactivity as the fourth leading cause of global mortality [20]. Traditionally, intervention studies are designed to test preexisting hypotheses using controlled pre-, post-, and follow-up measurements of a specific variable across an intervention and control group.

In recent years, a shift towards more sophisticated study setups that grow in complexity and scope has become apparent. Often, substantially more data than required is collected to test the initial hypotheses [2], resulting in complex datasets that are high-dimensional and heterogeneous, comprising variables that differ in type (e.g., numerical, ordinal, categorical), temporal resolution (e.g., discrete lab visits vs. continuous tracking), and measurement context (e.g., controlled laboratory tests vs. real-world observations). While this increased complexity imposes challenges for data analysis, the diversity of these datasets opens up new opportunities for data-driven hypothesis generation methods, offering a complement to the predominantly employed theory-driven approaches [24].

Visual Analytics (VA) approaches have demonstrated their potential for hypothesis generation by integrating computational methods with interactive visual representations [27, 15]. Involving PA researchers as domain experts in these human-in-the-loop analysis processes supports the discovery of new patterns and intervention strategies beyond established theoretical models.

A number of studies have already been conducted in the PA domain to investigate how behavioral, physiological, and environmental factors can increase activity levels and improve health outcomes [12, 11, 8]. Given the multifaceted nature of PA, its quantification remains challenging and a broad variety of measurement methods exist [30]. While interactive VA tools on intervention studies mostly rely on lab-based assessments [29, 3, 1], PA also plays a vital role outside the lab in everyday life activities. Understanding the interplay between laboratory-based measures and everyday life behavior provides a possibility to assess the real-world effectiveness of interventions, yet remains underexplored in current visual analytics solutions.

The study *Promoting Physical Exercise in Lab and Life* (ProPELL) [4] evaluates the effects of a lab-based jump training intervention on the promotion of PA in a 21-week randomized control trial (RCT). In addition to various physiological and psychological measurements conducted in the lab at three time points (baseline, post-intervention, and follow-up), participants answered weekly questionnaires and wore smartwatches throughout the study period. This study design enables the integration and assessment of both episodic lab measurements gathered from different disciplines and longitudinal, real-world behavioral data, offering a broader and more holistic perspective on the promotion of PA.

To support the exploration and analysis of such a heterogeneous, multidimensional dataset, we derive analysis tasks from existing literature in the public health domain and propose a visual analytics application tailored to support intervention researchers in PA. Our approach enables domain experts to explore the available data, assess the effectiveness of the intervention, investigate inter-variable relationships, and generate new hypotheses based on trends observed across both lab-based and real-world data. We evaluate our application in an expert study that illustrates how analysts interact with the visualizations and demonstrates insights that can be derived using our approach.

2 Related Work

The data analyzed in this paper stems from a RCT conducted to evaluate a public health intervention; hence, our work builds on VA approaches for analyzing and visualizing such data to support domain experts in drawing evidence-based conclusions.

In their survey on Visual Analytics in the public health domain, Preim and Lawonn [22] outline commonly used tasks, requirements, and visualization techniques, and they further indicate a lack of evaluation of public health interventions in their research agenda. To address the challenges of underlying heterogeneous data often encountered in cohort studies, Steenwijk et al. [29] propose a conceptual data structure of domains, features, mappers, and studies coupled with an interactive data exploration environment consisting of scatterplots, PCPs, and time plots. In a similar fashion, Angelelli et al. [2] structure heterogeneous study data in multiple OLAP cubes that can be aggregated for visual exploration in a prototype consisting of multiple coordinated views. Malik et al. [18] establish a taxonomy of metrics for comparing cohorts of temporal event sequences, and allow a visual exploration of these metrics in their VA interface dubbed *CoCo*. Klimm et al. [16] leverage 3D rendering techniques of MRI data to integrate image data in the analysis of cohort study data. Bernard et al. [3] propose a VA tool incorporating varying levels of detail to enable both individual record and cohort views to support physicians in analyzing patient data with regard to prostate cancer. The tool S-ADVIsed [1] integrates interactive subspace clustering methods to explain risk factors of diseases in cohort study data. Ritti-Dias et al. [26] highlight the importance of visual elements such as glyphs to communicate the effects of an intervention in randomized controlled trials. Brich et al. [5] leverage dimensionality reduction to visualize multivariate time series data as time curves recorded in intensive care units. Wang et al. [33] present ThreadStates, an approach for visualizing and identifying states of disease progressions in longitudinal cohort studies through scatter plots, feature matrices, and area charts. Li et al. [17] propose TrialView, a system that leverages a graph autoencoder to construct a patient similarity graph, enabling clustering and coordinated visualizations at both individual and cohort levels for the exploration of temporal event data in clinical trials.

Figure 2: Variable Selection Search Bar. Allows querying available measurements through keywords, and can be sorted by either *DiD* or *Cohen's d*. All currently selected measurements are pinned to the top of the list.

While existing VA approaches for intervention studies primarily focus on lab-based measurements, they often overlook the influence of real-world context and everyday life, highlighting the need for methods that integrate and visualize out-of-lab data to provide a more holistic view of intervention effects.

