Check for
updates

Wearable Based Calibration of Contactless In-home Motion Sensors for Physical Activity Monitoring in Community-Dwelling Older Adults

Narayan Schütz[1], Hugo Saner[1,2], Angela Botros[1], Philipp Buluschek[3], Prabitha Urwyler[1,4], René M. Müri[1,4] and Tobias Nef[1]*

[1] ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland, [2] Sechenov First Moscow State Medical University, Moscow, Russia, [3] DomoSafety S.A., Lausanne, Switzerland, [4] Department of Neurology, University Neurorehabilitation Unit, University Hospital Bern (Inselspital), University of Bern, Bern, Switzerland

Passive infrared motion sensors are commonly used in telemonitoring applications to monitor older community-dwelling adults at risk. One possible use case is quantification of in-home physical activity, a key factor and potential digital biomarker for healthy and independent aging. A major disadvantage of passive infrared sensors is their lack of performance and comparability in physical activity quantification. In this work, we calibrate passive infrared motion sensors for in-home physical activity quantification with simultaneously acquired data from wearable accelerometers and use the data to find a suitable correlation between in-home and out-of-home physical activity. We use data from 20 community-dwelling older adults that were simultaneously provided with wireless passive infrared motion sensors in their homes, and a wearable accelerometer for at least 60 days. We applied multiple calibration algorithms and evaluated results based on several statistical and clinical metrics. We found that using even relatively small amounts of wearable based ground-truth data over 7–14 days, passive infrared based wireless sensor systems can be calibrated to give largely better estimates of older adults' daily physical activity. This increase in performance translates directly to stronger correlations of measured physical activity levels with a variety of age relevant health indicators and outcomes known to be associated with physical activity.

Keywords: sensor calibration, pervasive computing, passive infrared, physical activity, older adults, outing imputation, ambient assisted living, telemonitoring

INTRODUCTION

Population aging poses unprecedented global challenges to modern health care systems, economies and last but not least, society as a whole (1, 2). Modern information and communication technology has the potential to contribute in overcoming some of these challenges (3–5). This includes the use of pervasive computing technology, such as microprocessor enhanced objects of everyday life. Small sensing devices like smartwatches or smart home appliances may be used to provide continuous remote monitoring of relevant health indicators and outcomes (4), increasingly referred to as digital biomarkers (6–8). These may allow for early detection of health deteriorations, enabling for instance better preventive measures or earlier interventions (9, 10). Additionally, monitoring

of relevant digital biomarkers by means of pervasive computing technologies could allow for continuous assessments of chronic conditions and help in evaluating intervention efficacy (9, 11).

Physical activity (PA) is associated with a wide range of health benefits, including lower rates of all-cause mortality, non-communicable diseases, cardiorespiratory and muscular fitness across all age groups. Regular PA also helps to protect against frailty, sarcopenia, and cognitive decline (12–14). Wearable technologies, known as wearables, that can track individual's PA behavior are popular consumer items with a worldwide distribution, particularly in younger and middle-aged populations. Also, wearable accelerometers are a well-accepted method to objectively measure PA in everyday life (15–17).

While wearable devices like smartwatches, smartphones or fitness trackers would be ideal to track a variety of health relevant markers like physical activity, post-implementation based experience, including our own, point toward a clear preference for unobtrusive contactless sensing devices (9). Reasons for that may include a certain social stigma associated with visibly wearing devices amongst peers (18, 19), difficulty in handling them, added discomfort of having to think about charging and wearing a device (20), as well as skin irritations related to long-term biosensor wear (intensified by sweat in summer). While some of the mentioned issues are related to the perception of the current generation of older adults toward technology, handling wearable devices that need regular maintenance, can also be problematic for older adults with motor, cognitive, and especially memory related, issues. However, the alternative, wireless ambient sensors, are oftentimes either less accurate (for instance infrared sensors or bed motion sensors) or overly intrusive (for instance video or audio-based recording devices).

The use of wearable devices for initial calibration of less accurate but unobtrusive ambient sensors for PA quantification is a novel approach that could minimize the burden of wearing a device, while improving the reliability and thus usefulness of unobtrusive ambient sensors for physical activity tracking significantly. A similar strategy was employed with passive infrared (PIR) sensor based gait-speed estimation, where calibration was performed using a sensor array as ground-truth, but as the authors state, another source, such as a wearable device, could have been used (21).

PIR motion sensors are rather inexpensive, contactless, and unobtrusive. Therefore they are commonly used in long-term in-home monitoring settings with older adults (9, 11, 22–26). We have previously shown that in-home physical activity, quantified by PIR motion sensors can be used to approximate physical activity in old and oldest-old community-dwelling adults (26). However, the PIR motion sensor-based approach has two main disadvantages: (1) baseline activity comparisons of absolute values between participants are difficult if apartments and sensor placements differ and (2) it is unclear how to address outings correctly. We aim to address both problems by using the much more accurate and well-validated accelerometer based physical activity, to initially calibrate the ambient sensor systems.

METHODS

Participants

The data used for this work stems from a study where modern pervasive computing systems were evaluated for telemonitoring in older adults (26). Participants were part of the StrongAge cohort in Olten (Switzerland) (27) and should represent a naturalistic population sample of community-dwelling, alone-living, old and oldest-old adults in Switzerland. We included all participants that had at least 60 days of wearable activity data recorded (first 30 days reserved for calibration and ≥30 days for evaluation) in this analysis, totalling 20 participants (age = 88 ± 8 years). The 60 days were chosen to include as many participants in the dataset as possible while guaranteeing a minimal number of data points.

The original study was conducted based on principles defined in the Declaration of Helsinki and approved by the Ethics Committee of the canton of Bern, Switzerland (KEK-ID: 2016-00406). All subjects signed and handed in an informed consent before study participation.

Pervasive Computing Systems

In this work we made use of the DomoCare® home monitoring system for older adults (DomoSafety S.A., Lausanne, Switzerland), the same as in (26). The system consists of PIR motion sensors (sampling at 0.5 Hz) placed in the participant's apartment. Kitchen, toilet, living-room, entrance, and bedroom were always equipped with at least one sensor, if a separate bathroom was present it was equipped with a sensor as well. In addition, a magnetic door sensor was placed on the entrance and fridge door, respectively. All sensing units communicate via the ZigBee protocol with a base unit, that then sends data to a secure cloud in real-time. The PIR system allows motion detection in individual rooms based on changes in infrared radiation caused by human activity (28). The door sensors allow outings to be calculated based on entrance door opening and closing events, as explained in (29).

