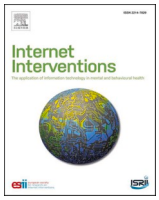
Internet Interventions 40 (2025) 100828

Contents lists available at ScienceDirect

Internet Interventions

journal homepage: www.elsevier.com/locate/invent



The promise and challenges of computer mouse trajectories in DMHIs – A feasibility study on pre-treatment dropout predictions

Kirsten Zantvoort [a,b], Jennifer J. Matthiesen [a,b,*], Pontus Bjurner [b], Marie Bendix [b,c], Ulf Brefeld [a], Burkhardt Funk [a], Viktor Kaldo [b,d]

[a] Institute of Information Systems, Leuphana University, Lüneburg, Germany
[b] Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, & Stockholm Health Care Services, Region Stockholm, Sweden
[c] Department of Clinical Sciences, Division of Psychiatry, Umeå University, Umeå, Sweden
[d] Department of Psychology, Faculty of Health and Life Sciences, Linnaeus University, Växjö, Sweden

ARTICLE INFO

Keywords:
E-mental health
Dropout
Machine learning
Prediction
ICBT

ABSTRACT

With the impetus of Digital Mental Health Interventions (DMHIs), complex data can be leveraged to improve and personalize mental health care. However, most approaches rely on a very limited number of often costly features. Computer mouse trajectories can be unobtrusively and cost-efficiently gathered and seamlessly integrated into current baseline processes. Empirical evidence suggests that mouse movements hold information on user motivation and attention, both valuable aspects otherwise difficult to measure at scale. Further, mouse trajectories can already be collected on pre-treatment questionnaires, making them a promising candidate for early predictions informing treatment allocation. Therefore, this paper discusses how to collect and process mouse trajectory data on questionnaires in DMHIs. Covering different complexity levels, we combine hand-crafted features with non-sequential machine learning models, as well as spatiotemporal raw mouse data with state-of-the-art sequential neural networks. The data processing pipeline for the latter includes task-specific preprocessing to convert the variable length trajectories into a single prediction per user. As a feasibility study, we collected mouse trajectory data from 183 patients filling out a pre-intervention depression questionnaire. While the hand-crafted features slightly improve baseline predictions, the spatiotemporal models underperform. However, considering our small data set size, we propose more research to investigate the potential value of this novel and promising data type and provide the necessary steps and open-source code to do so.

1. Introduction

Digital Mental Health Interventions (DMHIs) are pivotal in expanding much-needed psychological treatment (Andersson et al., 2019). However, high dropout and moderate success rates suggest that this treatment form may not be the optimal choice for all patients (Haller et al., 2023; Lipschitz et al., 2023). Machine Learning (ML) approaches are hoped to offer a possible remedy through individualizing care but require patient-level information to do so (Sajjadian et al., 2021).

Although successful approaches to predicting treatment outcomes have been reported (e.g., Bremer et al., 2020; Forsell et al., 2019; Zantvoort et al., 2023), these rely on features extracted from *already ongoing* treatments. The problem this presents is that, at this time, the costs of treatment start have already occurred for providers and patients. Identifying patients at risk *before* treatment starts would allow the allocation of more promising treatment instead of attempting to make a possibly inferior one work (Günther et al., 2023).

Researchers have attempted to predict whether patients will finish treatment based on pre-treatment features such as symptom scores, socio-demographic variables, or personality scores with little success (Bremer et al., 2020; Gonzalez Salas Duhne et al., 2022; Günther et al., 2023; Linardon et al., 2022; Zantvoort et al., 2024b). Further, meta-analyses such as Vieira et al. (2022) and Sajjadian et al. (2021) report a lack of robust evidence for consistently successful approaches to outcome predictions based on pre-treatment questionnaire data only. Other data types, such as medical imaging, genetic, or heart rate variability data, have not shown much better results, and many of them are resource-intensive to obtain (Hilbert et al., 2024; Hornstein et al., 2022; Sajjadian et al., 2021).

Self-reported measures from questionnaires are a central information

* Corresponding author at: Universitätsallee 1, 21335 Lüneburg, Germany.
E-mail address: jennifer.matthiesen@leuphana.de (J.J. Matthiesen).

https://doi.org/10.1016/j.invent.2025.100828
Received 3 September 2024; Received in revised form 5 April 2025; Accepted 8 April 2025
Available online 9 April 2025

source in psychological research. However, they suffer from human bias in their design, selection, and answer patterns (Breakwell et al., 2020). Further, increasing the number of items affects patients' time requirements, while it often offers a limited marginal informational gain due to high multicollinearity (Tomitaka and Furukawa, 2021). Standard data collection methods for questionnaire data yield no information about the user's decision process, such as indecisiveness, lack of understanding or attention manifesting in changes to an item (Bremer et al., 2020). The potentially lengthy decision processes and times spent on the page to answer the questions are missed when only recording the final choice for each questionnaire item. While methods such as eye-tracking have provided valuable insights into these decision processes (Aimone et al., 2016), they are costly and difficult to scale to real-world scenarios. However, computer mouse cursor movements are highly correlated with gaze and can be unobtrusively and automatically gathered, rendering them a low-cost and scalable alternative (Chen et al., 2001).

Whenever using a computer mouse, users move the cursor across the screen to reach the displayed option of choice. This dynamic decision-making process involves sensorimotor control subsystems to progress the hand toward the desired location. Previous research emphasized the cerebral connections between motor control and emotions, motivation, and decision-making (Yamauchi and Xiao, 2018). As a result, computer mouse dynamics are influenced by factors such as the users' state of confusion (Hucko et al., 2019), level of satisfaction or frustration (Matthiesen and Holte, 2019; Tzafilkou and Protogeros, 2018), and stress and anxiety (Yamauchi, 2013). Further, mouse trajectory data has previously been leveraged as a proxy for attention and motivation (dos Santos and Santana, 2022; Gledson et al., 2021; Leiva and Arapakis, 2020). Given this information value, mouse trajectories constitute a promising candidate for data-driven predictions in DMHIs. However, considering the complexity of spatiotemporal data, the question is whether and in what form such information can be efficiently extracted and leveraged in this clinical context.

