

Releasing eHealth Analytics into the Wild: Lessons Learnt from the SPHERE Project

Tom Diethé*
Amazon Research
Cambridge, UK
tdieth@amazon.com

Mike Holmes
University of Bristol
Bristol, UK
Mike.Holmes@bristol.ac.uk

Meelis Kull
University of Tartu
Tartu, Estonia
Meelis.Kull@ut.ee

Miquel Perello Nieto
University of Bristol
Bristol, UK
Miquel.PerelloNieto@bristol.ac.uk

Kacper Sokol
University of Bristol
Bristol, UK
K.Sokol@bristol.ac.uk

Hao Song
University of Bristol
Bristol, UK
Hao.Song@bristol.ac.uk

Emma Tonkin
University of Bristol
Bristol, UK
E.L.Tonkin@bristol.ac.uk

Niall Twomey
University of Bristol
Bristol, UK
Niall.Twomey@bristol.ac.uk

Peter Flach
University of Bristol
Bristol, UK
Peter.Flach@bristol.ac.uk

ABSTRACT

The SPHERE project is devoted to advancing eHealth in a smart-home context, and supports full-scale sensing and data analysis to enable a generic healthcare service. We describe, from a data-science perspective, our experience of taking the system out of the laboratory into more than thirty homes in Bristol, UK. We describe the infrastructure and processes that had to be developed along the way, describe how we train and deploy Machine Learning systems in this context, and give a realistic appraisal of the state of the deployed systems.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; • **Applied computing** → **Health informatics**; • **Hardware** → **Sensor applications and deployments**; • **Information systems** → *Sensor networks*; *Data streaming*; • **Social and professional topics** → Remote medicine;

KEYWORDS

Machine Learning; Smart homes; Sensor networks; eHealth; Remote Medicine; Ambient Assisted living; Internet of things

ACM Reference Format:

Tom Diethé, Mike Holmes, Meelis Kull, Miquel Perello Nieto, Kacper Sokol, Hao Song, Emma Tonkin, Niall Twomey, and Peter Flach. 2018. Releasing eHealth Analytics into the Wild: Lessons Learnt from the SPHERE Project. In *KDD '18: The 24th ACM SIGKDD International Conference on Knowledge*

*Work done prior to joining Amazon

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '18, August 19–23, 2018, London, United Kingdom

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-5552-0/18/08...\$15.00

<https://doi.org/10.1145/3219819.3219883>

Discovery & Data Mining, August 19–23, 2018, London, United Kingdom. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3219819.3219883>

1 INTRODUCTION

Many countries are now experiencing demographic challenges, and subsequently traditional regimes of health-care are in need of re-examination. According to the United Nations [21]:

By 2030 the world's population is projected to rise by more than 1 billion, bringing the total to over eight billion [...] the fastest growing segment of the population will be the over 65s - there will be 390 million more of them in 2030 than in 2015

This, coupled with a rise in chronic health conditions, is accelerating the trend towards the diagnosis, treatment, and management of a wide variety of health-related issues in the home. In this context, advances in Ambient Assisted Living (AAL) are providing resources to improve the experience of patients, as well as informing necessary interventions from relatives, carers and health-care professionals [29].

With the aim of addressing these issues, the UK-funded “Sensor Platform for HEalthcare in a Residential Environment (SPHERE)” project [34, 35, 38] has designed a multi-modal system driven by data analytics requirements. The system was first tested in a single test-bed house, now with deployment underway in a general population of 100 homes in Bristol (UK). The datasets collected will be made available to researchers in a variety of communities.

In order to create such a system, there are many challenges, including (but not limited to) hardware and software engineering, ethics, user acceptance, and data handling procedures. In this paper we will focus on the challenges around developing and deploying the data analytics and Machine Learning (ML) pipeline within the SPHERE project. We will cover the following:

- The general purpose and context of the SPHERE system and its data. (Section 2).

- Our approach to building reproducible ML workflows that are *compute-on-demand*, which led us to develop the HYPER-STREAM system (Section 3).
- General challenges when deploying an ML pipeline within the SPHERE system, including issues such as data and annotation collection, data integrity testing and pre-processing, coping with noise and abnormalities with the sensors (Section 4).
- Details of model deployment at three time points: before deployment, during data collection in the houses, analysis after data is received back (Sections 5.2, 5.3 and 6).
- Approaches to verifying the integrity of the results (Section 7).

1.1 Related Work

Early attempts at smart-home projects, such as PRIMA [3], HSAUDE [10], and Place Lab [19], whilst making some progress with approaches to data fusion and ML, essentially failed due to the immaturity of the supporting technologies, but also due to issues around acceptance [1]. In terms of the latter, there has been a general shift in attitudes towards smart-home technology (as evidenced by the rise of commercial products), but can also be further mitigated with careful user studies, as have taken place during the SPHERE project (e.g. [4, 6]).

The Centre for Advanced Studies in Adaptive Systems (CASAS) project, based at Washington State University, has instrumented and published data for more than 50 smart environments (homes and offices) using a re-usable methodology [20]. The CASAS research group focus on many aspects of Activity Recognition (AR) in smart environments, and provide a number of annotated datasets¹. In general, the research focuses on presence detection using Passive Infra-Red (PIR) sensors, ambient environmental sensors (light/ temperature), and appliance sensors (taps, doors, cupboards), but for the main part does not examine cameras or wearable devices. As such, not only are the potential insights from the data somewhat limited, but also the scale of the task in terms of data processing, storage, transmission and analysis is far less than we encounter in a fully instrumented setup.

2 THE SPHERE SYSTEM

The SPHERE system [34, 38], developed primarily at the University of Bristol, uses three sensing modalities: environmental, video, and wearable, as well as collecting contextual information through questionnaires.

The environmental sensors include humidity, temperature, air quality, noise level, luminosity, occupancy, door contacts, and utility (water, electricity) consumption, centrally and at appliance/ faucet level. The currently deployed system uses 40 nodes providing more than 90 data streams, all structured and time-stamped to establish context and temporal relationships.

