

## ARTICLE OPEN



# A systems approach towards remote health-monitoring in older adults: Introducing a zero-interaction digital exhaust

Narayan Schütz<sup>1</sup>✉, Samuel E. J. Knobel<sup>1</sup>, Angela Botros<sup>1</sup>, Michael Single<sup>1</sup>, Bruno Pais<sup>2</sup>, Valérie Santschi<sup>2</sup>, Daniel Gatica-Perez<sup>3,4</sup>, Philipp Buluschek<sup>5</sup>, Prabitha Urwyler<sup>1</sup>, Stephan M. Gerber<sup>1</sup>, René M. Müri<sup>1,6</sup>, Urs P. Mosimann<sup>1</sup>, Hugo Saner<sup>1,7,8</sup> and Tobias Nef<sup>1,6,8</sup>

Using connected sensing devices to remotely monitor health is a promising way to help transition healthcare from a rather reactive to a more precision medicine oriented proactive approach, which could be particularly relevant in the face of rapid population ageing and the challenges it poses to healthcare systems. Sensor derived digital measures of health, such as digital biomarkers or digital clinical outcome assessments, may be used to monitor health status or the risk of adverse events like falls. Current research around such digital measures has largely focused on exploring the use of few individual measures obtained through mobile devices. However, especially for long-term applications in older adults, this choice of technology may not be ideal and could further add to the digital divide. Moreover, large-scale systems biology approaches, like genomics, have already proven beneficial in precision medicine, making it plausible that the same could also hold for remote-health monitoring. In this context, we introduce and describe a zero-interaction digital exhaust: a set of 1268 digital measures that cover large parts of a person's activity, behavior and physiology. Making this approach more inclusive of older adults, we base this set entirely on contactless, zero-interaction sensing technologies. Applying the resulting digital exhaust to real-world data, we then demonstrate the possibility to create multiple ageing relevant digital clinical outcome assessments. Paired with modern machine learning, we find these assessments to be surprisingly powerful and often on-par with mobile approaches. Lastly, we highlight the possibility to discover novel digital biomarkers based on this large-scale approach.

*npj Digital Medicine* (2022)5:116; <https://doi.org/10.1038/s41746-022-00657-y>

## INTRODUCTION

Rapid population ageing is a global phenomenon that affects virtually every country on earth<sup>1</sup>. A major challenge in this regard is the increased disease burden and related disability associated with ageing, which is posed to further increase already high healthcare expenditures and significantly reduces older adults' quality of life<sup>1</sup>. While there is unlikely to be a single solution to those challenges, concepts from precision medicine could play an integral part<sup>2,3</sup>. One idea here is the systems medicine approach, aiming to move from reactive to predictive, preventive, personalized, and participatory (P4) medicine<sup>4,5</sup>, a concept that was initially introduced by Leroy Hood and colleagues, more than two decades ago. A key aspect of systems, or P4, medicine is to monitor a holistic view of a person's wellbeing in order to better manage their health and gain a deeper understanding of disease processes<sup>4,6</sup>. This is ought to be achieved by comprehensive systems approaches, like multi-omics profiling but more recently also by means of remote health-monitoring<sup>3,6–8</sup>. That is, by using connected sensing devices, such as smartphones, wearables or embedded internet of things sensing units, to continuously and objectively monitor health relevant information in everyday life<sup>9,10</sup>. This is in contrast to the currently often employed on-site visits that tend to merely provide a, potentially biased, snapshot of health states<sup>3,11</sup>. Today, such remote health-monitoring approaches are mostly focusing on a select few individual aspects of older adults' lives, instead of employing more comprehensive systems approaches as advocated by P4 medicine. This focus, likely results in many phenotypes of health and disease

to be missed and may limit the potential of large scale machine learning (ML) approaches. Furthermore, the used technologies tend to include sensing devices that are rather optimized towards younger demographics and may thus prove suboptimal for use by many (but certainly not all) older adults, particularly long-term. Something that could further increase the digital divide, potentially excluding seniors that would benefit the most from remote health-monitoring approaches. More comprehensive approaches towards remote health-monitoring in older adults, with a particular emphasis on ageing-inclusive sensing technologies, may thus be of high relevance for the future of remote health-monitoring in this growing demographic.

In the past decade an impressive number of studies have demonstrated the potential of digital measures, such as their use as digital analogs to biomarkers<sup>12–17</sup> (hence, digital biomarkers<sup>18</sup>) and clinical outcome assessments (COAs)<sup>19–22</sup> (hence, digital COAs<sup>18</sup>). Much of this research revolved around clinical research in neuropsychiatric disorders<sup>23–25</sup> which are often linked to ageing and act as major contributors to disease burden<sup>26</sup>. In view of the above, it is not far fetched to assume that long-term monitoring of digital measures may provide continuous and objective information about an older individual's functional status and health changes. This, in turn, could facilitate earlier and more personalized interventions<sup>2,3</sup>. Eventually, this may also help facilitate older adult's independence, allowing them to stay at home longer, and increase their quality of life<sup>9</sup>. For instance, Rantz et al. show how sensor technologies linked to early alert systems led to better health outcomes amongst older adults<sup>9</sup>. Another example of long-

<sup>1</sup>ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland. <sup>2</sup>LaSource School of Nursing Sciences, HES-SO University of Applied Sciences and Arts Western Switzerland, Lausanne, Switzerland. <sup>3</sup>Idiap Research Institute, Martigny, Switzerland. <sup>4</sup>School of Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. <sup>5</sup>DomoHealth SA, Lausanne, Switzerland. <sup>6</sup>Department of Neurology, Inselspital, Bern, Switzerland. <sup>7</sup>Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland. <sup>8</sup>These authors contributed equally: Hugo Saner, Tobias Nef. ✉email: narayan.schuetz@unibe.ch

term monitoring is presented by Austin et al., where they managed to assess loneliness using connected sensing technologies<sup>27</sup>. Finally, in one of the earliest long-term monitoring efforts related to older adults, Hayes et al. demonstrate that variations in sensor-derived gait speed and physical activity differ significantly between older adults with mild cognitive impairment (MCI) and healthy older adults<sup>28</sup>.

A large part of conducted research around digital measures, however, has focused on shorter-term studies in combination with mobile technologies, such as smartphones, smartwatches, as well as activity and fitness trackers. While there is no doubt that mobile technologies are a great way to derive health relevant information, they may not necessarily be ideal for long-term monitoring in the broader population of older adults. There are multiple reasons for this: (1) older adults tend to be more wary of novel technologies<sup>29</sup>; (2) since monitoring durations may become very long or even unlimited, it is ideal if there is no interaction (zero-interaction) with the system, as there is, for instance, evidence of wear-time-dependent compliance issues<sup>30</sup>; (3) there is a certain stigma attached to the use of wearable devices, whereby many older adults tend to fear being seen as frail if they wear a device – even if it is just an alarm clock<sup>29</sup>; (4) for seniors with potential memory issues, wearing and maintaining devices may not be feasible. As a result, many older adults that are affected by the digital divide<sup>31</sup> may be excluded, this may be even more problematic in those of lower socio-economic background<sup>32–34</sup>, those living in harder to access rural areas<sup>32</sup>, or those living with certain conditions like cognitive impairments or late-life depression<sup>33</sup>. Not too surprisingly, most successful real-world, long-term research using sensor technology with older adults has focused on contactless, zero-interaction approaches<sup>9,13,28,35–40</sup>. Such technologies include passive infrared (PIR) motion sensors that capture an individual's activity in a given room<sup>9,28,35,36,38,41,42</sup>, contact door sensors that can signal when a person leaves or enters the home<sup>36–38,43</sup>, pressure sensors on or under a mattress that capture sleep measures<sup>9,36,39</sup>, and electronic pillboxes to track medication adherence<sup>44</sup>, along with more obtrusive depth-sensing cameras that track silhouettes to detect falls and monitor gait parameters<sup>9</sup>.

Currently, digital measures are commonly used individually or by combining several specific ones based on concepts of interest. While this approach is entirely reasonable, it may limit the potential of digital measures as many potential characteristics of health and disease may simply go unnoticed. Furthermore, it is oftentimes not clear in advance, which, amongst correlated measures, may be the most relevant<sup>13</sup>. We therefore hypothesise that a more holistic systems approach — inspired by systems biology and applied to remote health-monitoring — may be highly promising. This involves deriving larger sets of digital measures, potentially in the hundreds to thousands, which may be particularly helpful in exploratory research and could also enable the creation of strong digital COAs by leveraging large-scale machine learning approaches<sup>20</sup>. This is in some ways analogous to more classical biological settings, where measurements can assess individual blood tests or genes (for instance, by means of single-nucleotide polymorphisms) but also whole sets, such as metabolomes or genomes, to identify new phenotypes of health and disease, such as is being proposed with the systems-oriented P4 medicine. In the context of zero-interaction, contactless technologies, a systems approach could also help to counteract some of the downsides of contactless technologies, such as lower accuracy as a result of the indirect measurement modality. In the context of digital measures, an extensive set of measures may be referred to as some sort of digital “ome”, such as a digital behaviorome<sup>45</sup>, or a digital exhaust<sup>46</sup> — where the basic measurement unit is a digital measure. In this work, we will use the latter term as it is likely less controversial<sup>47</sup>.

Two notable real-world examples of using extensive sets of digital measures are studies by Cook et al.<sup>42</sup> and Chen et al.<sup>20</sup>,

which demonstrate the feasibility of using a digital exhaust based on wearable and contactless sensors to predict multiple clinical scores (in the former) and MCI (in the latter). Building on their work, we aim to evaluate the potential of a systems oriented approach towards long-term remote health-monitoring in the demographics of older adults. To this end, we first introduce an extensive set of 1268 well-documented digital measures that are entirely obtainable with sensing technologies that demonstrate extensive long-term, real-world evidence in older adults. Thus, all of these measures are based on a small set of zero-interaction, contactless, and cost-effective sensors that, as shown by Baettie et al., scale well to large ageing-related remote-monitoring projects<sup>36</sup>; this also means that these sensors should be compatible with most long-term monitoring projects in community-dwelling older adults. Using the resulting comprehensive set of digital measures, or digital exhaust, we further demonstrate how powerful ML based ageing-relevant digital COAs for fall risk, frailty, late-life depression, and MCI can be created. Finally, we highlight the possibility to leverage a digital exhaust to discover new potential digital biomarkers, demonstrating how a comprehensive systems approach could also help in establishing new phenotypes of health and disease.