This article, therefore, studies the use of mouse trajectory data in DMHIs through the example of baseline online symptom questionnaires. First, we investigate a simple approach that combines hand-crafted features with traditional ML models. Second, we propose a task-specific time-series pre-processing method and apply a state-of-the-art Neural Network to leverage the spatiotemporal data. To illustrate the process, we apply the proposed methodology on mouse movement data from 183 patients filling out a pre-treatment depression symptom questionnaire to predict treatment dropout. As such, this paper aims to introduce for DMHIs novel data type and disseminates the mouse tracker and modelling code necessary to leverage it.

2. Methods

2.1. Interventions and participants

The data for this study was gathered from a subgroup of two ongoing studies on internet-based cognitive behavioural therapy (ICBT) at the Karolinska Institutet in Stockholm, Sweden. The data was collected between the beginning of March 2023 and March 2024, the latter of which is when the treatment platform was taken out of service. All participants gave informed consent to participate in the studies.

The first trial (SOPHIA) evaluates the clinical benefits of an ML-based decision support tool for ICBT for depression, social anxiety, and panic disorder. Patients receive twelve weeks of the respective treatment program, all three of which have previously been described in detail and were evaluated with positive results (El Alaoui et al., 2015; Hedman et al., 2013, 2014). The screening and randomization process is further described in the original study's pre-registration (Bjurner et al., 2024). The second trial (DANA) includes women in pregnancy weeks 8–29 with mild or moderate major depression. Patients receive 10-week therapist-guided pregnancy-adapted ICBT for depression, partly including supportive counselling sessions with perinatal health staff. The screening, randomization, and treatment processes are further detailed in Appendix A.3. As pooling different interventions has been shown to decrease overfitting and improve results (Zantvoort et al., 2024a), all patients are considered as one data set.

2.2. Dropout definition

The number of modules completed is used to operationalize dropout in this study as it has been found to be most relevant for symptom outcomes (Gan et al., 2021). The goal is to identify patients at risk of prematurely leaving the intervention before assumingly sufficiently benefiting (Beintner et al., 2019). For the treatments in the SOPHIA trial, seven modules have previously been identified as suitable dropout threshold (Zantvoort et al., 2024a). As the DANA trial comprised different content, target groups, and treatment design, the cutoff is set to six modules.

2.3. Baseline data

In order to evaluate the additional information value of mouse trajectory data, the prediction results of the mouse data are compared to a minimal baseline approach. The baseline approach uses the depression scores at screening and pre-intervention, age and gender, and target disorder to predict dropout before treatment start (Cabitza and Campagner, 2021). Several papers attempting similar and even more extended feature groups report no or very low predictive power (AUC = 0.50–0.58) (Bremer et al., 2020; Gonzalez Salas Duhne et al., 2022; Günther et al., 2023; Linardon et al., 2022).

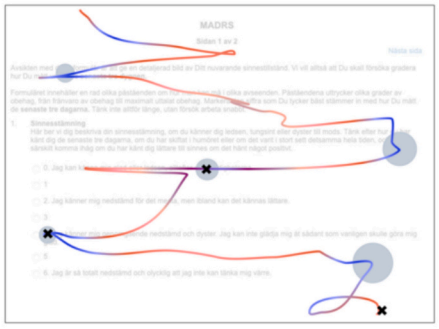
2.4. Mouse data

The tracker to gather mouse trajectories on web applications does not require additional hardware, does not affect the user experience, and is publicly available in the tracker GitHub repository of this paper. It is based on JavaScript and PHP and can be incorporated into any website by inserting the initialization script. The tracker records a mouse cursor's position on the page (x- and y-coordinates) every 40 milliseconds until the page is closed or a pre-set limit of seconds has passed. The tracker further records scroll speed and event names, i.e., 1) whether the mouse is moving, 2) if the user is scrolling and how fast, and 3) clicks as *mouse-up* and *mouse-down* events (Fig. 1). Next to the page coordinates, the client's screen coordinates are tracked in order to relate the mouse movement to the interface and the users' screens, which is needed for normalization.

As the care provider in this study routinely gathers the Montgomery–Åsberg Depression Rating Scale-Self Report (MADRS-S) (Montgomery and Asberg, 1979; Svanborg and Asberg, 1994) across interventions, all mouse dynamics were gathered on this questionnaire. The MADRS-S questionnaire comprises nine questions regarding depressive symptoms (0–6 scale) and, in the setting at hand, is spread over three pages. The inactivity limit to stop tracking is based on the 95 %-percentile (1000 s) of 2500 patients previously filling out the questionnaire (Zantvoort et al., 2024a).

Mouse trajectories are a time-ordered series of x- and y-coordinates such that the raw data holds the l information of a user's decision process. We study two different computational representations of this complex data: i) as aggregates in form of hand-crafted features, and ii) as time series.

Hand-crafted Features: Firstly, raw mouse trajectories can be aggregated into different hand-crafted features. Every feature is designed to capture a certain aspect of the trajectory (e.g., length, speed, area), and the entirety of features spans a vector space where every trajectory is represented by a position vector. Hence, geometric (e.g., Euclidian) distances are applicable to compare trajectories in the space, and non-sequential models such as Logistic Regression or tree-based models are



timestamp	x-coord (client)	y-coord (client)	mouseaction	scroll speed	
1612448647902	361	266	mousemove	0	movement
1612448647942	343	290	mousemove	0	
1612448647982	307	315	mousemove	0	
1612448648022	282	315	pause	0	pause
1612448648062	282	315	pause	0	
1612448648102	282	315	pause	0	
1612448648142	282	315	mousedown	0	click
1612448648182	282	315	pause	0	
1612448648222	282	315	mouseup	0	
1612448647783	282	315	scroll	16	scroll up & down
1612448647823	282	315	scroll	22	
1612448647863	282	315	scroll	-18	

Fig. 1. Example of mouse data representation of a single user on the interface and the dimensions described above. Page coordinates are not displayed.

directly compatible.