The video sensors are RGB-D devices which are placed in various locations, such as the living room, kitchen, corridor/hall and staircases. The video sensors allow obtaining of the cadence, gait and 3D trajectory of the residents throughout the smart environment. Through the use of a real-time tracker based on RGB-D [8],

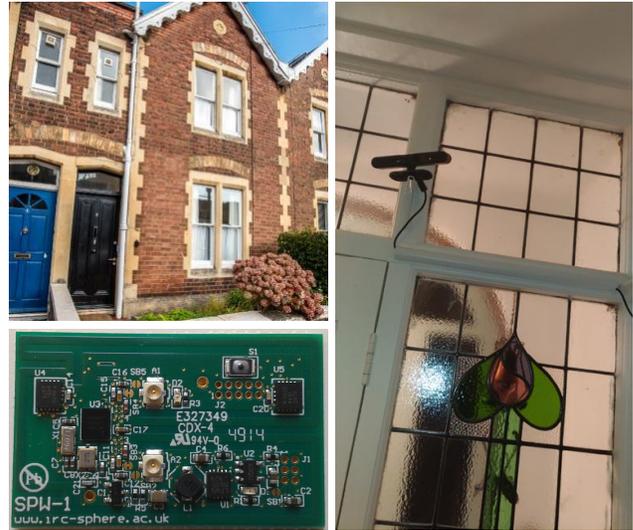


Figure 1: Top left: the first “SPHERE house”. Bottom left: top view of the SPW-1 wearable circuit-board. Right: an RGB-D camera placed inside a house.

these are capable of generating black-and-white silhouettes of any people found in the images that can be used for identification of people or the activities that they are undertaking, as well as calorific consumption [31] and quality of movement [23].

The wearable sensors are custom-developed low-maintenance Bluetooth Low Energy (BLE) devices with dual accelerometer data [18]. These support dual operation mode (connection-oriented and extra-low energy connection-less communication modes) to provide full 25Hz acceleration measurements in addition to a localisation service, using triangulation from the Received Signal Strength (RSS) as measured at each of the BLE access points within the house (this will be discussed in Section 5).

The data from each sensor cluster is collected in a SPHERE Home Gateway, which maintains time synchronisation in the system and, in addition, controls data access to ensure user privacy. The data from the SPHERE Home Gateway is collected by a heterogeneous data management platform (SPHERE Data Hub), which manages data access and will allow a dynamic library of data analytics services to be available for registered end users.

The current system is operational and before deployment underwent scripted validation experiments, where multiple-sensor data were processed to establish Activities of Daily Living (ADL) against external (manual or automatic) activity tagging. Since deployment, the data from the sensing sub-systems are being fused and processed in real-time for activity and health monitoring in longitudinal and focused studies.

One of the key objectives of the SPHERE project is to deliver datasets with a strong focus on the richness of meta-data annotations, as well as the experimental and user contexts in order to provide to the wider research community a platform for improved understanding of their roles in behavioural trends for healthcare. Examples of this include; the H130 dataset [30] which contains RGB-D

¹<http://casas.wsu.edu/datasets/>

images from the cameras along with data from two accelerometers for activity recognition; the “SPHERE Challenge” [33], hosted by <http://drivendata.org> and presented at ECML-PKDD 2016 (discussed further in Section 5.2); environmental sensor data of several people performing unscripted cooking activities in the SPHERE kitchen [37]; and the SPHERE-Calorie dataset containing RGB-D images, and the data from two accelerometers, together with ground truth calorie values from a calorimeter for calorie expenditure estimation in home environments [31].

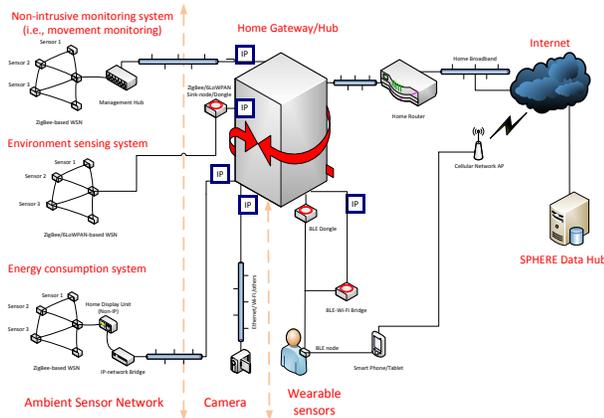


Figure 2: SPHERE system architecture.

From a data science perspective, given the goals described above and the technological solution proposed, some of the key questions for a system to be useful at the most fundamental level are:

- (1) Who is in the house?
- (2) Where are they?
- (3) What are they doing?

If those questions can be answered satisfactorily, then more subtle questions can be asked, such as:

- (1) What is the *quality* of that activity: are tasks being performed as they would normally, or is there something unusual?
- (2) Is this activity happening more or less frequently than we’d expect for this individual?
- (3) Are patterns of activity changing over time?
- (4) Is patient X with condition Y improving/declining at a rate that is to be expected given the social and medical context?

It stands to reason that in order to answer such questions, the sensor data will not be sufficient on its own. In order to deliver the greatest benefits, it is necessary to incorporate contextual information, such as demographics, medical history, socioeconomic status, and normal daily life patterns. In practice, incorporating data of this nature with the streaming sensor data is not a trivial task. Furthermore, such data is rarely useful in isolation: it is only when examining cohorts of patients with similar medical conditions and background context that meaningful patterns appear. While it is unreasonable to expect that the initial deployments being described herein will solve all of these problems, the SPHERE project has been designed such that full studies of this nature can be conducted in the future.

3 STREAM PROCESSING

In streaming data scenarios, such as when dealing with multi-sensor systems, dynamic data is generated on a continual basis. Stream processing solutions have been receiving increasing interest, a popular example being Apache Spark™ Streaming². In parallel, scientific workflow systems are designed to compose and execute a series of computational or data manipulation operations [11]. Workflows simplify the process of sharing and reusing such operations, and enable practitioners to track the provenance of execution results and the workflow creation steps. However, workflow managers are generally designed to work in offline (batch) mode and are not well suited to the streaming scenario.

When choosing how to create and deploy data analysis pipelines, due to the specific nature of the deployment, a need for a solution that had some specific characteristics was identified:

- (1) the capability to create complex interlinked workflows
- (2) an engine that is designed to be “compute-on-request”
- (3) to be capable of storing the history of computation
- (4) a plugin system for user extensions
- (5) to be able to operate in online and offline mode
- (6) to be lightweight, predictable in its use of computing resources, and have minimal requirements

Existing software stacks were considered, such as the Apache Kafka³ combined with the aforementioned Apache Spark™ Streaming (along with its ML library MLlib⁴). However, such stacks are generally designed for enterprise servers rather than low-powered consumer devices, such as the Intel® NUC⁵ used by the SPHERE project to ensure low-cost deployments.

