

Original Paper

Continuous Assessment of Function and Disability via Mobile Sensing: Real-World Data-Driven Feasibility Study

Emese Sükei¹, MSc; Lorena Romero-Medrano^{1,2}, MSc; Santiago de Leon-Martinez^{1,3,4}, MSc; Jesús Herrera López^{1,2}, MSc; Juan José Campaña-Montes², MSc; Pablo M Olmos^{1,5}, PhD; Enrique Baca-García^{2,6,7,8,9,10,11,12,13}, MD, PhD; Antonio Artés^{1,2,5}, PhD

¹Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Leganés, Spain

²Evidence-Based Behavior S.L., Leganés, Spain

³Kempelen Institute of Intelligent Technologies, Bratislava, Slovakia

⁴Faculty of Information Technology, Brno University of Technology, Brno, Czech Republic

⁵Grupo de Tratamiento de Señal, Gregorio Marañón Health Research Institute, Madrid, Spain

⁶Department of Psychiatry, University Hospital Rey Juan Carlos, Móstoles, Spain

⁷Department of Psychiatry, General Hospital of Villalba, Madrid, Spain

⁸Department of Psychiatry, University Hospital Infanta Elena, Madrid, Spain

⁹Department of Psychiatry, Madrid Autonomous University, Madrid, Spain

¹⁰Centro de Investigación en Salud Mental, Carlos III Institute of Health, Madrid, Spain

¹¹Department of Psychiatry, Universidad Católica del Maule, Madrid, Spain

¹²Department of Psychiatry, Centre Hospitalier Universitaire, Nîmes, France

¹³Department of Psychiatry, University Hospital Jimenez Diaz Foundation, Madrid, Spain

Corresponding Author:

Emese Sükei, MSc

Department of Signal Theory and Communications

Universidad Carlos III de Madrid

Ed. Torres Quevedo, 3rd Fl.

Av. de la Universidad, 30

Leganés, 28911

Spain

Phone: 34 916 248 741

Email: esukei@tsc.uc3m.es

Abstract

Background: Functional limitations are associated with poor clinical outcomes, higher mortality, and disability rates, especially in older adults. Continuous assessment of patients' functionality is important for clinical practice; however, traditional questionnaire-based assessment methods are very time-consuming and infrequently used. Mobile sensing offers a great range of sources that can assess function and disability daily.

Objective: This work aims to prove the feasibility of an interpretable machine learning pipeline for predicting function and disability based on the World Health Organization Disability Assessment Schedule (WHODAS) 2.0 outcomes of clinical outpatients, using passively collected digital biomarkers.

Methods: One-month-long behavioral time-series data consisting of physical and digital activity descriptor variables were summarized using statistical measures (minimum, maximum, mean, median, SD, and IQR), creating 64 features that were used for prediction. We then applied a sequential feature selection to each WHODAS 2.0 domain (cognition, mobility, self-care, getting along, life activities, and participation) in order to find the most descriptive features for each domain. Finally, we predicted the WHODAS 2.0 functional domain scores using linear regression using the best feature subsets. We reported the mean absolute errors and the mean absolute percentage errors over 4 folds as goodness-of-fit statistics to evaluate the model and allow for between-domain performance comparison.

Results: Our machine learning-based models for predicting patients' WHODAS functionality scores per domain achieved an average (across the 6 domains) mean absolute percentage error of 19.5%, varying between 14.86% (self-care domain) and 27.21%

(life activities domain). We found that 5-19 features were sufficient for each domain, and the most relevant being the distance traveled, time spent at home, time spent walking, exercise time, and vehicle time.

Conclusions: Our findings show the feasibility of using machine learning–based methods to assess functional health solely from passively sensed mobile data. The feature selection step provides a set of interpretable features for each domain, ensuring better explainability to the models’ decisions—an important aspect in clinical practice.

(*JMIR Form Res* 2023;7:e47167) doi: [10.2196/47167](https://doi.org/10.2196/47167)

KEYWORDS

WHODAS; functional limitations; mobile sensing; passive ecological momentary assessment; predictive modeling; interpretable machine learning; machine learning; disability; clinical outcome

Introduction

Background

Functional limitations are associated with poor clinical outcomes, higher mortality, and disability rates, especially in older adults [1]. Moreover, they are closely related and used for predicting transitions in activities of daily living or instrumental activities of daily living due to a disability, thereby significantly impacting the quality of life of older adults and other age groups [2]. COVID-19 has been associated with functional limitations in patients with post-COVID-19 sequelae, further increasing an already present problem for older adults [3-5]. Sarcopenia, progressive muscle loss due to aging, is one of the common functionality problems for older adults with an estimated cost of hospitalizations for adults in the United States of US \$40.4 billion [6]. Early detection of an increase in disability is essential for clinical practice. It can still be stabilized or even reversed in the early stages, as in the case of sarcopenia, which can be prevented, treated, and reversed by exercise. Moreover, one of the cornerstones of rehabilitation research is the reduction of disability and restoration of function [7].

There is a great need and much to be gained from defining a way to measure functioning and disability on a relevant scale, ideally daily. However, assessing everyday functioning and disability is complicated due to current measurement modalities (eg, self-report, proxy report, and clinician rating [8]). These reports are time-consuming and tedious to fill on follow-up visits. In addition, disciplines have disagreements about what constitutes a disability and the methods to measure this disability especially in a clinical setting [9]. Ecological momentary assessment (EMA) allows for more continuous assessment and monitoring of patients without face-to-face appointments and has the crucial advantage of providing data that is more relevant to daily life [10]; however, it still requires active patient input, leading to refusal and attrition. Developing adequate passive EMA tools may increase retention and help overcome the limitations of active EMA [11].

Patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs) [12] are increasingly recognized as tools providing valuable information about patients’ health statuses and perceptions of treatment at a particular time. Including such tools in the health care workflow aims to provide a patient-centered, value-based health care system [13]. A commonly used PROM for disability assessment is the second version of the World Health

Organization Disability Assessment Schedule (WHODAS 2.0) [14]. This 36-item questionnaire provides a generic tool to measure health and disability. It assesses difficulties due to health conditions, including diseases or illnesses, short or long-lasting health problems, injuries, mental or emotional problems, and substance use disorders [15]. WHODAS 2.0 captures the level of functioning in six domains of life: (1) cognition—understanding and communicating; (2) mobility—moving and getting around; (3) self-care—attending to one’s hygiene, dressing, eating, and staying alone; (4) getting along—interacting with other people; (5) life activities—domestic responsibilities, leisure, work, and school; and (6) participation—joining in community activities and participating in society. Respondents are asked to reflect over the last 30 days and answer a series of questions, thinking about how much difficulty they had doing the given activities. There is a possible maximum score of 5 points for all items, indicating a rising level of difficulty in performing the activity (1: none, 2: mild, 3: moderate, 4: severe, and 5: extreme). A higher final score value, calculated as a total score or score by domain, indicates a higher level of disability [16].

Mobile sensing offers various sources, such as GPS, accelerometer, gyroscope, and light sensor that can be used to implement behavioral measures [17]. Unlike traditional assessment tools, these technologies enable long-term passive and ecological measurement of patient function. While there has been some work in digital mental health and machine learning [18], no studies predict WHODAS 2.0 functionality score changes using smartphone sensor data. These approaches are particularly important since they may enable the analysis of individuals’ functioning and disability evolution and provide a clinical tool to monitor the progression and efficacy of treatment. In addition, they provide the opportunity to build targeted, just-in-time adaptive interventions [19] in a designated population. Such frameworks deliver interventions within the context of daily life. Including passive data-driven solutions as part of the typical PROM frameworks could enrich existing information and better inform decisions.

Goal of This Study

This work aims to provide a baseline analysis of the feasibility of using machine learning to predict patients’ WHODAS 2.0 functionality scores from passively gathered digital biomarkers. Furthermore, we aim to determine which behavioral features are the most important for predicting different WHODAS 2.0 domains. Using these features, we train a linear regressor for simplicity and interpretability of the biomarkers as predictors.

This paper presents a continuous passive assessment modality and its feasibility in predicting a particular disability scale.

Methods

Ethical Considerations

The data used in this study were collected from an ongoing study involving passive smartphone monitoring of clinical outpatients (reference number EO 46-2013). The study received approval from the institutional review board at the Psychiatry Department of Fundación Jimenez Diaz Hospital, and all participants provided written informed consent. The institutional review board approved the secondary analysis without additional consent.

The study recruited patients who were at least 18 years old and were clinical outpatients diagnosed with mental disorders or attending therapy groups at the institutions mentioned above. To participate, patients had to own a smartphone with either Android or iOS operating systems, which they used to connect to a Wi-Fi network at least once a week. The patients did not receive payment for their participation. To protect the patient's privacy, the study data were anonymized.

Data Set

A clinically validated eHealth platform, eB2 Mindcare [20,21], collected the participants' passive data. In addition, clinicians used the MEMind [22] electronic health tool to record the patients' WHODAS 2.0 scores. Patients with the option of electronic self-reporting filled in an equivalent electronic WHODAS 2.0 self-report. Patients were able to complete the WHODAS 2.0 in parts by domain, which allowed for flexibility and less disruption to the patients. This led to later difficulties in analyzing WHODAS 2.0 total scores (officially called WHODAS 2.0 summary score computed by summing the 6 domain scores and adjusting to a range from 0 to 100) due to uncompleted domains and timing differences of when a patient would complete each domain. Thus, our predictions are focused on domain scores that allow for a larger user pool and accurate time windows for mobile data.

Mobile-Sensed Data

The eB2 MindCare [20,21] mobile app (Evidence-Based Behavior) collects data from different sources (mobile phone sensors and wearables) at different frequencies. For this work, we focused on the data streams related to patient mobility (daily step count; distance traveled; the number of locations visited by the patient; time spent at home; and time spent performing activities such as walking, running, and exercising) and time

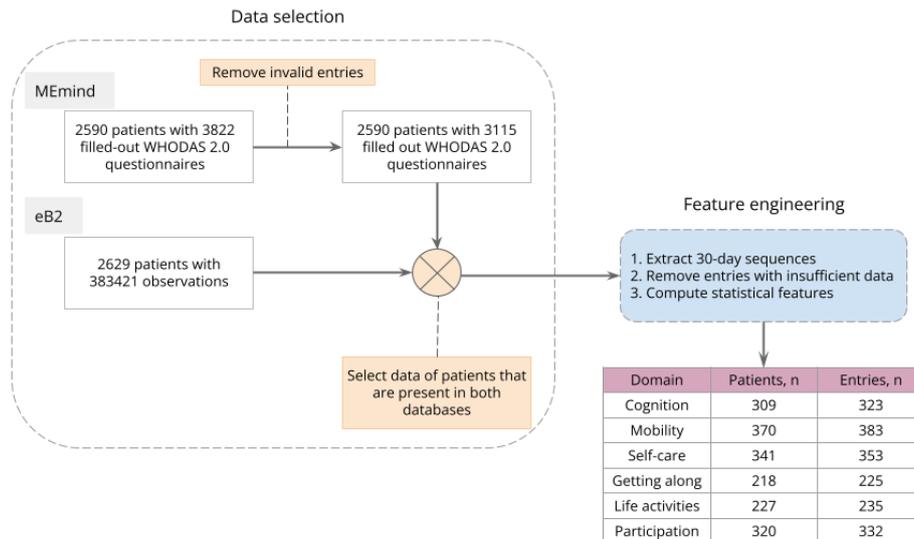
spent asleep. Daily summaries were calculated on the values of these variables, which were then used to extract 64 descriptive statistical features for characterizing the patient's behavior in a 30-day interval.

