Indirectly-Supervised Anomaly Detection of Clinically-Meaningful Health Events from Smart Home Data

JESSAMYN DAHMEN, Washington State University
DIANE J. COOK, Washington State University

Anomaly detection techniques can extract a wealth of information about unusual events. Unfortunately, these methods yield an abundance of findings that are not of interest, obscuring relevant anomalies. In this work, we improve upon traditional anomaly detection methods by introducing Isudra, an Indirectly-Supervised Detector of Relevant Anomalies from time series data. Isudra employs Bayesian optimization to select time scales, features, base detector algorithms, and algorithm hyperparameters that increase true positive and decrease false positive detection. This optimization is driven by a small amount of example anomalies, driving an indirectly-supervised approach to anomaly detection. Additionally, we enhance the approach by introducing a warm start method that reduces optimization time between similar problems. We validate the feasibility of Isudra to detect clinically-relevant behavior anomalies from over 2 million sensor readings collected in 5 smart homes, reflecting 26 health events. Results indicate that indirectly-supervised anomaly detection outperforms both supervised and unsupervised algorithms at detecting instances of health-related anomalies such as falls, nocturia, depression, and weakness.

• Computing methodologies➝**Machine learning algorithms** • Human-centered computing➝**Ubiquitous and mobile computing** • Applied computing➝**Life and medical sciences.**

Additional Key Words and Phrases: Anomaly detection, Bayesian optimization, Smart homes

1. INTRODUCTION

Anomaly detection techniques are used to identify unexpected or abnormal data patterns. One timely application of anomaly detection is to detect behavioral anomalies from sensor data that might indicate a critical health event. The world's population is aging, and 80% of these older adults are diagnosed with at least one chronic health condition (Lock et al. 2017). Chronic conditions are frequently accompanied by health events that have a sudden onset and last for a brief time. Traditionally, detection and diagnosis of these anomalies rely heavily on self-report. Automated detection of health-related behavior anomalies can result in more accurate diagnosis and timely treatment. At the same time, acceptance of the

methods depends on their ability to highlight relevant anomalies without detecting irrelevant events that only increase the analysis burden.

Finding anomalies that are valuable for a domain expert is equated with "finding a needle in a farm full of haystacks" (Almatrafi, Johri, and Rangwala 2018; Fridman et al. 2017). While many detection algorithms have been introduced, their efficacy depends on numerous data feature and algorithm hyperparameter choices. One way to improve on current methods is to let a small number of known relevant anomalies inform these choices. In this paper, we introduce Isudra, an algorithm that offers a new approach for anomaly detection by introducing indirect supervision. Using this approach, known anomalies indirectly inform an otherwise-unsupervised process. Isudra invokes Bayesian optimization to make choices that improve detection of application-relevant anomalies. We hypothesize that this strategy will reduce anomaly detection false positive rates while increasing true positive rates. Furthermore, we postulate that we can learn from past cases of indirectly-supervised anomaly detection to reduce the number of required Bayesian optimization evaluations for future cases, known as warm-starting the Bayesian optimization.

We validate these hypotheses by analyzing sensor data collected in smart homes housing older adults and labeled by nurse-clinicians with detected health events. We compare our proposed approaches to baseline techniques using both real and synthetically-generated sensor data.

2. RELATED WORK

Anomaly detection (and the related problem of outlier detection) consists of identifying abnormal instances in data. In anomaly detection, two assumptions are typically made, namely that anomalous instances are rare and that anomalies differ from typical instances with respect to their features. While unsupervised detection is the most common approach to anomaly detection and is popular for fraud detection and monitoring of patients, systems, and security (Ahmed, Mahmood, and Islam 2016; Jha, Raghunathan, and Zhang 2018; Mirsky et al. 2017; Muralidhar et al. 2018), supervised and semi-supervised methods can be employed when large amounts of training data are available for the target anomaly. Unfortunately, this constraint is not realistic for many applications due to the rarity assumption of anomalous instances and the resources involved in obtaining labeled data (Goldstein and Uchida 2017). Furthermore, with this approach a separate classifier must be trained for each anomaly class of interest.

Most anomaly detection techniques are unsupervised and therefore flag all abnormalities that are present in the data, whether or not they are relevant to a particular goal (Haley et al. 2009). They give analysts a "fire hose" approach, offering much more information than they actually need. Thus for each target type of anomaly, there is substantial effort in selecting the most effective features, time frames, and algorithm hyperparameters for each unsupervised detector to be sensitive to that anomaly class. One way to address this problem is to use expert guidance that can transform a purely unsupervised algorithm into one that is guided by indirect input from an expert.

Although it has not been explored for anomaly detection, indirect supervision has been investigated for reinforcement learning in which unintentional strategies yield serendipitous results (Guu et al. 2017). Researchers investigating structured prediction have also used indirect supervision to estimate probabilistic models in cases where the supervised variables are only partially accessible (Raghunathan et al. 2016). Others have used indirect supervision by receiving supervised training of a

problem that is a companion to the problem of interest and can, therefore, guide learning of the target challenge (Chang et al. 2010). While not directly related to these earlier works, our proposed indirect supervision approach to anomaly detection does share marked similarities. Specifically, we do not use supervised training to learn a target anomaly class from labeled examples. However, we make use of such labeled examples to guide refinement of a related problem, namely fine-tuning the representational and hyperparameter choices of a broader unsupervised learning algorithm. With indirect supervision, the supervisor provides examples of a category of unsupervised patterns (in this case, anomalies) that should ideally be discovered, and the unsupervised algorithm is adjusted to encompass the new category by increasing the likelihood of finding these highlighted cases.

Also related to our indirect supervision of unsupervised learning algorithms is weakly-supervised learning. Common types of weak supervision found in the literature are incomplete supervision, inexact supervision, and inaccurate supervision (Zhou 2017). In the case of incomplete supervision, a small amount of labeled data is available together with a large amount of unlabeled data. To address the need for additional training, active learning and semi-supervised approaches can be employed to obtain more labels for the unlabeled data (Li, Zhu, and Zhang 2018; Pohl, Bouchachia, and Hellwagner 2018). In the case of inexact supervision, labeled data is available but is not detailed enough to yield strong predictive performance. This case often arises in image processing and audio processing (Wang et al. 2018). Finally, in the case of inaccurate supervision, labeled data is available but many of the labels are known to be inaccurate. For example, Zhao et. al. (Zhao, Chu, and Martinez 2018) introduced a weakly supervised clustering algorithm to make use of freely available, but often inaccurate, annotations to analyze images of faces.

Researchers have also explored using weakly labeled data to improve performance without directly attempting to provide labels for unlabeled data. In these approaches, the data is used to help inform an indirect step that is then used to improve performance without acting on the data directly. For example, Gornitz et. al. (Gornitz et al. 2013) proposed a supervised approach to unsupervised anomaly detection using what they called "semi-supervised" anomaly detection. Here, unsupervised support vector data descriptions are augmented by incorporating known labels into the model that distinguishes normal from abnormal.

While we use anomaly detection to discover health events, there is a growing body of research to improve and augment manual detection of health events and health status. These approaches represent a technological replication of manual assessment techniques for sleep apnea, cardiac defibrillation, atrial fibrillation, chronic obstructive pulmonary disease, emotional and mental diseases, and post-op pain (Alvarez-Estevez and Moret-Bonillo 2015; Halcox et al. 2017; Just et al. 2017; Ross et al. 2017; Swaminathan et al. 2017). In contrast to these methods, we propose that data mining methods work alongside clinicians to analyze large amounts of data for monitoring of health events related to ongoing chronic conditions.