Consequently the decision was taken to develop an in-house solution, HYPERSTREAM [16], which differs from other related toolboxes in various respects: i) it depends only on a small set of requirements to ease deployment; ii) it focuses on streaming data sources unlike most workflow engines, and is suitable for limited resource environments such as found in Internet of Things (IoT) and Fog computing scenarios [2]; iii) it allows both online and offline computational modes unlike most streaming solutions; iv) it is distributed under the permissive MIT license, encouraging its use in both academic and commercial settings. Source code, installation instructions, tutorials, and documentation are provided online.⁶

HYPERSTREAM performs operations on streams using *tools* - small Python classes following a specific interface, and through the use of *channels* the author can choose the data of a stream is manifested (e.g. in a database, memory, file etc). A workflow Application Programming Interface (API) is provided, which allows for workflows to be executed in offline mode or in a continual online fashion, effectively bridging the functionality of stream processing systems and workflow engines.

We have taken advantage of the HYPERSTREAM plugin system meaning that SPHERE specific code could be effectively separated from the main code-base. Additionally, a Python/Flask web-app was developed called HYPERSTREAM-VIEWER⁷ which allows users

²<https://spark.apache.org/streaming/>

³<https://kafka.apache.org/>

⁴<https://spark.apache.org/docs/1.1.0/mllib-guide.html>

⁵<https://www.intel.co.uk/content/www/uk/en/products/boards-kits/nuc.html>

⁶<https://irc-sphere.github.io/HyperStream/>

⁷<https://github.com/IRC-SPHERE/HyperStreamViewer>

of the system to inspect the entirety of the HYPERSTREAM back-end database. Examples will be given in Section 6.

4 DEPLOYMENT CHALLENGES

The SPHERE system is designed to be deployed into participants' homes. For this reason, we had to impose certain restrictions on the system in order to be accepted by the general public. This acceptability has been addressed through a concerted research avenue of User-Centred Design (UCD) [4-7]. A key deliverable of this was a set of rules that need to be adhered to. i) the importance of good communication through a shared language with the participants [5]; ii) the system needs to foster comfort in the participants daily lives [6]; iii) ensure a level of data privacy and anonymisation for the participants' comfort; iv) the system must be robust and dependable in order to minimise the interaction with the participants.

In order to ensure the aforementioned points we had to adapt every technological aspect to the participants' needs; in some cases involving the design and creation of our own unobtrusive technology such as wireless environmental sensors with a battery autonomy of one year [17]. Other more intrusive sensors, such as the cameras, have been limited to only detect the boundaries of people (bounding boxes), and within these regions extract standard computer vision features and perform background subtraction to create silhouettes (stored as binary bit masks). To further foster the comfort of the participants, all the data is stored physically in the participants' household and only periodic summaries are sent to the project headquarters. Furthermore, participants are able to stop the recording or delete previously recorded data by means of a software interface in a tablet that is provided as part of the deployment. With respect to privacy concerns, all transferred data is initiated by the SPHERE deployments (i.e. all transactions are pull-only). Furthermore, all communication takes place within a Virtual Private Network (VPN) over a cellular network.

As mentioned, key non-functional requirements of relevance to participants include rapid, straightforward installation and maintenance. The installation process is done by two house visits:

A: measure room sizes and wireless signal transmission; ensure that participants are aware of every detail of the project; acquire participant consent.

B: complete installation and calibration of the system.

One of the calibration procedures consists of annotating all the rooms and corners while wearing the participants' wrist bands and recording all the data in order to train a room-level localisation model (this will be described in more detail in Section 5.2). After the model has been trained, various performance metrics are used to validate the data collection, and the model is deployed in order to perform quasi-real time location predictions (an example of predictions is shown in Figure 3). After these two visits, there is no need of interruption for the following 3 months when data is manually collected (visit **D**).

Given the need to minimise disruption of participants' daily lives, the ability to perform 'self-healing' – automated recovery from certain types of fault – is essential. Various mechanisms have been put in place for the different software components, including automated restarts and timeouts and process monitoring. Additionally, a diagnostic process periodically sends quasi-real time data

summaries and device status reports in order to act promptly to avoid any data losses, which (in rare cases) can if necessary be solved with a visit **C**.

A positive indication of the participants' comfort is their interest in providing additional feedback to the team, asking proactively to participate in the generation and collection of additional data to detect activities or provide annotations that are especially valuable to the ML researchers involved with the project. A team within SPHERE is in fact dedicated to engage the participants in such activities and to maximise the benefit of these activities.

Finally, due to the large number of deployments for a project of this kind, scalability is a significant concern; therefore, the central data hub on which the project depends is deployed on a private cloud provided by the University. Furthermore, the quasi-real time analysis of the collected data and sensor status is monitored in order to solve any possible issue minimising data losses.

5 MACHINE LEARNING IN THE WILD

Perhaps the most challenging part of developing and deploying a system such as described here is how to train and deploy predictive Machine Learning models effectively. Often ML research has focused on the (rarely realistic) scenario that there is a *gold-standard* set of labelled training data which is representative of the deployment scenario, and there are simple forms of noise. Naturally, the SPHERE setting presents many sources of uncertainty. Firstly, we are dealing with multiple sensor modalities (environmental, body-worn, video), each of which will have different noise profiles and failure modes. Secondly, we are dealing with a situation where annotated or labelled data is expensive and intrusive to acquire, and the resulting labels are potentially noisy and inaccurate – indeed in some cases, there may be no "ground truth" in the classical sense. Lastly, patterns of human behaviour are subject to many factors (internal and external) that may or may not be attributed to the particular health context of a given individual. In the following sections, we will cover firstly the details of the construction of predictive models within the SPHERE system, with emphasis on the collection of annotated ground truth (Section 5.2), followed in Section 5.3 by a discussion of issues surrounding the deployment of trained models.

5.1 Annotation Collection

As a fundamental step towards intelligent healthcare, activity recognition and indoor localisation lie at the core of the ML component within the SPHERE system. However, as described in the previous section, ground truth for activities and locations are particularly hard to obtain in a smart home environment [36]. The difficulties can be seen from two perspectives. First, precise activities and locations need to be reported by the users themselves, which is not practical in a real-life scenario. Second, the environment can vary widely among different houses (e.g. layout, sensor setup, number of residents), which further makes the transfer of knowledge (e.g. in the form of annotations) a difficult task.

As a result, in the SPHERE project, we adopt a range of approaches to obtain the ground truth to train the models for activity recognition and indoor localisation. These approaches include both

data collection within a controlled environment and during the deployments within real homes.

The first approach adopted was to collect data within a controlled environment. At the early stage of the project, a house was procured and instrumented with the SPHERE sensor system for the purposes of research and development. This house, known as the SPHERE house (shown in Figure 1), hence became one of the major venues for early data collection.

To collect detailed activities, a number of scripted experiments were performed. Volunteers were asked to perform Activities of Daily Living following a script which gives instructions to the participant regarding which room to go to and what to do. The participants had to follow the instructions in a particular order but could choose their own pace. They were asked to wear a head-mounted camera, and the videos were later used to carefully annotate all activities and ambulation, such as brushing teeth, walking, transitioning from standing to sitting, etc.