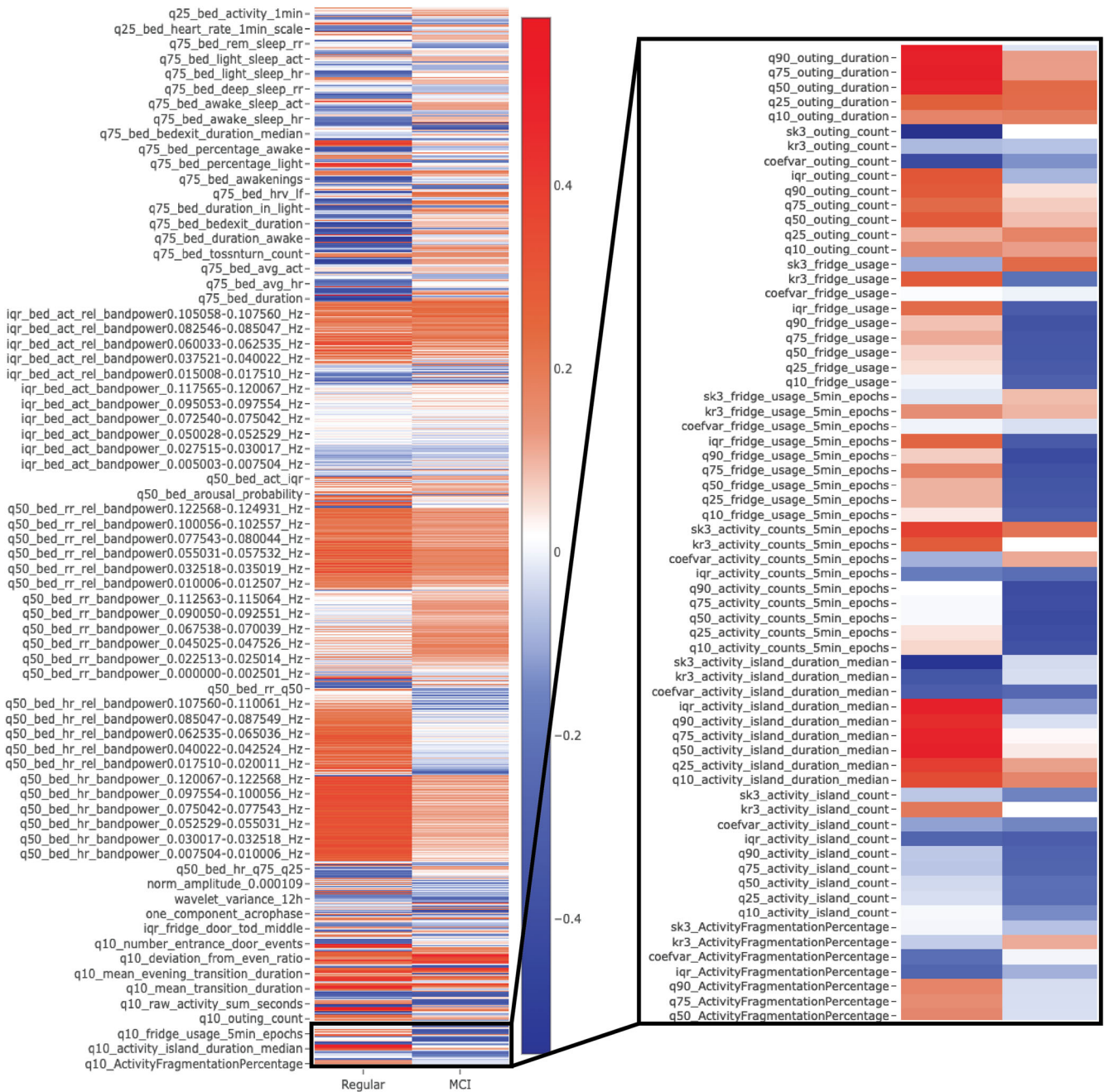
## RESULTS

### A zero-interaction digital exhaust

We introduce a set of 94 hypothesis-driven base measures, from which we further derive a total of 1268 digital measures using aggregation and frequency analysis. All measures are obtainable through zero-interaction and privacy-preserving (neither video nor audio) contactless (thus no direct physical contact) sensing devices, which do not require any user interaction. Of these 1268 digital measures, 224 were extracted by means of PIR motion sensors in essential rooms (the entrance, bathroom, living room, bedroom, and kitchen) and magnetic door sensors on the refrigerator and entrance door. An additional 1044 measures were extracted on the basis of sleep data from a quasi-piezoelectric bed sensor placed under the mattress. Detailed descriptions and derivations, as well as associated hypotheses, are provided in the Supplementary Methods, together with a high-level overview of all presented digital measures (see Dataset 1). Furthermore, an extensive online version with interactive visualisations, along with additional data, including measure distributions and correlations with various ageing-relevant health indicators and outcomes, is available on GitHub (<https://narayanschuetz.github.io/digital-exhaust/>) and serves as an online supplementary to this article. An example of averaged digital exhausts is shown in Fig. 1.

### Machine learning based digital clinical outcome assessments

Here, we demonstrate how the introduced digital exhaust could be useful for ageing and ageing related research. To this end, we created machine learning-derived digital COAs, aimed at automatically classifying ageing-relevant health outcomes. We created five datasets, one based on each clinical assessment, including fall risk, frailty, late-life depression, and MCI. This analysis is based on remote-monitoring data from two observational longitudinal pilot studies in Switzerland, where independently living, community-dwelling older adults were equipped with pervasive computing systems and monitored over the course of a year, while simultaneously being subject to regular visits and clinical assessments. The results on predicting ageing-relevant positive and negative health outcomes are summarised in Tables 1 and 2. The differences between using the digital exhaust alone versus using the exhaust in addition to demographics were minimal and, judging by overlapping 95% CIs, non-significant. The highest discriminative power, in terms of ROC AUC (area under the



**Fig. 1 Exemplary visualizations of averaged digital exhausts.** Depicts an example of z-normalised, averaged digital exhausts of participants with mild cognitive impairment (MCI) (based on a Montreal Cognitive Assessment screening < 23 points). Digital measures > 0 (in blue) indicate above-average values for that group, while < 0 (red) indicates below-average values. Many digital measures visually differ in both examples. It should be noted that this is a down-scaled visualisation, as not all measures would fit in the figure. For the complete and interactive version, see the [supplementary online version](#) (Note the zoom-in functionality).

receiver-operating characteristic curve), was achieved with the Tinetti Performance-Oriented Mobility Assessment (POMA)-based fall-risk dataset when including both demographic and digital exhaust information (ROC AUC = 0.805). Notably, however, demographic information alone was sufficient to produce good performance (ROC AUC = 0.777) in this particular case. Performances on the fall-risk related Timed Up and Go Test (TUG), the MCI-related Montreal Cognitive Assessment (MoCA), and the frailty-related Edmonton Frail Scale (EFS) datasets were also relatively high, with ROC AUC values of 0.786, 0.780, and 0.704, respectively, when using only the digital exhaust. The worst

performance was achieved with the dataset based on the Geriatric Depression Scale (GDS) (ROC AUC = 0.620), when using only the digital exhaust. Here, the difference between using only demographics versus using the digital exhaust was also minimal, with a slight but non-significant advantage in favour of the exhaust-only scenario. Overall, though, the addition of the digital exhaust resulted, in all cases, in higher ROC AUC and PrAUC (Area Under the Precision-Recall Curve) values than those obtained when using only demographic information. These differences were statistically significant in all but the POMA and GDS datasets (which were based on rather conservative, non-overlapping CI intervals). The



**Table 1.** Performance Evaluation based on ROC AUC.

Dataset	Demographics	Digital Exhaust	Digital Exhaust + Demographics
TUG	0.662 ± 0.041	0.786 ± 0.037	0.783 ± 0.036
POMA	0.777 ± 0.028	0.782 ± 0.035	0.805 ± 0.027
EFS	0.564 ± 0.034	0.704 ± 0.039	0.698 ± 0.038
GDS	0.602 ± 0.046	0.620 ± 0.048	0.616 ± 0.048
MoCA	0.430 ± 0.038	0.780 ± 0.039	0.757 ± 0.038

ROC AUC based results for the creation of machine learning-derived digital COAs based on clinical assessments for fall risk (TUG & POMA), frailty (EFS), late-life depression (GDS), and mild cognitive impairment (MoCA).  
TUG Timed Up and Go Test  
POMA Performance-Oriented Mobility Assessment  
EFS Edmonton Frail Scale  
MoCA Montreal Cognitive Assessment

**Table 2.** Performance Evaluation based on PrAUC.

Dataset	Demographics	Digital Exhaust	Digital Exhaust + Demographics
TUG	0.681 ± 0.035	0.816 ± 0.033	0.805 ± 0.034
POMA	0.615 ± 0.046	0.650 ± 0.047	0.678 ± 0.044
EFS	0.523 ± 0.041	0.625 ± 0.045	0.603 ± 0.044
GDS	0.397 ± 0.051	0.477 ± 0.054	0.458 ± 0.053
MoCA	0.653 ± 0.027	0.863 ± 0.028	0.839 ± 0.028

PrAUC based results for the creation of machine learning-derived digital COAs based on clinical assessments for fall risk (TUG & POMA), frailty (EFS), late-life depression (GDS), and mild cognitive impairment (MoCA).  
TUG Timed Up and Go Test  
POMA Performance-Oriented Mobility Assessment  
EFS Edmonton Frail Scale  
MoCA Montreal Cognitive Assessment

largest differences were found with the MoCA and TUG datasets, for which the objectives were to identify participants with an indication of MCI or increased fall risk, respectively.