Due to the complexity of human behavior data, researchers propose the application of anomaly detection techniques for health event detection. Anomaly detection can help inform and automate the analysis of large datasets that may be difficult and time consuming for human analysts to examine without support. Anomaly detection has been used in various studies related to detecting unusual behavior and health events (Aran et al. 2016). In some cases, these methods can identify changes and anomalies in behavior data, though these methods fall prey to the possibility of finding all anomalies, not just those that are due to a health

condition of concern (Bakar et al. 2015; Hela, Amel, and Badran 2018). In other cases, detection is constrained to a particular type of anomaly correlated with a specific health concern. Of particular interest in this area of research is fall detection in home environments. Falls represent a significant cause of mortality in older adults (Evans et al. 2015). A great deal of effort therefore focuses on this case, although some of these methods use more intrusive forms of wearable sensors rather than ambient sensors that are embedded in a residence (Cola et al. 2015; Han et al. 2014; Khan and Hoey 2017).

Many design choices have to be made when addressing the problem of detecting anomalies of interest. Changing any component can dramatically affect the types of anomalies that are highlighted. In this work we explore how different combinations of time scale, feature representation, base detector, and algorithm parameters affect the ability to detect health events. One way to help tune these choices is to exhaustively search the hyperparameter and data representation space, run multiple experiments, and obtain clinical feedback to select the ideal set of choices. This grid search method is very time consuming and resource intensive. Random search can improve upon exhaustive grid search by randomly sampling the space of choices (Bergstra and Bengio 2012). Yet another alternative which we employ in this work is Bayesian optimization, which has been shown to be more efficient than both grid and random search at selecting an optimal set of choices given expert guidance (Feurer and Hutter 2019).

3. CLINICAL HEALTH EVENTS IN SMART HOME DATA

We define our indirectly-supervised anomaly detection algorithm, Isudra, in the context of detecting clinically-meaningful health events from smart home data. Additionally, we use data from this problem domain, with ground truth provided by nurse-clinicians, to evaluate our methods. Smart home technologies can empower individuals in managing their own chronic health conditions by automating behavior monitoring, health assessment, and evaluation of health interventions (Cook et al. 2018).
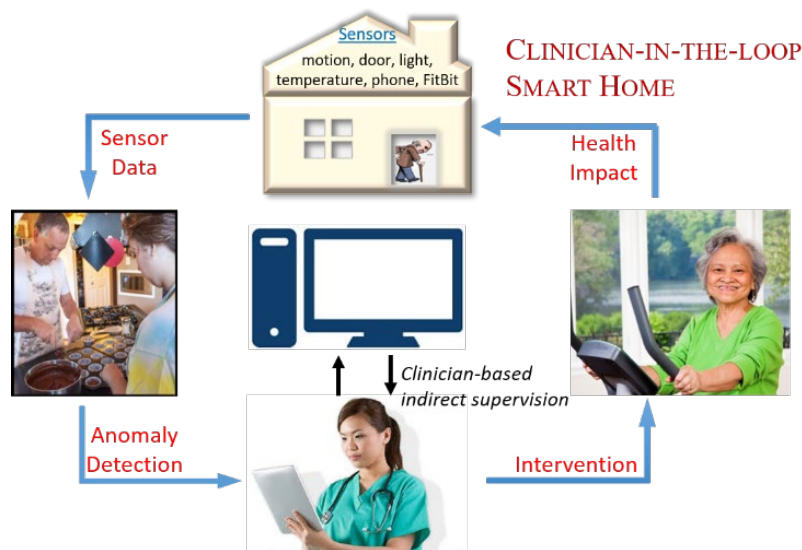

Fig. 1. The clinician-in-the-loop smart home collects ambient sensor data in the homes of adults with chronic health conditions. Isudra reports times when the data contains anomalous behavior, based on indirect clinician supervision. The clinician can then intervene in response to a detected health event.

As shown in Figure 1, behavior-driven sensor data can be collected from ambient sensors embedded in homes and other buildings. These data are collected while residents perform their daily routines, with no scripted tasks and no required deviations from their natural behavior. Health events are spotted by indirectly-supervised anomaly detectors. An event detection can trigger a response to call a care provider, and respond quickly with a treatment plan. While current research focuses primarily on a single class of health condition, we focus on detecting a variety behavioral aberrations, or anomalies, that are a consequence of one or more chronic health conditions. Being able to reliably detect diverse health events in one's home environment helps health care professionals better understand and respond to health-critical situations.

3.1 Smart Home Sensor Data

In smart homes, data is collected from ambient sensors continuously while residents perform their daily routines. For our experiments, we analyze data collected in CASAS smart homes (Cook, Crandall, et al. 2013). These homes are filled with passive infrared (PIR) motion sensors and magnetic door sensors (an average of 25 sensors in each home). Sensors generate readings when there is a change in state (rather than at a constant sampling rate). The smart home middleware collects readings from around the home, adds timestamps and sensor identifiers, and stores the data in a password-protected database. A floorplan for one of the equipped homes is shown in Figure 2 together with the locations of the installed sensors.

MA1: BathroomBArea
MA2: BedroomBArea
MA3: OfficeAArea
MA4: BedroomAArea
MA5: BathroomAArea
MA6: KitchenAArea
MA7: LivingRoomAArea
MA8: DiningRoomAArea
M01: BedroomBBed
M02: OfficeAChair
M03: BedroomBDoor
M04: BedroomADoor
M05: BedroomABed
M06: HallwayA
M07: KitchenAStove
M08: KitchenASink
M09: KitchenARefrigerator
M10: HallwayB
M11: BathroomASink
M12: BathroomAToilet
M13: EntrywayB
M14: KitchenADiningChair
M15: LivingroomAChair
M16: MainEntryway
M17: MainDoor
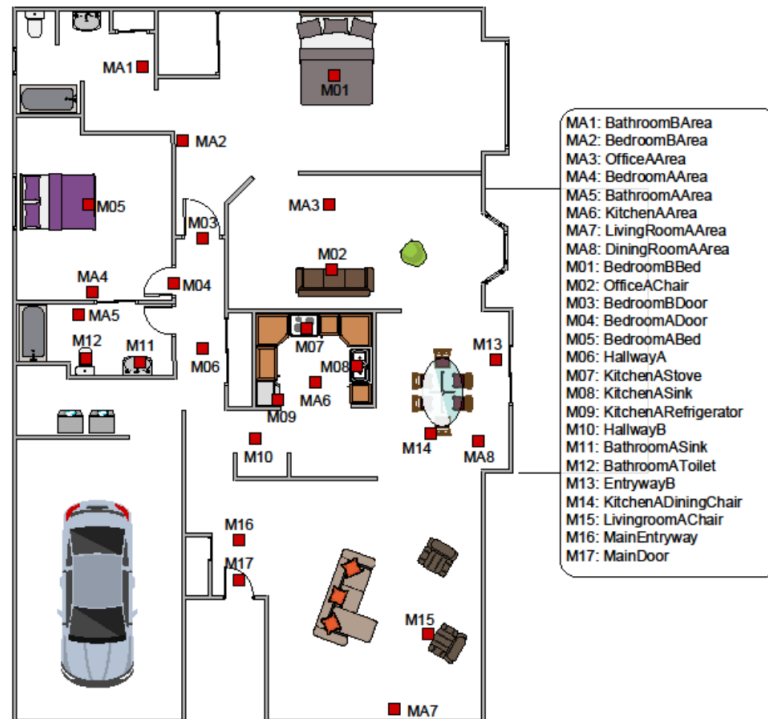
Fig. 2. A smart home floor plan. Motion sensors are coupled with ambient light sensors and are placed on ceilings to monitor a small focused region 1 meter in diameter or are angled to monitor an entire room ("area" sensor). Door sensors are coupled with ambient temperature sensors and are placed on room doors, doors to cabinets containing medicine and other key items, and exterior doors.