Data from the scripted experiments were used to set up a public machine learning competition called the “SPHERE Challenge” [33], the goal of which was to recognise a person’s ambulation and posture from the sensor data within the scripted experiments. The challenge was hosted by drivendata.org and presented at ECML-PKDD 2016⁸. The evaluation measure was chosen to favour solutions that not only predict an activity but also quantify their uncertainty about the prediction. Usually represented as probabilities, information about uncertainty becomes crucial when humans use predictive models as input for decision-making purposes.

Despite covering a wide range of activities, the scripted experiments cannot cover all aspects of daily living in their natural succession. Therefore, longer-term experiments were performed with a volunteer living in the SPHERE house over a period ranging from a few days up to a full month in one case. Collecting ground truth by manually annotating the video recorded by a head-mounted camera becomes unfeasible for such long-term experiments, due to the excessive effort on annotating the video as well as due to discomfort and invasion of privacy from wearing the camera. Therefore, the participant was instead asked to use a smartphone (or watch) to select the current activity using a dedicated annotation app.

While data collection in the SPHERE house gives detailed annotations for initial investigation and model training, the requirement of annotations for each deployment still exists to personalise the models for each house. Therefore, the second approach adopted was to design a scripted experiment to be performed within each deployed house. In contrast to the controlled environment of the SPHERE house, in real deployments we aim to capture daily-life related activities rather than precise positions and movements. Therefore, the script we use in the deployments is to reproduce a daily routine for each household, known as the “day in fast-forward” script. As in the previous scripted experiment, the participants are also required to wear a head-mounted camera, and the resulted videos are annotated after the deployment.

5.2 Predictive Model Training

As has been described previously in Section 2, one of the first steps to capturing the behaviour of residents in the house is to ascertain their location. Here we will make the following assumptions:

- When the resident is in the house, they wear their wearable.
- The residents always wear their own wearable.
- The position of the residents can be determined by the strength of the wearable signal at the different access points.

All of these assumptions may be invalidated in reality, but they give a starting point for training a ML model for localisation prediction. In addition, errors in the predictions arising from violations of these assumptions can possibly be mitigated in post-processing.

While activity recognition models can be to some extent be transferred across houses, the localisation model needs to be trained in each house individually. For this purpose, each deployment at a new house includes a short annotation procedure, where the deployment technician visits all the rooms of the house carrying the wearable accelerometer sensors and records the moments of being at each corner of the room using a smartphone application. The received signal strengths recorded by the SPHERE system and the annotations are then ready for training a localisation model which is to be used in this particular house for localising all residents.

Once the walk-around experiments have been completed the technician interacts with the SPHERE HYPERSTREAM system to launch the localisation model training and deployment workflow. The first part of the model deployment workflow searches for completed technician walk-around scripts in the database⁹. The technician is then presented with a list of these along with time-stamps and durations, and asked to select two. These two walk-around scripts will then be used in two-fold cross-validation (i.e. train on one, test on the other, and vice-versa). This proceeds as follows:

- (1) **Data windowing**: the time-stamps from the scripts are then used to generate windows for the sensor data and annotations, as provided by the annotation app.
- (2) **Data ingress**: the raw RSS values are pulled in from the SPHERE MongoDB and extracted along with the raw annotations for these time windows. These are collected into pandas data frames.
- (3) **Feature extraction**: these are bucketed into non-overlapping two-second windows: in the case of the RSS, the maximum value is taken over this time window; in the case of the annotations, the one with the greatest duration is taken.
- (4) **Merging**: these are merged to produce input and output examples for the classifier(s). Predictions will be generated for each 2s window.
- (5) **Classifier training**: The specific classification models for localisation prediction are described below. Here a HYPERSTREAM tool was developed that allowed various classifiers to be used simply by changing a parameter value. The deployed model, in this case, is a Logistic Regression model using scikit-learn [24], chosen for its interpretability, simplicity, and ability to produce well-calibrated probabilistic outputs.

⁸<http://www.irc-sphere.ac.uk/sphere-challenge/home>

⁹All scripts and the SPHERE-specific plugin can be found at <https://github.com/IRC-SPHERE/Sphere-HyperStream>

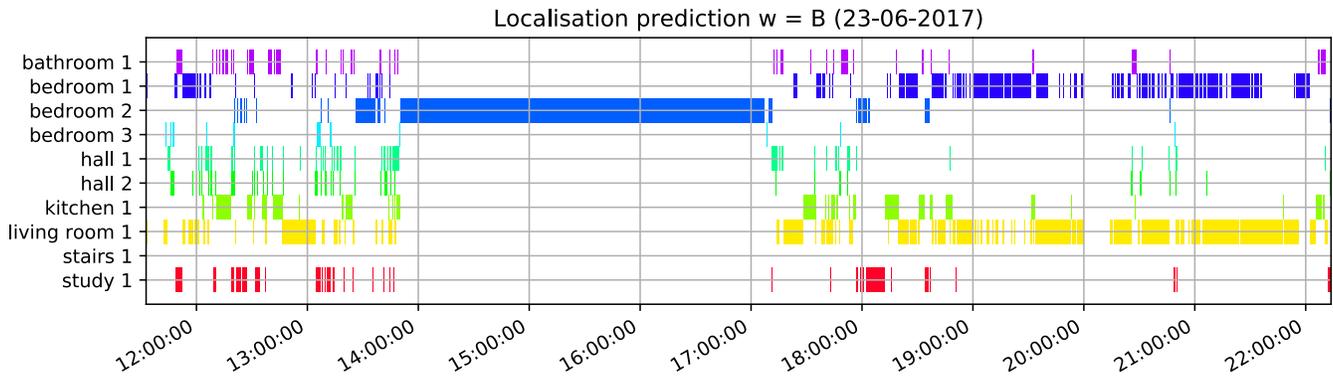


Figure 3: Localisation predictions made by HYPERSTREAM in one household for one participant

- (6) **Evaluation:** The script that the technician is presented with then outputs various metrics for the two-fold cross-validation, including accuracy, Brier score, F-measure, and a confusion matrix. If the models are deemed to be satisfactorily performing, the technician can then accept the model; if not, another data-collection walk-around is required.
- (7) **Deployment:** When a model is accepted, another HYPERSTREAM workflow is executed that stores all of these metrics to the database and deploys the model. This is physically manifested in another stream, meaning that it can be overwritten by further executions of the technician’s script.