Data Preprocessing

Figure 1 shows the data selection and feature extraction process. For each domain, users with the following were removed: incomplete answers (if some questions were not answered within a specific domain) and incorrect scores for the individual questions (if the registered score was out of the range of the possible scores). In addition, the particular questions' scores from 1 to 5 were rescaled following the World Health Organization guidelines [16] to either values from 0 to 4 or 0, 1, 1, 2, 2. Finally, the score by domain was computed as the sum of the scores of the respective questions. These scores served as our target for the supervised prediction problem.

To build the input data set, we cropped a 30-day window of the data sequences for each WHODAS 2.0 entry. For the baseline WHODAS 2.0 score completed at study enrollment, we consider the next 30 days of observations because no previous mobile-sensed data were collected. For follow-up scores, which are usually collected biannually, we centered the window on encapsulating 15 days before and after the score. It was necessary to transform the time-series data set to be modeled as a nonsequential supervised learning problem. Therefore, we extracted statistical summary features (count, minimum, maximum, mean, SD, and IQR) from the sequences for each variable and obtained a data set of 64 features. We filtered the sequences for comprehensive statistics calculation by requiring every feature to contain at least 2 counts (days) of data and removing missing values.

We divided the data sets for each domain into 2 independent subsets based on the patients, ensuring no overlap. The first subset is the training data set, consisting of 80% of the entries used to fit the feature selection and predictive models. The second data set was held out for testing the model performance. The training set was split into 4 equal folds of 20% for cross-validation. The train test and cross-validation splits were done by ensuring the grouping of entries of the same user within the same set or fold since the model is user independent and stratifying the IQR of the WHODAS 2.0 scores. The stratification ensures the model can train and test low, middle, and high population WHODAS 2.0 scores. Subsequently, the features were standardized by centering the values around zero mean with a unit SD. Moving features to a similar scale helps avoid feature weight problems and provides an interpretable bias in the case of linear regression.

Figure 1. Data selection and feature engineering flowchart. WHODAS: World Health Organization Disability Assessment Schedule.

Feature Selection and Predictive Modeling

We used sequential forward selection (SFS) [23], also known as a sequential feature selection or stepwise forward selection, a greedy search algorithm for feature selection, which reduces an initial d -dimensional feature space to a k -dimensional feature subspace, where $k < d$. SFS avoids the feature selection stability problems of the LASSO with a similar idea as best subset selection but on a reduced set of subsets, which is computationally feasible [24]. In SFS, features are added one by one to an empty set of features until reaching an upper bound set of features. At each step, a criterion or score (in our case mean absolute error [MAE]) is calculated and saved that is used to select the feature that provides the better score (lower MAE) before continuing on to the next step. To find the best set of features in the case of each domain, we performed a search by iterating from $k=1$ to 20 with 4-fold cross-validation over the training set and selecting the k with the highest average performance across folds with the same model design as our final model. Finding the best features for predicting each domain provides greater interpretability to our models, which is essential in a clinical setting where clinicians need reliable and straightforward decision rules [25].

Once the best features were found, we trained linear regression models to perform the prediction task. To better suit the ordinal classification problem, we performed a simple modification after the regressor by thresholding the predictions between the minimum (0) and maximum values of the specific domain by rounding. While linear regression has limitations in this specific setting, it does offer certain advantages. Linear regression models are less prone to overfitting, especially when dealing with the noise commonly encountered in real-world data. Although nonlinear models may perform better, they are more prone to overfitting and are less interpretable and explainable.

These latter factors are crucial considerations when applying machine learning in a clinical setting [26,27].

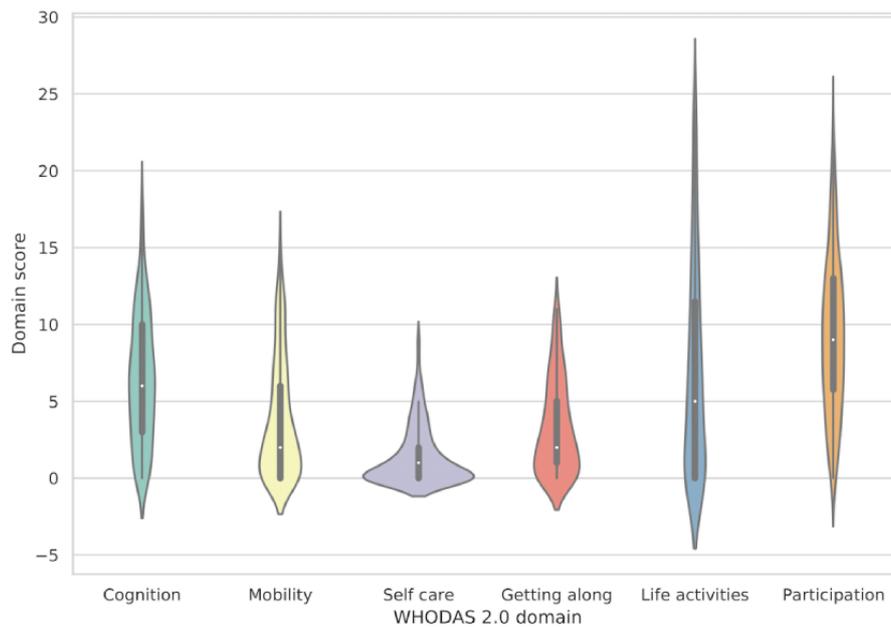
The final models underwent evaluation using the held-out test set. Test MAE and test mean absolute percentage error (MAPE) served as the evaluation metrics. MAE was chosen for its linearity, which aligns well with the context of this study. This is in contrast to root-mean-squared error, which tends to overemphasize larger errors due to the squaring of error values. Given the inherent variability in real-world data sets like the passive sensing data set in this study, metrics that are more robust against larger errors were preferred [28,29]. Furthermore, MAPE provides a metric to compare the different domains with a different number of questions and different total scores. We applied this approach separately for the different WHODAS 2.0 domains, using all the extracted features and the best subset of features.