3 MotiVAtor

The following section describes the utilized data, outlines the task elicitation process based on existing literature, and provides an overview of the prototype along its key components.

3.1 Data

In the *ProPELL* study setup, 74 participants (age 23 $\pm$ 3 years) were divided into a control and training group, where the training group underwent an 8-week jump training, performing 15 minutes of exercise 3 days per week. In the presented prototype, we analyze up to 109 variables that were collected from different domains, such as physiology (e.g., cardiovascular, neuromuscular, and endocrine) and psychology (e.g., motivational, emotional, and behavioral) across varying temporal granularities. Most lab measurements were recorded three times, once before the intervention (**m1**), once after the intervention (**m2**), and once in a follow-up (**m3**). Every participant was further equipped with a smartwatch and was instructed to wear it as often as possible, recording exercise- and activity-related variables in the form of active steps or sedentary, light, moderate, or vigorous physical activity.

3.2 Analysis Tasks

In conjunction with two experts involved in the project from the domains of sports science and psychology, we elicited analysis tasks central to the goals of our visual analytics approach in supporting interdisciplinary data exploration in lab and life, derived from Preim and Lawonn [22]:

- `[T1]` **Exploration**: Gaining an overview and explore the data.
- `[T2]` **Assessment and Pattern Identification**: Drill down on specific measurements and assess their relevance in promoting PA.
- `[T3]` **Associations**: Finding and analyzing associations between interdisciplinary measurements.
- `[T4]` **Comparisons**: Support comparisons between subgroups, such as the control and training group.

Other available tasks from Preim and Lawonn were deemed as not relevant. For *Verification* of hypotheses, domain experts use their established analysis methods and rather rely on statistical

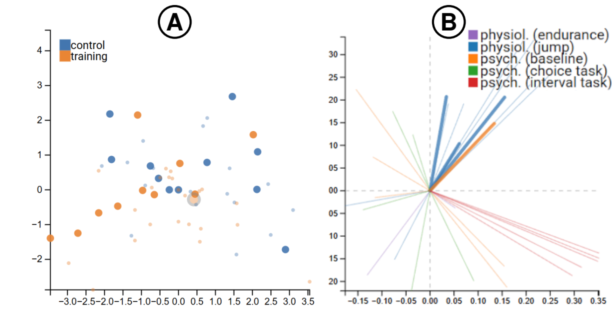

Figure 3: PCA BiPlot showing information about participants (A) and measurements (B). Filtered participants by brushing from other visualizations are highlighted. Measurements are colored by discipline and generally show strong groupings.

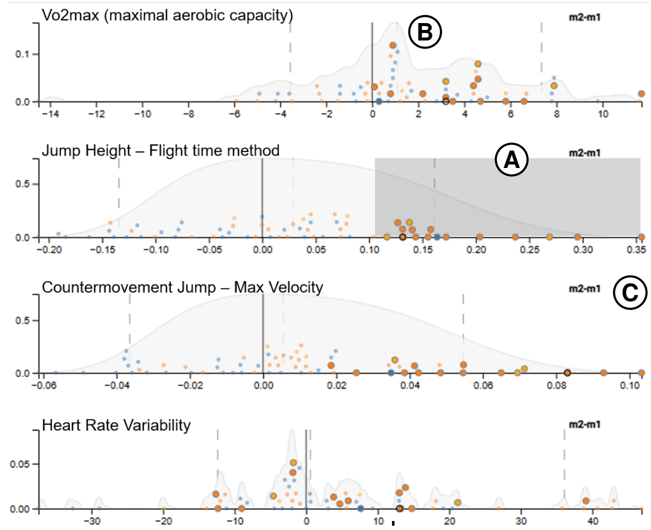

Figure 4: Distribution plots showing the distribution of participants for several variables concerning the improvement of physical activity (m2-m1). Brushing on one plot will highlight the filtered participants in the other plots (A). The shift of the main KDE peak in the Vo2max measurement indicates an overall improvement (B); however, the control and training groups are not well separated. The Countermovement Jump maximum velocity measurement shows a better split (C).

methods instead of interactive VA applications. While the interpretation of the study results allows reflecting on and potentially refining the study design and intervention strategy, *Policy Development* is not a primary target of this application. Likewise, educating the public about the importance of PA is an overarching goal of the performed intervention study, but the developed VA application is not designed for *Dissemination* purposes.

3.3 Prototype

The VA prototype consists of multiple linked views, namely a Variable Selection Search Bar, a PCA Biplot, Variable Distribution Plots, and an activity PixelPlot (see Figure 1). Each plot is brushable, allowing users to create a subgroup that is simultaneously highlighted across all other plots. Participants are represented in all views of the prototype and are visualized using a consistent color scheme. By default, the coloring encodes study group affiliation (control vs. training), but it can be adjusted to reflect other attributes such as metadata, selected lab measurements, or aggregated life-based indicators. The following section explains the individual components in more detail and outlines how the aforementioned tasks are supported.