### Discovering novel digital biomarkers

The importance of individual digital measures (that may be interesting as digital biomarker candidates) with respect to the COAs were evaluated by means of SHapley Additive exPlanations (SHAP) values<sup>48,49</sup>. In Fig. 2, we present the most important digital measure, based on global SHAP values, across all 100 simulations of each single dataset - corresponding to the different COAs. A more detailed table, highlighting the top 10 highest-ranked measures based on global SHAP values, is available in Supplementary Table 3. Furthermore, in Fig. 3, we display *beeswarm* plots of SHAP values for the individual COAs. These show how the nine most important digital measures - as well as the sum of all other measures combined - influence the log odds ratio of having a negative health outcome on the various datasets. Values shown in Fig. 3 mostly align with the global SHAP importance rankings, although there are minor differences. Since the global SHAP rankings are based on 100 iterations (and thus 100 models), and since the importances shown in Fig. 3 are based on a single model, we generally place greater emphasis on the global SHAP values. Nonetheless, the *beeswarm* plots are still useful, as they provide insights into the direction of effects.

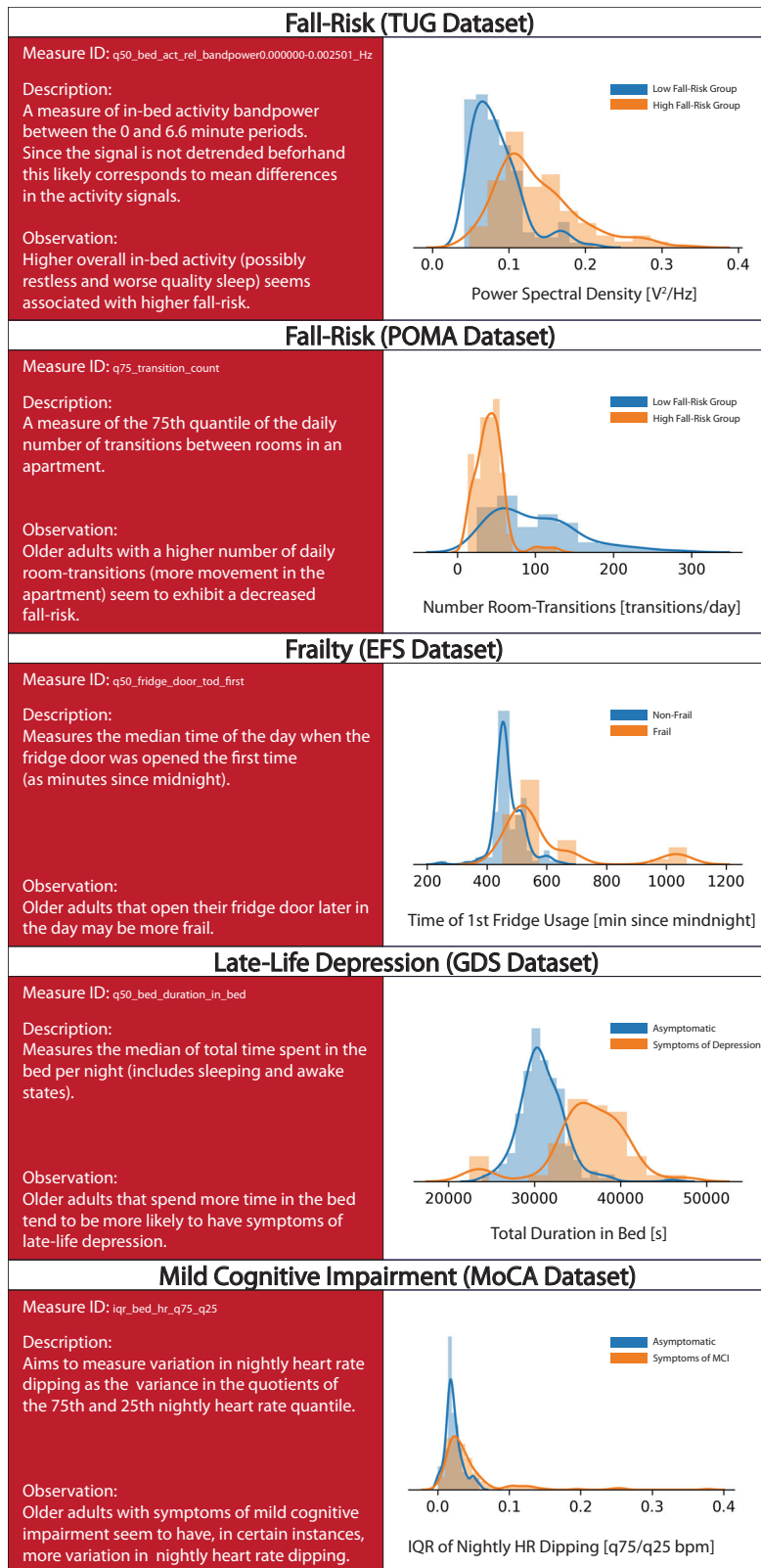
### DISCUSSION

The present study evaluated the idea of a comprehensive digital exhaust for long-term remote monitoring in older adults. To this end, we introduced 1268 well-documented digital measures that aim to cover a large part of a person's activity, behavior and physiology (extensive online documentation can be found on <https://narayanschuetz.github.io/digital-exhaust/>). Since most successful long-term remote-monitoring projects in older adults have employed zero-interaction, contactless sensing technologies, we based all introduced digital measures on such technologies. Using the resulting digital exhaust, in combination with real-world data, we could successfully create large-scale ML derived digital COAs for common ageing-relevant outcomes, including fall risk, frailty, MCI, and to a lesser degree late-life depression. Furthermore, we were able to showcase the discovery of potentially interesting new digital biomarkers related to the created digital COAs.

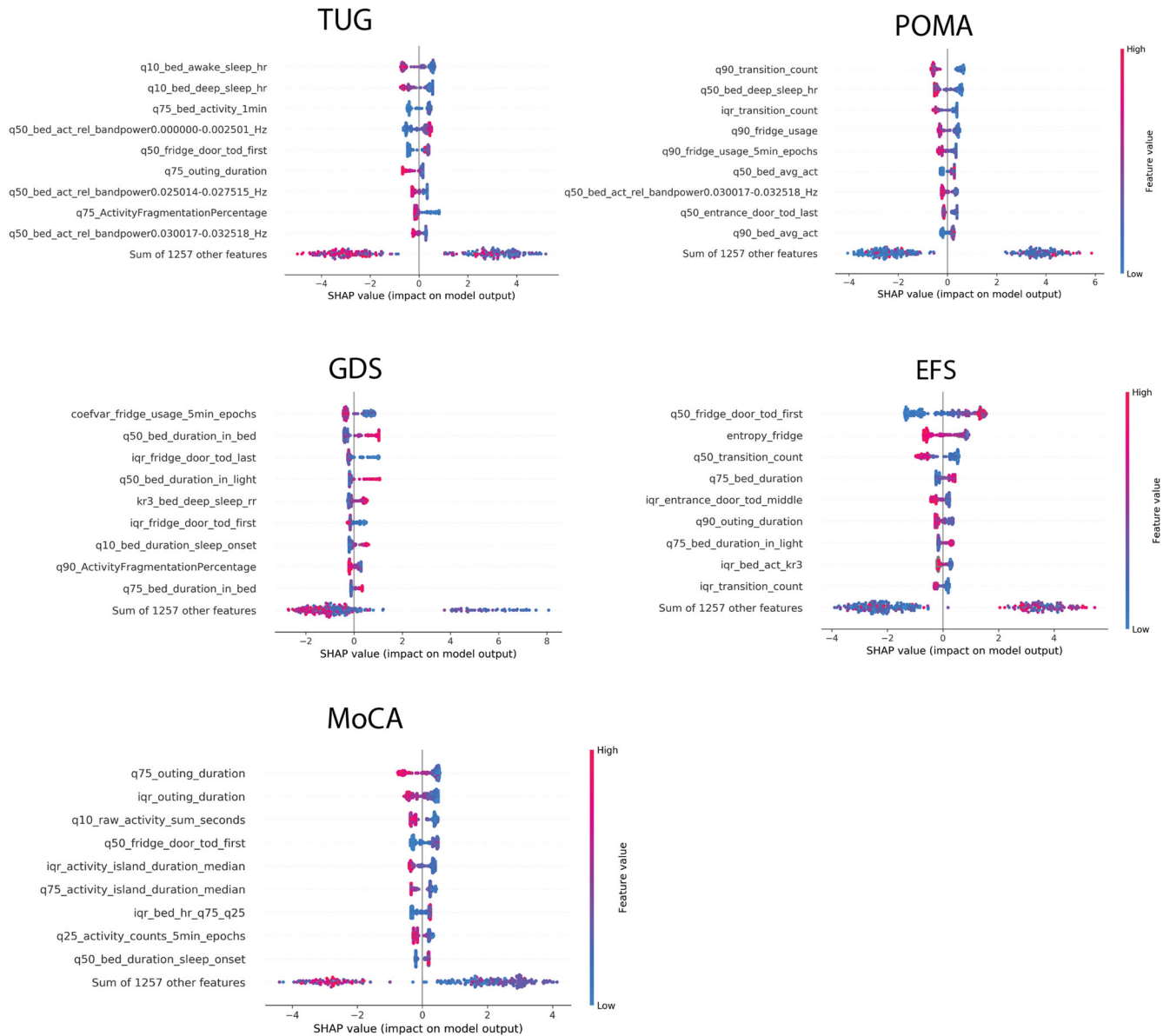
Beginning with the digital COAs, we found good discriminative performances across all but late-life depression, resulting in ROC AUC values of  $\geq 0.7$ . Notably, these results are based on a very limited sample size, which makes it probable that this is a rather conservative estimate of what could be possible. In all cases, digital COAs based on the digital exhaust led to higher ROC AUC (see Table 1) and PrAUC (see Table 2) values than those obtained from only using demographics information. These differences were significant in three (TUG, EFS, and MoCA) out of the five outcome datasets, indicating that a digital exhaust captures information beyond just simple demographics. It is also notable that adding demographic information to the digital exhaust did not result in significantly better performance across outcomes, which may indicate that this type of information was already latently captured by the measures making up the exhaust. While we used digital COAs here more as an example for feasibility purposes, those could in fact be highly useful for continuously assessing an older adult's health and functional status with respect to specific outcomes and may allow for the implementation of early preventive interventions, fitting in well with the proactive nature of precision medicine. For instance, if an older adult exhibits increased fall risk, it may be reasonable for them to see a fall prevention specialist, as opposed to taking action after a fall has already occurred.