Results

User Statistics

After our data filtering (Figure 1), 1526 WHODAS 2.0 domain entries of 396 participants collected between January 2017 and April 2021 were selected for our analysis. The cohort of patients had a median age of 44 (IQR 33-53) years at baseline, 63.13% (250/396) were female participants, and 29.04% (115/396) were male participants. Age and gender information were unavailable for 8.08% (32/396) and 7.83% (31/396) of the participants, respectively. Sociodemographic information was not included as input to the model. Figure 2 depicts an overview of the WHODAS 2.0 domain score distributions in the overall population in the form of violin plots, encompassing the summary statistics and the density of each domain score. The white dot in the middle is the median value, and the thick black bar in the center represents the IQR. Table S1 in Multimedia Appendix 1 provides the numerical statistics on the distributions.

Figure 2. The distribution of World Health Organization Disability Assessment Schedule 2.0 functionality scores per domain in the patient cohort. WHODAS: World Health Organization Disability Assessment Schedule.



Results of Feature Selection and Predictive Modeling

We performed the feature selection using SFS, followed by the training of unregularized linear regressors with the best feature subset for each domain. We then compared the performance to linear regressors trained on the entire feature set. Table 1 shows the domain prediction errors as MAE and MAPE for both experimental set-ups. Note that these scores are inversely oriented, meaning that lower values indicate better performance.

Figure 3 summarizes the best feature subsets selected across all 6 domains (refer to Table S2 in Multimedia Appendix 1 for a detailed overview). In each case, the models consistently discarded a majority of input features, resulting in a reduction of the feature space to 19, 19, 5, 6, 17, and 13 features from 64 for the respective domains. Figure 4 illustrates the feature weights of the linear regressor trained with all features per domain, with the best subset features indicated by a grid overlay. To facilitate comparison, the absolute value of regression coefficients per domain model was normalized within the range of 0 to 1. While each feature statistic is not shared, it can be seen that the models coincide in the feature groups important to all feature models. Both models capture the relevant data from each feature group but use different statistics. Since the extracted statistics (count, mean, min, max, and quartiles) are interrelated, choosing a subset of these statistics can summarize the relevant information captured by the regressor.

Across all domains, at least 1 statistic related to distance traveled and time spent at home was consistently selected. These features

impart information on daily movement patterns, which may indicate many elements of an individual's lifestyle, including work (or lack of work), socialization (in and out of the home), and isolation, among many others. This focus on movement patterns was further reinforced by including vehicle time, step count, and walking time statistics across multiple domains. Step count and walking time are additional physical activity descriptors, along with exercise time. Physical activity biomarkers proved important in the cognition, mobility, life activities, and participation domains. Sleep-related biomarkers, however, were exclusively selected for the cognition and participation domains. Physical well-being in both exercise and sleep is reasonably related to these domains. The self-care domain and getting along domain were described with the lowest amount of important features compared to the 4 other domains, indicating a different pattern of feature relevance.

The performance results presented in Table 1 demonstrate that the regression models trained on the reduced feature space slightly outperformed those that were trained on all features. Regression models estimate parameters for each term in the model, and the presence of noninformative variables can introduce uncertainty, thereby diminishing the overall performance. However, the marginal difference observed suggests that both models can capture relevant information. Despite the model trained on all features being burdened with a larger number of features and lacking regularization, it still manages to retain some level of effectiveness in regression.

Table 1. Regression evaluation metrics over the 4 folds.

WHODAS ^a 2.0 domain	Score range	Predicting with all 80 features		Predicting with selected features		
		MAE ^b , mean (SD)	MAPE ^c (%), mean (SD)	Features selected, n	MAE, mean (SD)	MAPE (%), mean (SD)
Cognition	0-20	3.76 (3.21)	18.84 (16.09)	19	3.55 (2.90)	17.76 (14.54)
Mobility	0-16	3.50 (2.44)	21.91 (15.26)	19	3.40 (1.98)	21.26 (12.42)
Self-care	0-10	1.56 (1.75)	15.69 (17.54)	5	1.48 (1.50)	14.86 (15.09)
Getting along	0-12	3.13 (2.72)	26.11 (22.74)	6	2.44 (1.79)	20.37 (14.96)
Life activities	0-24	7.57 (6.40)	31.56 (26.66)	17	6.53 (5.40)	27.21 (22.52)
Participation	0-24	3.88 (3.14)	16.16 (13.10)	13	3.73 (2.69)	15.54 (11.23)

^aWHODAS: World Health Organization Disability Assessment Schedule.