Variable Selection As an initial starting point, analysts might want to inspect a number of specific variables and investigate how they relate to other measurements **[T1]**. To facilitate variable subselection, we provide three different approaches: (1) Analysts, especially those who designed the study, often want to make use of their domain knowledge and already have a set of variables they want to examine in mind. We facilitate this selection task by a keyword-based **Search Bar** that returns a list of matches from the descriptions of every variable (see Figure 2). These variables can be added to the analysis process by clicking on the checkbox that is present on the left-hand side of each entry. (2) For a supported data-driven selection, we provide two effect size measures that quantify the magnitude of change between the training and control groups: The Difference in Differences (*DiD*) [10], which captures the relative change in outcomes over time between groups, and *Cohen's d* [6], which measures the standardized difference in means between groups. The search results can be ordered by both the *DiD* and the *Cohen's h* to highlight variables exhibiting larger changes between the groups. (3) Furthermore, we provide automated feature selection through feature importance values obtained from machine learning models trained on the study data using the internal feature importance techniques, namely a *Support Vector Machine (SVM)*, *Random Forest (RF)*, and *Extreme Gradient Boosting (XGBoost)*. We train the classifier using the control or training association as labels and return the top ten features with the highest importance threshold. These features can be further evaluated for relevance using the Variable Distribution Plots.

PCA BiPlot To provide an overview of the lab-measured variables **[T1]**, we provide a Principal Component Analysis (PCA) BiPlot, which is capable of assessing similarities both on a variable and participant level (see Figure 3). PCA analysis is commonly deployed by researchers in PA [14, 9], which reduces the learning curve and cognitive load for the application for the analysts. To capture the performance increase among the available measurements, we calculated the relative increase after the intervention $\frac{m2-m1}{(m1+m2)\cdot 0.5}$, resulting in a single, normalized score for every variable. From these scores, we can obtain so-called loadings and scores using PCA and visualize them in a BiPlot. Every line represents a loading of a variable where the length and direction encode the variance, with small angles between lines indicating correlations. Such correlations can indicate associations between interdisciplinary factors **[T3]**. To facilitate this, we employ a categorical color scheme to visually discern loadings from different domains (see Figure 3 (B)). On the other hand, every point represents a study participant, where close proximity indicates similar performance increases, allowing for inspection of possible cluster structures. The points can be interactively colored according to any attribute to support various comparisons, e.g., between the control and training group or any currently selected measurements **[T4]**. Variables can be added or removed from the analysis by clicking on the loading line, which triggers a recomputation of the PCA according to the selected variables, fostering an iterative analysis loop. As with all plots in MotiVAtor, brushing the plot is supported and will act as a filter, highlighting the included participants in all other visualizations.

Variable Distributions To allow a deeper understanding of one specific variable and permit pattern identification **[T2]**, we provide a **Distribution Plot** (see Figure 4) for every variable that was selected through the Search Bar or the BiPlot loadings. As the focus of the study is to assess the *promotion* of PA, we show the difference between the measurements before and after the intervention $(m2-m1)$ rather than absolute values. This captures the perfor-

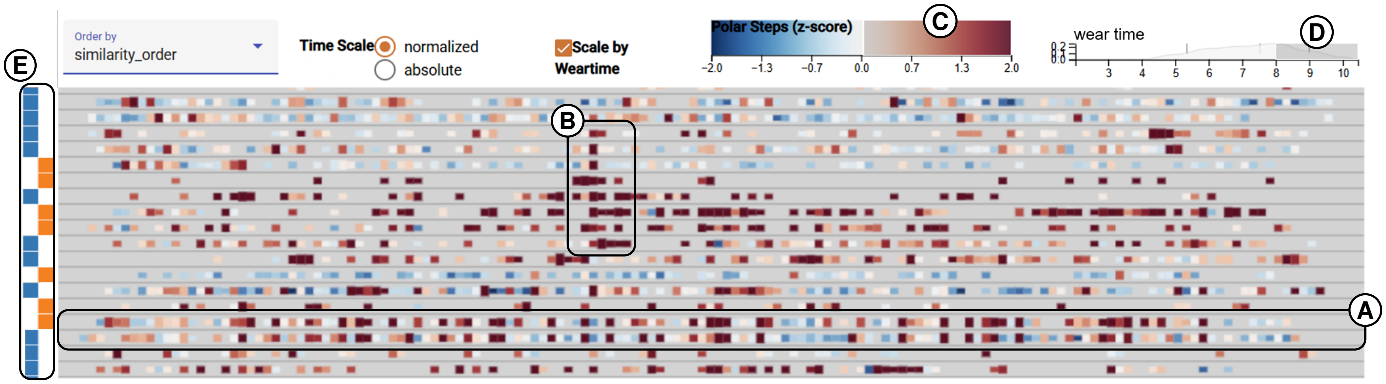

Figure 5: Pixel Plot giving an overview of PA in life situations. Scaling the rectangle height by wear time shows large deviations. Reordering techniques to make patterns salient, such as training together (A), or especially active days (B). The rectangles can be filtered by value (C) or wear time (D). Reference view indicating how the control/training group is aligned with the current ordering (E).