For the calibration of the PIR sensor system we used the medical grade Everion® biosensor worn on the upper arm (Biovotion AG, Zürich, Switzerland). Amongst other sensors, the device contains a 3-axis accelerometer that samples at 50 Hz and outputs/stores aggregated and standardized activity (vector magnitude) at 1 Hz. The participants wore the device throughout the daytime and put it on an inductive charger overnight. While charging the device, data was transmitted to a smartphone via Bluetooth Low Energy which was then encrypted and automatically transferred to a secure cloud. Data from DomoCare® systems was first stored on cloud instances from DomoSafety S.A. located in Switzerland and data from the Everion® was initially stored on instances at the University of Bern. Post collection, all data was subsequently transferred to local servers and ingested into an OmniSci (OmniSci, San Francisco, CA, United States) analytics database instance after quality control. A schematic including the data structure is available in the **Appendix**. To initially ensure accelerometer validity, we compared values from the Everion® with the widely used and validated (30) Axivity AX3 (Axivity

Ltd., Newcastle, UK), 3-axis accelerometer [calibrated to local gravity and temperature, as described in (31)] and found good overall agreement.

Problem Definition

There are three major limitations related to the use of PIR sensors for PA quantification: (1) Motion measured by the commonly used simple PIR motion sensors is converted to a binary response, zero if there was no change in infrared radiation above the sensor's sensitivity threshold and one otherwise. It is thus apparent, that simple PIR motion sensors cannot differentiate between the intensity of the motion, unlike a body attached accelerometer; (2) the angle and distance to a sensor can influence if and how long motion is being detected; (3) the size of equipped rooms and the apartment layout in general can lead to different results for the same amount of physical activity exerted by a person. As a result, even if the same person performed the exact same finite set of activities $A = \{a_1, \ldots, a_n\}$ in different PIR motion sensor equipped apartments, measures of these activities between the PIR motion sensor measurement functions $f_{PIR} : A \rightarrow M_{PIR}; M_{PIR} \in \mathbb{R}_+$ and the accelerometer $f_{acc} : A \rightarrow M_{acc}, M_{acc} \in \mathbb{R}_+$ would likely differ widely. Now in reality, this simplification is not exactly true, because certain activities a_i will be measured by the accelerometer but not by the PIR sensors— for instance when a person is outside the apartment, outside the field of view of the PIR sensors or in a non-equipped room. This gives rise to a subset of all measured activities $\tilde{A} \subseteq A = \left\{ a_i \mid a_i \in A, a_i \in dom(f_{PIR}) \right\}$. We will henceforth refer to f_{acc} that is only defined over this subset as $\tilde{f}_{acc} : \tilde{A} \rightarrow \tilde{M}_{acc}$.

The idea of initial calibration is then to find a mapping $\hat{f}_{PIR} : \tilde{A} \rightarrow \hat{M}_{PIR}$, such that for a given activity a_i, the Euclidean distance between the calibrated PIR motion measurement function $\hat{f}_{PIR}$ and the domain restricted accelerometer measurement function $\tilde{f}_{acc}$ is minimized, which can be thought of as a classic regression objective:

$$\min \sqrt{\left(\hat{f}_{PIR}(a_i) - \tilde{f}_{acc}(a_i) \right)^2} \; \forall \, i = 1 \ldots n$$

Well-calibrated $\hat{f}_{PIR}$ functions from different PIR sensor equipped apartments should then allow that somewhat similar results are obtained for a given activity, since f_{acc}, given a certain activity a_i should be similar across apartments. This assumption is only true if the difference between $\tilde{f}_{acc}$ and f_{acc} is not too large and the accelerometer intensity measurements between participants is mostly comparable. The latter assumption is likely true as for instance described in (32), while the former is largely apartment and person specific but may be improved upon by including an estimate for activity while outside.

Learning Calibration Function

To find a suitable function $\hat{f}_{PIR}$, or in this case a distribution over $\hat{f}_{PIR}$ we propose to use Gaussian process regression, such that $\hat{f}_{PIR} \sim GP(\mu, k)$, where $\mu(A) = 0$ is the standardized activity mean and $k(A, A')$ the activity covariance function. Gaussian process regression (GPR) provides various characteristics that

are likely useful in our calibration scenario. First, it allows non-linear relationships to be modeled and is non-parametric (33). In addition, GPR is known to work well with relatively little data and allows a predictive distribution to be obtained, which can help in detecting model uncertainty (33, 34). The included epistemic model uncertainty could be helpful post calibration as it could allow for quantification when patterns not seen during calibration occur, and give respective warnings if total uncertainty increases.

In the shown experiments we ended up using $k(a_i, a_j) = \sigma_0^2 + a_i \cdot a_j + \sigma_n^2 \delta_{ij}$, as kernel defining the covariance function, where δ_{ij} is a Kronecker delta, σ_n^2 is a learnable bias term and σ_0^2 is a learnable noise constant representing additional homogenous aleatoric uncertainty in the activity measurements (33). To give a comparison how other, more traditional algorithms might perform, we additionally evaluated calibration performance with a regular linear regression (LR) algorithm and the popular XGBoost (XGB) implementation (35) of a tree boosting algorithm. The GPR kernel and hyperparameters for the other algorithms were selected by means of 3-fold cross-validation (splitting at the participant level) and random search (36). It is rather difficult to assess the usefulness of the predictive distribution, obtained by the marginal normal of the GP, in a realistic manner. We try to quantify its utility by calculating the linear correlation between the daily average MAE_p^d (see below) and the daily average uncertainty estimate (the σ of the marginal Gaussian distribution).

Data Pre-processing and Representation

We represent individual activities a_i as activity bouts/islands and describe their characteristics in vector space. The activity islands are extracted by first applying a simple moving average low pass filter, with 1-min length, to the total PIR motion activity signal (the sum of the duration where all PIR motion sensors in an apartment were active) and then extracting the activity islands (stretches where low-pass filtered activity is constantly > 0). Based on these islands we calculate the following features that can be used to summarize the islands in vector space: the total duration of the island, the hour of the day, the duration PIR sensors detected activity for each equipped room and the relative activity of each room with respect to the total island duration. Corresponding activity from the wearable accelerometer was also extracted and summed over the island duration, giving us the target activity f_{acc}. The feature matrix was standardized to have zero mean and unit variance across column.