Putting the above mentioned results into perspective, passive sensor-based fall-risk assessments were shown to yield AUC values in the range of 0.65–0.89<sup>50</sup>. It should be noted that these values were obtained using wearables by means of accelerometry and predominantly with very few digital measures of gait (and sometimes with accelerometer signal characteristics). Furthermore, none of the studies mentioned in this paper used long-term data, and some were performed under laboratory conditions far removed from real life. Our results, with ROC AUC values of 0.786 (TUG dataset) and 0.805 (POMA dataset), are thus in line and satisfactory by comparison. Meanwhile, in terms of relevant digital measures, it is notable that not only physical activity and broadly gait-related measures (such as the number of room transitions) but also sleep and rhythmicity measures, such as activity in bed, bed-exit count, or activity fragmentation, were of major importance in discriminating between participants with high and low fall risk (see Fig. 3). Although prospective data would be necessary to form clear conclusions, this may suggest that behavioural data beyond just gait and physical activity may be relevant for fall-risk assessments in older adults. Moreover, results from Piau et al. show that PIR array-based gait speed may help identify future fallers<sup>15</sup>. In this regard, the inclusion of gait-speed information thorough zero-interaction approaches would likely further increase performance on fall-risk related COAs.

In terms of frailty, comparable studies report ROC AUC values between 0.72 and 0.86, based on wearable sensors<sup>51–53</sup>. These results were primarily obtained on the basis of gait and physical



**Fig. 2 Most important digital measure for each outcome.** Displays descriptions and density plots of the most important digital measure for each outcome. Across all density plots, blue indicates a positive/neutral outcome, while orange indicates a negative outcome. It should be noted that the proposed associations reflect correlation and not causation and should be interpreted accordingly.



**Fig. 3 Beeswarm plot indicating digital measure importances across outcomes.** Shows beeswarm plots of the 9 most important digital measures based on SHAP values on all outcome datasets: TUG (Timed Up and Go) & POMA (Performance Oriented Mobility Assessment = fall risk, GDS (Geriatric Depression Scale) = late-life depression, EFS (Edmonton Frail Scale) = frailty, MoCA (Montreal Cognitive Assessment) = mild cognitive impairment. Finally, the contributions of the sum of the remaining measures is displayed. Digital measures are ordered according to their importance, from top to bottom. The x-axis represents log odds, where values above zero indicate relevance for a negative outcome. Colouring further shows the direction of this association, where blue indicates lower values of a given measure and red indicates higher respective measure values. Detailed explanations of the individual measure names are given in the supplementary material or on the supplementary website. Note that these plots are based on models trained on the whole respective dataset and are therefore slightly different from the global importances shown in Supplementary Table 3, which are based on 100 simulation iterations.

activity measures. Additionally, frailty definitions, study type, and participant characteristics differ quite widely, so at best this gives a broad idea of what is possible. With a ROC AUC value of 0.704, our results are on the lower end of this spectrum. However, given the types of sensors we employed, this seems realistic. Indeed, some of the most important measures related to frailty were related to fridge usage, physical activity (room-transition counts), and sleep duration (see Fig. 3), all of which seem plausible as potential digital biomarkers for frailty.

For late-life depression, comparable studies are lacking. Several studies have demonstrated the utility of using wearable-based digital measures in assessing general depression<sup>54–56</sup>. Furthermore, one instance reported on the assessment of late-life

depression using PIR-derived information on activities of daily living (ADL)<sup>57</sup>. However, it is unclear whether their methodology prevented data leakage, judging by the unusually high ROC AUC values  $\geq 0.95$ . Our own results, by contrast, show modest performance in assessing late-life depression, with a ROC AUC value of 0.620. While this may be due to the low number of participants with a GDS score above 5 in our cohorts, it could also indicate some inherent difficulties in measuring this outcome. Many of the most important individual measures for late-life depression assessments are related to sleep duration (see Fig. 3), which is known to be associated with depression. More interestingly, variations in fridge usage and behaviour complexity

were relevant; however, due to the relatively low discriminative power, further interpretation may not be meaningful.

Regarding the distinction between healthy older adults and those with MCI, recent work has shown ROC AUC values of 0.62–0.80, based on comparable time intervals<sup>20,58</sup>. These values were achieved with wearable devices but also when using additional modalities such as smartphone and computerised assessments. Our result, with a ROC AUC value of 0.780, is thus aligned with similar research and shows that good discriminative performance may potentially be achieved through an entirely passive, zero-interaction set of sensors here as well. Further supporting the plausibility of our results, a respectable body of literature shows how individual digital measures based on PIR and door sensors — such as variability in PIR array-based gait speed<sup>59–61</sup>, ADL regularity<sup>41</sup>, regularity in physical activity<sup>28</sup>, sleep disturbances<sup>62</sup>, and outing duration<sup>63</sup> — differed between older adults with MCI and healthy controls. Finally, with regards to MCI, highly important measures include those related to physical activity, such as the number of room-transitions or the total amount of PIR-based activity. Moreover, sleep-related measures such as sleep duration, activity in bed, and variation in in-bed activity were found to be important (see Fig. 3). Regarding MCI, however, the most noteworthy finding is the inclusion of various sleep-related heart-rate measures - most importantly, the variation in nightly heart-rate dipping behaviour, where unusually high variation seems indicative of MCI (see Fig. 2). This is especially interesting, as it has not been previously reported in connection with digital measures. However, it is known that heart-rate dipping is associated with cardiovascular disease<sup>64</sup> and that cardiovascular risk factors may be involved in cognitive decline<sup>65</sup>. As such, it could be beneficial to further investigate the relationship between nightly heart rate and mild cognitive impairment.

Overall, our findings not only suggest that a more comprehensive systems approach towards remote health-monitoring may be promising for long-term clinical care and research, particularly when combined with modern ML approaches, but also demonstrate a potential alternative to commonly employed wearable monitoring of digital measures. As such, although this should be seen as early evidence, employing a digital exhaust, as opposed to using few individual measures, could enable powerful ML derived digital COAs and help to profile and discover novel characteristics of health and disease, eventually empowering the idea of precision (or systems) medicine. The value of a digital exhaust in creating ML derived digital COAs is also supported by the observation that, across all outcomes, the sum of the remaining SHAP values — that is, all digital measures except for the 9 most important ones combined — was highly important in explaining model outcomes (see Fig. 3). Since zero-interaction technologies could be very relevant for long-term remote health-monitoring in older adults, the proposed digital exhaust would also give seniors, that are not comfortable, or not able, to use wearable devices (anymore), a promising alternative. This may be particularly important when considering the still existing digital divide, that was recently shown not to narrow for older adults with serious conditions<sup>66</sup>. Nonetheless, it must be mentioned that most, if not all, digital measures presented could also, at least in theory, be derived by wearable devices, which may be suitable for some but likely not all older adults.

Furthermore, by relying entirely on contactless, zero-interaction technologies, large sets of digital measures can be derived and used without major ethical concerns related to burdening subjects with unnecessary sensing modalities. Best practices laid out by Goldsack et al., for instance, discourage efforts towards sensor-symptoms mapping, which is, to some degree, what a systems approach is doing<sup>67</sup>. However, since in this case sensor technologies respect privacy (no video or audio recordings, for instance) and do not add any additional burden (hence zero-

interaction), there are scarcely any downsides, as would occur with adding additional wearables or even active tasks that require interactions. Despite the positives, our findings also suggest that, at least with the presented digital measures and zero-interaction technologies, some modalities may not be easily assessed, such as in our case the assessment of late-life depression. Eventually, we thus believe that the presented digital exhaust has the potential to serve as a baseline set of measures that may be calculated over long time frames (ranging from years to potentially decades), but which could also be supplemented (potentially over shorter time periods) with digital measures based on more specific sensors, such as pillbox sensors, wearables, smartphones, or even non-invasive biomolecular sensors (for instance on the basis of sweat<sup>68</sup> or saliva<sup>69</sup>), depending on the specific needs, circumstances, and conditions. In clinical care, a baseline set of digital measures could make for a first defense, a basic monitoring layer that helps to indicate when more elaborate, but also more obtrusive and potentially expensive, measurement modalities are necessary (be it based on specific sensing devices, such as a Holter electrocardiogram, or more biological modalities like blood panels or even multi-omics profiling).

Future research should emphasise further analytical and prospective, clinical validation of the included digital measures<sup>70</sup>. Here it will also be of high relevance, to implement clinical grade software infrastructure to support robust long-term collection of the proposed measures as well as integration of new ones. While this will likely require industry participation, a solid first effort has been made with the recently established Collaborative Aging Research Using Technology (CART) initiative, which seeks to make ageing-related digital health approaches more accessible to the broader research community<sup>36</sup>. Moreover, analysing long-term temporal dynamics will be essential, as it would enable the identification of trajectories of certain digital measures or even whole groups thereof. Evaluating trajectories, could be extremely valuable, as was shown by Akl et al., who showcased impressive results regarding MCI classification based on long-term trajectories of several individual digital measures<sup>61</sup>. When considering longitudinal aspects, also concepts around digital resilience biomarkers<sup>69</sup> may be of interest, by, for instance, monitoring how certain measures change as a result of disturbances to others. For instance, how a night of restless sleep influences certain physiological or activity measures the next day. In addition, future research may seek to combine a digital exhaust, such as the one we utilised, with traditional multi-omics profiling and mobile bio molecule sensing in a deep phenotyping effort. This may enable a wide range of new research insights into ageing and ageing-related conditions, as it adds a new layer of objective information for characterising phenotypes of health and disease. As a final caveat, a digital exhaust such as this should never be assumed to be complete or fixed. Future research will add new digital measures while old measures may be merged if they exhibit closely correlated behaviour. As such, in the immediate future, it may be of major value to add more accurate gait-related digital measures to the introduced set, as these have consistently been shown to be highly important across many ageing-relevant health outcomes. Consequently, they are likely to add significant value. Also the addition of novel contactless sensing technologies that support a zero-interaction approach, could be promising. Good candidates here would be sensors based on radio signal technologies.