^bMAE: mean absolute error.

^cMAPE: mean absolute percentage error.

Figure 3. Feature selection results per World Health Organization Disability Assessment Schedule 2.0 domain.

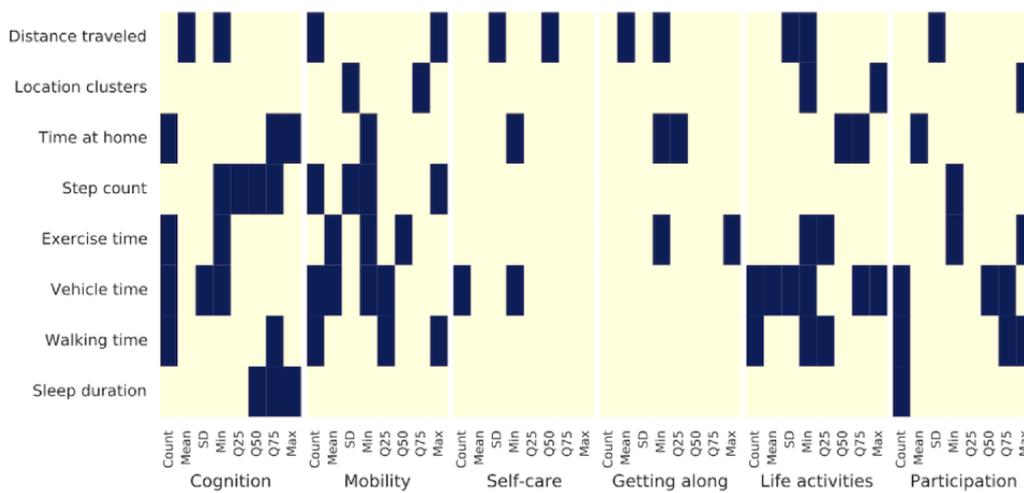
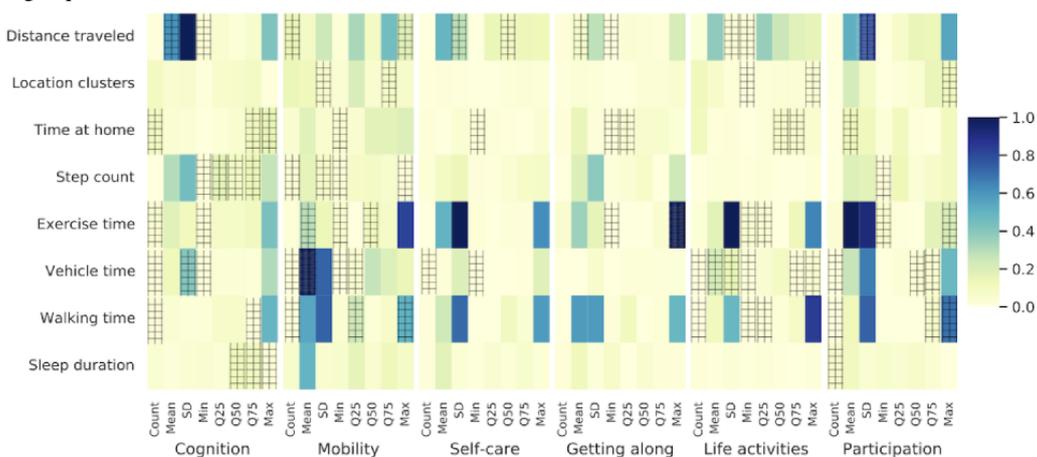


Figure 4. The feature importance of the linear regressor with all features per World Health Organization Disability Assessment Schedule 2.0 domain (shaded boxes) overlaid with the selected ones (hatched area). The shading indicates the size of the regression coefficient, while the hatching marks the selected feature using sequential forward selection.



Discussion

Principal Findings

In this study, our objective was to predict WHODAS 2.0 functionality scores for each domain using only passively

collected digital biomarkers. To achieve this, we used statistical feature engineering techniques, followed by a simple machine learning approach that involved selecting features through SFS for linear regression. This approach demonstrated the potential to predict functionality from passively sensed data. Additionally,

by using a straightforward linear regression model, we ensured the interpretability of the model's decisions.

Extracting statistical measures of the time-series sequences allowed dealing with missing data without applying imputation techniques; however, certain entries had to be excluded due to their limited information content (eg, sequences with completely missing observations per feature). Our focus was to identify the most relevant features from an ample feature space and eliminate noninformative or redundant predictors from the model for each domain. We found that 5-19 features were sufficient for each domain, and the most relevant being the distance traveled, time spent at home, time spent walking, exercise time, and vehicle time. These features are biomarkers for daily movement patterns and individuals' physical activity that were most informative for linear regression.

The models using feature selection outperformed those using the entire feature space across all domains. They achieved accurate predictions of patients' WHODAS 2.0 functionality scores per domain, with a maximum MAPE of 27.21% on the life activities domain and a minimum MAPE of 14.86% on the self-care domain. These errors are reasonable for a linear regression performing a complicated ordinal classification task.

The overall model performance can be explained by the distribution of the outcome variable within the respective data sets: mid-range values dominate in each domain; therefore, it becomes harder for the models to predict the more extreme scores. This issue is exacerbated by the fact that the target variable's distribution is skewed rather than normally distributed, potentially affecting the performance of a regression model. Additionally, both the presence of missing values and a target variable that is better suited for ordinal classification presented challenges that impacted performance.