There are a few points of note here. Firstly, although the current ML model is a relatively straightforward implementation using a standard toolbox, HYPERSTREAM allows users to use any arbitrary model (either written in Python, or using a Python wrapper). In particular, we have implemented models using the Bayesian inference toolbox Infer.NET produced by Microsoft® Research [22], such as the probabilistic sensor fusion model described in [15], and the Bayesian Dictionary Learning model applied to the accelerometer data, and described in [14]. Secondly, we should note here that in this paradigm, there is an individually trained predictive model operating in each of the households. Whilst we have investigated the possibility of having models that can be transferred between houses [13], the complexity of such approaches precludes them from the initial deployments for the time being, although they may be used in future. Finally, although not yet implemented, there are two options for acting on decisions made by the classification algorithms in (near) real-time: (i) messages can be sent to what is known as the “SPHERE Genie”, a app that is running continually on an Android™ tablet that is provided to each SPHERE household¹⁰ (ii) messages can be sent directly back to the headquarters¹¹ bypassing the summarisation system described in Section 5.3 below.

5.3 Predictive Model Deployment

Once the localisation model has been trained, HYPERSTREAM keeps the localisation workflow running constantly, recording the predictions into the HYPERSTREAM MongoDB instance and thus providing

the SPHERE system with residents’ locations. An example of the localisation predictions can be seen in Figure 3.

Currently, the localisation model uses only the BLE RSS as input. In principle, the location predictions could be made more accurate when using additional information from the PIR sensors and cameras. HYPERSTREAM has been developed in a way which supports a simple deployment of updates to its tools and workflows without requiring a major software update. In fact, all that is needed is for a new HYPERSTREAM workflow to be serialised and inserted into the database, which can be performed *on-the-fly*.

With a huge number of deployed sensors and other devices across a large number of houses, it is inevitable that faults occur. Due to almost completely uni-directional communication, it is crucial that the deployed SPHERE system itself would report enough diagnostic information, so that faults could be diagnosed in the SPHERE headquarters and maintenance visits arranged accordingly. For this purpose HYPERSTREAM is running a data summarisation workflow on a regular basis. The workflow collects all information recorded from all sensors and calculates hourly and 3-hourly aggregates for each sensor, including the average reported values as well as the number of recorded packets. Additionally, similar summaries are produced from the localisation predictions. All the created summaries are communicated once per day to the SPHERE headquarters for diagnostic purposes.

The diagnosis can be made according to specific prior knowledge on these summaries. For instance, if the summaries show the predicted location from a wearable hasn’t been changed for several days, then the researchers can infer that the wearable has been dropped somewhere. Similarly, if the summaries from a particular sensor show some abnormal pattern, such as the temperature sensor has been reporting a constant value, further investigation can also be initiated to check whether there is a failure on the corresponding sensor.

5.4 Unsupervised Approaches

Thus far we have focused on predictive machine learning. However, given the difficulty with gathering labelled data, our research group have also conducted promising research on the use of unsupervised learning and data mining approaches in this context. We have explored methods to replace or augment supervised training of

¹⁰See a demonstration of the Genie at <https://youtu.be/G4dWxPrOE-Q>

¹¹i.e. currently the University of Bristol, but ultimately to clinicians

activity recognition models through the discovery of patterns of activities using topic models [9], and similarly for indoor localisation models, by automatically inferring the layout of the home [32]. Further, given inferred activity labels, we have investigated the temporal nature of activities using circular statistics, and in particular by performing Bayesian inference with circular distributions [12]. Regarding data mining, we have also investigated approaches directly using the sensor data to detect potential behaviour changes of the households, without explicitly predicting variables like activities and locations. In [28] we proposed a method to find statistically abnormal subgroups by summarising different probabilistic models, and later in [27] we further demonstrated the proposed approach can be adopted to find abnormal spatio-temporal patterns from long-term sensor data within a home.

6 POST-HOC ANALYSIS

When the daily summaries about sensor readings and localisation predictions come down from all the houses, another summarisation workflow HYPERSTREAM is launched in the SPHERE lab. This produces meta-summaries (summaries of summaries) about the state of all sensors within all houses, making it easier to get an overview of all deployments and the faults that occurred, if any.

HYPERSTREAM-VIEWER is a companion web application, built with Python Flask using HighCharts¹² for charting, which is used to search and visualise summaries and meta-summaries returned from deployed HYPERSTREAM instances. Summaries are available per home and per device and can be searched for keywords. Figure 4 shows the HYPERSTREAM-VIEWER summary search interface.

Individual summaries can be displayed in JSON, tabular or graphical display modes. Figure 5 shows a close-up screen-shot of room temperature over a selected time period.

Case study: labelling participant data with *post-hoc* analysis of video and location data

The SPHERE datasets, consisting of raw time-stamped sensory data and HYPERSTREAM’s *in-situ* aggregate statistics and classifications, provide a rich source of information about the domestic environment and behaviour of inhabitants. Attributing visual behavioural information to specific participants in homes of multiple occupancy is one example of the *post-hoc* analyses made possible by combining sensor data and classification streams.

Re-identification (Re-ID) is the task of recognising a person that has already been observed [26]. In applications such as CCTV monitoring, this allows an unknown individual to be tracked across multiple scenes and cameras by matching a given target representation to previously observed representations. Re-ID for SPHERE must also correctly attribute a true identity label to clusters of unlabelled Re-ID features. Doing so allows for alignment between visual information, captured from cameras, and other sensory data such as the wrist-worn accelerometer, environmental sensors or device-attached smart meters on an individual level.

As described in Section 5, HYPERSTREAM uses wrist-wearable RSS indication to classify the room-level location of each participant. Location predictions are made at regular short intervals for

each participant in range of the network. Elsewhere in the sensor network, Re-ID features are extracted from people present in view of cameras deployed within the home. By temporally aligning Re-ID features for each camera with location predictions for each participant, it is possible to identify when a known participant was located in view of a camera at the same time a Re-ID feature was generated. Using this approach, it is possible to propagate labels within clusters of Re-ID features.

A sample set of data was gathered from a single multi-occupancy home during the visit B ‘day in fast forward’ activity. Six instances were found in which location prediction, Re-ID feature and timestamp overlapped and where only a single participant was in view of the camera at the time. By propagating the participant wearable label across the video *tracklets* (see [25]) it is possible to assign ground truth to a further 714 unlabelled Re-ID features. Figure 6 shows two participant Re-ID clusters, coloured to indicate their respective propagated ground-truth participant label.