2019-07-20	20:34:49.052126	LivingRoom	Room	OFF	Relax
2017-07-20	20:36:05.927267	LivingRoom	Chair	ON	Relax
2012-07-20	20:36:07.032472	LivingRoom	Chair	OFF	Relax
2012-07-20	20:36:07.599722	LivingRoom	Chair	ON	Relax
2012-07-20	20:36:08.738449	LivingRoom	Chair	OFF	Relax
2012-07-20	20:38:11.169706	LivingRoom	Chair	ON	Relax
2012-07-20	20:39:59.548221	LivingRoom	Chair	OFF	Relax
2012-07-20	20:40:01.184403	LivingRoom	Chair	ON	Relax
2012-07-20	20:40:05.321364	LivingRoom	Chair	OFF	Relax
2012-07-20	20:41:48.847038	Bathroom	Sink	ON	Personal hygiene
2012-07-20	20:41:55.175907	Bathroom	Sink	OFF	Personal hygiene
2012-07-20	20:41:56.278471	Bathroom	Sink	ON	Personal hygiene
2012-07-20	20:41:57.409996	Bathroom	Sink	OFF	Personal hygiene

Fig. 3. Activity-labeled smart home time-series sensor data.

Figure 3 provides a sample of collected smart home sensor data. In addition to a date, time, sensor locator, and sensor message, each reading is assigned an activity label. This activity information together with the sensor readings is analyzed to detect behavioral anomalies. The activity labels are provided by AR (Cook, Krishnan, and Rashidi 2013), a real-time activity recognition algorithm. AR is trained from human-annotated ground truth data to label sensor events using the label set *A={Cook, Eat, Sleep, Personal hygiene, Take medicine, Work, Leave home, Enter home, Bathe, Relax, Bed-toilet transition, Wash dishes, Other activity}*. Because the data is not pre-segmented, a sliding window of 30 sensor readings is processed at a time. Features are extracted from a single window of data and are mapped onto a corresponding activity label. Details of the feature extraction and learning process are described in the literature (Aminikhanghahi and Cook 2019).

The activity model we use for our experiments was trained on two months of annotated data from 30 smart homes. Based on 3-fold cross validation for the entire dataset, AR recognized the activities with an accuracy of 99.3%. Leave-one-home-out validation yielded an accuracy of 86.8%. This provides a basis for expected label accuracy as we perform anomaly detection evaluation.

Table I. Smart home resident demographic and chronic health condition information.

Smart Home	Age	Sex	Health Conditions	Health Events
1	89	F	Hypoxia secondary to thoracic collapse from osteoporosis	Depression, Weakness
2	83	M	Parkinson's disease and Sjogren's syndrome	Nocturia, Falls, Weakness
3	88	F	Chronic obstructive pulmonary disease, oxygen dependent	Depression
4	75	M	Parkinson's disease	Falls
5	89	F	Congestive heart failure, resulting in swelling of the lower legs due to water retention	Nocturia

3.2 Participant Chronic Health Conditions and Related Health Events

We collected data from 5 smart homes, each housing a single older adult resident. Data are collected for a minimum of six months in each home. Each resident was experiencing at least one chronic health condition. These conditions are associated

with common health events that should be detectable from smart home sensor data. In these homes, such health events were detected by clinicians and confirmed by smart home residents. Table 1 provides a summary of smart home resident demographics, health conditions, and related health events.

Smart home resident 1 suffers from hypoxia, a reduction in the oxygen level for body tissue. Common symptoms include shortness of breath, sweating, anxiety, and confusion. Residents in smart homes 2 and 4 were diagnosed with Parkinson's disease (PD), a neurodegenerative disorder. PD is characterized by motor features such as rigidity and tremor. Other symptoms may appear, such as cognitive impairment, autonomic dysfunction, sleep disorders, depression, and hyposmia. Due to changes in posture and difficulty transitioning between lying, sitting, and standing, falls are frequent symptoms of PD (Bloem et al. 2004). Additionally, smart home resident 2 experiences Sjogren's syndrome (SS), an autoimmune disorder that results in dryness of the mouth and eyes. Likewise, SS can also be associated with symptoms of constant thirst, overactive bladder, and nocturia. This resident makes frequent trips to the kitchen to get a drink of water in the middle of the night, and he has experienced falls during these trips.

Smart home resident 3 is oxygen-dependent due to chronic obstructive pulmonary disease (COPD). COPD is an inflammatory lung disease that is accompanied by frequent cough and shortness of breath. The resident also experienced situational depression due to loss of a family member. Finally, smart home resident 5 is diagnosed with congestive heart failure (CHF), an inability of the heart to sufficiently pump blood to metabolizing tissues. CHF can present with fatigue, palpitations, shortness of breath, swelling in the feet and ankles, nocturia, and chest pain. This resident takes a diuretic to reduce water retention, which causes an increase in urination frequency.

During our data collection, smart home residents experienced several varied health events. These include falls, nocturia, weakness-related mobility reduction, and behavior changes related to depression. Falls are frequent among older adults – more than one-third fall each year (Stevens et al. 2006). Of these falls, 10-20% result in severe injury and subsequent decrease of functional independence. However, current research is limited in sensor-based fall detection because there is a lack of non-scripted examples. In our data, falls are sensed as they occur and they were all related to PD. A fall health event is evidenced in a smart home by several minutes or hours of abnormal behavior (e.g., extended lack of movement in an unusual place at an unusual time). Another observed health event is nocturia, voluntary urination that occurs after a person goes to sleep. Nocturia impacts sleep quality, which in turn impacts other health-related functions (Umlauf et al. 2004). In a smart home, nocturia can be sensed by frequent nighttime bathroom trips.

Weakness, often referring specifically to muscle weakness, is common in older adults and can be a consequence of multiple causes, including disease process and injury (Moreland et al. 2004). Similarly, depression is a complex health condition with numerous causes. As many as 10-15% older adults experience significant depression symptoms (Kok and Reynolds 2017). Because depression and weakness have many possible manifestations, we focus on a subset of depression and weaknesses symptoms. Detected depression health events include less time spent out of home (isolation and loss of interest in activities), more time spent in one place (fatigue), poor sleep (insomnia and hypersomnia), and slower walking speed (fatigue). Weakness is evidenced by fatigue and decreased mobility. In a smart home, these are evidenced by slower walking speed and longer time spent sitting in one place. For

both weakness and depression, we identify behavior anomalies that are associated with the conditions but occur on smaller, time scales lasting several minutes or hours at a time.

To provide ground truth for Isudra, clinicians label sensor data with detected health events. These clinicians are trained to interpret the sensor readings and each detected event is confirmed by the smart home residents. The clinicians label the start and end sensor readings corresponding to each health event and offer possible interpretations of the event and their contexts.

4. INDIRECTLY-SUPERVISED ANOMALY DETECTION

The Isudra algorithm offers a way to improve standard unsupervised anomaly detection methods. Using indirect supervision, a small amount of labeled anomaly instances guides the selection of unsupervised learning parameter choices. Bayesian optimization selects these parameters using available labeled data. Here, we focus specifically on time series sensor data. We let S denote a time series which is comprised of a sequence of ordered events such that $S=\{s_1 .. s_t ..\}$, where s_t represents a data point that is observed at timestamp t. We decompose the time series into sliding windows of fixed size. An anomaly θ_t is a subsequence of S, of arbitrary length, that begins at time t in the time series.

Using indirect supervision, each window, or subsequence $w \in S$, is considered separately. Descriptive features are extracted from the window and input to an objective function $f(x)$. The function outputs a score, y, representing the anomaly detection performance using a particular set of parameter choices, x. Our indirectly-supervised algorithm chooses four types of parameters: the unsupervised anomaly detector D, the window size (number of sensor readings) ws, the feature set fs, and detector hyperparameters h. The objective function consists of using the choice of detector (together with the selected window size, feature set, and hyperparameters) to find anomalies within the current window, then comparing the findings with ground truth to calculate the score. The process of computing the score for a particular set of parameter choices is summarized in Algorithm 1 and illustrated in Figure 4.