It is worth noting that the domains with the fewest selected features and the smallest range of possible scores exhibited the lowest and highest MAPE values, respectively. Feature selection did not alter the error ranking of the domains when compared to the model that used all features.

Regarding feasibility, an average MAPE (across the 6 domains) of 19.5% was a reasonable error for a simple yet interpretable model. WHODAS 2.0 domain or total scores cut-offs tend to be used clinically based on population norms [14]. A total score of 50, which is similar to a percentile score in how it is rescaled from 0 to 100, almost represents the 95th population percentile [14]. Our model could be feasibly used to determine cases of severe disability even with an error margin of around 20%. As mentioned previously, the most important use case would be passive continuous assessment of patients. With the current level of error, using this simple yet effective model would be feasible for detecting moderate or severe changes in functionality. Additionally, the experiments presented herein do not take into account the continuous streaming of data that can further help the model to be more accurate globally and for a specific user. We believe that the framework and approach presented, especially with room to improve, show promising results for creating a continuous assessment and monitoring

system that could alert physicians to patient loss of functionality and allow for just-in-time interventions.

Limitations

Although our approach showed promising results, it also has limitations. The automatically generated wearable device data were passively collected in a real-world setting. This is a strength in ecological validity; however, it also introduces challenges, such as significant missing data and noise, which are commonly encountered in real-world passive data acquisition. Moreover, there are also data quality problems: missing data due to users not wearing the device or incorrect data arising from the device malfunctioning. These issues may have adversely affected the predictive performance of our model and biased the importance of certain variables. Missing data also posed a problem in the features used from the eB2 database, as other helpful features, such as app usage and phone unlocks, which are directly related to social domains, were filtered out, causing a severely reduced data set. Consequently, this led to a substantial reduction in our data set. The data set presents many challenges because it is the result of merging 2 real-world databases with missing values, erroneous entries, and noise.

Another limitation arises from the fact that, in many cases, individuals have only a single score, which does not allow for training personalized models that could better account for the intraindividual variations. While step data can provide relevant information for determining mobility, factors such as an individual's lifestyle, work conditions, and other variables greatly influence step count. We hypothesize that a model that learns individual variability and patterns and then examines the population would better suit this task. However, obtaining longitudinal large population data sets that combine clinical and wearable data poses considerable challenges.

Conclusions

Our work is the first to suggest a machine learning-based approach for assessing WHODAS 2.0 functionality from passively sensed data. The results demonstrate the feasibility of designing a pipeline that passively monitors patients' functionality over time. However, it is important to note that predicting scores for each WHODAS 2.0 domain equally well poses a significant challenge. Nevertheless, the feature selection approach provides insight into relevant behavioral measures contributing to improved prediction performance. This aspect of our approach enhances the interpretability of the results, which is important for real-world and clinical applications.

Moving forward, an interesting avenue to explore is deep learning-based temporal methods that operate on the data at a more refined time scale. Instead of relying on statistical feature engineering and selection, these models can learn representations directly from raw input sequences that are most relevant to the prediction problem. We will further investigate the possibilities of specialized models for individual patients to understand the functionality evolution better, improve prediction accuracy, and provide interpretable model outputs to clinicians. This line of investigation holds promise for advancing the field and addressing the challenges associated with predicting WHODAS 2.0 functionality using passive data collection methods.

Acknowledgments

ES received funding from the European Union Horizon 2020 research and innovation program (Marie Skłodowska-Curie grant 813533). AA and PMO are supported by the Spanish Ministerio de Ciencia, Innovación y Universidades and Agencia Estatal de Investigación (RTI2018-099655-B-I00, PID2021-123182OB-I00, and TED2021-131823B-I00), and the Comunidad de Madrid (Y2018/TCS-4705 PRACTICO-CM and IND2018/TIC-9649).

Data Availability

The data sets generated and analyzed during this study are not publicly available due to their personally identifiable and sensitive health information content but are available from the corresponding author on reasonable request.

Authors' Contributions

LRM, JHL, and SdLM analyzed the data, developed the machine learning pipeline, and generated the results. ES and SdLM wrote the manuscript and analyzed the results. JJCM acquired and interpreted data for the work. EBG provided expert advice on clinical and patient-related matters. PMO and AA supervised and guided the work and provided important intellectual input.

Conflicts of Interest

AA, EBG, and JJCM are cofounders of eB2. EBG designed the MEmind app.

Multimedia Appendix 1

Tables of the WHODAS 2.0 functionality scores per domain in the patient cohort and the lists of selected features per domain after sequential feature selection.

[\[PDF File \(Adobe PDF File\), 39 KB-Multimedia Appendix 1\]](#)