7 DEPLOYMENT MONITORING

Each deployed home sends status reporting information for each sensor in its network. The system architecture of each SPHERE deployment involves both Internet of Things (IoT) and control systems; hence, the monitoring network itself is multi-layered to allow technology stacks to coexist. This also has the corresponding benefit that, should a key software component fail in one stack, the other stacks, which operate in parallel, will continue to report. Three monitoring channels are returned: MQTT data, typically containing event-driven reports; system monitoring data is returned using the Nagios¹³ NRDP (XML-based) interface; finally, ‘deep-packet inspection’ level summaries of sensor performance are returned through the HYPERSTREAM database.

An indicative subset of this information is returned through Nagios, an infrastructure monitoring tool. Information from Nagios is visualised in a web application built on the application’s JSON API and historical data stored through the NDO2DB data storage interface, allowing for easy appraisal of sensor status and performance. Figure 7 shows a screen-shot of sensor network visualisation. Installation identifiers have been removed. The integration of these distinct monitoring sources into a coherent image of system health is non-trivial. Diagnostics are addressed through a rule-based approach, which allows for problem classes and hence potential solutions to be identified.

8 DISCUSSION

The SPHERE project has been an ambitious attempt to take a research project into the wild. From a data science perspective, on the whole this has been a successful exercise: it has been demonstrated that, starting more or less from scratch, it is possible to construct a sensor system that is robust, enables the storage and transmission of data, and produces insights from that data from *in-situ* ML models. There do, however, remain open questions.

As a research project, taking place under the auspices of a large-scale University-led research programme, the SPHERE deployed system had to serve two simultaneous functions. On the one hand it had to be a “real” test of deploying an Ambient Assisted Living

¹²<https://www.highcharts.com/>

¹³<https://www.nagios.org/>

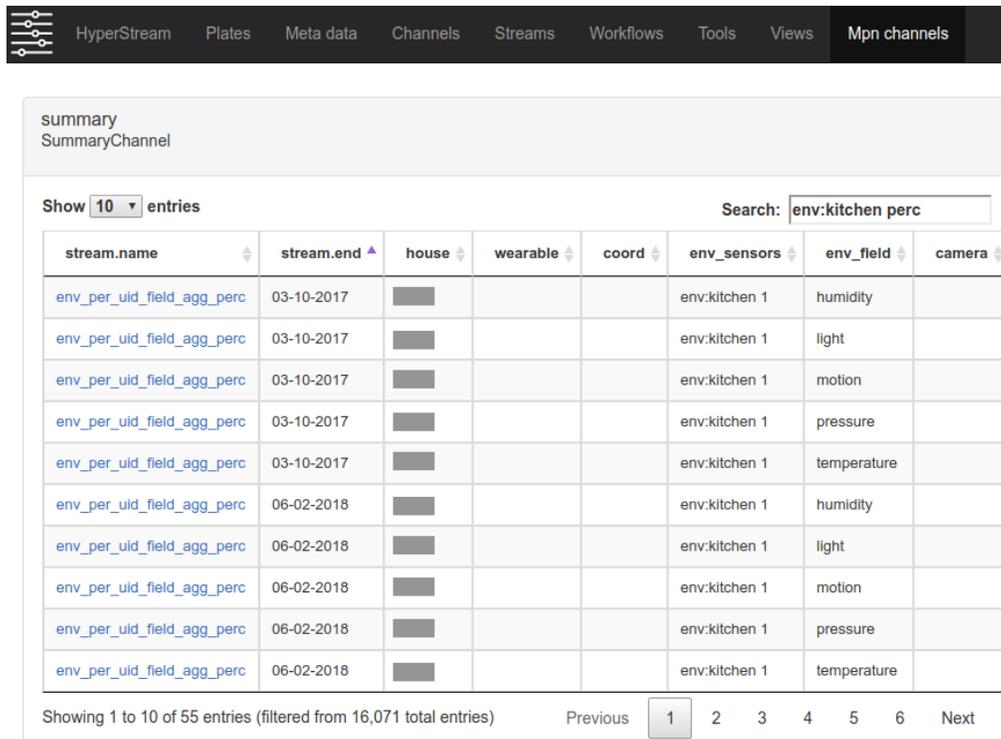


Figure 4: HYPERSTREAM-VIEWER with table of summaries beginning with search term showing aggregates of percentiles of environmental sensors in the kitchens of all of the houses.

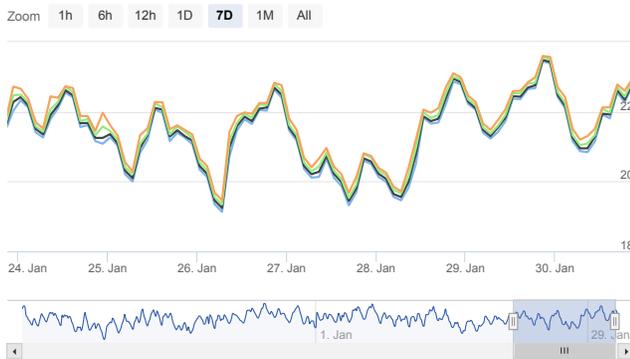


Figure 5: Interactive summary of the temperature for a period of one week in a room of a household.

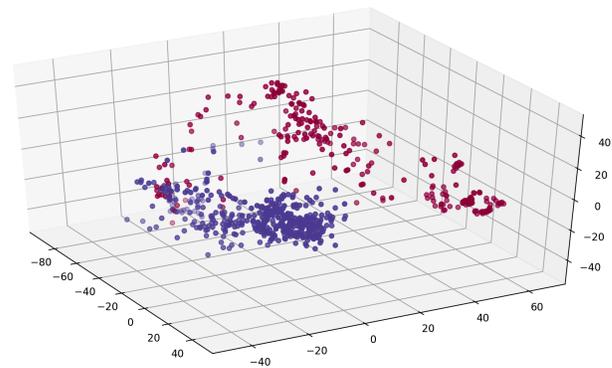


Figure 6: Isomap clustering of label-propagated participant Re-ID features

system, meaning that, for example, researchers had no access to the homes (whether physically, electronically, or otherwise), apart from specific agreed visits. On the other hand, it was deemed to be of great importance that the full collated dataset from all of the homes could be made available in full to researchers from the project and in some form to other researchers. In true deployment settings, this second need would disappear, and only a minimal amount of diagnostic and summary data would need to leave the homes for the lifetime of its operation. This has led to somewhat complicated procedures for recovering data, including manually

replacing hard-drives at fixed intervals, that would not otherwise have occurred. In this sense, one might consider that the SPHERE system is not yet truly *deployment-ready*.