As shown in Fig. 2, ample hand-crafted feature options are possible and have proven beneficial in producing insights and predictions (e.g., Arapakis and Leiva, 2016; Matthiesen and Brefeld, 2020; Yamauchi, 2013).

A first frequently used feature group encodes temporal information of a given event, for example pauses occurring due to hesitation (Arapakis and Leiva, 2016; Matthiesen and Brefeld, 2020). Second, events, such as the number of clicks may reflect aspects like a patient changing their answer many times. Third, speed features can indicate the distances crossed by the mouse, where fast movement could, for example, be an indication of frustration or lack of patience (Yamauchi, 2013). Lastly, this can be further enriched by more complex movement patterns including angles, curvature or jitter, which can, for example, reflect a user's stress level (Martín-Albo et al., 2016).

Parallel to intervention login data, hand-crafted mouse features can be aggregated in different ways, such as sums, means, or extreme values (Bremer et al., 2020). Which of these options makes the most sense depends on the nature of the feature at hand. For example, average and maximum could be interesting for values as speed, while the minimum will likely be 0 for all users.

A key drawback of hand-crafted features is the time-intensive manual preprocessing and consequent bias in their selection. However, a critical upside is the transformation into meaningful, humanly interpretable features that require less computational resources.

While ML models can theoretically handle large and complex feature groups, in DMHI settings, fewer hand-crafted features have been shown to improve results, especially for small data sets (Bremer et al., 2020; Hentati Isacsson et al., 2023; Zantvoort et al., 2024b). Following related work (Arapakis and Leiva, 2016; Feher et al., 2012; Matthiesen and Brefeld, 2020), we, therefore, focus on two subsets of three (F3) and ten (F10) of the most promising features. The small group (F3) only focuses on the basic temporal aspects: Average speed, total time of pauses, and scrolling speed. The larger (F10) extends the previous one by the moved distance, the number of data points, the average change in angle, the amount of acute and obtuse angles, and jitter). A detailed overview of the feature calculations can be found in Appendix A.1 and the code to produce them on data sets from the above-mentioned tracker is available in this study's GitHub repository.

Time-series representation: Secondly, mouse trajectory data are usually modelled as time series as they naturally form a sequence of coordinates representing the change of x and y positions of the mouse pointer over time. In this representation, every mouse trajectory is thus given by a one-dimensional time series with two-dimensional elements (x_t, y_t), where t indexes time. To aggregate spatial and temporal aspects of the data, derivatives $\delta x/\delta t$ and $\delta y/\delta t$ can be calculated to inform about directions and velocities of movements, respectively. As tracking pauses do not serve a purpose in the time series, they are removed. Such time series representations of mouse data have been found to lead to reasonable outcomes while not being particularly demanding in terms of computational infrastructure (Antal et al., 2021; Chong et al., 2019).

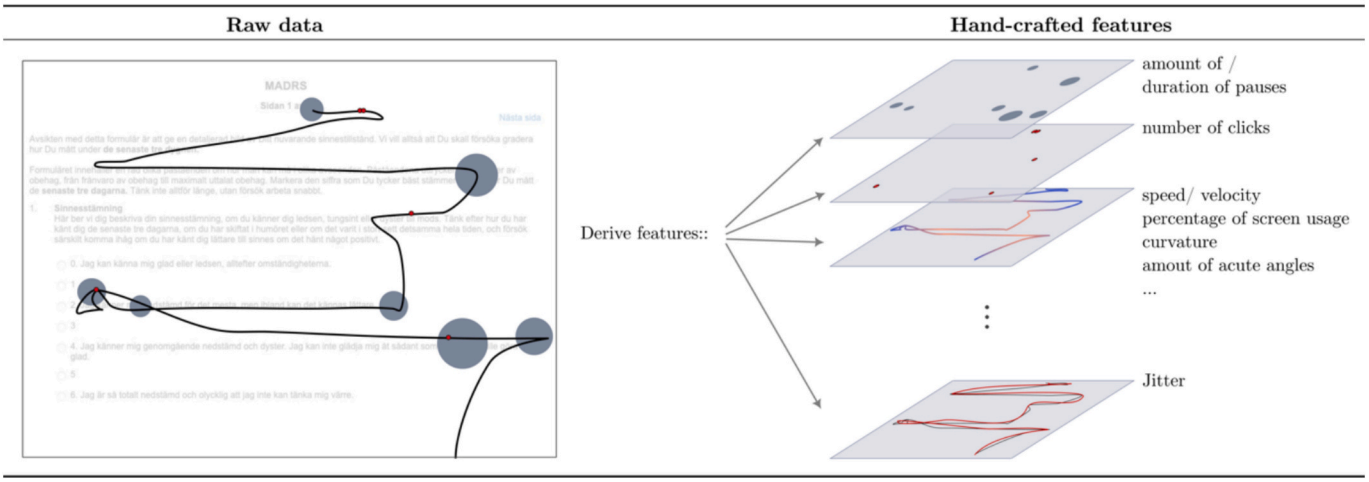


Fig. 2. Depiction of a) the plotted raw data including the trajectory, pauses, and clicks and b) the derived hand-crafted features.

2.5. Prediction models and experiment setup

To mitigate the risk of overfitting on the small data set (Bates et al., 2022), the models are trained and evaluated via a 5 × 5-fold nested-cross validation (CV) with five different seeds. Furthermore, both pipelines include a standard scaler. Regarding evaluation, more sophisticated features need to outperform simpler ones to warrant the additional effort needed. As a baseline, we compare our results to the reported area under the curve (AUC) of 0.57 for dropout predictions by Günther et al. (2023).

Non-sequential Models: To analyse the hand-crafted features, a Random Forest model is trained in a trade-off between simple but flexible enough to detect non-linear and interaction effects. Missing depression scores are imputed by assuming the mean change between screening and pre-treatment in the training data. The random forest is parameterized by the number of estimators {3, 5, 10, 25}, minimum samples for a split {5, 15, 25, 50}, the maximum depths {3, 5, 10, 25, 50, and a binary bootstrapping activation flag.