References

1. Chamberlain AM, Rutten LJJ, Jacobson DJ, Fan C, Wilson PM, Rocca WA, et al. Multimorbidity, functional limitations, and outcomes: interactions in a population-based cohort of older adults. *J Comorb* 2019;9:2235042X19873486 [FREE Full text] [doi: [10.1177/2235042X19873486](https://doi.org/10.1177/2235042X19873486)] [Medline: [31523633](https://pubmed.ncbi.nlm.nih.gov/31523633/)]
2. Miller ME, Rejeski WJ, Reboussin BA, Ten Have TR, Ettinger WH. Physical activity, functional limitations, and disability in older adults. *J Am Geriatr Soc* 2000;48(10):1264-1272 [doi: [10.1111/j.1532-5415.2000.tb02600.x](https://doi.org/10.1111/j.1532-5415.2000.tb02600.x)] [Medline: [11037014](https://pubmed.ncbi.nlm.nih.gov/11037014/)]
3. Torres-Castro R, Solis-Navarro L, Sitjà-Rabert M, Vilaró J. Functional limitations post-COVID-19: a comprehensive assessment strategy. *Arch Bronconeumol* 2021;57:7-8 [FREE Full text] [doi: [10.1016/j.arbres.2020.07.025](https://doi.org/10.1016/j.arbres.2020.07.025)] [Medline: [34629627](https://pubmed.ncbi.nlm.nih.gov/34629627/)]
4. Roberts P, Wertheimer J, Park E, Nuño M, Riggs R. Identification of functional limitations and discharge destination in patients with COVID-19. *Arch Phys Med Rehabil* 2021;102(3):351-358 [FREE Full text] [doi: [10.1016/j.apmr.2020.11.005](https://doi.org/10.1016/j.apmr.2020.11.005)] [Medline: [33278363](https://pubmed.ncbi.nlm.nih.gov/33278363/)]
5. Fernández-de-Las-Peñas C, Martín-Guerrero JD, Navarro-Pardo E, Rodríguez-Jiménez J, Pellicer-Valero OJ. Post-COVID functional limitations on daily living activities are associated with symptoms experienced at the acute phase of SARS-CoV-2 infection and internal care unit admission: a multicenter study. *J Infect* 2022;84(2):248-288 [FREE Full text] [doi: [10.1016/j.jinf.2021.08.009](https://doi.org/10.1016/j.jinf.2021.08.009)] [Medline: [34375711](https://pubmed.ncbi.nlm.nih.gov/34375711/)]
6. Goates S, Du K, Arensberg MB, Gaillard T, Guralnik J, Pereira SL. Economic impact of hospitalizations in US adults with sarcopenia. *J Frailty Aging* 2019;8(2):93-99 [FREE Full text] [doi: [10.14283/jfa.2019.10](https://doi.org/10.14283/jfa.2019.10)] [Medline: [30997923](https://pubmed.ncbi.nlm.nih.gov/30997923/)]
7. Frontera WR, DeLisa JA, Gans BM, Robinson LR. *DeLisa's Physical Medicine and Rehabilitation: Principles and Practice*. Philadelphia: Lippincott Williams & Wilkins Health; 2019.
8. Patterson TL, Goldman S, McKibbin CL, Hughs T, Jeste DV. UCSD performance-based skills assessment: development of a new measure of everyday functioning for severely mentally ill adults. *Schizophr Bull* 2001;27(2):235-245 [FREE Full text] [doi: [10.1093/oxfordjournals.schbul.a006870](https://doi.org/10.1093/oxfordjournals.schbul.a006870)] [Medline: [11354591](https://pubmed.ncbi.nlm.nih.gov/11354591/)]
9. McKibbin C, Patterson TL, Jeste DV. Assessing disability in older patients with schizophrenia: results from the WHODAS-II. *J Nerv Ment Dis* 2004;192(6):405-413 [doi: [10.1097/01.nmd.0000130133.32276.83](https://doi.org/10.1097/01.nmd.0000130133.32276.83)] [Medline: [15167403](https://pubmed.ncbi.nlm.nih.gov/15167403/)]
10. Miguelez-Fernandez C, de Leon SJ, Baltasar-Tello I, Peñuelas-Calvo I, Barrigon ML, Capdevila AS, et al. Evaluating attention-deficit/hyperactivity disorder using ecological momentary assessment: a systematic review. *Atten Defic Hyperact Disord* 2018;10(4):247-265 [doi: [10.1007/s12402-018-0261-1](https://doi.org/10.1007/s12402-018-0261-1)] [Medline: [30132248](https://pubmed.ncbi.nlm.nih.gov/30132248/)]
11. Porrás-Segovia A, Díaz-Oliván I, Barrigón ML, Moreno M, Artés-Rodríguez A, Pérez-Rodríguez MM, et al. Real-world feasibility and acceptability of real-time suicide risk monitoring via smartphones: a 6-month follow-up cohort. *J Psychiatr Res* 2022;149:145-154 [doi: [10.1016/j.jpsychires.2022.02.026](https://doi.org/10.1016/j.jpsychires.2022.02.026)] [Medline: [35276631](https://pubmed.ncbi.nlm.nih.gov/35276631/)]
12. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. Hoboken, NJ: Wiley Blackwell; 2019.