ALGORITHM 1.   Indirectly-Supervised Anomaly Detection

Input:

      S # Sensor time series

      ws # Window size

      fs # Feature set

      D_h # Anomaly detector with hyperparameters h

Output: y # Score

$y = 0;$ $d = |S|/ws;$

for $1 \leq i \leq d$:   # Consider all windows

    $w = \{s_i .. s_{i+ws}\} \subseteq S$ # Current window i

    $X = $ getFeatures(fs, w)   # Extract features fs from current window

    $a_i = $ containsAnomaly(w)   # True if window contains $\geq$25% anomalous sensor readings

    $\hat{a}_i = $ findAnomaly(D, X)   # Use D to detect anomaly in current window

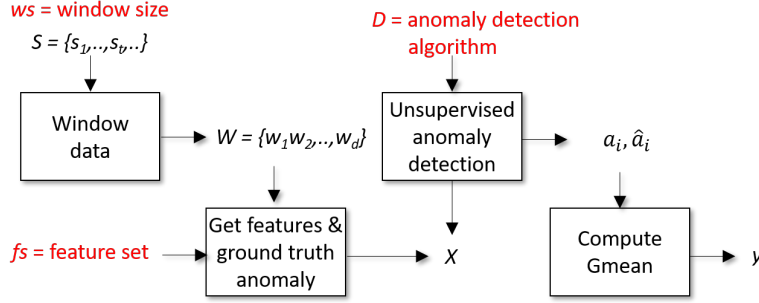$y = $ calcGmean($a, \hat{a}$)   # Calculate gmean for the entire set of windows

---

Fig. 4. Illustration of Isudra's indirectly-supervised algorithm for clinically-relevant anomaly detection.

Bayesian optimization identifies x^*, the most successful combination of parameter choices based on performance measure y. Because anomalies are rare occurrences, we utilize a gmean score for the performance measure. This measure is popular for assessing learning problems on imbalanced data. The gmean score is defined as $\sqrt{(TP\,/\,TP+FP) \times (TP\,/\,TP+FN)}$. Here, positive instances are those windows containing ≥25% sensor readings tagged as anomalous by a clinician, and negative instances are all other windows. This threshold was determined empirically with clinician guidance and can be modified without changing the nature of the algorithm.

While applying Bayesian optimization to unsupervised anomaly detection is an untapped area of research, this technique has been used to tune hyperparameters for numerous supervised learning algorithms (Lancaster et al. 2018; Snoek, Larochelle, and Adams 2012; Zhang et al. 2015), including some in high-dimensional parameter spaces (Li et al. 2017). Bayesian optimization performs n iterations, each of which evaluates a particular set of parameter choices. The parameter search space is characterized by χ, a bounded subset in real values $\mathbb{R}$, integer values $\mathbb{Z}$, and (in our approach) categorical values $\mathbb{C}$.

ALGORITHM 2.   Bayesian Optimization

Input:
n	# Number of iterations (evaluations)
x	# Array of observed parameters in χ
$f(x)$	# Objective scoring function (Algorithm 1)
$\mathcal{M} = p(\text{y}\,|\,\text{x})$	# Probabilistic model serving as surrogate for $f(x)$
α	# Acquisition function
Q	# Set of evaluated parameters x and corresponding scores y

Output: x^* # Most successful set of parameter choices
$\mathcal{M} = \text{initModel}()$ # Evaluate $f(x)$ on random values of x, use to initialize model
$Q = [\,]$

for $1 \leq i \leq n$:
    $\hat{x} = \text{argmax}_{\text{x}\in\chi}\,\alpha(x,\mathcal{M})$ # Maximize acquisition function to obtain next set of
                                        #   candidate parameters
    $\hat{y} = f(\hat{x})$ # Calculate score for $\hat{x}$ using objective function
    $\mathcal{M} = \text{update}(\mathcal{M}, \hat{x}, \hat{y})$ # Update the surrogate model
    $Q = \text{append}(Q, \hat{x}, \hat{y})$ # Extract features fs from current window
    $x = \hat{x}$
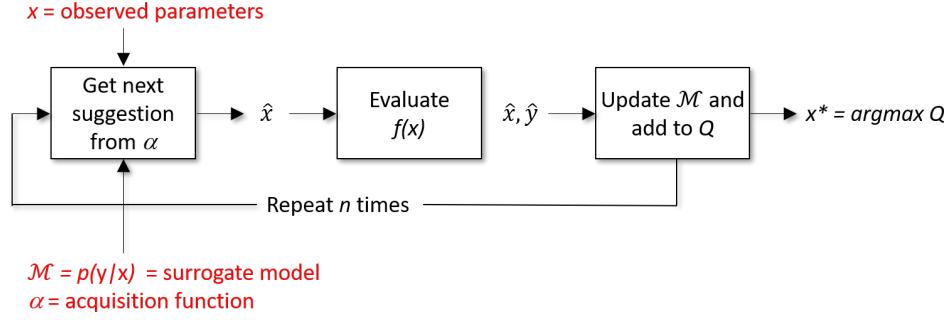$x^* = \text{argmax}\ Q$

Fig. 5. Illustration of Bayesian optimization process.

Figure 5 illustrates the Bayesian optimization procedure that is employed by Isudra, and Algorithm 2 illustrates how the Bayesian approach optimizes parameter choices based on Algorithm 1's scoring process. Bayesian optimization computes a probability model, $\mathcal{M}$, of the objective function. Model represents a mapping of parameter values x to the probability of achieving a corresponding objective function score, y. This model acts as a surrogate for the actual objective function in Algorithm 1 and it is simpler to optimize. Rather than performing grid search or random sampling, Bayesian updates the surrogate model based on past information and arrives at an optimized result by reasoning about which parameter combinations to try next. The algorithm builds the surrogate probability model of the objective function, finds the hyperparameters that perform best on the surrogate, applies the hyperparameters to the true objective function, and updates the surrogate model.

As Figure 5 shows, the update procedure is repeated for n iterations. The surrogate model is constructed by building a probability model of $f(x)$ using a prior belief distribution over $f(x)$. Once the surrogate model is built, an acquisition function α guides exploration of the parameter space and choose the next candidate to evaluate. Scores of the parameter candidates refine the model. We employ a Tree Parzen estimator (TPE) surrogate model (Shahriari et al. 2016; Snoek et al. 2012)[1]. TPEs are well-suited to conditional spaces. They build models by applying Bayes' rule, thus instead of directly representing $p(y|x)$, they employ the formula $p(x|y)p(y)/p(x)$. The term $p(x|y)$ represents the probability of hyperparameters x given the objective function score y.

The acquisition function we use in Algorithm 2 is Expected Improvement, which can be expressed as $EI_{y^*}(x) = \int_{-\infty}^{y^*}(y^* - y)p(y|x)\,dy$. Here, y^* is a threshold value of the objective function, x is the proposed set of hyperparameters, y is the actual value of the objective function using hyperparameters x, and $p(y|x)$ is the surrogate probability model expressing the probably of outputting score y given hyperparameters x. The goal is to maximize the expected improvement with respect to x, or find the best hyperparameters based on the simplified surrogate model. Expected improvement assures that parameter combinations are chosen based on their anticipated ability to maximize performance.

---

[1] Code for the Isudra indirectly-supervised anomaly detection algorithm can be downloaded from https://github.com/jb3dahmen/indirectsupervision.

5. EXPERIMENTAL RESULTS

5.1 Experimental Conditions

The goal of indirectly-supervised anomaly detection is to improve the detection of true positive anomalies while reducing the detection of false positives. We evaluate Isudra's performance using gmean, which reflects these two measures. For these experiments, we focus on smart home time series data. Table 2 summarizes the data used for our experiments. For each anomaly, we analyze two weeks of sensor data that encompass the event as well as readings occurring before and after the event. The actual number of sensor readings vary depending on the amount of activity that occurs in the home during the two week period. In each case, nurse-clinicians confirmed that the data surrounding the anomaly represent typical baseline behavior for the resident. We note that sensor data resulting from complex human behavior can contain many irregularities. Because any of these irregularities could be considered as an anomaly, gmean scores which indicate the relevance of discovered anomalies will likely be low. Our goal is to increase the clinical relevance of discovered anomalies over existing methods.