mance during the intervention period better, where a value above 0 indicates an increase, and a value below 0 a decrease. The distribution can be viewed on two different levels of granularity, with a kernel density estimation (KDE) showing the global distribution represented by a single line and a beeswarm plot that shows the frequency of participants (on the y-axis) across the value range (on the x-axis) on a finer level with every circle uniquely representing one participant. As pointed out by Yang et al. [21], beeswarm plots provide a more lucid representation of the distribution, where we apply the global coloring scale to every point individually and facilitate a comparison of different groups across the data range **[T4]**. The KDE serves as the static baseline across all participants, whereas the beeswarm points adapt to the filtering and depict how the filtered selection diverges from this baseline. A solid vertical line at 0 visually separates performance decrease from increase, while dashed lines indicate the 10th, 50th, and 90th percentiles of the distribution. Hovering a beeswarm circle will display the absolute values of the measurements via a tooltip. The plot itself can also function as an interactive filter, where brushing over a selected range creates a subgroup. The participants within the brushed interval are then highlighted in all other plots.

PixelPlot The activity data collected from smartwatches throughout the entire study duration serves as the basis for assessing PA in everyday life situations, as opposed to the lab measurements predominant in the other views. The data is available as the number of steps or broken down into the 4 levels of activity (sedentary, light, moderate, vigorous) on a daily resolution. To provide a compact overview **[T1]**, we provide a dense pixel visualization where the rows represent participants and the columns sequentially reveal time (see Figure 5).

Since the base level of activity can vary among the participants, absolute values are susceptible to a visual dominance of high-activity participants. To mitigate this, we provide the option to visualize a z-Score normalized $z = \frac{X-\mu}{\sigma}$ value instead. The mean (μ) and standard deviation (σ) are obtained from the first 2 weeks of data, allowing changes in PA to be interpreted relative to a pre-intervention baseline. The values are encoded using a diverging blue-red color scale, visually emphasizing deviations from the baseline. Following a cold–hot metaphor, blue indicates lower, red higher activity levels, while white represents no deviation from the baseline. A brushable legend is shown alongside the visualization (see Figure 5 (C)), allowing users to filter participants and days based on their activity changes. This reduces visual clutter and supports targeted exploration of selected deviation levels, such as sustained increases in PA over time.

Another notable impact on the performance measurement is the wear time of the smartwatch, as participants who wear the watch longer most likely accumulate more PA. We indicate the wear time for every day by scaling the height of every pixel accordingly, resulting in larger rectangles for days with longer wear durations. The PixelPlot can also be filtered according to the wear time (see Figure 5 (D)), allowing analysts to focus their analysis on days when the watch was worn for a minimum number of hours.

We further provide the option for an *absolute* or *relative* temporal arrangement on the columns. In the absolute arrangement, the pixels are ordered according to the actual day the activity was recorded, while in the relative the pixels are placed according to the day in the study. The absolute arrangement allows users to identify the influence of outside factors such as holidays (see Figure 1), whereas the relative arrangement enables a comparison of participants according to the progress of the study **[T4]**.

Applying 1-D ordering strategies is a common approach to make patterns and structures in dense pixel visualizations salient [28, 23], and hence facilitate pattern identification **[T2]**. The implemented strategies fall into three main categories: (1) group-based orderings leverage known participant attributes such as study condition (training/control), sex, or adherence status to cluster participants by metadata. (2) To provide a similarity-based ordering on the underlying time series, we apply the normalized Euclidean distance metric from van Wijk and van Selow [31] and obtain an ordering from the leaf nodes of the dendrogram from complete-linkage hierarchical clustering. We further use trend decomposition to obtain the slope value of the decomposed trend, which reflects the long-term progression of the series and gives an indication of whether an increase in PA has been recorded. (3) variable-based orderings rank participants based on any given selected lab measurement, enabling direct visual comparisons along known outcome measures. On the left-hand side of the PixelPlot, a reference view visualizes the global coloring across all participants. It serves as a visual guide to assess how the current ordering aligns with the selected color encoding. For instance, users can evaluate how clearly an ordering separates training and control groups.

4 Evaluation

To assess the applicability of our approach, we conducted an expert user study with a professor in sports science (55 years old) with more than 25 years of experience in the field of promoting PA, who is also actively involved in the ProPELL project.

Setup After inquiring about the participant's expectations for a prototype aimed at supporting hypothesis generation in the domain of PA, each component of the prototype was introduced. The participant was given time to familiarize himself with the environment and was initially asked to gain an overview of the dataset **[T1]**. Next, he was prompted to select a measurement of interest and assess the influence of the intervention on this variable, guided by visual cues provided in the interface **[T2]**. Following this, the participant was asked to identify and interpret a potential association between variables from different disciplinary domains, such as physiological and psychological data **[T3]**. To assess the prototype's support for comparative analysis, the participant was then tasked with comparing subgroups, for example, the control and intervention groups, concerning selected variables **[T4]**. The session concluded with a think-aloud interview, in which the participant shared his impressions of the application's usability, limitations, and its potential usage in real-world research settings.