Sequential Neural Network: In order to leverage the sequentiality of a time-series representation, a one-dimensional convolutional neural network (1D-CNN) is implemented in this study (Antal et al., 2021; Chong et al., 2019). The architecture comprises the concatenation of two towers of each two convolutional filters, one dropout layer and two fully connected layers, and a sigmoid activation function (Fig. 3).

Since neural networks accept only inputs of the same length, a key challenge for the time-series approach is posed by the (very) different lengths of the trajectories across users. For example, some users may be familiar with the questionnaire and click through it within seconds, whereas others take several minutes to answer. Typical approaches employ padding for short and truncating for longer sequences (Arapakis and Leiva, 2020). However, this approach is wasteful when the variance of the lengths of the trajectories is as large as in the case at hand. Moreover, in DMHI settings, sequences are rather long in general and need to be classified in their entirety. Strategies involving a single prediction per equal-length segmentation of the trajectories (Antal et al., 2021; Chong et al., 2019) are not applicable, as they risk target leakage (Leiva and Arapakis, 2020). To address these two problems, we propose to first split the trajectories into equally sized small blocks but rejoin them before updating the network (see further details in Appendix A.2). As a result, we avoid target leakage while ensuring adequate handling of differently sized trajectory lengths. Additionally, using a sliding window approach retains the context of each data point when fed into the model (Chong et al., 2019).

In terms of hyperparameters, the sliding window has a fixed size and is used with an overlap of 50 % with the previous sequence. We experiment with different block sizes {100, 128, 265} to split the trajectories per user are tested within the inner fold of the nested CV. Further, the model is trained for 35 epochs with a learning rate of 0.001. Lastly, we use the Adam optimizer and binary cross entropy as a loss function.

3. Results

3.1. Final data sets

The final baseline data set including the questionnaire data before treatment comprises 408 patients, 55 % of which used a mobile device and hence, did not produce mouse trajectories. Therefore, the mouse trajectory analysis includes 184 desktop computer or laptop users. One user had insufficient mouse data points (<10) and was, therefore, excluded. The final descriptive numbers across studies can be seen in Table 1. For the screening and pre-intervention MADRS-S scores, there were 13 (2.5 %) and 17 (4.2 %) missing values, but no patient missed

Table 1
Number of participants and dropout rates per intervention for a) baseline data and b) mouse data.

	Target disorder	N	Dropout in %	Mean age (SD)	Females (%)
a) Mouse data					
Dana	Depression	75	48 %	34.67 (3.83)	100 %
Sophia	Depression	52	44 %	42.25 (10.51)	54 %
	Social anxiety	36	50 %	35.97 (10.30)	78 %
	Panic disorder	20	31 %	42.40 (12.58)	50 %
b) Baseline data					
Dana	Depression	144	56 %	34.28 (3.95)	100 %
Sophia	Depression	115	54 %	41.54(10.59)	66 %
	Social anxiety	85	63 %	36.21 (9.72)	74 %
	Panic disorder	64	38 %	41.09 (10.98)	58 %

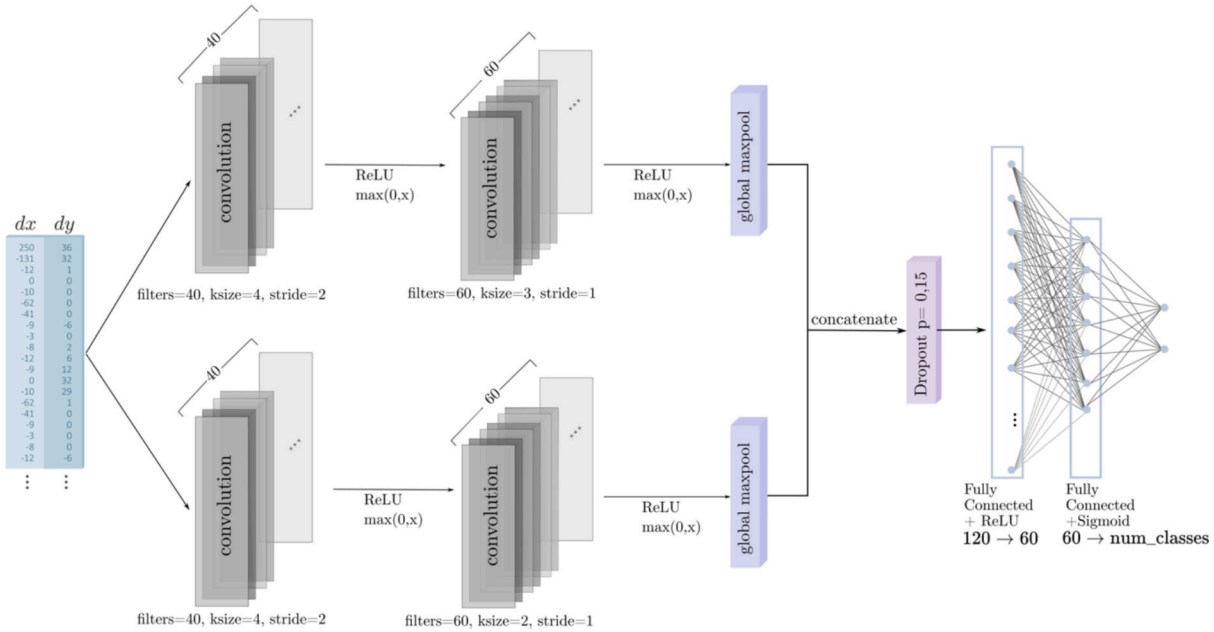


Fig. 3. The architecture of the 1-dimensional convolutional neural network.

both time points. Two-sided Student's tests comparing the mean values of the mouse and touch device groups showed no statistically significant differences ($p > 0.05$) and the detailed analyses and results can be found in Appendix A.4.