Although this work provides promising results we would like to point out some of the major shortcomings. First, some of the introduced measures have not been validated beyond the scope of this research (the use or validation of a digital measure in other studies is indicated in the Supplementary Information). This implies that some measures may not quantify what we hypothesise, which could lead to inaccurate interpretations and conclusions. Here it is also important to stress that associations



revealed by the employed model explainability approach do not imply causality. One potential way to overcome this would be to apply approaches around computational causal discovery<sup>71</sup>. Another limitation is the relatively small number of participants used to demonstrate the potential of the presented digital exhaust. Therefore, our results should be treated as early evidence and interpreted with caution, although general tendencies are likely to be valid. Furthermore, in the feasibility demonstration, we use a cross-sectional approach that fails to leverage the temporal trajectories of sequentially collected slices. Indeed, we strongly believe that data collected over multiple years would be necessary to fully explore the utility of digital exhaust based approaches. Here, further longitudinal (ideally over multiple years) studies will be necessary to evaluate this potential. Finally, one drawback of simple PIR and door sensors is that not all digital measures based on this technologies can be calculated when more than one person is living in an apartment. Here a potential solution may be found in wireless radio signal technologies<sup>14,72,73</sup>, that could not only provide more accurate digital measures, compared to simple PIR motion sensors, but should also be able to differentiate between multiple persons, by detecting specific signatures (such as gait characteristics)<sup>74</sup>.

To concluded, we introduce a comprehensive set of digital measures, what may be referred to as a digital exhaust, for long-term remote health-monitoring in the older adult demographic. Overall, the digital exhaust consists of 1268 digital measures derived from 94 hypothesis-driven base measures, covering large parts of a person's daily activity, behavior and physiology. All included digital measures are derived from a small set of zero-interaction, contactless sensing devices that have been successfully used in numerous ageing-related, long-term, remote-monitoring projects around the world. For each measure, we provide a detailed description, background information, and additional real-world data as supplementary online material (<https://narayanschuetz.github.io/digital-exhaust/>). While use cases for the introduced digital exhaust are diverse, we

demonstrate the case of creating multiple large-scale machine learning-derived digital COAs and evaluate their discriminative performance. To this end, we show how ageing-relevant outcomes such as fall risk, frailty, and MCI may be assessed. Our results with this systems approach not only show that combined information from the digital exhaust significantly outperformed basic demographic information, but also that the digital exhaust based digital COAs could often match the performance reported in studies employing more obtrusive wearable sensors. Finally, we highlight the possibility of using the digital exhaust to discover novel digital biomarker candidates, using a model explainability approach on the basis of the ML models used to create the aforementioned digital COAs. The respective results show that the most important digital measures are reasonable digital biomarker candidates, while also revealing two potentially relevant insights. The first being, that while fall risk may be primarily associated with gait and physical activity, it also potentially exhibits strong associations with sleep-related measures. The second indicating that unusually high variation in nocturnal heart-rate dipping may be uniquely related to MCI.

## METHODS

### Creating a zero-interaction digital exhaust

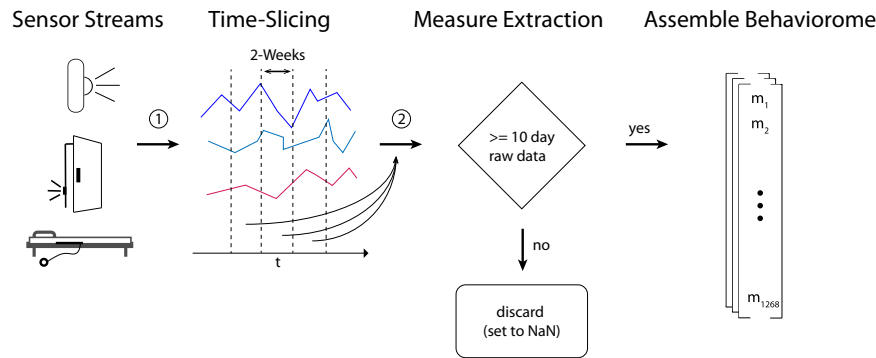
The introduced digital measures are based on three sensor types that have been commonly used in remote-monitoring projects with older adults: PIR sensors, contact door sensors, and a sleep sensor. The PIR motion sensors were placed in the essential rooms of older adults' apartments. Essential rooms included the living room, bedroom, entrance, bathroom/toilet, and kitchen. The employed PIR sensors sampled with 0.5 Hz, and thus reported activity on or off states every 2 s. The reed switch-based door sensors, meanwhile, were placed at the entrance and the refrigerator door. Both PIR and door sensors were part of the DOMO Care® (DomoHealth SA, Lausanne, Switzerland) home-monitoring system. Finally, for the sleep sensor, we used an EMFIT QS ferroelectret sensor (Emfit Ltd, Vaajakoski, Finland), which was fixed beneath the mattress at approximately chest height. A summary of these three devices, as well as their respective source data streams, is given in Table 3.

**Table 3.** Sensing devices data stream summary.

Sensor	Device	Location	Device type	Data stream	Sampling rate	Channels
PIR Motion	DomoCare	Wall/ Ceiling	Contactless	Activity	0.5 Hz	Entrance
						Toilet
						Bedroom
						Livingroom
						Kitchen
Door Sensor	DomoCare	Door	Contactless	Open/ Close	Events	Others
						Entrance
Bed Sensor	EMFIT QS	Beneath Mattress	Contactless	Heart Rate	0.25 Hz	N/A
				Respiration Rate	0.25 Hz	N/A
				Activity	0.25 Hz	N/A
				Toss & Turns	Events	N/A
				Sleep Phases	0.25 Hz	Awake
						REM
						Deep
						Light
						N/A
				Raw Lowband	50 Hz	N/A
Raw Highband	100 Hz	N/A				
Bed Exits	Events	N/A				
Statistics	Daily	N/A				

A summary of the contactless pervasive computing devices used, including the respective data streams they provide.





**Fig. 4 Digital measure extraction flowchart.** Shows a broad summary of how digital measures were calculated, starting with raw sensor data from PIR sensors, door sensors, and bed sensors. Raw data streams were first segmented into non-overlapping bi-weekly segments. Then, for each bi-weekly segment, digital measures were calculated. If, for a given measure, less than 10 days of data were present, the measure was encoded as missing, which eventually left 1268 dimensional vectors - one per bi-weekly segment.

The creation of the digital 94 base measures was mostly hypothesis-driven or based on measures from the existing literature. The majority of base measures was calculated on a daily or nightly basis (for instance, daily total activity, daily outing duration, or average heart rate during a night). Subsequently, we calculated derivatives of those measures by means of descriptive statistics over non-overlapping bi-weekly segments, resulting in the final number of 1268 digital measures. While bi-weekly segments may be somewhat arbitrary, two weeks is a sufficient period to capture variation in behaviour that often follows daily or weekly cycles while still being short enough to capture temporally limited behaviours. Additionally, it serves to increase the number of data points (data augmentation) and facilitates the process of working with sensor recordings of various lengths or with data gaps. For cases of certain behavioural or rhythmicity measures, such as Cosinor regression-based measures, raw data from the whole bi-weekly segment was used directly. To avoid the inclusion of measures with insufficient data, we set a minimum number of 10 days for which raw source data was available throughout a given bi-weekly segment; otherwise, the measure was set as missing. The criteria for including a day's worth of data for each sensor type are explained in detail in the supplementary material (Supplementary Note 1).

For all daily or sub-daily base measures, derivatives based on summary statistics were calculated over the bi-weekly segments. Summary statistics include various quantiles, denoted as  $q_n$  (e.g.,  $q_{10}$ ), the interquartile range (iqr), the mean, median ( $=q_{50}$ ), coefficient of variation (coefvar), and robust measures of kurtosis and skewness (kr3 and sk3, respectively), following the naming convention proposed by Kim and White<sup>75</sup>. Figure 4 summarises this workflow visually. Eventually, this left 1268 dimensional vectors (one per bi-weekly segment). Of those, 224 dimensions are related to PIR and door sensors, while the remaining 1044 are based on sleep-sensor data. Detailed information regarding the exact calculation of each measure, as well as individual distributions across our cohort, can be found at [online](#).

### Example visualisation

We provide an example of averaged digital exhausts with respect to MCI. These visualisations were created by first averaging the digital exhausts on a per-participant basis, followed by z-normalisation. After that, the exhausts were split into the positive and negative outcome groups and averaged once more. Finally, heatmaps for both conditions were created. As a result, the values of individual digital measures  $> 0$  indicate above-average values, while those  $< 0$  indicate below-average values.