Results Regarding his expectations, the expert stated that he would like to explore the full extent of available measurements and examine how different interdisciplinary variables are connected. In order to generate novel hypotheses, PA researchers often rely on existing theories and perform a methodological transfer where theoretical constructs from one context are applied to another. A data-driven approach to explore such associations was considered valuable when theoretical foundations are limited or not yet established.

To get an overview of PA in life, he first examines the PixelPlot and observes the high deviation in terms of wear time. As low wear times are associated with high uncertainty, he filters the plot to only show days where the watch was at least worn for eight hours. To determine possible clusters among participants, he orders the pixel plot by time series similarity. This revealed a period shortly after the second lab measurement (**m2**) during which a group of participants recorded unusually high step counts (see Figure 5 (B)). Furthermore, he detects a pair of participants with very similar activity patterns, which led to speculation that they may have exercised together (see Figure 5 (A)). When reordering by mean step count, it becomes evident that these two were among the most active participants overall ($\sim$20k steps/day). This aligns with physiological research stating that shared activity can increase motivation and, in turn, promote higher levels of PA.

For assessing lab performances, he starts with selecting the *Vo2max* (maximal aerobic capacity) measurement, a parameter known to quantify endurance fitness in established PA theories. The Distribution Plot visible in Figure 4 indicates a general increase in performance across all participants, as the major peak of the KDE distribution is shifted rightward relative to the zero indicator. However, the expert inferred that repeated Vo2max measurements are often subject to familiarization effects, where participants may perform better in a second trial not because of improved fitness, but due to increased comfort with the procedure and more accurate self-pacing. When exploring individual-level data via the beeswarm plot, no clear separation between the control and training groups could be observed. Using the PixelPlot, the expert further observed that many participants already exhibited high baseline activity levels in everyday life ($>$10k steps/day), indicating a ceiling effect in the Vo2max measurement. To find variables that better capture a difference between the groups, he sorts by the entries in the Search Bar by *DiD* and selects the *Countermovement Jump Max Velocity* and *Jump Height using Flight Time Method*. As seen in Figure 4, most training group participants increased their performance in contrast to the control group, which is unsurprising as these measurements capture properties related to the jump-training of the intervention. When inspecting the loadings in the BiPlot (see Figure 3), the expert notices that the *Heart Rate Variability* measured in the psychology lab exhibits a correlation to the selected jump-related variables. Increased heart rate variability is seen as an indicator of reduced stress, leading to the interdisciplinary hypothesis that jump training can reduce stress levels.

Feedback The expert was overall very satisfied with the capabilities of the prototype. Although the range of interaction options introduces a learning curve, the expert noted that the system becomes intuitive after a short period of use. Especially the interactive brushing functionality on all visualizations, coupled with the instantly filtered feedback, was valued highly and enabled rapid comparisons of arbitrary subgroups sharing common characteristics. In conjunction with the adaptive coloring scheme and various available ordering strategies for the PixelPlot, the prototype was perceived as a powerful toolbox for investigating PA-related phenomena in lab and life scenarios. Since the main focus of the analysis consisted of comparing the training and control groups, this group assignment was used as the predominant coloring scheme. However, applying other variables as the coloring dimension, such as the mean number of steps or the trend slope obtained from the smart watch activity, helped connect lab-based assessments to everyday life activities. In reverse, applying other variables as the coloring dimension, such as psychological scores or physiological measurements, particularly in the PCA and Distribution Plots, helped connect real-world behavior with lab-based assessments.

The automated feature selection functionality, on the other hand, was rated as less helpful. The rationale behind the selected features was not always clear to the expert, limiting interpretability and trust in the provided methods. Ordering the available variables in the Search Bar by *DiD* or *Cohen's d* provided more intuitive support in the selection of variables. To retrieve additional associated variables, the expert found the BiPlot loadings helpful, but noted that confirmation via the Distribution Plots was necessary to ensure interpretability.

5 Discussion

The expert evaluation in Section 4 sheds light on how potential users solve the tasks that we elicited in Section 3. In the following, we review the applicability of our approach and discuss the generalizability beyond the used dataset.

[T1] Exploration As outlined in the expert feedback, the prototype already supports users in examining the data before conducting their domain-specific analysis. The Distribution Plots provide a general overview of the measurement distributions, and the PixelPlot offers a compact overview of PA in everyday life situations. Through the rectangle height in Figure 5, it is easily observable that the wear time of the smart watch is subject to high deviations within and between participants. The large differences in wear time introduce uncertainty and skew the results, possibly limiting the expressiveness of the recorded activity data in life. Various imputation strategies [25] could be applied to mitigate this discrepancy, at the risk of possibly introducing bias. While large deviations from wear time and PA are easily detectable, low-level activity patterns are visually underrepresented due to the chosen divergent color scale and the scaled rectangle height.

[T2] Assessment and Pattern Identification To assess the relevance of lab measurements on PA, the Distribution Plots offer a compact perspective on performance increases or decreases amongst the participants, and have proven beneficial in identifying patterns as outlined in the expert study in Section 4. By using different coloring schemes on the beeswarm plot, these results can be interpreted with respect to different contexts, such as the impact of the intervention (by control/training coloring), or the performance in everyday life situations (by coloring according to the trend slope or the mean number of steps). Figure 4 (Vo2max) exposes that such plots can be susceptible to outliers where large deviations between the **m1** and **m2** measurements were recorded. While this facilitates the detection of outliers and possible measurement errors, the overall distribution is squished, limiting the interpretability of the visualization.