Statistical Evaluation Metrics

First, it should be noted that evaluation was always performed on all available data beyond the initial 30 days that were reserved for calibration. Throughout this work we refer to multiple evaluation metrics that are explained here. First, the mean absolute error MAE_p between the activity estimate for each activity island and its corresponding accelerometer activity was calculated for each participant p. MAE_p^d refers to the same but averaged over a day d. Second, to measure the proportionality of calibration, the Pearson correlation coefficients ρ_p between the sum of estimated post calibration activity $\hat{F}_{PIR}^d = \sum_{a_i \in d} \hat{f}_{PIR}(a_i)$ per day d and the

sum of total accelerometer activity $F_{acc}^d = \sum_{a_i \in d} f_{acc}(a_i)$ per day d were calculated. Similarly to ρ_p, we calculate $\tilde{\rho}_p$ the Pearson correlation coefficient between the sum of daily calibration activity $\hat{F}_{PIR}^d$ and domain restricted sum of accelerometer activity $\tilde{F}_{acc}^d = \sum_{a_i \in d, a_i \in \tilde{A}} \tilde{f}_{acc}(a_i)$. For all metrics, the sample average over all participants can be calculated, resulting in the global MAE, ρ and, $\tilde{\rho}$.

Determining the Amount of Wearable Ground-Truth Data

To assess the relationship between wear-time and calibration performance, we performed calibration with 1 day, 7 days, 14 days 21 days, and 30 days of accelerometer data and calculated $\tilde{\rho}$ and MAE for each wear-time point and each learning algorithm (as results may be algorithm dependent).

Evaluation of Post-calibration Performance Evolution

One of the main concerns about this kind of calibration procedure is the potential degradation that calibration quality could be subjected to over time as a result of a shift in the data generating distribution (e.g., as a result of changing behavior or seasonal patterns). To assess the potential for degradation, we calculated the weekly average MAE for all participants over thirty consecutive weeks (if data was available). To ensure similar scales, we first standardize weekly averages by removing the median and scaled data with respect to the interquartile range. Finally, for each week, the global average was taken, and a regression line estimated. The p-value of the slope coefficient was then used to determine whether the parameter differed significantly (based on $\alpha = 0.05$) from 0, allowing one to decide whether any relevant trend might be present.

Effect of Calibration on Correlations With Clinical Assessments

To evaluate how calibration influences overall relationships with health indicators and outcomes, we calculate the non-parametric Spearman's rank correlation coefficients r between median daily total activity and the mean of the respective clinical assessments (if multiple were taken per participant over the same duration). Clinical assessments include: the fall-risk focused Timed Up and Go (TUG) (37), the balance and gait focused Tinetti Performance Oriented Mobility Assessments (POMA-b and POMA-g) (38), the late life depression focused Geriatric Depression Scale (GDS) (39), the cognition focused Montreal Cognitive Assessment (40), the frailty focused Edmonton Frail Scale (EFS) (41) as well as muscle force focused handgrip, hip flexor and knee extensor strength. To assess whether there are statistical differences between pre- and post-calibration, we apply the non-parametric Wilcoxon signed rank test to the absolute correlation values under the alternative hypothesis that post-calibration values are on average greater compared to pre-calibration values.