3.1.1. Prediction results

Providing the model with only questionnaire answers (baseline) resulted in an average inner CV (training) score of 0.59 and outer (test) score of 0.56 (95 %-CI 0.54–0.57) AUC. Only adding the average speed, pause time, and scroll speed (F3), increased the inner CV to 0.66 and the outer CV score to 0.58 (95 %-CI 0.55–0.62). However, adding the ten hand-crafted mouse features (F10) increased the inner CV score to 0.65 but did not improve the outer CV score at 0.56 AUC (95 %-CI 0.52–0.61). Finally, the neural network with the time series representation produced an inner CV score of 0.51 and an outer CV score of 0.50 (95 %-CI 0.44–0.54).

4. Discussion

Mouse trajectories' information power regarding users' mindset and emotional state (Fu et al., 2017; Yamauchi, 2013) renders them a promising candidate for predictions in mental health care. However, they are complex in nature and no insights into their use and value in DMHIs exist so far. Therefore, the study at hand discusses how to gather and process mouse trajectory data and provides an exemplary use case for intervention dropout predictions. Specifically, the present work includes detailed steps accompanied by source code to gather raw trajectories, extract selected hand-crafted features from them, leverage them as time series, and pair them with appropriate prediction models.

The baseline prediction reached 0.56 AUC, which is competitive with findings in related work (Gonzalez Salas Duhne et al., 2022; Günther et al., 2023). Our small-scale intervention dropout data set, however, did not suffice to exploit the potential of mouse data. Compared to the baseline, the three basic hand-crafted features (F3) increased the prediction results only marginally (+0.01 AUC) and the ten extended features (F10) and time-series approach did not achieve convincing results at all. As such, this paper is in line with several works finding more sophisticated models to fail on small-sized e-mental health data (Gogoulou et al., 2021; Smink et al., 2021; Zantvoort et al., 2023; Zantvoort et al., 2024b).

Although our data is not considered small for DMHIs with median data sets around 155–350 patients (Hornstein et al., 2023; Karyotaki et al., 2021), neural networks are data-hungry and require much larger data sets (Alzubaidi et al., 2023; Piccialli et al., 2021). This holds particularly true for complex tasks such as predicting user behaviour several weeks in the future (Gogoulou et al., 2021). Further, small data sets risk overfitting, especially when paired with inadequate validation methods and flexible models (Bates et al., 2022). Single test set and even CV results vary widely, with higher, improbable results likely being overrepresented in publications (Hullman et al., 2022; Piccialli et al., 2021; Zantvoort et al., 2024b). Consequently, concerns about the generalizability of prediction results in mental health are increasing (Hilbert et al., 2024; Sajjadian et al., 2021; Zantvoort et al., 2024b).

To mitigate overfitting, we deployed a nested CV where the inner loop optimizes hyperparameters and the outer loop estimates model performance on independent data (Bates et al., 2022). We further repeated the experiment and reported averages and confidence intervals to account for the randomness in the models due to the small data (Cabitza and Campagner, 2021). We found that overfitting is a substantial risk for the mouse features as the inner CV score increased by up to 0.09 AUC but this performance gain did not generalize to the outer CV score. Overfitting is presumably also the issue of the neural networks, as the high training performance of 0.90 AUC did not generalize to the test scores (0.50 AUC). As such, the complex models seemed to memorize the data but never learned to generalize. Relying on an inadequate validation setup with the data set size at hand would have certainly led to a vast overestimation of the performance (Hilbert et al., 2024; Sajjadian et al., 2021; Zantvoort et al., 2024b). At the same time, our simplest model with only three handcrafted features (F3) slightly improved the performance of the baseline and related work, indicating the possibility of information value. Therefore, future research needs be conducted on larger mouse trajectory datasets to investigate the power of these complex features (Zantvoort et al., 2024b).

Beyond the data set size, this study has several limitations. Firstly, the recorded data neither accounted for different devices (e.g., touchpad, vertical, or standard mouse) nor artifacts caused by the hardware. Secondly, albeit proposed to increase prediction accuracy and stability (Zantvoort et al., 2024a), the heterogeneity of the different interventions could be too pronounced considering the small data set size at hand. Thirdly, we only explore the prediction power of mouse trajectories on the pre-treatment depression questionnaires. MADRS-S is a very limited part of the assessment, especially as it is not the primary symptom for all patients. Future research could explore the second proposed use case of leveraging mouse trajectories to improve adherence measures, for example, through pauses, the explored content, or predicted attention (Leiva and Arapakis, 2020). Fourthly, other ways of processing the mouse data are possible, most noticeably via spatial approaches (e.g., shape and patterns such as loops) by plotting the trajectories on a two-dimensional positioning plane, thus representing them as images. Accordingly, image processing models such as two-dimensional convolutional neural networks (2D-CNN) could be paired with this form of data representation (Arapakis and Leiva, 2020; Chong et al., 2019). A vital drawback of this method is the lack of consideration of the underlying user interface, such that differently sized screens yield varying images of the same trajectories. Additionally, creating images from the mouse trajectories is a resource-intensive step, because of which it was out of the scope of this study. Further, even within time-series approaches, other options such as Long-Term-Short-Term memory models (Arapakis and Leiva, 2020; Hochreiter and Schmidhuber, 1997) may yield different results. Lastly, more than half of the patients accessed the questionnaire via mobile devices, emphasizing the importance of also investigating touch data (Yang et al., 2021). Only by combining both input channels can predictions for all patients be made. However, desktop and laptop devices still hold a significant share of web traffic in Europe (StatCounter, 2025). Further, the intervention provider informs participants that they are required to use a desktop computer or laptop to complete the intervention. As it makes access to such a device already a requirement for participants, they could be asked to fill out the questionnaire with it in the future.

In conclusion, the paper at hand introduces mouse trajectory data for DMHIs and provides interested researchers with the code to gather and use the obtained representations for machine learning predictions via a robust experimental framework. Our dropout prediction results showed that hand-crafted features slightly improved results while the more complex models suffered from the lack of data. Based on related work we conclude that 183 patients are arguably too few to leverage complex feature and model combinations (Zantvoort et al., 2024a; Zantvoort et al., 2024b). Considering their successful implementation in other domains, more research on larger data sets is necessary to determine the predictive value of mouse trajectories as an unobtrusive and rich data type in DMHI predictions.