Table II. Summary statistics of data used in experiments.

Home	Event type	#Sensor events	#Anomalous sensor events
2	nocturia	86,100	833 (0.97%)
2	nocturia	94,457	891 (0.94%)
2	nocturia	99,062	1,956 (1.97%)
2	nocturia	94,148	1,857 (1.97%)
5	nocturia	112,449	543 (0.48%)
5	nocturia	113,533	186 (0.16%)
2	fall	110,088	110 (0.01%)
2	fall	135,112	65 (0.05%)
4	fall	170,482	36 (0.02%)
4	fall	165,035	96 (0.06%)
3	Depression (event: not sleeping well)	45,311	114 (0.25%)
3	Depression (event: not sleeping well)	50,708	41 (0.08%)
3	Depression (event: less time out of home)	51,926	231 (0.44%)
3	Depression (event: less time out of home)	51,926	15,895 (30.61%)
3	Depression (event: less time out of home)	54,004	157 (0.29%)
3	Depression (event: less time out of home)	220,415	12 (0.01%)
1	Depression (event: slower walking)	63,980	394 (0.62%)
1	Depression (event: slower walking)	48,718	264 (0.54%)
1	Depression (event: more time in chair)	65,196	511 (0.78%)
1	Depression (event: more time in chair)	47,100	271 (0.58%)
1	Weakness	72,532	640 (0.88%)
1	Weakness	57,294	2001 (3.49%)
1	Weakness	52,129	761 (1.46%)
2	Weakness	80,284	7,903 (9.84%)
2	Weakness	79,222	711 (0.90%)
2	Weakness	66,648	673 (1.01%)
Total		2,287,859	373,152 (0.02%)

Numerous choices exist for detector algorithms, time scales, feature sets, and algorithm hyperparameters. In these experiments, we select five alternative anomaly detection algorithms. These are principal components analysis (PCA) (Lee, Yeh, and Wang 2013), isolation forest (iForest) (Liu, Ting, and Zhou 2008), k-nearest neighbors (KNN) (Haq et al. 2015), local outlier factor (LOF) (Breunig et al. 2000), and one-class SVM (SVM) (Amer, Goldstein, and Abdennadher 2013). We consider window sizes ranging from 10 to 150 events. This range is defined based on the minimum and

maximum duration of the annotated anomalies. Instead of evaluating each possible feature individually and in combination, which would dramatically increase computational cost, we consider alternative groups of features. These include Sensor, Activity, and Baseline methods that have been used for other smart home analysis tasks including activity recognition and automated health assessment (Alberdi Aramendi et al. 2018; Krishnan and Cook 2014). Additionally, they include Location, Bathroom usage, and Social/sleep behavior feature categories that were suggested by nurse-clinicians. Details of these feature groups are provided in Table 3. Hyperparameters are specific to each detector and are specified by those individual algorithms.

Table III. Features, grouped into categories, that are evaluated by Isudra. These features are extracted from each window of sensor readings.

Feature category	Description	Features
Sensor	Basic sensor information, also used for activity recognition	*Time* (time of day, day of week, elapsed time since most recent sensor reading); *Sensor* (most common sensor ID in window, ID of most recent sensor reading, location of most recent sensor reading, location of most recent motion sensor reading, reading counts for each sensor, elapsed time since most recent reading for each sensor); *Window* (time duration of window, sequence complexity, change in activity level between first and second half of sequence, number of transitions between locations in window, number of distinct sensors in window)
Activity	Information about activity context and frequency	*Time* (time of day, day of week, elapsed time since most recent sensor reading); *Activity* (number of activities in window, number of readings for each activity in window)
Baseline	All non-clinical features	Sensor features + Activity features
Location	Clinician-indicated features representing prolonged periods of time in one location	*Time* (time of day, day of week, elapsed time since most recent sensor reading); *Location* (most recent sensor location, time spent in each location within window)
Bathroom usage	Clinician-indicated features representing bathroom use frequency and time of day	*Time* (time of day, day of week, elapsed time since most recent sensor reading); *Bathroom frequency* (number of sensor readings in bathroom within window, duration of time in bathroom)
Social / sleep	Clinician-indicated features associated with slowed movement, disturbed sleep, and changes related to increased isolation	*Time* (time of day, day of week, elapsed time since most recent sensor reading); *Mobility* (walking speed, overall activity level); *Sleep* (amount of sleep time in window, number of sleep interruptions in window, amount of sleep movement in window); *Socialization* (amount of time with visitor in window, amount of time out of home in window)

5.2 Comparison of Unsupervised, Supervised, and Indirectly-Supervised Anomaly Detection

Because Isudra utilizes a small amount of labeled data to guide parameter selection, we compare its performance with both unsupervised and supervised algorithms. For the unsupervised detection algorithm, we select iForest because it provided the best overall results on the smart home data. Similarly, we select one-class SVM as a supervised detector. Table 4 reports the average gmean scores for each class of health event. Performance is compared for the unsupervised, supervised, and indirectly-supervised methods. For each health event class, scores are averaged

over all event occurrences for the two participants who experienced the type of health event. Averaging scores over all of these instances analyzes the generalizability of each method in detecting the event type. Additionally, Isudra results are averaged over 30 runs of Bayesian optimization, each with a different random initialization. We report results for *n=30* iterations of Bayesian optimization. While performance does improve with more evaluations, the computational cost for identifying successful parameter combinations also increases. We choose this number because it represents a point at which performance plateaued for a majority of the cases.

The results summarized in Table 4 indicate that indirect supervision yields improved detection of clinically-relevant anomalies for the real smart home data. Isudra outperforms both unsupervised and supervised anomaly detection for fall, nocturia, weakness, and some depression-related events. Furthermore, the performance difference for these classes of health events is statistically significant. Higher gmean scores indicate higher true positive and true negative detection with lower false positive and false negative detection. For example, for nocturia detection in home 2, Isudra's gmean score is 0.1187. with a true positive rate (tpr) of 0.3656 and a false positive rate (fpr) of 0.9614. For the same case, iForest yields a gmean score of 0.0614 with tpr=0.1946 and fpr=0.9806, while the one-class SVM results in a gmean score of 0.0324 with tpr=0.1084 and fpr=0.9892. Thus, the proposed method is achieving our goal of decreasing false positive rates while increasing true positive anomaly detections.

The improvement over unsupervised detection suggests that taking advantage of labeled anomalies is effective because of the number and diversity of parameters that govern unsupervised approaches. One possible explanation for the improvement over a supervised learner is that the data contains a large amount of diverse normal data in comparison with the anomalous data. This situation can be difficult for a classifier to model.

Table IV. Gmean scores of indirectly-supervised (Isudra), unsupervised (iForest), and one-class (SVM) anomaly detectors for synthetic and real data sets. * = the difference between Isudra and the alternative approach is statistically significant (p<0.05).

Health event	Isudra	iForest	SVM
fall	**0.0468**	0.0252*	0.0176*
nocturia	**0.0635**	0.0291*	0.0308*
depression (not sleeping well)	**0.1146**	0.0695*	0.0769*
depression (less time out of home)	0.4120	0.1810*	**0.4670**
depression (slower walking speed)	**0.0466**	0.0162*	0.0344*
depression (more time in chair)	**0.1840**	0.0900*	0.0742*
weakness	**0.0618**	0.0402*	0.0493*

In the case of the depression-related "less time out of home" health event, the one-class SVM generates a higher gmean score than Isudra (although the difference is not statistically significant). The SVM may prove to be an ideal choice for this type of event. While the gmean scores are fairly low for most health events, as is consistent with anomaly detection on complex multivariate data, the gmean scores were high for all detectors on this event. The strong performance for this health event suggests that the event may be easily detectable by a classifier even with a limited amount of labeled data.