Time Spent Outside

To assess overall PA, time spent outside the home needs to be considered. In our case, using PIR in-home sensors, PA outside the home can be seen as blocks of missing data.
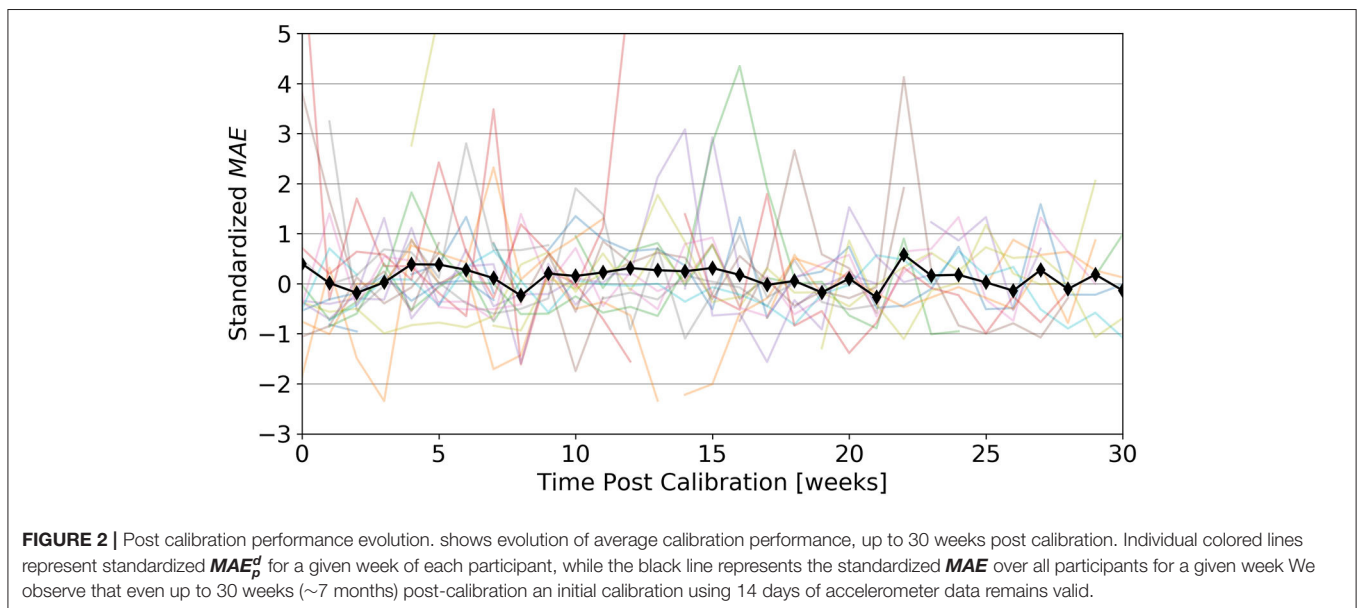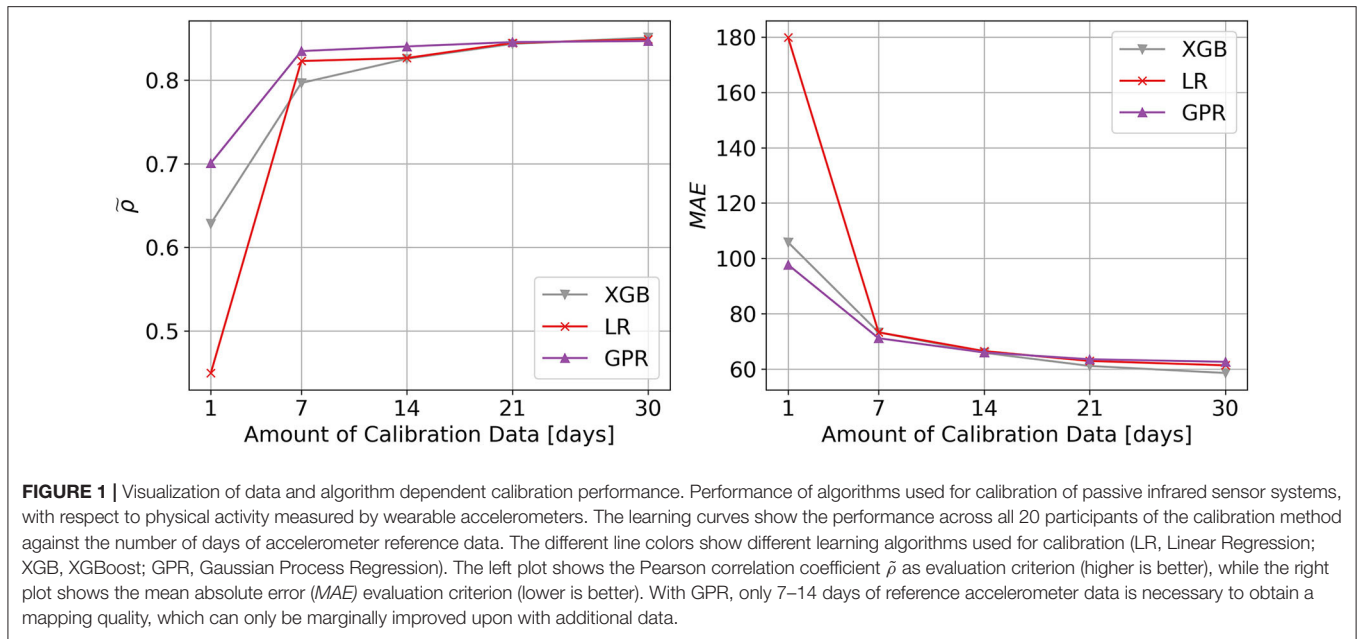
A common strategy to deal with missing data is called imputation, which refers to replacing missing data with substitutes (for instance a variable's mean over all observed values) (42). Imputation can often work reasonably well, if the data is "missing completely at random" or "missing at random" (42). Given that outing likely involves more physical activity than being inside, it may be impossible to correctly impute physical activity of outing periods. Fortunately, access to calibration data from a wearable (given the wearable is also worn outside, which is true in our case), allows us to estimate a factor τ_p (for each participant p) by which the expected inside activity should be multiplied with. To calculate this factor, we first calculate outings according to (29) and then for each outing we divide the physical activity measured by the accelerometer with the average activity of the accelerometer during the same time of day, when the person was at home. Eventually, the median of these ratios gives us τ_p. A global factor τ can then be calculated by averaging over all individual participant's τ_p. As we are dealing with missing time blocks, we use temporal means—similar to what has been used for imputing non-wear time intervals with accelerometers (17, 43). That is, the expected activity sum for the given time-interval (when the outing occurred) over all observed days. To evaluate the effect of this imputation procedure on overall calibration, the evaluation metric $\tilde{\rho}$ is calculated using (1) temporal mean imputation, (2) temporal mean imputation with factor τ_p and (3) temporal mean imputation with factor τ.

All data processing, analyses and plotting have been performed with the Python (Python Software Foundation) scripting language (version 3.7). For the LR and GPR algorithm implementations from Scikit-learn library (44) were used. In case of the XGB algorithm, the official Python implementation was used.

RESULTS

Calibration Results With Differing Amounts of Data and Learning Algorithms

In **Figure 1**, we visualized evaluation metric $\tilde{\rho}$ and MAE for 1, 7, 14, 21, and 30 days of calibration data in combination with the proposed GPR based calibration as well as LR and XGB based calibration. It should be noted, that for both evaluation metrics, the largest increase in performance can be seen between one and seven days of calibration data (from wearable accelerometer). Beyond 14 days, more data leads to increasingly smaller improvements. In case of the correlation coefficient ρ performance saturation seems to be reached by 21 days, while in case of the MAE saturation is not completely evident, even after 30 days. In terms of the learning algorithms used to approximate the activity calibration function $\hat{f}_{PIR}$, it is visible how GPR shows the best performance with little data up to 14 days. After that, GPR is mostly on par with LR and starts losing in comparison to XGB. Correlation values ρ show an average of 0.84 after 14 days. Note that all results displayed downstream were based on the 14 days calibration data case.

FIGURE 1 | Visualization of data and algorithm dependent calibration performance. Performance of algorithms used for calibration of passive infrared sensor systems, with respect to physical activity measured by wearable accelerometers. The learning curves show the performance across all 20 participants of the calibration method against the number of days of accelerometer reference data. The different line colors show different learning algorithms used for calibration (LR, Linear Regression; XGB, XGBoost; GPR, Gaussian Process Regression). The left plot shows the Pearson correlation coefficient $\tilde{\rho}$ as evaluation criterion (higher is better), while the right plot shows the mean absolute error (*MAE*) evaluation criterion (lower is better). With GPR, only 7–14 days of reference accelerometer data is necessary to obtain a mapping quality, which can only be marginally improved upon with additional data.

FIGURE 2 | Post calibration performance evolution. shows evolution of average calibration performance, up to 30 weeks post calibration. Individual colored lines represent standardized MAE_p^d for a given week of each participant, while the black line represents the standardized *MAE* over all participants for a given week We observe that even up to 30 weeks (~7 months) post-calibration an initial calibration using 14 days of accelerometer data remains valid.

Post-calibration Performance Evolution

Visually, it is difficult to discern any sort of overall deterioration throughout 30 weeks post calibration, beyond some short-term variation (see **Figure 2**). Regression analysis of *MAE* against time, further reveals that the slope is not statistically significant ($p = 0.262$).