Ethical approval and consent

The Swedish Ethical Review Authority, the national governmental body responsible for all research-related ethical applications in Sweden, has approved the study Stockholm (SOPHIA-study DNR: 2011/2091-31/3, amendment 2016/21-32, 2017/2320-32 and 2018/2550-32), (DANA-study DNR: 2020-06049, amendments: DNR 2022-04575-02 and 2024-00796-02). In accordance with Swedish data privacy laws, the opt-out consent routine applies to all patients at the routine health care service i.e., the Internet Psychiatry Unit.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Marie Bendix reports financial support was provided by Soderstrom Konigska Hospital Foundation. Marie Bendix, Viktor Kaldo, Pontus Bjurner report financial support was provided by Fredrik and Ingrid Thurings Foundation. Marie Bendix reports financial support was provided by FORTE research council. Marie Bendix reports financial support was provided by Foundation Professor Bror Gadelius Memorial Fund. Pontus Bjurner, Viktor Kaldo report financial support was provided by The Swedish Research Council (VR). Pontus Bjurner, Viktor Kaldo report financial support was provided by The Erling Persson family foundation (EP-Stiftelsen). Pontus Bjurner, Viktor Kaldo report was provided by Swedish government. Pontus Bjurner, Viktor Kaldo report financial support was provided by Psykiatrifonden. Jennifer J. Matthiesen, Kirsten Zantvoort report financial support was provided by German Research Foundation (DFG). Jennifer J. Matthiesen reports a relationship with Mouseflow ApS that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We want to thank Mathias Leblanc for the support in implementing the mouse tracker in the study platform.

Code availability

The underlying code for the tracker, mouse feature generation and model training are available without restrictions in the projects GitHub repositories.

Tracker: https://github.com/jjmatthiesen/evtracker

Features and Analysis: https://github.com/jjmatthiesen/MouseTrajectoriesICBT

Funding

This work was mainly supported by The Swedish Research Council (VR), The Erling Persson family foundation (EP-Stiftelsen), and The Swedish ALF-agreement between the Swedish government and the county councils, with additional funding by Psykiatrifonden and Thuring's Foundation. The Dana study was (additionally) funded by Söderström Königska Foundation, FORTE research council, and Gadelius Foundation. The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript. This open-access publication was funded by the German Research Foundation (DFG).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.invent.2025.100828.

Data availability

The datasets analysed in the current study are not publicly available due to the sensitivity of health data and data privacy requirements.

References

Aimone, J.A., Ball, S., King-Casas, B., 2016. It's not what you see but how you see it: using eye-tracking to study the risky decision-making process. J. Neurosci. Psychol. Econ. 9 (3–4), 137–144. https://doi.org/10.1037/npe0000061.

Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A.S., Al-dabbagh, B.S.N., Fadhel, M.A., Manoufali, M., Zhang, J., Al-Timemy, A.H., Duan, Y., Abdullah, A., Farhan, L., Lu, Y., Gupta, A., Albu, F., Abbosh, A., Gu, Y., 2023. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. J. Big Data 10 (1), 46. https://doi.org/10.1186/s40537-023-00727-2.

Andersson, G., Carlbring, P., Rozental, A., 2019. Response and remission rates in Internet-based cognitive behavior therapy: an individual patient data meta-analysis. Front. Psych. 10, 749. https://doi.org/10.3389/fpsyt.2019.00749.

Antal, M., Fejér, N., Buza, K., 2021. SapiMouse: mouse dynamics-based user authentication using deep feature learning. In: 2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI), pp. 61–66.

Arapakis, I., Leiva, L.A., 2016. Predicting user engagement with direct displays using mouse cursor information. In: Proceedings of the 39th Int. ACM SIGIR Conf. on Research and Development in Inform, pp. 599–608. Retrieval.

Arapakis, I., Leiva, L.A., 2020. Learning efficient representations of mouse movements to predict user attention. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1309–1318. https://doi.org/10.1145/3397271.3401031.

Bates, S., Hastie, T., Tibshirani, R., 2022. Cross-validation: what does it estimate and how well does it do it? (arXiv:2104.00673). arXiv. http://arxiv.org/abs/2104.00673.

Beintner, I., Vollert, B., Zarski, A.-C., Bolinski, F., Musiat, P., Jacobi, C., 2019. Adherence reporting in randomized controlled trials examining manualized multisession online interventions: systematic review of practices and proposal for reporting standards. J. Med. Internet Res. 21 (8), e14181. https://doi.org/10.2196/14181. Aug 15.

Bjurner, P., Isacsson, N.H., Abdesslem, F.B., Boman, M., Forsell, E., Kaldo, V., 2024. Study protocol for a triple-blind randomised controlled trial evaluating a machine learning-based predictive clinical decision support tool for internet-delivered cognitive behaviour therapy (ICBT) for depression and anxiety. https://osf.io/cs4bx/.

Breakwell, G.M., Barnett, J., Wright, D.B., 2020. Research Methods in Psychology, 5th ed. Sage Publications https://www.torrossa.com/en/resources/an/5019108.

Bremer, V., Chow, P.I., Funk, B., Thorndike, F.P., Ritterband, L.M., 2020. Developing a process for the analysis of user journeys and the prediction of dropout in digital health interventions: machine learning approach. J. Med. Internet Res. 22 (10). https://doi.org/10.2196/17738.

Cabitza, F., Campagner, A., 2021. The need to separate the wheat from the chaff in medical informatics. Int. J. Med. Inform. 153. https://doi.org/10.1016/j.ijmedinf.2021.104510.

Chen, M.C., Anderson, J.R., Sohn, M.H., 2001. What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing. In: CHI '01 Extended Abstracts on Human Factors in Computing Systems, pp. 281–282. https://doi.org/10.1145/634067.634234.

Chong, P., Elovici, Y., Binder, A., 2019. User authentication based on mouse dynamics using deep neural networks: a comprehensive study. IEEE Trans. Inf. Forensics Secur. 15, 1086–1101.

dos Santos, T., Santana, V., 2022. Identifying distractors for people with computer anxiety based on mouse fixations. Interact. Comput. 35. https://doi.org/10.1093/iwc/iwac025.