To determine how well Bayesian-optimized decisions generalize to unlabeled data, we repeated the experiment by optimizing two weeks of data for each anomaly class and testing it on a second two weeks of data from the same home. In this case, the gmean performance over all detected anomalies was 0.0876 for Isudra, 0.0489 for iForest, and 0.0756 for the one-class SVM.

Additionally, we tested the ability of traditional binary classifiers to recognize each class of anomalies using 3-fold cross validation for each home. For this experiment, we selected decision tree (DT) and Gaussian naïve Bayes (GNB) classifiers. This is a very difficult classification problem for a standard technique to tackle because of the subtle nature of the anomalies and the extreme class imbalance. For each of the cases, both DT and GNB are unable to generalize the anomaly concept to holdout data and do not detect any true positives. The gmean performances of DT and GNB in the homes are thus 0.0000 for each case.

5.3 Analysis of Choices Made by Indirect Supervision

Next, we examine the parameters that were chosen by Isudra. This lets us determine the diversity of choices that are made for each type of health event. The results may also provide insights on the particular algorithms and other parameter choices that are best suited to different classes of events.

As shown in Table 5, all parameters vary considerably across the different health events. In general, larger window sizes are preferred. This observation may support further expanding the range of window sizes in future analyses. Additionally, the listed feature sets in Table 5 offer insights on the relationship between sensor features and detection of alternative health events. For example, features related to disturbed sleep may represent stronger indicators of nocturia than features related only to bathroom usage. For other health events, more general features describing sensor and activity statistics were preferred over clinician-suggested feature categories. The large variation in selected parameters provides evidence to support our hypothesis that these choices influence the quality of detected anomalies. The indirect supervision employed by Isudra offers one strategy to identify successful choices for clinical applications.

Table V. Indirectly-supervised selection of detector, window size, feature set, and hyperparameter values that obtained the gmean scores reported in Table IV. For the hyperparameters, *ne* = number of iForest estimators, *ms* = iForest maximum number of samples, *k* = one-class SVM kernel, *nu* = one-class SVM regularization parameter, *kn* = number of KNN neighbors, and *ln* = number of LOF neighbors. The PCA implementation did not employ tunable hyperparameters.

Health event	Detector	Window size	Features	Hyper-parameter values
fall	KNN	65	Activity	*kn*: 12
nocturia	iForest	52	Social/sleep	*ne*: 133 *ms*: 51,189
depression (not sleeping well)	iForest	57	Activity	*ne*: 91 *ms*: 22,429
depression (less time out of home)	SVM	49	Activity	*k*: rbf *nu*: 0.690
depression (slower walking speed)	SVM	59	Sensor	*k*: polynomial *nu*: 0.480
depression (more time in chair)	PCA	23	Social/sleep	N/A
weakness	iForest	59	Sensor	*ne*: 111 *ms*: 33,359

6. WARM STARTING INDIRECT SUPERVISION

The goal of indirectly-supervised anomaly detection is to improve the detection of true positive anomalies while reducing the detection of false positives. We evaluate Isudra's performance using gmean, which reflects these two measures. For these experimentsEmploying Bayesian optimization for indirectly-supervised anomaly detection often means restarting the parameter search process for each new type of anomaly. Once parameters are selected, the resulting anomaly detector is no more computational expensive than the chosen unsupervised method. However, the initial search can be time consuming. While the Bayesian optimization algorithm repeats the evaluate-and-update process a user-specified number of iterations, each evaluation can itself be costly. For example, the one-class SVM required over 4 minutes to perform a single evaluation on the smart home data using a 2.5 GHz computer with 16GB of RAM. In some cases, there are underlying similarities between some anomaly classes that can exploited to speed up the optimization process. Because Bayesian optimization relies on Bayesian reasoning and can learn from past results, a natural extension is to use a solution from a previous task to jumpstart a related optimization task. The result is a reduction in the number of evaluations, n, that are needed to find an acceptable solution.

This problem, known as warm starting Bayesian optimization (Poloczek, Wang, and Frazier 2016), has sparked research by the community. However, many of the existing methods rely on the availability of a large number of related examples (Alaa and van der Schaar 2018; Kim, Kim, and Choi 2017). Given sufficient use cases, a meta-learner can map parameters of the task to Bayesian optimization parameters (Feurer, Letham, and Bakshy 2018; Kim et al. 2017). A more complex approach proposed by Perrone et al. (Perrone et al. 2017) creates multiple adaptive Bayesian linear regression models. The models are offered to a feedforward neural network that learns a joint representation. Yet another approach suggested by Swersky (Swersky, Snoek, and Adams 2013) reengineers the Bayesian optimization framework to incorporate multi-task Gaussian processes. Some proposed methods

require a large number of optimization parameters to be evaluated. While the extra evaluations incur more computational cost, average performance can be computed across different configurations and used to warm start the optimization process (Brecque 2018).

In our proposed approach, a surrogate model is shared across related examples. Examples from only one other home are used to guide selection of parameters for a new home. This reduces the parameter search space to focus on historically well-performing parameter combinations. This technique is uniquely beneficial when only a small amount of labeled information is available. Specifically, warm-start Isudra guides Bayesian more rapidly to optimal regions, even when labeled data is limited. Because obtaining direct observations of health events is rare, this enhancement is well suited to detection of clinically-relevant anomalies.

In the warm start method, let $\mathcal{M}^*$ represent a surrogate model that has been constructed over several related anomaly examples. $\mathcal{M}^*$ can be used in place of a newly-constructed $\mathcal{M}$ for each example. This starts the search process closer to success parameter combinations than a random initial model. To further reduce the complexity for problems that require a large number of expensive evaluations, warm-start Isudra considered a reduced search space, χ^*, that is limited to parameters which performed well on past examples. For example, the range of considered window sizes is trimmed based on past optimization experiences. The new range is defined by the smallest and largest high-performing window sizes in prior examples. As each new optimization example is encountered, χ^* can replace the original full space χ to perform warm started Bayesian optimization in combination with $\mathcal{M}^*$. We hypothesize that warm starting Bayesian optimization can achieve the same (or better) gmean performance as a non-warm started procedure, in fewer iterations. Warm-start Isudra's goal is thus to lower the number of evaluations, n.

To evaluate our proposed warm start Bayesian optimization method, we compare Isudra's performance with and without warm start using the same data as in the previous experiments. In our experiments, we group data by health event type. Isudra is used with and without warm start to identify anomalies for one smart home. While original Isudra starts the process from scratch for a new home, warm-start Isudra jumpstarts the process using learned information from the first home with the same health event. Results are averaged over every possible ordering of homes.

Table VI. Gmean performance for cold-start and warm-start Isudra. In each case, the choice of parameters is made based on one home and used to warm start search in a second home, then is repeated in the reverse direction. * = the difference between the warm start approach and the cold start approach is statistically significant (p<0.05).

Home pairs	Health event	Cold start (30 evaluations)	Warm start (15 evaluations)	Warm start (20 evaluations)	Warm start (30 evaluations)
2 & 4	fall	0.0642	0.0566	0.0717*	0.0721*
2 & 5	nocturia	0.0676	0.0664	0.0761*	0.0793*
1 & 3	depression (less time out of home)	0.1713	0.1501	0.1519	0.1576
1 & 3	weakness	0.0798	0.0805*	0.0760*	0.0846*

As Table 6 shows, a similar performance can be achieved with 20 iterations of warm start optimization as opposed to 30 iterations with cold start. Continuing the

search process for a total of 30 iterations actually yields performance that is superior to 30 iterations of cold start search.