Impact of Calibration on Age Relevant Health Indicators and Outcomes

Results showing correlations of clinical assessments using calibrated and uncalibrated activity from the ambient sensor system as well as the accelerometer, demonstrate how calibration

leads to increases in correlation for all assessments except hip extensor strength. Oftentimes post-calibration correlations reach strengths close to the accelerometer gold standard (see **Table 1**). Results based on the Wilcoxon signed-rank test, additionally suggest that the differences in correlations between pre- and post-calibration are statistically significant ($n = 8$, $p = 0.004$).
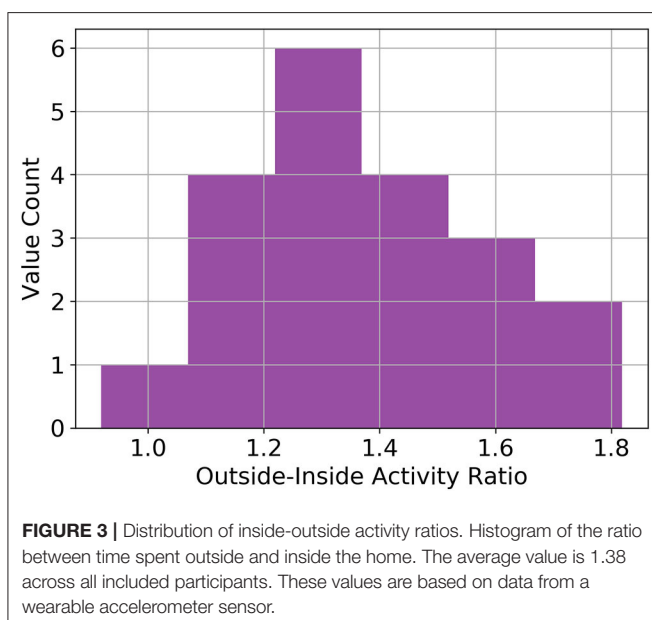
Handling Outings

We found that for most participants time spent outside the house leads to more activity compared to the average of the same time they spent at home. On average the ratio of activity outside vs. inside was found to be 1.38. However, depending on the person

TABLE 1 | Participant characteristics and demographics.

	Pre-calibration	Post-calibration	Accelerometer
TUG	−0.42	−0.57	−0.54
POMA–b	0.51	0.66	0.66
POMA-g	0.49	0.56	0.65
GDS	−0.43	−0.61	−0.64
MoCA	0.68	0.85	0.80
EFS	−0.56	−0.64	−0.65
Handgrip right	0.39	0.49	0.48
Hip right	0.37	0.36	0.34

Shows Spearman's rank correlation coefficient r_s between clinical assessment results and median daily activity sum of the ambient system, pre- and post-calibration, as well as from the worn accelerometer sensor.

FIGURE 3 | Distribution of inside-outside activity ratios. Histogram of the ratio between time spent outside and inside the home. The average value is 1.38 across all included participants. These values are based on data from a wearable accelerometer sensor.

this ratio can be quite a bit different, ranging from 0.92 up to 1.82. The distribution is visualized as a histogram shown in **Figure 3**.

We further found that by temporal mean imputing, ρ (the correlation to overall daily accelerometer activity) increases in most cases. Regarding the type of temporal mean imputation, using no coefficient seems to lead to significantly lower correlation values compared to using a person specific coefficient ($p = 0.0007$) or a static coefficient value ($p = 0.0007$), between which no significant difference ($p = 0.8$) was found (see **Figure 4**).

Predictive Distribution

To evaluate the potential usefulness of a predictive distribution we assessed how well it correlates with the daily MAE_p^d for each participant. The median correlation coefficient across all participants was 0.49 ± 0.15 (min $= 0.1$, max $= 0.67$). An example of a decent correlation is given in **Figure 5**.

DISCUSSION

We found that using even relatively small amounts of wearable based ground-truth data, PIR based wireless sensor systems can be calibrated to considerably improve estimates of older adults' daily physical activity. We could additionally verify, that this increase in performance directly translates to stronger correlations of the measured physical activity levels with a variety of age relevant health indicators and outcomes, known to be associated with physical activity. This indicates that the performance gained by calibration is not only present on paper but also manifests itself in physical activity readings that capture relations to health significantly better than would be the case without calibration.

Deciding on the necessary amount of wearable data, sufficient for calibration, is a rather subjective and task specific matter, as it is a trade-off between calibration performance and wear-time. In our case, calibrating a PIR motion sensor system, 7–14 days seem to give reasonable results, with diminishing additional benefit employing longer calibration periods. We also observed that the optimal type of algorithm, approximating the calibration function $\hat{f}_{PIR}$, seems dependent on the amount of available calibration data. For small amounts of calibration data, GPR may be considered the best choice—which is a known property of GP based approaches (45). As a side note, in our case a linear kernel proved to be the best parametrization, which would be equivalent to using a Bayesian linear regression algorithm, but the GP view might still be more effective given little data (46). This also explains why the GPR results largely converge to the LR results given more data. On the other hand, the XGB algorithm leads to slightly better performance, given more than 14 days of calibration data, which would be the expected behavior for an algorithm with much more learning capacity. Now, since we want to restrict the necessary wear-time to a minimum, GPR is, as we initially assumed, a suitable algorithm for the task. An additional benefit of GPR's Bayesian nature, is the included predictive uncertainty, which we think can be quite useful as it often indicates a simultaneous increase in model error and may thus be used to diagnose when a calibration model's performance degrades. For our data (see **Figure 5**), however, we found no significant degradation in calibration performance up to 30 weeks post calibration, indicating that calibration is overall relatively stable and resilient toward smaller potential perturbations.