Secondly, the project has not yet got to the point where it can be demonstrated whether the data science and ML pipeline is capable of conferring the health benefits that it set out to. This is in part due to a fundamental design choice: when building a system such as this, one can choose to build a general-purpose platform that can tackle many health-care issues, or one can focus on specific issues from the start. Recognising its greater potential for impact, SPHERE chose

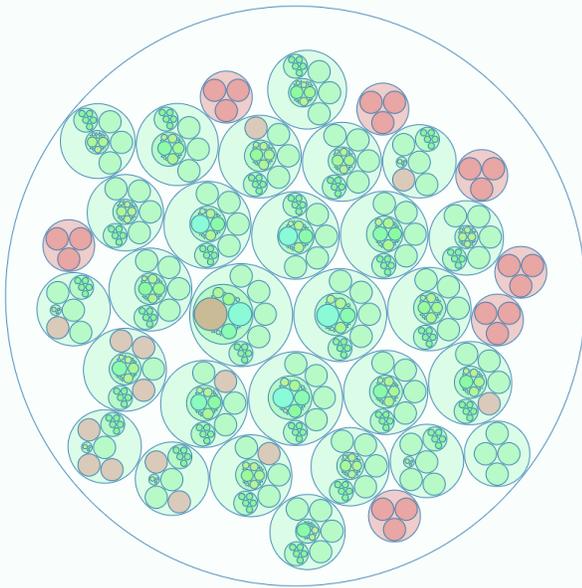


Figure 7: Screen-shot of sensor network status visualisation. Each big circle corresponds to a household while each smaller circle symbolises the status of a device or service. Abnormal behaviours are symbolised with the red colour.

to take the former route, but this means that true validation from a medical perspective is still ongoing. There are now two projects that are building technologies developed in SPHERE. The first is the “HEmiSPHERE” project¹⁴ funded as part of SPHERE, examining post-operative hip/knee surgery. The second is an Medical Research Council (MRC) funded project “CUBOID” aimed at applying the technologies developed in SPHERE to dementia patients, aiming to detect the first subtle signs of this life-changing condition in order to develop new treatments that target the disease before it causes irreversible changes to the brain.¹⁵

Other projects are likely to follow – for example, discussions are ongoing about the possibilities of employing a SPHERE-like system for monitoring animal health and welfare in farms. It is also worth noting that SPHERE has been evaluated positively by the research council and awarded follow-on funding to become a self-sustaining centre of excellence in sensing systems for twenty-first century healthcare¹⁶. From an applied data science perspective the journey so far has been utterly fascinating and worthwhile, as the interplay between powerful methods and challenging applications has resulted both in practical solutions and in new theory and algorithms in areas where practical solutions did not yet exist.

We close with the following points of note for practitioners:

- (1) Houses need to be measured carefully for both communication and localisation.
- (2) User-Centred Design is critical to gaining user acceptance.
- (3) Real-time data transfer is not practical for such a sensing system, so the overall system must be designed accordingly.
- (4) As a result, a data summary pipeline is required for diagnosis and monitoring purposes.
- (5) Fault-tolerant software is required to mitigate potential issues after detecting any anomalies.
- (6) We have developed a set of software tools, based around the stream processing engine HYPERSTREAM.
- (7) Detailed and clean ground truth is particularly hard to obtain. Solutions include:
 - (a) Establish a controlled environment for repeated experiments.
 - (b) Design short scripts that can be carried out during the deployment process.

ACKNOWLEDGMENTS

The SPHERE Interdisciplinary Research Collaboration (IRC) includes 50+ researchers working together and contributing to the SPHERE vision. The authors acknowledge fully the valued contribution, discussions and exchange of ideas that have influenced the analysis in this paper. A full list of SPHERE collaborators can be found at <http://www.irc-sphere.ac.uk>.

This work was performed under the SPHERE IRC funded by the UK Engineering and Physical Sciences Research Council (EPSRC), Grant EP/K031910/1.

REFERENCES

- [1] Muhammad Raisul Alam, Mamun Bin Ibne Reaz, and Mohd Alauddin Mohd Ali. 2012. A review of smart homes - Past, present, and future. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 6 (2012), 1190–1203.
- [2] Flavio Bonomi, Rodolfo Milito, Jiang Zhu, and Sateesh Addepalli. 2012. Fog computing and its role in the internet of things. In *Proceedings of the 1st MCC workshop on Mobile cloud computing*. ACM, 13–16.
- [3] Oliver Brdiczka, James L Crowley, and Patrick Reignier. 2009. Learning situation models in a smart home. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39, 1 (2009), 56–63.
- [4] Alison Burrows, Rachael Goberman-Hill, and David Coyle. 2015. Empirically derived user attributes for the design of home healthcare technologies. *Personal and Ubiquitous Computing* 19, 8 (dec 2015), 1233–1245.
- [5] Alison Burrows, Rachael Goberman-Hill, and David Coyle. 2016. Shared Language and the Design of Home Healthcare Technology. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3584–3594.
- [6] Alison Burrows, Rachel Goberman-Hill, Ian Craddock, and David Coyle. 2014. Sensors for healthcare: Would you want them in your home?. In *CHI 2014 workshop on Enabling Empathy in Health and Care: Design Methods and Challenges*.
- [7] Alison Burrows, Jan Noyes, Rachael Goberman-Hill, and David Coyle. 2015. Investigating Contexts of Use for the Development of Domestic Healthcare Technology: An Ethnographic Study. In *Proceedings of the Seventh Int. Conf. on Information and Communication Technologies and Development (ICTD '15)*. ACM.
- [8] Massimo Camplani, Sion L Hannuna, Majid Mirmehdi, Dima Damen, Adeline Paiement, Lili Tao, and Tilo Burghardt. 2015. Real-time RGB-D Tracking with Depth Scaling Kernelised Correlation Filters and Occlusion Handling.. In *BMVC*. 145–1.
- [9] Yu Chen, Tom Diethe, and Peter Flach. 2016. ADLTM: A Topic Model for Discovery of Activities of Daily Living in a Smart Home. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*.
- [10] Alessandro Copetti, Orlando Loques, Julius CB Leite, Thais PC Barbosa, and Antonio CL da Nobrega. 2009. Intelligent context-aware monitoring of hypertensive