7. CONCLUSIONS

Anomaly detection algorithms are valuable for alerting users to data that potentially represent threats to health, security, or smooth system operation. However, existing methods can produce an overwhelming number of irrelevant, or false positive, anomalies. In this paper, we introduced Isudra, an approach to anomaly detection that is based on indirect supervision. We explored how indirect supervision can be applied to unsupervised methods to detect anomalies of greater relevance for a target application. As part of a clinician-in-the-loop smart home project, we employed our method to detect fall, nocturia, weakness, and depression-related behavior anomalies from ambient sensor data with clinician-supplied indirect supervision. Experiments using data based on smart home sensor events showed that indirect supervision of anomaly detection can outperform both unsupervised anomaly detection and supervised learning of clinically-relevant anomalies.

There are many possibilities to consider in expanding this direction of research. For the clinician-in-the-loop smart home, we will explore the use of indirect supervision to learn additional health events such as urinary tract infections. We will enhance our algorithm to generalize to a wider variety of similar but distinct new health event instances. An interesting additional future step would be to examine the interaction between multiple overlapping health events. For example, an individual with depression may experience a fall. Future versions of Isudra will consider detecting such co-occurring health events. Additionally, we will expand the set of features for Bayesian optimization to consider, rather than pre-grouping them by type. Other parameters can be optimized based on anomaly type as well, including the anomaly threshold value. Another possible improvement is to employ ensembles of anomaly detectors, including some detection algorithms that are experts in distinct classes of anomalies.

ACKNOWLEDGMENTS

REFERENCES

Ahmed, M., A. N. Mahmood, and M. R. Islam. 2016. "A Survey of Anomaly Detection Techniques in Financial Domain." *Future Generation Computer Systems* 55:278–88.

Alaa, Ahmed M., and Mihaela van der Schaar. 2018. "AutoPrognosis: Automated Clinical Prognostic Modeling via Bayesian Optimization with Structured Kernel Learning." *ArXiv:1802.07207*.

Alberdi Aramendi, A., A. Weakley, M. Schmitter-Edgecombe, D. J. Cook, A. Aztiria Goenaga, A. Basarab, and M. Barrenechea Carrasco. 2018. "Smart Home-Based Prediction of Multi-Domain Symptoms Related to Alzheimer's Disease." *IEEE Journal of Biomedical and Health Informatics*.

Almatrafi, O., A. Johri, and H. Rangwala. 2018. "Identifying Learner Posts That Require Urgent Response in MOOC Discussion Forums." *Computers and Education* 118:1–9.

Alvarez-Estevez, Diego, and Vicente Moret-Bonillo. 2015. "Computer-Assisted Diagnosis of the Sleep Apnea-Hypopnea Syndrome: An Overview of Different Approaches." in *2015 37th Annual International Conference of the {IEEE} Engineering in Medicine and Biology Society ({EMBC})*.

IEEE.

Amer, Mennatallah, Markus Goldstein, and Slim Abdennadher. 2013. "Enhancing One-Class Support Vector Machines for Unsupervised Anomaly Detection." in *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description*.

Aminikhanghahi, S., and Diane J. Cook. 2019. "Enhancing Activity Recognition Using CPD-Based Activity Segmentation." *Pervasive and Mobile Computing* 53(75–89).

Aran, O., D. Sanchez-Cortes, M. T. Do, and D. Gatica-Perez. 2016. "Anomaly Detection in Elderly Daily Behavior in Ambient Sensing Environments." Pp. 51–67 in *Human Behavior Understanding*.

Bakar, U. A. B. U. A., Hemant Ghayvat, S. F. Hasanm, and S. C. Mukhopadhyay. 2015. "Activity and Anomaly Detection in Smart Home: A Survey." Pp. 191–220 in *Smart Sensors, Measurement and Instrumentation*. Springer International Publishing.

Bergstra, J., and Y. Bengio. 2012. "Random Search for Hyper-Parameter Optimization." *Journal of Machine Learning Research* 13:281–305.

Bloem, B. R., J. M. Hausdorff, J. E. Visser, and N. Giladi. 2004. "Falls and Freezing of Gait in Parkinson's Disease: A Review of Two Interconnected, Episodic Phenomena." *Movement Disorders* 19(8):871–84.

Brecque, Charles. 2018. "Warm Starting Bayesian Optimization." *Towards Data Science*.

Breunig, Markus M., Hans-Peter Kriegel, Raymond T. Ng, and Jorg Sander. 2000. "LOF: Identifying Density-Based Local Outliers." *ACM SIGMOD Record* 29(2):93–104.

Chang, M. W., V. Srikumar, D. Goldwasser, and D. Roth. 2010. "Structured Output Learning with Indirect Supervision." Pp. 199–206 in *International Conference on Machine Learning*.

Cola, Guglielmo, Marco Avvenuti, Alessio Vecchio, Guang-Zhong Yang, and Benny Lo. 2015. "An On-Node Processing Approach for Anomaly Detection in Gait." *{IEEE} Sensors Journal* 15(11):6640–49.

Cook, D. J., A. S. Crandall, B. L. Thomas, and N. C. Krishnan. 2013. "CASAS: A Smart Home in a Box." *Computer* 46(7).

Cook, D. J., N. C. Krishnan, and P. Rashidi. 2013. "Activity Discovery and Activity Recognition: A New Partnership." *IEEE Transactions on Cybernetics* 43(3).

Cook, Diane J., Maureen Schmitter-Edgecombe, Linus Jonsson, and Anne V Morant. 2018. "Technology-Enabled Assessment of Functional Health." *IEEE Reviews in Biomedical Engineering* to appear.

Evans, Daniel, Jonathan Pester, Luis Vera, Donald Jeanmonod, and Rebecca Jeanmonod. 2015. "Elderly Fall Patients Triaged to the Trauma Bay: Age, Injury Patterns, and Mortality Risk." *The American Journal of Emergency Medicine* 33(11):1635–38.

Feurer, M., B. Letham, and E. Bakshy. 2018. "Scalable Meta-Learning for Bayesian Optimization Using Ranking-Weighted Gaussian Process Ensembles." in *International Workshop on Automatic Machine Learning*.

Feurer, Matthias, and Frank Hutter. 2019. "Hyperparameter Optimization." Pp. 3–33 in *Automated Machine Learning*. Springer International Publishing.

Fridman, N., D. Amir, I. Schvartzman, O. Stawitzky, I. Kleinerman, S. Kligsberg, and N. Agmon. 2017. "Finding a Needle in a Haystack: Satellite Detection of Moving Objects in Marine Environments." Pp. 1541–43 in *International Conference on Autonomous Agents and Multiagent Systems*.

Goldstein, M., and S. Uchida. 2017. "A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data." *PLoS One* 11(4).

Gornitz, N., M. Kloft, K. Rieck, and U. Brefeld. 2013. "Toward Supervised Anomaly Detection." *Journal of Artificial Intelligence Research* 46:235–62.

Guu, K., P. Pasupat, E. Z. Liu, and P. Liang. 2017. "From Language to Programs: Bridging Reinforcement Learning and Maximum Marginal Likelihood." Pp. 1051–62 in *Annual Meeting of the Association for Computational Linguistics*.

Halcox, Julian P. J., Kathie Wareham, Antonia Cardew, Mark Gilmore, James P. Barry, Ceri Phillips, and Michael B. Gravenor. 2017. "Assessment of Remote Heart Rhythm Sampling Using the AliveCor Heart Monitor to Screen for Atrial Fibrillation." *Circulation* 136(19):1784–94.

Haley, G. E., N. Landauer, L. Renner, A. Weiss, K. Hooper, H. F. Urbanski, S. G. Kohama, M. Neuringer, and J. Raber. 2009. "Circadian Activity Associated with Spatial Learning and Memory in Aging Rhesus Monkeys." *Experimental Neurology* 217(1):55–62.