### Machine learning based digital clinical outcome assessments

To test the feasibility of using the previously described digital exhaust to create ageing-relevant digital COAs, we used real-world remote-monitoring data from two cohorts of older Swiss adults (pooled age years =  $87 \pm 7$ ; sex 67% [30/45] female). The original studies were both pilots designed to assess novel computing technologies for ageing-in-place scenarios in the German- and French-speaking cantons of Switzerland<sup>76,77</sup>. They were conducted between 2017 and 2018 and monitored participants over one year with a set of pervasive computing devices and clinical assessments<sup>76,77</sup>. The inclusion criteria between cohorts were similar in the sense that both aimed to recruit a natural sample of community-dwelling older

**Table 4.** Participant Characteristics.

Characteristic	Cohort 1 ( $n = 24$ )	Cohort 2 ( $n = 21$ )	Pooled ( $n = 45$ )	Differences ( $p$ -value)
Age (years)	$88 \pm 7$	$86 \pm 7$	$87 \pm 7$	0.50
Sex female (%)	79	52	67	0.06
TUG	$14.0 \pm 10.0$	$15.3 \pm 4.8$	$14.5 \pm 8.2$	0.58
POMA	$21.4 \pm 7.4$	$24.1 \pm 3.3$	$22.5 \pm 6.1$	0.13
EFS	$4.6 \pm 3.6$	$5.0 \pm 1.7$	$4.7 \pm 2.9$	0.65
GDS	$2.3 \pm 2.3$	$4.2 \pm 3.2$	$3.1 \pm 2.9$	0.04
MoCA	$20.5 \pm 5.1$	$20.6 \pm 3.7$	$20.5 \pm 4.7$	0.95

Details characteristics of the participants from the two pilot studies we used in the presented real-world evaluations.

adults (aged  $\geq 70$  years) who lived alone and without pets. On the other hand, the exclusion criteria between cohorts differed. For cohort 1, the only exclusion criterion was an unwillingness to comply with the study protocol. But, for cohort 2, the exclusion criteria were as follows: (1) severe cognitive impairment rendering the individual unable to follow study protocol (clock-drawing score  $\geq 4$ ); (2) skin problems such as irritations, itching, or serious redness; (3) undergoing dialysis; (4) unwillingness to comply with the study protocol; (5) an inability to understand the study aim; or (6) hospitalisation planned within a short period of time<sup>76</sup>. Both studies were conducted based on principles declared in the Declaration of Helsinki and approved by the Ethics Committees of the cantons of Bern and Vaud (KEK-ID: 2016-00406 and CER-VD ID: 2016-00762, respectively). All subjects signed and returned informed consent forms before participating in the study. Detailed participant characteristics and cohort differences are shown in Table 4. The differences between cohorts were statistically examined on the basis of unpaired, two-sided, two-sample  $t$ -tests ( $\alpha = 0.05$ ). In every analysis involving participant data, all participants with any available data (depending on sensor data and the availability of clinical assessments) were included; this also applies to participants that dropped out of the studies.

Participants in both cohorts were subject to an overlapping set of standardised clinical assessments. These include the following six assessments: (1) the Timed Up and Go Test (TUG), which is often used in geriatrics to assess fall risk<sup>78</sup>; (2) the Tinetti Performance-Oriented Mobility Assessment (POMA), which, as with the TUG, also measures balance and gait characteristics that are often indicators for elevated fall risk among older adults<sup>79</sup>; (3) the Edmonton Frail Scale (EFS), a frequently used measure of frailty among older adults<sup>80</sup>; (4) the short version (15-item) of the Geriatric Depression Scale (GDS), a commonly used late-life depression screening tool<sup>81</sup>; (5) the Montreal Cognitive Assessment (MoCA), which measures cognitive function and is often used as a brief screening tool for the detection of MCI in older adults<sup>82</sup>. In each cohort, these assessments were planned to be conducted at least once during the one-year study duration. Detailed assessment intervals are summarised in Table 5.

**Table 5.** Clinical assessments and employed cut-off points.

Assessment	Interval cohort 1	Interval cohort 2	Cut-off points	Outcome
TUG	6 weeks	6 weeks	$\geq 12$ s <sup>88</sup>	Fall-risk
POMA	half year	half year	$< 19$ points <sup>89</sup>	Fall-risk
EFS	begin/end	begin/end	$> 5$ points <sup>90</sup>	Frailty
GDS	begin/end	begin/end	$\geq 5$ points <sup>91</sup>	Depression
MoCA	begin/end	end	$< 23$ points <sup>92</sup>	MCI

Displays an overview of clinical assessments used to evaluate the health status of study participants, including the measurement interval and cut-off points used to divide participants into groups with positive/neutral or negative health outcomes.

**Table 6.** Dataset Characteristics.

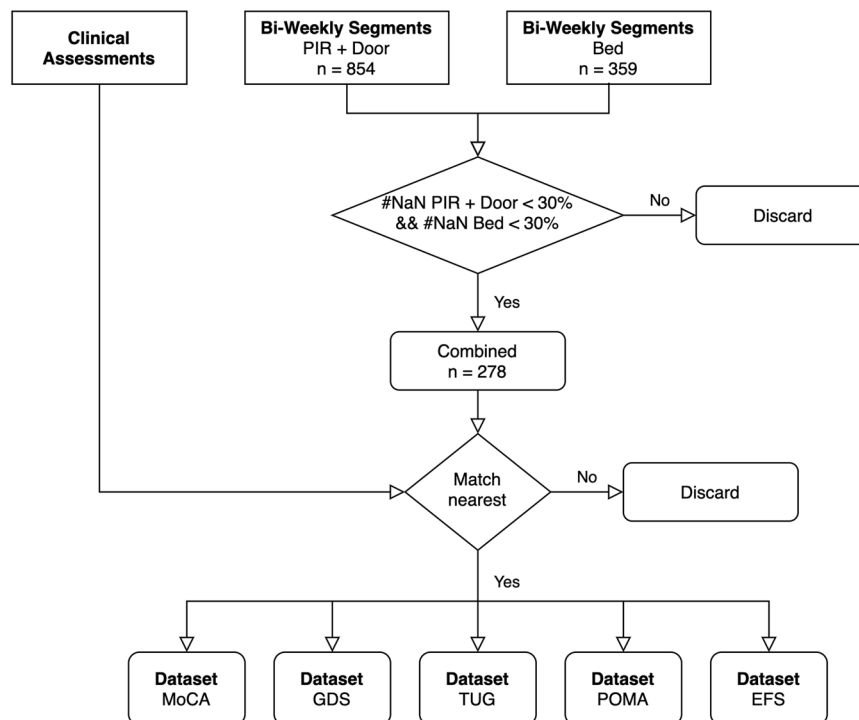
Dataset	Participants total	Participants with negative outcome	Number Bi-weekly segments
TUG	28	14	277
POMA	28	10	277
EFS	28	10	277
GDS	28	7	277
MoCA	25	16	260

Displayed are the detailed characteristics of the individual outcome datasets. This includes the total number of participants per dataset, the number of participants exhibiting a negative health outcome, as well as the total number of resulting bi-weekly segments (data samples).

To evaluate the potential for creating digital COAs that may help differentiate between positive and negative ageing-relevant health outcomes based on the proposed digital exhaust, we categorise participants into one of the two categories for each clinical assessment. This was done on the basis of validated cut-offs for each assessment, as described before. The respective cut-off values for the negative groups are shown in Table 5. Next, we calculated the digital exhaust for all participants. For each assessment, we then combined the positive/negative labels with the bi-weekly segments of a given participant. If multiple records of the same clinical assessments were obtained throughout the study, we assigned the target label corresponding to the assessment closest in time. After this procedure, we obtained one dataset per assessment.

Note that measures derived from PIR and door sensors stem from one sensor system (meaning that technical failure usually affect both sensor types, except for instances where an individual sensor unit failed, which happened rarely), while sleep stems from another sensor; thus, for a bi-weekly segment to be valid, at least 30% of measures from both sensor systems must be valid. This led to a significant reduction in the number of bi-weekly segments, as a large number of sleep sensor data were missing due to technical issues, as has been discussed in prior work<sup>39</sup>. These two issues — lacking sensor data from both PIR/door and bed sensors and the unavailability of respective assessments — are responsible for the generally lower numbers of participants who could be included in this analysis (the exact numbers with regards to each assessment are given in Table 6). In Fig. 5, we present the high-level flowchart of dataset creation.

A small forenote aimed at a more technically oriented audience: what we call digital measures throughout this work can be seen as synonymous to the more abstract and general term “features”. To evaluate digital COAs based on the digital exhaust, we largely followed the approach set out by Chen et al., albeit with some minor changes<sup>20</sup>. As such, we use the gradient boosting-based XGBoost algorithm<sup>83</sup> as a classifier, since it generally performs impressively on tabular data, tends to deal reasonably well with high-dimensional feature spaces (even in  $p \gg n$ -type scenarios, as here), and can inherently deal with missing values, all of which means it is close to being the gold standard for this kind of application<sup>20,54,84</sup>. Furthermore, gradient boosting-based tree approaches tend to be more easily explainable than modern neural network approaches such as



**Fig. 5 Dataset creation overview.** Highlights the workflow of creating datasets, subsequently used for the creation of digital clinical outcome assessments. First, digital measures were separately calculated for the PIR + door and bed sensors and segmented into non-overlapping, bi-weekly segments. After that, the measures from bi-weekly segments, where the percentage of missing digital measures from either sensor system was  $< 30\%$ , were combined. Next, the clinical assessments from each participant were matched with the respective bi-weekly digital measure vectors to combine 5 datasets --- one for each assessment.

convolutional neural networks, while also retaining high accuracy, especially on tabular data structures<sup>49</sup>. To better account for stochasticity in participant selection, we further adapted the simulation strategy of Chen et al.<sup>20</sup>, in which 70% of participants were repeatedly drawn from the entire participant pool to form a training set, while the remaining 30% were used as a test set (the splits are stratified for the respective clinical assessment labels). This procedure was repeated for 100 iterations. Note, that this way each new draw represents a shuffling of the dataset without introducing data leakage between training and test splits. Throughout each iteration, hyperparameters were first optimised within the training split by means of stratified 3-fold cross-validation coupled with random search (consisting of 50 search trials). For more detailed explanation of this strategy, we refer to the original article by Chen et al. where it is demonstrated in detail<sup>20</sup>.

Eventually, for each iteration, we calculated the Area Under the Receiver-Operating Curve (ROC AUC) and the Area Under the Precision Recall Curve (PrAUC) on the test set, where multiple bi-weekly segments from a single participant were combined into one score by averaging their predictions (soft voting), as was done in<sup>20</sup>. Likewise, if multiple assessment results were available, they were first averaged, which should have also reduced some of the inherent noise; these results were then dichotomised on the basis of the previously introduced cut-off points (see Table 5) to yield a single label per participant. We removed three digital measures (Measure IDs: *iqr\_entrance\_door\_tod\_first*, *q50\_entrance\_door\_tod\_first*, and *q50\_fridge\_door\_tod\_middle*) from the full set for this portion of the analysis, as they were biased towards identifying one of the two cohorts (to account for further less-obvious biases in this regard, we included cohort information in the demographics). Note that the PrAUC is sensitive to label distribution, which means it only lends itself to comparisons within the same dataset. For each assessment, we ran three different scenarios, one with only demographic information (age, sex, and cohort membership) as baseline, one with only the digital exhaust, and one with both the exhaust and demographic information combined. Differences between these scenarios were deemed statistically significant if the 95% CIs of two conditions do not overlap. Model hyperparameter ranges are given in Supplementary Table 1. We used the original Python (version 3.6) implementation of XGBoost (version 1.3.3). Model training was performed on UBELIX (<http://www.id.unibe.ch/hpc>), the HPC cluster at the University of Bern.