[T3] Associations Linking interdisciplinary measurements is supported by the BiPlot loadings, in which correlations between variables from different domains are visually identifiable by lines that are closely aligned with a small angle between them but originate from different disciplines, as indicated by their distinct colors (see Figure 3). The PixelPlot can further support the analysis by confirming whether performance increases in the lab are also recorded in everyday life situations of the participants. Ordering the PixelPlot by the trend slope allows easy filtering of participants with increased PA by brushing the plot, setting the focus on this subgroup in all other visualizations. While these associations visually suggest correlations, they do not directly imply causality. Since the primary purpose of the prototype is to facilitate hypothesis generation, the causality can subsequently be validated through specifically designed, rigorous follow-up studies.

[T4] Comparisons The color scale applied in all plots enables a comparison of given subgroups, e.g., training or control. The Beeswarms in the Distribution Plots support comparisons on an individual level, and coloring according to both numerical and categorical attributes can be applied. While using separate KDE curves for each group could offer a clearer visual distinction between control and training groups, this approach does not scale well when more than two groups are involved or when comparing against continuous variables. The PixelPlot further supports comparisons through the available ordering strategies, i.e., by revealing patterns such as training buddies with very similar activity patterns (see Figure 5). These dense pixel visualizations scale well with larger datasets, as observable in other domains [28]. In confirmatory scenarios, comparisons between subgroups are typically verified with statistical significance (e.g., via t-tests). However, since our approach emphasizes exploratory hypothesis generation, such formal tests are outside our immediate scope but may follow in subsequent follow-up studies.

5.1 Limitations and Future Work

Limitations While our approach relies on PCA, alternative nonlinear projection methods such as t-SNE or UMAP could potentially capture local neighborhood structures more effectively. However, these methods do not provide loadings, which are essential for interpreting and assessing variable similarities. Furthermore, they are more difficult to interpret and less familiar to researchers in the PA domain, which further motivated our decision to use PCA.

Recording activity in life throughout a study duration of 21 weeks inherently introduces patterns through seasonal trends, such as weekends or holidays. These effects are clearly visible in the absolute view (see Figure 1), and the relative view can even show such trends relative to the study progress (see Figure 5 (B)), sparking speculations about their cause from the study setup. However, such effects directly impact the assessment of physical activity in life and can also be viewed as undesired biases. Cyclic biases such as weekend-related activity peaks could be mitigated through basing the PixelPlot visualizations on trend decomposition (e.g., separating seasonal, residual, and overall trends). However, patterns like increased weekend activity may themselves be relevant findings, depending on the research question, and should not naively be dismissed as mere noise.

The evaluation of our prototype was carried out with only one expert user, which limits the generalizability of the findings. Although the expert provided valuable feedback, this perspective may not capture the full range of needs and preferences within the physical activity research community. Involving multiple experts from diverse backgrounds will help validate the application's usability and effectiveness more comprehensively.

Future Work Our automated feature selection functionality, which was not well received in the expert user study, is limited to the built-in feature importance of the provided classification models, which is largely influenced by the classification's performance concerning the target feature or the closeness of the learned knowledge to the expert's knowledge. As future work, we aim to deploy improved classification models, either with better performance or closer expert and model knowledge alignment. Further, more advanced XAI methods, such as Shapley Additive Global importancE (SAGE) [19] are available and have shown promising results in other application areas, identifying relevant features for feature selection. Including interpretable clustering approaches [13] towards the dense pixel visualization could also enable further insights outside of the domain knowledge of the experts grouping similar samples. However, we acknowledge that the abstraction level of such methods may pose challenges for domain experts without a strong technical background. Ensuring the usability and accessibility of these techniques remains an open issue.

We leveraged the scores and loadings from a PCA BiPlot to investigate similarities between participants and associations between measurements, respectively. While effective for linear patterns, this approach may overlook more complex structures in the data. As future work, we aim to investigate how Dual analysis methods [7], which simultaneously consider relationships across participants and features, can offer a complementary perspective and reveal latent patterns that are not easily captured by PCA alone.

The full extent of the data recorded in the study includes additional data sources such as weekly survey data or brain imaging in the form of MRI data. Previous studies have shown how such data can successfully be integrated in the analysis of cohort study data [32, 16], and incorporating them will contribute to a more holistic view on the interdisciplinary analysis of PA promotion.

6 Conclusion

We presented MotiVAtor, a VA application to support the analysis of complex, heterogeneous data from interdisciplinary PA intervention studies that combine lab-based and real-world measurements. By bridging controlled and everyday settings, the system offers researchers a more comprehensive view of intervention effects. Our approach is grounded in task categories identified from prior research, namely exploration, pattern identification, association, and comparison. Feedback from an expert user study highlights the value of the tool for exploratory analysis and hypothesis generation, particularly in contexts where theoretical guidance is limited. Future work will focus on improving model interpretability, integrating additional data sources, and extending the analytical capabilities to better support dynamic, multigranular data analysis in PA research.