Han, Chunmei, Kaishun Wu, Yuxi Wang, and Lionel M. Ni. 2014. "WiFall: Device-Free Fall Detection by Wireless Networks." in *IEEE Conference on Computer Communications*. IEEE.

Haq, N. F., A. R. Onik, M. A. K. Hridoy, M. Rafni, F. M. Shah, and D. M. Farid. 2015. "Application of Machine Learning Approaches in Intrusion Detection System: A Survey." *International Journal of Advanced Research in Artificial Intelligence* 4(3):9–18.

Hela, Sfar, Bouzeghoub Amel, and Raddaoui Badran. 2018. "Early Anomaly Detection in Smart Home: A Causal Association Rule-Based Approach." *Artificial Intelligence in Medicine* 91:57–71.

Jha, N. K., A. Raghunathan, and M. Zhang. 2018. "MedMon: Securing Medical Devices through Wireless Monitoring and Anomaly Detection." *IEEE Transactions on Biomedical Circuits and Systems* 7(6):871–81.

Just, Marcel Adam, Lisa Pan, Vladimir L. Cherkassky, Dana L. McMakin, Christine Cha, Matthew K.

Nock, and David Brent. 2017. "Machine Learning of Neural Representations of Suicide and Emotion Concepts Identifies Suicidal Youth." *Nature Human Behaviour* 1(12):911–19.

Khan, Shehroz S., and Jesse Hoey. 2017. "Review of Fall Detection Techniques: A Data Availability Perspective." *Medical Engineering & Physics* 39:12–22.

Kim, J., S. Kim, and S. Choi. 2017. "Learning to Warm-Start Bayesian Hyperparameter Optimization." in *Neural Information Processing Systems*.

Kok, R. M., and C. F. Reynolds. 2017. "Management of Depression in Older Adults: A Review." *JAMA* 317(20):2114–22.

Krishnan, N. C., and D. J. Cook. 2014. "Activity Recognition on Streaming Sensor Data." *Pervasive and Mobile Computing* 10(PART B).

Lancaster, J., R. Lorenz, R. Leech, and J. H. Cole. 2018. "Bayesian Optimization for Neuroimaging Pre-Processing in Brain Age Classification and Prediction." *Frontiers in Aging Neuroscience* 10:28.

Lee, Yuh-Jye, Yi-Ren Yeh, and Yu-Chiang Frank Wang. 2013. "Anomaly Detection via Online Oversampling Principal Component Analysis." *{IEEE} Transactions on Knowledge and Data Engineering* 25(7):1460–70.

Li, C., J. Zhu, and B. Zhang. 2018. "Max-Margin Deep Generative Models for (Semi-) Supervised Learning." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40(11):2762–75.

Li, Cheng, Sunil Gupta, Santu Rana, Vu Nguyen, Svetha Venkatesh, and Alistair Shilton. 2017. "High Dimensional Bayesian Optimization Using Dropout." Pp. 2096–2102 in *International Joint Conference on Artificial Intelligence*.

Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. 2008. "Isolation Forest." Pp. 413–22 in *IEEE International Conference on Data Mining*.

Lock, Sarah Lenz, Matthew Baumgart, Grace Whiting, and Lisa McGuire. 2017. *Healthy Aging: Promoting Well-Being in Older Adults*.

Mirsky, Y., A. Shabtai, B. Shapira, Y. Elovici, and L. Rokach. 2017. "Anomaly Detection for Smartphone Data Streams." *Pervasive and Mobile Computing* 35:83–107.

Moreland, J. D., J. A. Richardson, C. H. Goldsmith, and C. M. Clase. 2004. "Muscle Weakness and Falls in Older Adults: A Systematic Review and Meta-Analysis." *Journal of the American Geriatrics Society* 52(7):1121–29.

Muralidhar, Nikhil, Chen Wang, Nathan Self, Marjan Momtazpour, Kiyoshi Nakayama, Ratnesh Sharma, and Naren Ramakrishnan. 2018. "No Titleilliad: InteLLigent Invarant and Anomaly Detection in Cyber-Physical Systems." *ACM Transactions on Intelligent Systems and Technology* 9(3):35.

Perrone, V., R. Jenatton, M. Seeger, and C. Archameau. 2017. "Multiple Adaptive Bayesian Linear Regression for Scalable Bayesian Optimization with Warm Start." in *Advances in Neural Information Processing Systems*.

Pohl, D., A. Bouchachia, and H. Hellwagner. 2018. "Batch-Based Active Learning: Application to Social Media Data for Crisis Management." *Expert Systems with Applications* 93:232–44.

Poloczek, M., J. Wang, and P. I. Frazier. 2016. "Warm Starting Bayesian Optimization." Pp. 770–81 in *Winter Simulation Conference*.

Raghunathan, A., R. Frostig, J. Duchi, and P. Liang. 2016. "Estimation from Indirect Supervision with Linear Moments." Pp. 2568–77 in *International Conference on Machine Learning*.

Ross, Joseph S., Jonathan Bates, Craig Parzynski, Joseph Akar, Jeptha Curtis, Nihar Desai, James Freeman, Ginger Gamble, Richard Kuntz, Shu-Xia Li, Danica Marinac-Dabic, Frederick Masoudi, Sharon-Lise Normand, Isuru Ranasinghe, Richard Shaw, and Harlan Krumholz. 2017. "Can Machine Learning Complement Traditional Medical Device Surveillance? A Case-Study of Dual-Chamber Implantable Cardioverter Defibrillators." *Medical Devices: Evidence and Research* Volume 10:165–88.

Shahriari, B., K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas. 2016. "Taking the Human out of the Loop: A Review of Bayesian Optimization." *Proceedings of the IEEE* 104(1):2016.

Snoek, J., H. Larochelle, and R. P. Adams. 2012. "Practical Bayesian Optimization of Machine Learning Algorithms." *Advances in Neural Information Processing Systems* 2951–59.

Stevens, J. A., P. S. Corso, E. A. Finkelstein, and T. R. Miller. 2006. "The Costs of Fatal and Non-Fatal Falls among Older Adults." *Injury Prevention* 12(5):290–95.

Swaminathan, Sumanth, Klajdi Qirko, Ted Smith, Ethan Corcoran, Nicholas G. Wysham, Gaurav Bazaz, George Kappel, and Anthony N. Gerber. 2017. "A Machine Learning Approach to Triaging Patients with Chronic Obstructive Pulmonary Disease." *PLoS One* 12(11):e0188532.

Swersky, K., J. Snoek, and R. P. Adams. 2013. "Multi-Task Bayesian Optimization." Pp. 2004–12 in *Advances in Neural Information Processing Systems*.

Umlauf, M. G., E. R. Chasens, R. A. Greevy, J. Arnold, K. L. Burgio, and D. J. Pillion. 2004. "Obstructive Sleep Apnea, Nocturia and Polyuria in Older Adults." *Sleep* 27(1):139–44.

Wang, Xi, Hao Chen, Caixia Gan, Huangjing Lin, Qi Dou, Qitao Huang, Muyan Cai, and Pheng-Ann Heng. 2018. "Weakly Supervised Learning for Whole Slide Lung Cancer Image Classification." in *Medical Imaging with Deep Learning*.

Zhang, Y., K. Sohn, R. Villegas, G. Pan, and H. Lee. 2015. "Improving Object Detection with Deep
    Convolutional Networks via Bayesian Optimization and Structured Prediction." Pp. 249–58 in *IEEE
    Conference on Computer Vision and Pattern Recognition.*
Zhao, K., W. S. Chu, and A. M. Martinez. 2018. "Learning Facial Action Units from Web Images with
    Scalable Weakly Supervised Clustering." Pp. 2090–99 in *IEEE Conference on Computer Vision and
    Pattern Recognition.*
Zhou, Z. H. 2017. "A Brief Introduction to Weakly Supervised Learning." *National Science Review* 5(1):44–
    53.