### Discovering novel digital biomarkers

To better understand the role of individual digital measures in machine learning-based COAs, we used SHapley Additive exPlanations (SHAP), a game-theoretic approach for explaining complex machine learning models. With this approach, exact solutions can be found in the case of tree-based models<sup>48,49</sup>. SHAP values have been used fairly extensively in recent biomedical applications<sup>20,85–87</sup>. For each of the assessments, we provide overall global SHAP values across all 100 simulations. That is, we give the mean absolute value of the SHAP values for a given digital measure  $m$  in a single simulation, summed up over all simulations, as depicted in equation (1).

$$SHAP_m^{global} = \sum_{i=1}^{100} \text{mean}(|SHAP_m^i|). \quad (1)$$

While global SHAP values reveal the overall importance of a given digital measure, they do not say anything about the direction in which the digital measure influences a model. Therefore, we additionally calculate beeswarm plots of the SHAP values. These are based on a model trained over the entire respective dataset, with manually set hyperparameters (reported in the supplementary material). SHAP values were calculated using Python (version 3.6) with the *shap* package (version 0.39.0).

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

Original data used in this manuscript may be obtained upon request but will require ethical approval from the responsible authorities. Limited aggregated data are available online (<https://narayanschuetz.github.io/digital-exhaust/>).

### CODE AVAILABILITY

Code used to derive the respective digital measures may be obtained upon request but will not be useful without the exact same sensor setup. A detailed rigorous description is therefore provided on GitHub and in the supplementary Information, including mathematical depictions and pseudo-code (where applicable), such that it is easily possible to replicate each digital measure (<https://narayanschuetz.github.io/digital-exhaust/>).

Received: 3 January 2022; Accepted: 13 July 2022;

Published online: 16 August 2022

### REFERENCES

- Bloom, D. E., Canning, D. & Lubet, A. Global population aging: Facts, challenges, solutions & perspectives. *Daedalus* **144**, 80–92 (2015).
- Morley, J. E. & Vellas, B. Patient-centered (p4) medicine and the older person. *J. American Med. Directors Asso.* **18**, 455–459 (2017).
- Au, R., Ritchie, M., Hardy, S., Ang, T. F. A. & Lin, H. Aging well: Using precision to drive down costs and increase health quality. *Adv Geriatric Med. Res.* **1**, e190003 (2019).
- Hood, L. Systems biology and p4 medicine: past. *Rambam Maimonides Med. J.* **4**, e0012 (2013).
- Flores, M., Glusman, G., Brogaard, K., Price, N. D. & Hood, L. P4 medicine: How systems medicine will transform the healthcare sector and society. *Personalized Med.* **10**, 565–576 (2013).
- Hood, L. P4 medicine and scientific wellness: catalyzing a revolution in 21st century medicine. *Molecular Front. J.* **1**, 132–137 (2017).
- Gibbs, W. W. et al. Medicine gets up close and personal. *Nature* **506**, 144–145 (2014).
- Maron, J. L. & Jones, G. B. How sensors, devices, and biomarkers can transform precision medicine: Perspectives from a clinical and translational science institute. *Clinical Therapeutics* **40**, 345–348 (2018).
- Rantz, M. J. et al. A new paradigm of technology-enabled ‘vital signs’ for early detection of health change for older adults. *Gerontology* **61**, 281–290 (2015).
- Al-Khafajiy, M. et al. Remote health monitoring of elderly through wearable sensors. *Multimedia Tools Applications* **78**, 24681–24706 (2019).
- Lyons, B. E. et al. Pervasive computing technologies to continuously assess alzheimer’s disease progression and intervention efficacy. *Front. Aging Neurosci.* **7**, 102 (2015).
- Jacobson, N. C., Weingarden, H. & Wilhelm, S. Digital biomarkers of mood disorders and symptom change. *NPJ Digital Med.* **2**, 1–3 (2019).
- Piau, A., Wild, K., Mattek, N. & Kaye, J. Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive function for mild cognitive impairment to mild alzheimer disease and implications for clinical care: Systematic review. *J. Medical Internet Res.* **21**, e12785 (2019).
- Kabelac, Z. et al. Passive monitoring at home: A pilot study in Parkinson’s disease. *Digital Biomarkers* **3**, 22–30 (2019).
- Piau, A. et al. When will my patient fall? sensor-based in-home walking speed identifies future falls in older adults. *J. Gerontol.: Series A* **75**, 968–973 (2020).
- Zhang, H., Deng, K., Li, H., Albin, R. L. & Guan, Y. Deep learning identifies digital biomarkers for self-reported parkinson’s disease. *Patterns* **1**, 100042 (2020).
- Evers, L. J. et al. Real-life gait performance as a digital biomarker for motor fluctuations: The parkinson@ home validation study. *J. Medical Internet Res.* **22**, e19068 (2020).
- Coravos, A. et al. Digital medicine: A primer on measurement. *Digital Biomarkers* **3**, 31–71 (2019).
- Zhan, A. et al. Using smartphones and machine learning to quantify parkinson disease severity: the mobile parkinson disease score. *JAMA Neurol.* **75**, 876–880 (2018).
- Chen, R. et al. Developing measures of cognitive impairment in the real world from consumer-grade multimodal sensor streams. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2145–2155 (2019).
- Vieira, F. G. et al. A machine-learning based objective measure for als disease severity. *Npj Digit Med.* **5**, 45. <https://doi.org/10.1038/s41746-022-00588-8> (2022).
- Servais, L. et al. Stride velocity 95th centile: Insights into gaining regulatory qualification of the first wearable-derived digital endpoint for use in duchenne muscular dystrophy trials. *J. Neuromuscular Dis.* **9**, 335–346 (2021).
- Kourtis, L. C., Regele, O. B., Wright, J. M. & Jones, G. B. Digital biomarkers for alzheimer’s disease: the mobile/wearable devices opportunity. *NPJ Digital Med.* **2**, 1–9 (2019).
- Dillenseger, A. et al. Digital biomarkers in multiple sclerosis. *Brain Sci.* **11**, 1519 (2021).