It comes as no surprise that it is important to somehow factor in the time spent outside, else, physical activity of people spending a lot of time outside would be vastly underestimated. The question, as how to best deal with outings in this scenario does however remain open and we did not find any work assessing this in community-dwelling old and oldest old adults. Our findings suggest that just replacing time spent outside with the average activity throughout a given time-interval is a valid strategy, leading to significant calibration improvements but does in most cases underestimate physical activity as old and oldest-old adults tend to be more physically active when outside. We found our participant population to be, on average, 1.38 times as physically active when outside, compared to if they were inside at the same time of the day (see **Figure 3**). Using this
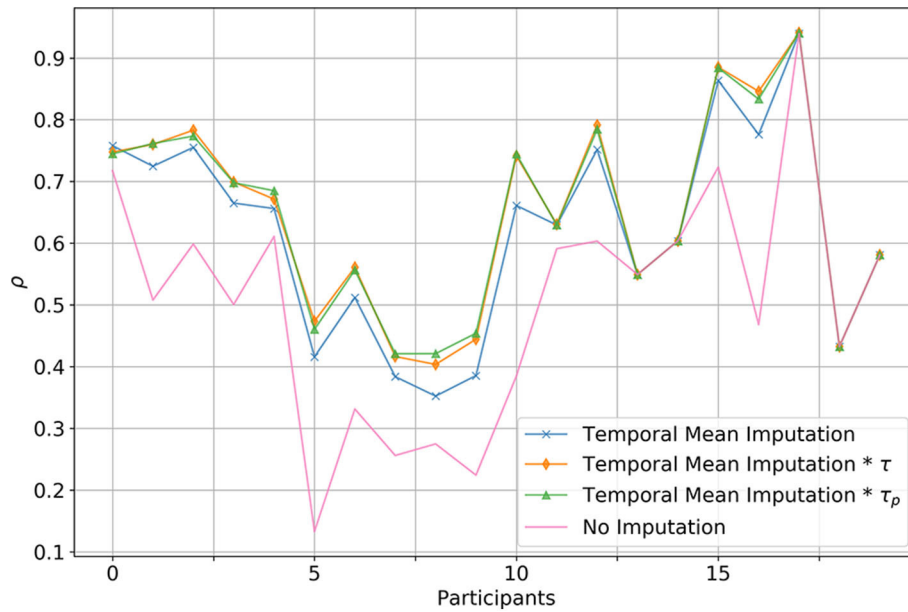
FIGURE 4 | Comparison between multiple imputation strategies to handle outing. Displayed is the correlation between the total daily calibrated activity and the total accelerometer activity. In the case of the blue line, simple temporal imputation has been used to substitute missing physical activity due to outings. The orange line denotes the case where in addition to temporal mean imputation a global correction factor was added, whereas with the green line a person specific correction factor was used. Red denotes the baseline, where outings where not imputed at all.
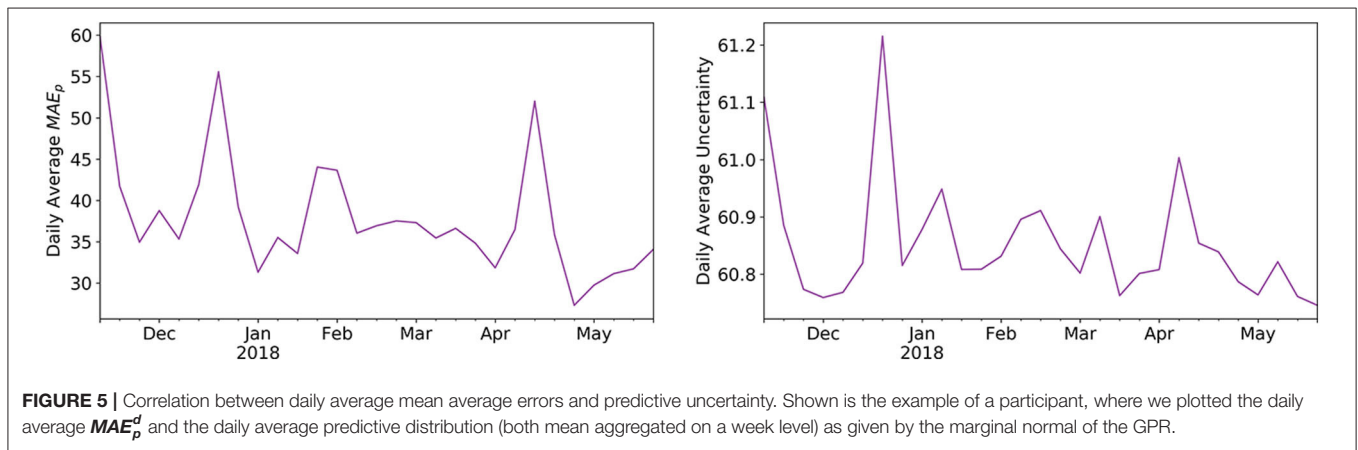
FIGURE 5 | Correlation between daily average mean average errors and predictive uncertainty. Shown is the example of a participant, where we plotted the daily average MAE_p^d and the daily average predictive distribution (both mean aggregated on a week level) as given by the marginal normal of the GPR.

knowledge, it is possible to further improve outing imputation, correcting somewhat for the bias caused by outing. Interestingly, no improvements were seen between using a static global factor and employing a person specific factor, suggesting, that even if no accelerometer ground-truth was available, outings may be corrected by a factor of around 1.4. We are not exactly sure why this is, but it may be due to the fact that we are using a very rough estimate anyways and the exact factor would only have an effect if our estimates were more accurate. However, this finding merits further investigation in different populations and under varying circumstances.

Using short-term data from a more accurate wearable device seems to work well for calibrating wireless PIR ambient sensor systems. Given that previous research on the calibration of PIR sensor systems to measure gait-speed also led to very promising results (21), such relatively simple initial calibration procedures should be considered in future long-term telemonitoring applications and research employing wireless PIR sensors.

After all, our calibration procedure has its obvious limitations and problems. In general, it should be noted that due to the relatively small sample size, generalization of our results involving statistical inference may be limited. Regarding the calibration procedure, most PIR sensors have relatively low sampling rates due to the having a refractory period and a restricted field of view. This makes it virtually impossible to get a completely accurate estimate of the real physical activity,

as we would get by using a high-frequency accelerometer. This means that there will likely always be a certain underestimation of physical activity even after calibration, as certain activities are just missed by the PIR system. Further, we should add that the approach can only function if someone is living alone. Although some work suggests PIR installations may be usable in a multi-person setting, this is likely not the case with physical activity quantification. Another problem is variance in results between participants (as can be easily seen in **Figure 4**). For certain people it did not seem possible to get a good calibration (although still slightly better than baseline), and even after in-depth manual investigation, in two instances we did not find any reasonable explanation for this behavior. Possible explanations could be that there were not enough sensors in a room, that the sensors were not placed ideally, or that the person's behavior makes it inherently difficult to capture physical activity using PIR sensors—for instance someone that is regularly taking care of the neighbor's pet. This is another important argument in favor of using reliable data for calibration of wireless systems. By employing cross-validation it is straight forward to identify installations for which there is a large disagreement before and after calibration, this also allows to manually check for potential biases using Bland-Altman plots. Considering medical applications, the validity of data coming from non-invasive ambient motion sensors is of particular importance for building up trust with this new technology, and may in that way allow for broader application. We would thus advice work related to contactless health monitoring to use more accurate and validated wearable devices for initial calibration and sanity checking of wireless sensors. Future work might evaluate similar calibration procedures applied to other modalities like contactless heart rate or breathing rate sensing. In addition, it would be very interesting to further investigate the found activity outside to activity inside ratio in larger populations of community-dwelling older adults.