13. Baumhauer JF, Bozic KJ. Value-based healthcare: patient-reported outcomes in clinical decision making. *Clin Orthop Relat Res* 2016;474(6):1375-1378 [FREE Full text] [doi: [10.1007/s11999-016-4813-4](https://doi.org/10.1007/s11999-016-4813-4)] [Medline: [27052020](https://pubmed.ncbi.nlm.nih.gov/27052020/)]
14. Üstün TB, Chatterji JR, Rehm J, editors. *Measuring Health and Disability: Manual for WHO Disability Assessment Schedule WHODAS 2.0*. Geneva, Switzerland: World Health Organization; 2010.
15. ICHOM. 2018. URL: <https://www.ichom.org/> [accessed 2021-12-07]
16. Paton M, Lane R. Clinimetrics: World Health Organization disability assessment schedule 2.0. *J Physiother* 2020;66(3):199 [FREE Full text] [doi: [10.1016/j.jphys.2020.03.002](https://doi.org/10.1016/j.jphys.2020.03.002)] [Medline: [32291217](https://pubmed.ncbi.nlm.nih.gov/32291217/)]
17. Majumder S, Deen MJ. Smartphone sensors for health monitoring and diagnosis. *Sensors (Basel)* 2019;19(9):2164 [FREE Full text] [doi: [10.3390/s19092164](https://doi.org/10.3390/s19092164)] [Medline: [31075985](https://pubmed.ncbi.nlm.nih.gov/31075985/)]
18. Hornstein S, Forman-Hoffman V, Nazander A, Ranta K, Hilbert K. Predicting therapy outcome in a digital mental health intervention for depression and anxiety: a machine learning approach. *Digit Health* 2021;7:20552076211060659 [FREE Full text] [doi: [10.1177/20552076211060659](https://doi.org/10.1177/20552076211060659)] [Medline: [34868624](https://pubmed.ncbi.nlm.nih.gov/34868624/)]
19. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAs) in mobile health: key components and design principles for ongoing health behavior support. *Ann Behav Med* 2018;52(6):446-462 [FREE Full text] [doi: [10.1007/s12160-016-9830-8](https://doi.org/10.1007/s12160-016-9830-8)] [Medline: [27663578](https://pubmed.ncbi.nlm.nih.gov/27663578/)]
20. Berrouguet S, Ramírez D, Barrigón ML, Moreno-Muñoz P, Carmona Camacho R, Baca-García E, et al. Combining continuous smartphone native sensors data capture and unsupervised data mining techniques for behavioral changes detection: a case series of the Evidence-Based Behavior (eB2) study. *JMIR Mhealth Uhealth* 2018;6(12):e197 [FREE Full text] [doi: [10.2196/mhealth.9472](https://doi.org/10.2196/mhealth.9472)] [Medline: [30530465](https://pubmed.ncbi.nlm.nih.gov/30530465/)]
21. Evidence-Based Behavior (eB2). URL: <https://eb2.tech/?lang=en> [accessed 2023-10-05]
22. Barrigón ML, Berrouguet S, Carballo JJ, Bonal-Giménez C, Fernández-Navarro P, Pfang B, et al. MEmind study group. User profiles of an electronic mental health tool for ecological momentary assessment: MEmind. *Int J Methods Psychiatr Res* 2017;26(1):e1554 [FREE Full text] [doi: [10.1002/mpr.1554](https://doi.org/10.1002/mpr.1554)] [Medline: [28276176](https://pubmed.ncbi.nlm.nih.gov/28276176/)]
23. Ferri FJ, Pudil P, Hatf M, Kittler J. Comparative study of techniques for large-scale feature selection. *Mach Intell Pattern Recognit* 1994;16:403-413 [doi: [10.1016/b978-0-444-81892-8.50040-7](https://doi.org/10.1016/b978-0-444-81892-8.50040-7)]
24. Hastie T, Tibshirani R, Tibshirani R. Best subset, forward stepwise or lasso? Analysis and recommendations based on extensive comparisons. *Statist Sci* 2020;35(4):579-592 [doi: [10.1214/19-sts733](https://doi.org/10.1214/19-sts733)]
25. Orrù G, Monaro M, Conversano C, Gemignani A, Sartori G. Machine learning in psychometrics and psychological research. *Front Psychol* 2019;10:2970 [FREE Full text] [doi: [10.3389/fpsyg.2019.02970](https://doi.org/10.3389/fpsyg.2019.02970)] [Medline: [31998200](https://pubmed.ncbi.nlm.nih.gov/31998200/)]
26. Li P, Liang X, Song H. A survey on implicit bias of gradient descent. : IEEE; 2022 Presented at: 14th International Conference on Computer Research and Development (ICCRD); January 7-9, 2022; Shenzhen, China
27. Hawkins DM. The problem of overfitting. *ChemInform Wiley* 2004;44(1):1-12 [doi: [10.1002/chin.200419274](https://doi.org/10.1002/chin.200419274)]
28. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature. *Geosci Model Dev* 2014;7(3):1247-1250 [doi: [10.5194/gmd-7-1247-2014](https://doi.org/10.5194/gmd-7-1247-2014)]
29. Brassington G. Mean absolute error and root mean square error: which is the better metric for assessing model performance? In: EGU General Assembly. 2017 Presented at: Proceedings From the Conference 19th EGU General Assembly, EGU2017; April 23-28, 2017; Vienna, Austria URL: <https://ui.adsabs.harvard.edu/abs/2017EGUGA..19.3574B/abstract>

Abbreviations

- EMA:** ecological momentary assessment
MAE: mean absolute error
MAPE: mean absolute percentage error
PREM: patient-reported experience measure
PROM: patient-reported outcome measures
SFS: sequential forward selection
WHO: World Health Organization
WHODAS: World Health Organization Disability Assessment Schedule

Edited by A Mavragani; submitted 17.03.23; peer-reviewed by J Hinman, A Abdel-Hafez; comments to author 08.06.23; revised version received 22.07.23; accepted 15.08.23; published 30.10.23

Please cite as:

Sükei E, Romero-Medrano L, de Leon-Martinez S, Herrera López J, Campaña-Montes JJ, Olmos PM, Baca-García E, Artés A
Continuous Assessment of Function and Disability via Mobile Sensing: Real-World Data-Driven Feasibility Study

JMIR Form Res 2023;7:e47167

URL: <https://formative.jmir.org/2023/1/e47167>

doi: [10.2196/47167](https://doi.org/10.2196/47167)

PMID: [37902823](https://pubmed.ncbi.nlm.nih.gov/37902823/)

©Emese Sükei, Lorena Romero-Medrano, Santiago de Leon-Martinez, Jesús Herrera López, Juan José Campaña-Montes, Pablo M Olmos, Enrique Baca-Garcia, Antonio Artés. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 30.10.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.