25. Gold, M. et al. Digital technologies as biomarkers, clinical outcomes assessment, and recruitment tools in Alzheimer's disease clinical trials. *Alzheimer's & Dementia: Translational Res. Clinical Interventions* **4**, 234–242 (2018).
26. Dorsey, E. R., Papapetropoulos, S., Xiong, M. & Kiebertz, K. The first frontier: digital biomarkers for neurodegenerative disorders. *Digital Biomarkers* **1**, 6–13 (2017).
27. Austin, J. et al. A smart-home system to unobtrusively and continuously assess loneliness in older adults. *IEEE J. Translational Eng. Health Med.* **4**, 1–11 (2016).
28. Hayes, T. L. et al. Unobtrusive assessment of activity patterns associated with mild cognitive impairment. *Alzheimer's & Dementia* **4**, 395–405 (2008).
29. Peek, S. T. et al. Factors influencing acceptance of technology for aging in place: a systematic review. *Int. J. Medical Informatics* **83**, 235–248 (2014).
30. Murphy, S. L. Review of physical activity measurement using accelerometers in older adults: Considerations for research design and conduct. *Preventive Med.* **48**, 108–114 (2009).
31. Rogers, E. M. The digital divide. *Convergence* **7**, 96–111 (2001).
32. Cullen, R. Addressing the digital divide. Online information review (2001).
33. Choi, N. G. & DiNitto, D. M. The digital divide among low-income homebound older adults: Internet use patterns, health literacy, and attitudes toward computer/internet use. *J. Medical Internet Res.* **15**, e2645 (2013).
34. Eruchalu, C. N. et al. The expanding digital divide: Digital health access inequities during the covid-19 pandemic in new york city. *J. Urban Health* **98**, 183–186 (2021).
35. Kaye, J. et al. Unobtrusive measurement of daily computer use to detect mild cognitive impairment. *Alzheimer's & Dementia* **10**, 10–17 (2014).
36. Beattie, Z. et al. The collaborative aging research using technology initiative: An open, sharable, technology-agnostic platform for the research community. *Digital Biomarkers* **4**, 100–118 (2020).
37. Kaye, J. A. et al. Intelligent systems for assessing aging changes: Home-based, unobtrusive, and continuous assessment of aging. *J. Gerontol. Series B: Psychol. Sci. Soc. Sci.* **66**, i180–i190 (2011).
38. Goonawardene, N., Tan, H.-P. & Tan, L. B. Unobtrusive detection of frailty in older adults. In International Conference on Human Aspects of IT for the Aged Population, 290–302 (Springer, 2018).
39. Schütz, N. et al. Contactless sleep monitoring for early detection of health deteriorations in community-dwelling older adults: Exploratory study. *JMIR mHealth uHealth* **9**, e24666 (2021).
40. Schutz, N. et al. A sensor-driven visit detection system in older adults homes: Towards digital late-life depression marker extraction. *IEEE J. Biomed Health Inform* **26**, 1560–1569 (2021).
41. Urwyler, P. et al. Cognitive impairment categorized in community-dwelling older adults with and without dementia using in-home sensors that recognise activities of daily living. *Scientific Rep.* **7**, 1–9 (2017).
42. Cook, D. J. & Schmitter-Edgecombe, M. Fusing ambient and mobile sensor features into a behaviorome for predicting clinical health scores. *IEEE Access* **9**, 65033–65043 (2021).
43. Schütz, N. et al. Validity of pervasive computing based continuous physical activity assessment in community-dwelling old and oldest-old. *Scientific Rep.* **9**, 1–9 (2019).
44. Hayes, T. L., Hunt, J. M., Adami, A. & Kaye, J. A. An electronic pillbox for continuous monitoring of medication adherence. In 2006 international conference of the IEEE engineering in medicine and biology society, 6400–6403 (IEEE, 2006).
45. Rashidisabet, H. et al. A systems biology approach to the digital behaviorome. *Curr. Opin. Sys. Biol.* **20**, 8–16 (2020).
46. Wright, J. M. & Jones, G. B. Harnessing the digital exhaust: incorporating wellness into the pharma model. *Digital Biomarkers* **2**, 31–46 (2018).
47. Baker, M. The omes puzzle. *Nature* **494**, 416 (2013).
48. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. In Guyon, I. et al. (eds.) *Advances in Neural Information Processing Systems* 30, 4765–4774 (Curran Associates, Inc., 2017). <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>.
49. Lundberg, S. M. et al. From local explanations to global understanding with explainable ai for trees. *Nat. Machine Intelligence* **2**, 2522–5839 (2020).
50. Sun, R. & Sosnoff, J. J. Novel sensing technology in fall risk assessment in older adults: a systematic review. *BMC Geriatrics* **18**, 1–10 (2018).
51. Kumar, D. P. et al. Sensor-based characterization of daily walking: a new paradigm in pre-frailty/frailty assessment. *BMC Geriatrics* **20**, 1–11 (2020).
52. Schwenk, M. et al. Wearable sensor-based in-home assessment of gait, balance, and physical activity for discrimination of frailty status: Baseline results of the arizona frailty cohort study. *Gerontology* **61**, 258–267 (2015).
53. Park, C., Mishra, R., Golledge, J. & Najafi, B. Digital biomarkers of physical frailty and frailty phenotypes using sensor-based physical activity and machine learning. *Sensors* **21**, 5289 (2021).
54. Makhmutova, M. et al. Prediction of self-reported depression scores using person-generated health data from a virtual 1-year mental health observational study. In Proceedings of the 2021 Workshop on Future of Digital Biomarkers, 4–11 (2021).
55. Meyerhoff, J. et al. Evaluation of changes in depression, anxiety, and social anxiety using smartphone sensor features: Longitudinal cohort study. *J. Medical Internet Res.* **23**, e22844 (2021).
56. Vahia, I. V. & Sewell, D. D. Late-life depression: A role for accelerometer technology in diagnosis and management. *American J. Psychiatry* **173**, 763–768 (2016).
57. Kim, J.-Y., Liu, N., Tan, H.-X. & Chu, C.-H. Unobtrusive monitoring to detect depression for elderly with chronic illnesses. *IEEE Sensors J.* **17**, 5694–5704 (2017).
58. Li, J. et al. Tatc: predicting Alzheimer's disease with actigraphy data. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 509–518 (2018).
59. Hayes, T. L., Hagler, S., Austin, D., Kaye, J. & Pavel, M. Unobtrusive assessment of walking speed in the home using inexpensive pir sensors. In 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 7248–7251 (IEEE, 2009).
60. Dodge, H., Mattek, N., Austin, D., Hayes, T. & Kaye, J. In-home walking speeds and variability trajectories associated with mild cognitive impairment. *Neurology* **78**, 1946–1952 (2012).
61. Akl, A., Taati, B. & Mihailidis, A. Autonomous unobtrusive detection of mild cognitive impairment in older adults. *IEEE Transac. Biomedical Eng.* **62**, 1383–1394 (2015).
62. Hayes, T. L., Riley, T., Mattek, N., Pavel, M. & Kaye, J. A. Sleep habits in mild cognitive impairment. *Alzheimer Dis. Asso. Disorders* **28**, 145 (2014).
63. Petersen, J., Austin, D., Mattek, N. & Kaye, J. Time out-of-home and cognitive, physical, and emotional wellbeing of older adults: A longitudinal mixed effects model. *PLoS One* **10**, e0139643 (2015).
64. Eguchi, K. et al. Nocturnal non-dipping of heart rate predicts cardiovascular events in hypertensive patients. *J. Hypertens.* **27**, 2265 (2009).
65. Ciobica, A., Padurariu, M., Bild, W. & Stefanescu, C. Cardiovascular risk factors as potential markers for mild cognitive impairment and Alzheimer's disease. *Psychiatria Danubina* **23**, 340–346 (2011).
66. Frydman, J. L., Gelfman, L. P., Goldstein, N. E., Kelley, A. S. & Ankuda, C. K. The digital divide: do older adults with serious illness access telemedicine? *J. Gen. Intern Med.* **37**, 984–986 (2022).
67. Goldsack, J. C. & Clay, I. It takes a village: Development of digital measures for computer scientists. In Proceedings of the 2021 Workshop on Future of Digital Biomarkers, 38–44 (2021).
68. Brasier, N. & Eckstein, J. Sweat as a source of next-generation digital biomarkers. *Digital Biomarkers* **3**, 155–165 (2019).
69. van den Brink, W. et al. Digital resilience biomarkers for personalized health maintenance and disease prevention. *Front. Digital Health* **2**, 54 (2021).
70. Goldsack, J. C. et al. Verification, analytical validation, and clinical validation (v3): the foundation of determining fit-for-purpose for biometric monitoring technologies (biomets). *NPJ Digital Med.* **3**, 1–15 (2020).
71. Eberhardt, F. Introduction to the foundations of causal discovery. *Int. J. Data Sci. Analytics* **3**, 81–91 (2017).
72. Su, B. Y., Ho, K., Rantz, M. J. & Skubic, M. Doppler radar fall activity detection using the wavelet transform. *IEEE Transac. Biomedical Eng.* **62**, 865–875 (2014).
73. Saho, K., Shioiri, K. & Inuzuka, K. Accurate person identification based on combined sit-to-stand and stand-to-sit movements measured using doppler radars. *IEEE Sensors J.* **21**, 4563–4570 (2020).
74. Vandersmissen, B. et al. Indoor person identification using a low-power fmcw radar. *IEEE Transac. Geosci. Remote Sensing* **56**, 3941–3952 (2018).
75. Kim, T.-H. & White, H. On more robust estimation of skewness and kurtosis. *Finance Res. Lett.* **1**, 56–73 (2004).
76. Pais, B. et al. Evaluation of 1-year in-home monitoring technology by home-dwelling older adults, family caregivers, and nurses. *Front. Public Health* **8**, 538 (2020).
77. Saner, H. et al. Potential of ambient sensor systems for early detection of health problems in older adults. *Front. Cardiovascular Med.* **7**, 110 (2020).
78. Podsiadlo, D. & Richardson, S. The timed "up & go": A test of basic functional mobility for frail elderly persons. *J. American Geriatrics Soc.* **39**, 142–148 (1991).
79. Tinetti, M. E., Williams, T. F. & Mayewski, R. Fall risk index for elderly patients based on number of chronic disabilities. *American J. Med.* **80**, 429–434 (1986).
80. Rolfsen, D. B., Majumdar, S. R., Tsuyuki, R. T., Tahir, A. & Rockwood, K. Validity and reliability of the edmonton frail scale. *Age Ageing* **35**, 526–529 (2006).
81. Leshner, E. L. & Berryhill, J. S. Validation of the geriatric depression scale-short form among inpatients. *J. Clinical Psychol.* **50**, 256–260 (1994).
82. Nasreddine, Z. S. et al. The montreal cognitive assessment, moca: A brief screening tool for mild cognitive impairment. *J. American Geriatrics Soc.* **53**, 695–699 (2005).
83. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and



- Data Mining, KDD '16, 785–794 (ACM, New York, NY, USA, 2016). <https://doi.org/10.1145/2939672.2939785>.
84. Cook, D. Digitally mapping the human behaviorome. In 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom), 1–1 (IEEE, 2020).
  85. Lundberg, S. M. et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomedical Eng.* **2**, 749 (2018).
  86. Johnsen, P. V., Riemer-Sørensen, S., DeWan, A. T., Cahill, M. E. & Langaas, M. A new method for exploring gene–gene and gene–environment interactions in gwas with tree ensemble methods and shap values. *BMC Bioinformatics* **22**, 1–29 (2021).
  87. Li, R. et al. Machine learning–based interpretation and visualization of nonlinear interactions in prostate cancer survival. *JCO Clinical Cancer Informatics* **4**, 637–646 (2020).
  88. Bischoff, H. A. et al. Identifying a cut-off point for normal mobility: a comparison of the timed 'up and go' test in community-dwelling and institutionalised elderly women. *Age Ageing* **32**, 315–320 (2003).
  89. Faber, M. J., Bosscher, R. J. & van Wieringen, P. C. W. Clinimetric properties of the performance-oriented mobility assessment. *Phys. Ther.* **86**, 944–954 (2006).
  90. Perna, S. et al. Performance of Edmonton Frail Scale on frailty assessment: its association with multi-dimensional geriatric conditions assessed with specific screening tools. *BMC Geriatr.* **17**, 1–8 (2017).
  91. da Costa Dias, F. L. et al. Accuracy of the 15-item Geriatric Depression Scale (GDS-15) in a community-dwelling oldest-old sample: the Pietà Study. *Trends Psychiatry Psychother.* **39**, 276–279 (2017).
  92. Carson, N., Leach, L. & Murphy, K. J. A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *Int. J. Geriatr. Psychiatry* **33**, 379–388 (2018).

## ACKNOWLEDGEMENTS

We would like to thank all people who contributed and made this work possible, in particular also the older adults who participated in the original studies. This study was partially funded by the Swiss Commission for Technology and Innovation (CTI) through the SWISKO project (17662.2 PFES-ES).

## AUTHOR CONTRIBUTIONS

N.S. came up with the original idea and design of the article. All authors designed and planned the study. N.S., S.K., A.B., and M.S. helped in writing and correcting the manuscript. N.S. wrote the draft for the manuscript. N.S. and S.K. performed the data analysis. N.S., S.K., A.B., and M.S. helped creating a detailed description of each

included measure. N.S., H.S., B.P., and P.B. helped collect the used data. All authors reviewed and approved the final manuscript.

## COMPETING INTERESTS

Dr. Philipp Buluscek and Dr. Narayan Schütz (effectively only after writing the article) were employed by DomoHealth SA, which is the manufacturer of one of the used sensor systems (PIR and door sensors). The remaining authors declare no competing financial or non-financial interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41746-022-00657-y>.

**Correspondence** and requests for materials should be addressed to Narayan Schütz.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022