CONCLUSION

We found that using calibration data from a wearable accelerometer, collected over 7–14 days, significantly improves physical activity estimates of wireless passive infrared sensor systems. This leads also to significantly stronger correlations with health indicators and outcomes, known to be associated with physical activity. Bayesian methods like Gaussian process regression, that work well with small datasets and provide an inherent predictive distribution, which can help in diagnosing when a calibration function deteriorates over time—for instance due to changes in a person's behavior. Time-spent outside should

be imputed with the average activity throughout the same time period at home, multiplied by an individual outing factor. If an individual outing factor is not available, a factor of $\sim$1.4 may be used.

We conclude that using even relatively small amounts of wearable based ground-truth data over 7–14 days, PIR based wireless sensor systems can be calibrated to give largely better estimates of older adults' daily physical activity. This increase in performance translate directly to stronger correlations with a variety of age relevant health indicators and outcomes known to be associated with physical activity.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Kantonale Ethikkommission des Kantons Bern, Murtenstrasse 31, 3010 Bern (KEK-ID: 2016-00406). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NS, HS, PB, PU, RM, and TN designed and planned the study. NS and HS installed and maintained the system and measured the participants. NS and AB analyzed the data. NS, AB, and HS wrote the manuscript. All authors reviewed and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

1. Knickman JR, Snell EK. The 2030 problem: caring for aging baby boomers. *Health Serv Res.* (2002) 37:849–84. doi: 10.1034/j.1600-0560.2002.56.x

2. Bloom DE, Canning D, Lubet A. Global population aging: facts, challenges, solutions & perspectives. *Daedalus.* (2015) 144:80–92. doi: 10.1162/DAED_a_00332

3. Koch S. Healthy ageing supported by technology - A cross-disciplinary research challenge. *Informatics Heal Soc Care.* (2010) 35:81–91. doi: 10.3109/17538157.2010.528646

4. Beard JR, Bloom DE. Towards a comprehensive public health response to population ageing. *Lancet.* (2015) 385:658–61. doi: 10.1016/S0140-6736(14)61461-6

5. Conti M, Orcioni S, Madrid NM, Gaiduk M, Seepold R. A review of health monitoring systems using sensors on bed or cushion. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence*

and *Lecture Notes in Bioinformatics*) (Berlin: Springer Verlag). (2018). p. 347–58. doi: 10.1007/978-3-319-78759-6_32

6. Meister S, Deiters W, Becker S. Digital health and digital biomarkers – enabling value chains on health data. *Curr Dir Biomed Eng.* (2016) 2:577–81. doi: 10.1515/cdbme-2016-0128

7. Coravos A, Goldsack JC, Karlin DR, Nebeker C, Perakslis E, Zimmerman N, et al. Digital Medicine: A Primer on Measurement. *Digit Biomarkers.* (2019) 3:31–71. doi: 10.1159/000500413

8. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *npj Digit Med.* (2019) 2:14. doi: 10.1038/s41746-019-0090-4

9. Rantz MJ, Skubic M, Popescu M, Galambos C, Koopman RJ, Alexander GL, et al. A new paradigm of technology-enabled "vital signs" for early detection of health change for older adults. *Gerontology.* (2015) 61:281–90. doi: 10.1159/000366518

10. Skubic M, Guevara RD, Rantz M. Automated health alerts using in-home sensor data for embedded health assessment. *IEEE J Transl Eng Heal Med.* (2015) 3:2700111. doi: 10.1109/JTEHM.2015.2421499

11. Lyons BE, Austin D, Seelye A, Petersen J, Yeargers J, Riley T, et al. Pervasive computing technologies to continuously assess Alzheimer's disease progression and intervention efficacy. *Front Aging Neurosci.* (2015) 7:102. doi: 10.3389/fnagi.2015.00102

12. Powell KE, Paluch AE, Blair SN. Physical activity for health: What kind? How much? How intense? On top of what? Annu *Rev Public Health.* (2011) 32:349–65. doi: 10.1146/annurev-publhealth-031210-101151

13. Buchman AS, Boyle PA, Yu L, Shah RC, Wilson RS, Bennett DA. *Total Daily Physical Activity and the Risk of AD and Cognitive Decline in Older Adults.* (2012). Available online at: http://n.neurology.org/content/neurology/78/17/1323.full.pdf (accessed September 16, 2018).

14. Taylor D. Physical activity is medicine for older adults. *Postgrad Med J.* (2014) 90:26–32. doi: 10.1136/postgradmedj-2012-131366

15. Garatachea N, Luque GT, Gallego JG. Physical activity and energy expenditure measurements using accelerometers in older adults. *Nutr Hosp.* (2010) 25:224–30. doi: 10.3305/nh.2010.25.2.4439

16. Lee IM, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: Issues and challenges. *Br J Sports Med.* (2014) 48:197–201. doi: 10.1136/bjsports-2013-093154

17. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large scale population assessment of physical activity using wrist worn accelerometers: the UK Biobank Study. *PLoS ONE.* (2017) 12:e0169649. doi: 10.1371/journal.pone.0169649

18. Porter EJ. Wearing and using personal emergency respone system buttons. *J Gerontol Nurs.* (2005) 31:26–33. doi: 10.3928/0098-9134-2005 1001-07

19. Wu X, Choi YM, Ghovanloo M. Design and fabricate neckwear to improve the elderly patients' medical compliance. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Berlin: Springer Verlag). (2015). p. 222–34. doi: 10.1007/978-3-319-20913-5_21

20. Peek STM, Wouters EJM, van Hoof J, Luijkx KG, Boeije HR, Vrijhoef HJM. Factors influencing acceptance of technology for aging in place: a systematic review. *Int J Med Inform.* (2014) 83:235–48. doi: 10.1016/j.ijmedinf.2014.01.004