El Alaoui, S., Hedman, E., Kaldo, V., Hesser, H., Kraepelien, M., Andersson, E., Rück, C., Andersson, G., Ljótsson, B., Lindefors, N., 2015. Effectiveness of Internet-based cognitive-behavior therapy for social anxiety disorder in clinical psychiatry. J. Consult. Clin. Psychol. 83 (5), 902–914. https://doi.org/10.1037/a0039198.

Feher, C., Elovici, Y., Moskovitch, R., Rokach, L., Schclar, A., 2012. User identity verification via mouse dynamics. Inform. Sci. 201, 19–36.

Forsell, E., Jernelöv, S., Blom, K., Kraepelien, M., Svanborg, C., Andersson, G., Lindefors, N., Kaldo, V., 2019. Proof of concept for an adaptive treatment strategy to prevent failures in internet-delivered CBT: a single-blind randomized clinical trial with insomnia patients. Am. J. Psychiatry 176 (4), 315–323. https://doi.org/10.1176/appi.ajp.2018.18060699.

Fu, E.Y., Kwok, T.C.K., Wu, E.Y., Leong, H.V., Ngai, G., Chan, S.C.F., 2017. Your mouse reveals your next activity: towards predicting user intention from mouse interaction. In: 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), 1, pp. 869–874. https://doi.org/10.1109/COMPSAC.2017.270.

Gan, D.Z.Q., McGillivray, L., Han, J., Christensen, H., Torok, M., 2021. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Frontiers in Digital Health 3, 764079. https://doi.org/10.3389/fdgth.2021.764079.

Gledson, A., Apaolaza, A., Barthold, S., Günther, F., Yu, H., Vigo, M., 2021. Characterising student engagement modes through low-level activity patterns. In: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, pp. 88–97. https://doi.org/10.1145/3450613.3456818.

Gogoulou, E., Boman, M., Ben Abdesslem, F., Hentati Isacsson, N., Kaldo, V., Sahlgren, M., 2021. Predicting treatment outcome from patient texts:the case of internet-based cognitive behavioural therapy. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, pp. 575–580. https://doi.org/10.18653/v1/2021.eacl-main.46.

Gonzalez Salas Duhne, P., Delgadillo, J., Lutz, W., 2022. Predicting early dropout in online versus face-to-face guided self-help: a machine learning approach. Behav. Res. Ther. 159, 104200. https://doi.org/10.1016/j.brat.2022.104200.

Günther, F., Yau, C., Elison-Davies, S., Wong, D., 2023. On the difficulty of predicting engagement with digital health for substance use. Stud. Health Technol. Inform. 302, 967–971. https://doi.org/10.3233/SHTI230319.

Haller, K., Becker, P., Niemeyer, H., Boettcher, J., 2023. Who benefits from guided internet-based interventions? A systematic review of predictors and moderators of treatment outcome. Internet Interv. 33, 100635. https://doi.org/10.1016/j.invent.2023.100635.

Hedman, E., Ljótsson, B., Rück, C., Bergström, J., Andersson, G., Kaldo, V., Jansson, L., Andersson, E., Andersson, E., Blom, K., El Alaoui, S., Falk, L., Ivarsson, J., Nasri, B., Rydh, S., Lindefors, N., 2013. Effectiveness of Internet-based cognitive behaviour therapy for panic disorder in routine psychiatric care. Acta Psychiatr. Scand. 128 (6), 457–467. https://doi.org/10.1111/acps.12079.

Hedman, E., Ljótsson, B., Kaldo, V., Hesser, H., El Alaoui, S., Kraepelien, M., Andersson, E., Rück, C., Svanborg, C., Andersson, G., Lindefors, N., 2014. Effectiveness of Internet-based cognitive behaviour therapy for depression in routine psychiatric care. J. Affect. Disord. 155, 49–58. https://doi.org/10.1016/j.jad.2013.10.023.

Hentati Isacsson, N., Abdesslem, F.B., Forsell, E., Boman, M., Kaldo, V., 2023. Machine Learning Predictions of Outcome in Internet-based Cognitive Behavioral Therapy: Methodological Choices and Clinical Usefulness. https://doi.org/10.21203/rs.3.rs-2751455/v1.

Hilbert, K., Böhnlein, J., Meinke, C., Chavanne, A., Langhammer, T., Stumpe, L., Winter, N., Leenings, R., Adolph, D., Arolt, V., Bischoff, S., Cwik, J., Deckert, J., Domschke, K., Fydrich, T., Gathmann, B., Hamm, A., Heinig, I., Herrmann, M., Lueken, U., 2024. Lack of evidence for predictive utility from resting state fMRI data for individual exposure-based cognitive behavioral therapy outcomes: a machine learning study in two large multi-site samples in anxiety disorders. NeuroImage 295, 120639. https://doi.org/10.1016/j.neuroimage.2024.120639.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Hornstein, S., Seiler, M., Hoffman, V., Nelson, B., Aschbacher, K., Ritter, K., Hilbert, K., 2022. Association of Depressive Symptoms with Resting Heart Rate Variability Recorded From a Wearable Device under Naturalistic Conditions: A Machine Learning Study. https://doi.org/10.31234/osf.io/9z3pr.

Hornstein, S., Zantvoort, K., Ulrike, L., Funk, B., Hilbert, Kevin, 2023. Personalization strategies in digital mental health interventions: a systematic review and conceptual framework for depressive symptoms. Front. Digit. Health 5. https://doi.org/10.3389/fdgth.2023.1170002.

Hucko, M., Gazo, L., Simun, P., Valky, M., Moro, R., Simko, J., Bielikova, M., 2019. YesElf: personalized onboarding for web applications. In: Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, pp. 39–44. https://doi.org/10.1145/3314183.3324978.

Hullman, J., Kapoor, S., Nanayakkara, P., Gelman, A., Narayanan, A., 2022. The worst of both worlds: a comparative analysis of errors in learning from data in psychology and machine learning. In: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, pp. 335–348. https://doi.org/10.1145/3514094.3534196.