¹⁴<http://www.irc-sphere.ac.uk/hemisphere>

¹⁵See <http://www.bristol.ac.uk/news/2016/october/sphere-mrc-award.html>

¹⁶<https://www.epsrc.ac.uk/newsevents/news/healthcareresearchcentres/>

- patients. In *3rd International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. IEEE, 1–6.
- [11] Ewa Deelman, Dennis Gannon, Matthew Shields, and Ian Taylor. 2009. Workflows and e-Science: An overview of workflow system features and capabilities. *Future generation computer systems* 25, 5 (2009), 528–540.
- [12] Tom Diethe, Niall Twomey, and Peter Flach. 2015. Bayesian modelling of the temporal aspects of smart home activity with circular statistics. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 279–294.
- [13] Tom Diethe, Niall Twomey, and Peter Flach. 2016. Active transfer learning for activity recognition. In *24th European Symposium on Artificial Neural Networks*.
- [14] Tom Diethe, Niall Twomey, and Peter Flach. 2016. BDL.NET: Bayesian dictionary learning in Infer.NET. In *Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on*. IEEE, 1–6.
- [15] Tom Diethe, Niall Twomey, Meelis Kull, Peter Flach, and Ian Craddock. 2017. Probabilistic sensor fusion for ambient assisted living. *arXiv preprint arXiv:1702.01209* (2017).
- [16] Tom Diethe, Niall Twomey, Meelis Kull, Miquel Perelló Nieto, Kacper Sokol, Hao Song, Mike Holmes, and Peter Flach. 2018. IRC-SPHERE/HyperStream: 0.3.8 Release. (Jan. 2018). <https://doi.org/10.5281/zenodo.1137803>
- [17] Xenofon Fafoutis, Atis Elsts, Antonis Vafeas, George Oikonomou, and Robert Piechocki. 2017. Demo: SPES-2 – A Sensing Platform for Maintenance-Free Residential Monitoring. In *Proceedings of the 2017 International Conference on Embedded Wireless Systems and Networks*. EWSN.
- [18] Xenofon Fafoutis, Balazs Janko, Evangelos Mellios, Geoffrey Hilton, R Simon Sherratt, Robert Piechocki, and Ian Craddock. 2017. SPW-1: A low-maintenance wearable activity tracker for residential monitoring and healthcare applications. In *eHealth 360°*. Springer, 294–305.
- [19] Jeffrey Hightower, Anthony LaMarca, and Ian E Smith. 2006. Practical lessons from place lab. *IEEE Pervasive Computing* 5, 3 (2006), 32–39.
- [20] Yang Hu, Dominique Tilke, Taylor Adams, Aaron S. Crandall, Diane J. Cook, and Maureen Schmitter-Edgecombe. 2016. Smart home in a box: usability study for a large scale self-installation of smart home technologies. *Journal of Reliable Intelligent Environments* 2, 2 (01 Jul 2016), 93–106.
- [21] Jeffrey Melorose, Ryan Perroy, and Sylvana Careas. 2015. World population prospects. *United Nations* 1, 6042 (2015), 587–92.
- [22] T. Minka, J.M. Winn, J.P. Guiver, Y. Zaykov, D. Fabian, and J. Bronskill. 2018. Infer.NET 2.7. (2018). Microsoft Research Cambridge. <http://research.microsoft.com/infernet>.
- [23] Adeline Paiement, Lili Tao, Sion Hannuna, Massimo Camplani, Dima Damen, and Majid Mirmehdi. 2014. Online quality assessment of human movement from skeleton data. In *British Machine Vision Conference*. Citeseer, 153–166.
- [24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [25] Jan Prokaj, Mark Duchaineau, and Gérard Medioni. 2011. Inferring tracklets for multi-object tracking. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. IEEE, 37–44.
- [26] Riccardo Satta. 2013. Appearance descriptors for person re-identification: a comprehensive review. *arXiv preprint arXiv:1307.5748* (2013).
- [27] Hao Song. 2017. *Model-Based Subgroup Discovery*. Ph.D. Dissertation. University of Bristol.
- [28] Hao Song, Meelis Kull, Peter Flach, and Georgios Kalogridis. 2016. Subgroup discovery with proper scoring rules. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part II*, Paolo Frasconi, Niels Landwehr, Giuseppe Manco, and Jilles Vreeken (Eds.). Springer International Publishing, Cham, 492–510. https://doi.org/10.1007/978-3-319-46227-1_31
- [29] Prabha Sundaravadivel, Elias Kougianos, Saraju P Mohanty, and Madhavi K Ganapathiraju. 2018. Everything You Wanted to Know about Smart Health Care: Evaluating the Different Technologies and Components of the Internet of Things for Better Health. *IEEE Consumer Electronics Magazine* 7, 1 (2018), 18–28.
- [30] Lili Tao, Tilo Burghardt, S Hannuna, et al. 2015. A comparative home activity monitoring study using visual and inertial sensors. In *E-health Networking, Application & Services (HealthCom), 17th Int. Conf. on*. IEEE, 644–647.
- [31] Lili Tao, Tilo Burghardt, Majid Mirmehdi, Dima Damen, Ashley Cooper, Sion Hannuna, Massimo Camplani, Adeline Paiement, and Ian Craddock. 2016. Calorie counter: RGB-depth visual estimation of energy expenditure at home. In *Asian Conference on Computer Vision*. Springer, 239–251.
- [32] Niall Twomey, Tom Diethe, Ian Craddock, and Peter Flach. 2017. Unsupervised Learning of Sensor Topologies for Improving Activity Recognition in Smart Environments. *Neurocomputing* 234, C (April 2017), 93–106.
- [33] Niall Twomey, Tom Diethe, Meelis Kull, Hao Song, Massimo Camplani, Sion Hannuna, Xenofon Fafoutis, Ni Zhu, Pete Woznowski, Peter Flach, et al. 2016. The SPHERE challenge: Activity recognition with multimodal sensor data. *arXiv preprint arXiv:1603.00797* (2016).
- [34] Przemyslaw Woznowski, Alison Burrows, Tom Diethe, et al. 2017. SPHERE: A Sensor Platform for Healthcare in a Residential Environment. In *Designing, Developing, and Facilitating Smart Cities*. Springer International Publishing, Cham, 315–333.
- [35] Przemyslaw Woznowski, Xenofon Fafoutis, et al. 2015. A multi-modal sensor infrastructure for healthcare in a residential environment. In *IEEE Int. Conf. on Communication Workshop (ICCW)*.
- [36] Przemyslaw Woznowski, Emma Tonkin, Pawel Laskowski, et al. 2017. Talk, text or tag?. In *Pervasive Computing and Communications Workshops (PerCom Workshops), IEEE International Conference on*. 123–128.
- [37] Kristina Yordanova, Samuel Whitehouse, Adeline Paiement, et al. 2017. What’s cooking and why? Behaviour recognition during unscripted cooking tasks for health monitoring. In *Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on*. IEEE, 18–21.
- [38] Ni Zhu, Tom Diethe, Massimo Camplani, et al. 2015. Bridging e-Health and the Internet of Things: The SPHERE Project. *Intelligent Systems, IEEE* 30, 4 (2015), 39–46.