Acknowledgments

The authors gratefully acknowledge funding from the Committee on Research (AFF) at the University of Konstanz for the research initiative "ProPELL: Promoting Physical Exercise in Lab and Life", the Federal Ministry for Economic Affairs and Climate Action (BMWK, grant No. 03EI1048D), and the Munich Center for Machine Learning (MCML).

References

[1] S. Alemzadeh, T. Hielscher, U. Niemann, L. Cibulski, T. Ittermann, H. Völzke, M. Spiliopoulou, and B. Preim. Subpopulation Discovery and Validation in Epidemiological Data. In *EuroVis Workshop on Visual Analytics (EuroVA)*, 2017. doi: 10.2312/eurova.20171118 2

[2] P. Angelelli, S. Oeltze, J. Haász, C. Turkay, E. Hodneland, A. Lundervold, A. J. Lundervold, B. Preim, and H. Hauser. Interactive Visual Analysis of Heterogeneous Cohort-Study Data. *IEEE Computer Graphics and Applications*, 34(5):70–82, 2014. doi: 10.1109/MCG.2014.40 1, 2

[3] J. Bernard, D. Sessler, T. May, T. Schlomm, D. Pehrke, and J. Kohlhammer. A Visual-Interactive System for Prostate Cancer Cohort Analysis. *IEEE Computer Graphics and Applications*, 35(3):44–55, 2015. doi: 10.1109/MCG.2015.49 2

[4] M. Bieleke. Promoting Physical Exercise in Lab and Life. *German Clinical Trials Register*, Aug 2022. 2

[5] N. Brich, C. Schulz, J. Peter, W. Klingert, M. Schenk, D. Weiskopf, and M. Krone. Visual analytics of multivariate intensive care time series data. In *Computer Graphics Forum*, vol. 41, pp. 273–286. Wiley Online Library, 2022. doi: 10.1111/cgf.14498 2

[6] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, 1988. 3

[7] F. L. Dennig, M. Miller, D. A. Keim, and M. El-Assady. FS/DS: A Theoretical Framework for the Dual Analysis of Feature Space and Data Space. *IEEE Trans. on Visualization and Computer Graphics*, 30(8):5165–5182, 2024. doi: 10.1109/TVCG.2023.3288356 6

[8] M. Di Pumpo, A. Miatton, M. T. Riccardi, E. A. Graps, V. Baldo, A. Buja, and G. Damiani. Digital Health Interventions to Promote Physical Activity in Community-Dwelling Older Adults: A Systematic Review and Semiquantitative Analysis. *Int. J. of Public Health*, 69, 2025. doi: 10.3389/ijph.2024.1607720 1

[9] A. Feraco, S. Gorini, E. Camajani, T. Filardi, S. Karav, E. Cava, R. Strollo, E. Padua, M. Caprio, A. Armani, and M. Lombardo. Gender differences in dietary patterns and physical activity: an insight with principal component analysis (PCA). *J. of Transl. Med.*, 22(1), Dec 2024. doi: 10.1186/s12967-024-05965-3 3

[10] A. Fredriksson and G. M. d. Oliveira. Impact evaluation using Difference-in-Differences. *RAUSP Management Journal*, 54(4):519–532, Jan 2019. doi: 10.1108/RAUSP-05-2019-0112 3

[11] L. Gormley, C. A. Belton, P. D. Lunn, and D. A. Robertson. Interventions to increase physical activity: An analysis of candidate behavioural mechanisms. *Preventive Medicine Reports*, 28:101880, 2022. doi: 10.1016/j.pmedr.2022.101880 1

[12] G. W. Heath, D. C. Parra, O. L. Sarmiento, L. B. Andersen, N. Owen, S. Goenka, F. Montes, and R. C. Brownson. Evidence-based intervention in physical activity: lessons from around the world. *The Lancet*, 380(9838):272–281, 2012. doi: 10.1016/S0140-6736(12)60816-2 1

[13] L. Hu, M. Jiang, J. Dong, X. Liu, and Z. He. Interpretable Clustering: A Survey. *arXiv preprint arXiv:2409.00743*, 2024. 6

[14] E. Iglesias-Soler, M. Rúa-Alonso, J. Rial-Vázquez, J. R. Lete-Lasa, I. Clavel, M. A. Giráldez-García, J. Rico-Díaz, M. R.-D. Corral, E. Carballeira-Fernández, and X. Dopico-Calvo. Percentiles and Principal Component Analysis of Physical Fitness From a Big Sample of Children and Adolescents Aged 6-18 Years: The DAFIS Project. *Front. in Psychol.*, 12 - 2021, 2021. doi: 10.3389/fpsyg.2021.627834 3

[15] X. Jing, V. L. Patel, J. J. Cimino, J. H. Shubrook, Y. Zhou, B. N. Draghi, M. A. Ernst, C. Liu, and S. De Lacalle. A Visual Analytic Tool (VIADS) to Assist the Hypothesis Generation Process in Clinical Research: Mixed Methods Usability Study. *JMIR Hum Factors*, 10:e44644, Apr 2023. 1