21. Rana R, Austin D, Jacobs PG, Karunanithi M, Kaye J. Gait velocity estimation using time-interleaved between consecutive passive IR Sensor Activations. *IEEE Sens J.* (2016) 16:6351–8. doi: 10.1109/JSEN.2016.2577708

22. Kaye J. Home-based technologies: a new paradigm for conducting dementia prevention trials. *Alzheimers Dement.* (2008) 4:S60–S66. doi: 10.1016/j.jalz.2007.10.003

23. Rantz MJ, Skubic M, Miller SJ, Galambos C, Alexander G, Keller J, et al. Sensor technology to support Aging in Place. *J Am Med Dir Assoc.* (2013) 14:386–91. doi: 10.1016/j.jamda.2013.02.018

24. Urwyler P, Stucki R, Rampa L, Müri R, Mosimann UP, Nef T. Cognitive impairment categorized in community-dwelling older adults with and without dementia using in-home sensors that recognise activities of daily living. *Sci Rep.* (2017) 7:42084. doi: 10.1038/srep 42084

25. Pais Bruno, Philipp B, Tobias N, Narayan S, Hugo S, Daniel G, et al. De nouvelles technologies au service du maintien à domicile des personnes âgées. *Rev Med Suisse.* (2019) 15:1407–11.

26. Schütz N, Saner H, Rudin B, Botros A, Pais B, Santschi V, et al. Validity of pervasive computing based continuous physical activity assessment in community-dwelling old and oldest-old. *Sci Rep.* (2019) 9:1–9. doi: 10.1038/s41598-019-45733-8

27. Saner H, Schütz N, Botros A, Urwyler P, Buluschek P, du Pasquier G, et al. Potential of ambient sensor systems for early detection of health problems in older adults. *Front Cardiovasc Med.* (2020) 7:110. doi: 10.3389/fcvm.2020.00110

28. Song B, Choi H, Lee HS. Surveillance tracking system using passive infrared motion sensors in wireless sensor network. In: *2008 International Conference on Information Networking* (Busan: IEEE) (2008). p. 1–5. doi: 10.1109/ICOIN.2008.4472790

29. Aran O, Sanchez-Cortes D, Do MT, Gatica-Perez D. Anomaly detection in elderly daily behavior in ambient sensing environments. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Berlin: Springer Verlag). (2016). p.51–67. doi: 10.1007/978-3-319-46843-3_4

30. Rowlands AV, Mirkes EM, Yates T, Clemes S, Davies M, Khunti K, et al. Accelerometer-assessed physical activity in epidemiology: Are monitors equivalent? *Med Sci Sports Exerc.* (2018) 50:257–65. doi: 10.1249/MSS.0000000000001435

31. van Hees VT, Fang Z, Langford J, Assah F, Mohammad A, da Silva ICM, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *J Appl Physiol.* (2014) 117:738–44. doi: 10.1152/japplphysiol.00 421.2014

32. White T, Westgate K, Hollidge S, Venables M, Olivier P, Wareham N, et al. Estimating energy expenditure from wrist and thigh accelerometry in free-living adults: a doubly labelled water study. *Int J Obes.* (2019) 43:2333–42. doi: 10.1038/s41366-019-0352-x

33. Williams CKI, Rasmussen CE. *Gaussian Processes for Machine Learning.* Cambridge, MA: MIT Press (2006).

34. Wang JM, Fleet DJ, Hertzmann A. Gaussian process dynamical models for human motion. *IEEE Trans Pattern Anal Mach Intell.* (2008) 30:283–98. doi: 10.1109/TPAMI.2007.1167

35. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* San Francisco, CA: Association for Computing Machinery (2016). doi: 10.1145/2939672.2939785

36. Bergstra J, Ca JB, Ca YB. *Random Search for Hyper-Parameter Optimization Yoshua Bengio.* (2012). Available online at: http://scikit-learn.sourceforge.net (accessed April 22, 2020).

37. Podsiadlo D, Richardson S. The timed "up & go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc.* (1991) 39:142–8. doi: 10.1111/j.1532-5415.1991.tb01616.x

38. Tinetti ME. Performance-oriented assessment of mobility problems in elderly patients. *J Am Geriatr Soc.* (1986) 34:119–26. doi: 10.1111/j.1532-5415.1986.tb05480.x

39. Yesavage JA, Sheikh JI. Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clin Gerontol.* (1986) 5:165–73. doi: 10.1300/J018v05n01_09

40. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* (2005) 53:695–9. doi: 10.1111/j.1532-5415.2005.53221.x

41. Rolfson DB, Majumdar SR, Tsuyuki RT, Tahir A, Rockwood K. Validity and reliability of the edmonton frail scale. *Age Ageing.* (2006) 35:526–9. doi: 10.1093/ageing/afl041

42. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values. *J Clin Epidemiol.* (2006) 59:1087–91. doi: 10.1016/j.jclinepi.2006.01.014

43. van Hees VT, Renström F, Wright A, Gradmark A, Catt M, Chen KY, et al. Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. *PLoS ONE.* (2011) 6:e22922. doi: 10.1371/journal.pone.0022922

44. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-Learn: machine learning in Python. *J Mach Learn Res.* (2011) 12:2825–30. doi: 10.5555/1953048.2078195

45. Faul S, Gregorči,č G, Boylan G, Marnane W, Lightbody G, Connolly S. Gaussian process modeling of EEG for the detection of neonatal seizures. *IEEE Trans Biomed Eng.* (2007) 54:2151–62. doi: 10.1109/TBME.2007.895745

46. Griffiths TL, Lucas CG, Williams JJ, Kalish ML. *Modeling Human Function Learning With Gaussian Processes.* Vancouver, BC (2009).

APPENDIX

While most fields are self-explanatory, we describe some details regarding fields used in the calibration procedure. The "duration" attribute of the PirMotions table refers to how many seconds a given sensor was reporting motion. The location attribute of the same table describes the room the sensor was in and the time_ the exact time of the firing (in UTC). The activity field of the Biovotion1 table represents the normalized activity values stemming from the device's accelerometer and the time_ describes the exact measurement time (in UTC). The DoorSensors table's location field refers to the location the sensor was placed—in this work only entrance sensors were relevant. The status attribute describes whether the door was opened or closed and the time_ attribute denotes the exact time of this event (in UTC).

FIGURE A1 | Schematic of sensor data acquisition and final data structure.