Karyotaki, E., Efthimiou, O., Miguel, C., Bermpohl, F.M. genannt, Furukawa, T.A., Cuijpers, P., Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, 2021. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 78 (4), 361–371. https://doi.org/10.1001/jamapsychiatry.2020.4364.

Leiva, L., Arapakis, I., 2020. The attentive cursor dataset. Front. Hum. Neurosci. 14. https://doi.org/10.3389/fnhum.2020.565664.

Linardon, J., Fuller-Tyszkiewicz, M., Shatte, A., Greenwood, C.J., 2022. An exploratory application of machine learning methods to optimize prediction of responsiveness to digital interventions for eating disorder symptoms. Int. J. Eat. Disord. 55 (6), 845–850. https://doi.org/10.1002/eat.23733.

Lipschitz, J., Pike, C., Hogan, T., Murphy, S., Burdick, K., 2023. The engagement problem: a review of engagement with digital mental health interventions and recommendations for a path forward. Curr. Treat. Options Psychiatry 10. https://doi.org/10.1007/s40501-023-00297-3.

Martín-Albo, D., Leiva, L.A., Huang, J., Plamondon, R., 2016. Strokes of insight: user intent detection and kinematic compression of mouse cursor trails. Inf. Process. Manag. 52 (6), 989–1003. https://doi.org/10.1016/j.ipm.2016.04.005.

Matthiesen, J.J., Brefeld, U., 2020. Assessing user behavior by mouse movements. In: Stephanidis, C., Antona, M. (Eds.), HCI International 2020—Posters. Springer International Publishing, pp. 68–75.

Matthiesen, J., Holte, M.B., 2019. Simple mouse attribute analysis: HCIBGO: 6th international conference on HCI in business, government and organizations under HCI international 2019. In: HCI in Business, Government and Organizations. Information Systems and Analytics, pp. 95–113. https://doi.org/10.1007/978-3-030-22338-0_8.

Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. Br. J. Psychiatry J. Ment. Sci. 134, 382–389. https://doi.org/10.1192/bjp.134.4.382.

Piccialli, F., Somma, V.D., Giampaolo, F., Cuomo, S., Fortino, G., 2021. A survey on deep learning in medicine: why, how and when? Information Fusion 66, 111–137. https://doi.org/10.1016/j.inffus.2020.09.006.

Sajjadian, M., Lam, R.W., Milev, R., Rotzinger, S., Frey, B.N., Soares, C.N., Parikh, S.V., Foster, J.A., Turecki, G., Müller, D.J., Strother, S.C., Farzan, F., Kennedy, S.H., Uher, R., 2021. Machine learning in the prediction of depression treatment outcomes: a systematic review and meta-analysis. Psychol. Med. 51 (16), 2742–2751. https://doi.org/10.1017/S0033291721003871.

Smink, W.A.C., Sools, A.M., Postel, M.G., Tjong Kim Sang, E., Elfrink, A., Libbertz-Mohr, L.B., Veldkamp, B.P., Westerhof, G.J., 2021. Analysis of the emails from the Dutch web-based intervention "alcohol de baas": assessment of early indications of drop-out in an online alcohol abuse intervention. Front. Psych. 12, 575931. https://doi.org/10.3389/fpsyt.2021.575931.

StatCounter, 2025. Desktop vs Mobile vs Tablet Market Share Europe 2024 [Dataset]. https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet/europe.

Svanborg, P., Asberg, M., 1994. A new self-rating scale for depression and anxiety states based on the Comprehensive Psychopathological Rating Scale. Acta Psychiatr. Scand. 89 (1), 21–28. https://doi.org/10.1111/j.1600-0447.1994.tb01480.x.

Tomitaka, S., Furukawa, T.A., 2021. The GAD-7 and the PHQ-8 exhibit the same mathematical pattern of item responses in the general population: analysis of data from the National Health Interview Survey. BMC Psychology 9, 149. https://doi.org/10.1186/s40359-021-00657-9.

Tzafilkou, K., Protogeros, N., 2018. Mouse behavioral patterns and keystroke dynamics in end-user development: what can they tell us? Computers in Human Behavior 83, 288–305.

Vieira, S., Liang, X., Guiomar, R., Mechelli, A., 2022. Can we predict who will benefit from cognitive-behavioural therapy? A systematic review and meta-analysis of machine learning studies. Clin. Psych. Rev. 97, 102193.

Yamauchi, T., 2013. Mouse trajectories and state anxiety: feature selection with random forest. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, pp. 399–404. https://doi.org/10.1109/ACII.2013.72.

Yamauchi, T., Xiao, K., 2018. Reading emotion from mouse cursor motions: a effective computing approach. Cognit. Sci. 42 (3), 711–819. https://doi.org/10.1111/cogs.12557.

Yang, W., Wang, M., Zou, S., Peng, J., Xu, G., 2021. An implicit identity authentication method based on deep connected attention CNN for wild environment. In: Proceedings of the 2021 9th International Conference on Communications and Broadband Networking, pp. 94–100. https://doi.org/10.1145/3456415.3457222.

Zantvoort, K., Scharfenberger, J., Boß, L., Lehr, D., Funk, B., 2023. Finding the best match—a case study on the (text-)feature and model choice in digital mental health interventions. Journal of Healthcare Informatics Research 7 (4), 447–479. https://doi.org/10.1007/s41666-023-00148-z.

Zantvoort, K., Hentati Isacsson, N., Funk, B., Kaldo, V., 2024a. Data set size vs homogeneity – a machine learning study on pooling intervention data in E-mental health dropout predictions. SAGE Digital Health 10, 1–11. https://doi.org/10.1177/20552076241248920.

Zantvoort, K., Nacke, B., Görlich, D., Hornstein, S., Jacobi, C., Funk, B., 2024b. Estimation of minimal data sets sizes for machine learning predictions in digital mental health interventions. Npj Digital Medicine 7 (1), 1–10. https://doi.org/10.1038/s41746-024-01360-w.