[16] P. Klemm, S. Oeltze-Jafra, K. Lawonn, K. Hegenscheid, H. Völzke, and B. Preim. Interactive Visual Analysis of Image-Centric Cohort Study Data. *IEEE Trans. on Visualization and Computer Graphics*, 20(12):1673–1682, 2014. doi: 10.1109/TVCG.2014.2346591 2, 6

[17] Z. Li, X. Liu, Z. Cheng, Y. Chen, W. Tu, and J. Su. TrialView: An AI-powered Visual Analytics System for Temporal Event Data in Clinical Trials. *Hawaii Int. Conf. on System Sciences*, 2024:1169–1178, 2023. 2

[18] S. Malik, F. Du, M. Monroe, E. Onukwugha, C. Plaisant, and B. Shneiderman. Cohort Comparison of Event Sequences with Balanced Integration of Visual Analytics and Statistics. In *Proc. of the Int. Conf. on Intelligent User Interfaces (IUI)*, 12 pages, pp. 38–49, 2015. doi: 10.1145/2678025.2701407 2

[19] E. Mill, W. Garn, N. Ryman-Tubb, and C. T. and. The SAGE Framework for Explaining Context in Explainable Artificial Intelligence. *Applied Artificial Intelligence*, 38(1):2318670, 2024. doi: 10.1080/08839514.2024.2318670 6

[20] W. H. Organization. Global health risks: mortality and burden of disease attributable to selected major risks, 2009. 1

[21] Y. Ouyang, Y. Wu, H. Wang, C. Zhang, F. Cheng, C. Jiang, L. Jin, Y. Cao, and Q. Li. Leveraging Historical Medical Records as a Proxy via Multimodal Modeling and Visualization to Enrich Medical Diagnostic Learning . *IEEE Trans. on Visualization and Computer Graphics*, 30(01):1238–1248, Jan. 2024. doi: 10.1109/TVCG.2023.3326929 4

[22] B. Preim and K. Lawonn. A Survey of Visual Analytics for Public Health. *Computer Graphics Forum*, 39(1):543–580, 2020. doi: 10.1111/cgf.13891 2

[23] J. Rauscher, F. L. Dennig, U. Schlegel, D. A. Keim, and J. Fuchs. Visually Assessing 1-D Orderings of Contiguous Spatial Polygons. *Computer Graphics Forum*, 44(3), 2025. doi: 10.1111/cgf.70100 4

[24] R. E. Rhodes, D. McEwan, and A. L. Rebar. Theories of physical activity behaviour change: A history and synthesis of approaches. *Psychol. of Sport and Exercise*, 42:100–109, 2019. doi: 10.1016/j.psychsport.2018.11.010 1

[25] A. Richter, J. Ijaradar, U. Wetzker, V. Jain, and A. Frotzscher. A Survey on Multivariate Time Series Imputation Using Adversarial Learning. *IEEE Access*, 12:148167–148189, 2024. doi: 10.1109/ACCESS.2024.3473540 5

[26] R. M. Ritti-Dias, M. d. A. Correia, and B. Q. Farah. Visual data: a new tool to improve the presentation of clinical trial results. *Einstein (Sao Paulo)*, 18:eAE4729, 2019. doi: 10.31744/einstein_journal/2020AE4729 2

[27] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge Generation Model for Visual Analytics. *IEEE Trans. on Visualization and Computer Graphics*, 20(12):1604–1613, 2014. doi: 10.1109/TVCG.2014.2346481 1

[28] U. Schlegel and D. A. Keim. Interactive Dense Pixel Visualizations for Time Series and Model Attribution Explanations. In *Workshop on Machine Learning Methods in Visualisation for Big Data (MLVis)*, 2023. doi: 10.2312/mlvis.20231113 4, 6

[29] M. D. Steenwijk, J. Milles, M. Buchem, J. Reiber, and C. P. Botha. Integrated visual analysis for heterogeneous datasets in cohort studies. In *IEEE VisWeek workshop on Visual Analytics in Health Care*, vol. 3, 2010. 2

[30] L. G. Sylvia, E. E. Bernstein, J. L. Hubbard, L. Keating, and E. J. Anderson. Practical Guide to Measuring Physical Activity. *J. of the Academy of Nutrition and Dietetics*, 114(2):199–208, 2014. doi: 10.1016/j.jand.2013.09.018 2

[31] J. Van Wijk and E. Van Selow. Cluster and calendar based visualization of time series data. In *Proc. IEEE Symp. on Information Visualization (InfoVis'99)*, pp. 4–9, 1999. doi: 10.1109/INFVIS.1999.801851 4

[32] H. Wang, Q. Liang, J. T. Hancock, and T. M. Khoshgoftaar. Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. *Journal of Big Data*, 11(1):44, 2024. 6

[33] Q. Wang, T. Mazor, T. A. Harbig, E. Cerami, and N. Gehlenborg. ThreadStates: State-based Visual Analysis of Disease Progression. *IEEE Trans. on Visualization and Computer Graphics*, 28(1):238–247, 2022. doi: 10.1109/TVCG.2021.3114840 2