Published in final edited form as:

Curr Opin Biomed Eng. 2019 March ; 9: 28–46. doi:10.1016/j.cobme.2019.01.001.

Windows Into Human Health Through Wearables Data Analytics

Daniel Witt^{1,4}, Ryan Kellogg², Michael Snyder², Jessilyn Dunn^{2,3,4}

¹Mayo Clinic Alix School of Medicine, Mayo Clinic, Rochester, Minnesota, USA

²Department of Genetics, Stanford University School of Medicine, Stanford, California, USA

³Department of Biomedical Engineering, Duke University, Durham, North Carolina, USA

⁴Department of Biostatistics & Bioinformatics, Duke University, Durham, North Carolina, USA

Abstract

Background: Wearable sensors (wearables) have been commonly integrated into a wide variety of commercial products and are increasingly being used to collect and process raw physiological parameters into salient digital health information. The data collected by wearables are currently being investigated across a broad set of clinical domains and patient populations. There is significant research occurring in the domain of algorithm development, with the aim of translating raw sensor data into fitness- or health-related outcomes of interest for users, patients, and health care providers.

Objectives: The aim of this review is to highlight a selected group of fitness- and health-related indicators from wearables data and to describe several algorithmic approaches used to generate these higher order indicators.

Methods: A systematic search of the Pubmed database was performed with the following search terms (number of records in parentheses): Fitbit algorithm (18), Apple Watch algorithm (3), Garmin algorithm (5), Microsoft Band algorithm (8), Samsung Gear algorithm (2), Xiaomi MiBand algorithm (1), Huawei Band (Watch) algorithm (2), photoplethysmography algorithm (465), accelerometry algorithm (966), ECG algorithm (8287), continuous glucose monitor algorithm (343). The search terms chosen for this review are focused on algorithms for wearable devices that dominated the commercial wearables market between 2014-2017 and that were highly represented in the biomedical literature. A second set of search terms included categories of algorithms for fitness-related and health-related indicators that are commonly used in wearable devices (e.g. accelerometry, PPG, ECG). These papers covered the following domain areas: fitness; exercise; movement; physical activity; step count; walking; running; swimming; energy expenditure; atrial fibrillation; arrhythmia; cardiovascular; autonomic nervous system; neuropathy; heart rate variability; fall detection; trauma; behavior change; diet; eating; stress detection; serum glucose monitoring; continuous glucose monitoring; diabetes mellitus type 1; diabetes mellitus type 2. All studies uncovered through this search on commercially available device algorithms and

Corresponding author Jessilyn Dunn, Stanford University.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

pivotal studies on sensor algorithm development were summarized, and a summary table was constructed using references generated by the literature review as described (Table 1).

Conclusions: Wearable health technologies aim to collect and process raw physiological or environmental parameters into salient digital health information. Much of the current and future utility of wearables lies in the signal processing steps and algorithms used to analyze large volumes of data. Continued algorithmic development and advances in machine learning techniques will further increase analytic capabilities. In the context of these advances, our review aims to highlight a range of advances in fitness- and other health-related indicators provided by current wearable technologies.

Keywords

digital health; wearables; algorithms; physiologic monitoring

Introduction: Mapping the Digitome

The use of wearable biosensors, referred to as “wearables”, has been encompassed in the broader interdisciplinary health services effort to utilize mHealth (mobile health) to enhance data collection, diagnoses, treatment monitoring, and health insights.[1, 2] While definitions vary, the World Health Organization’s Global Observatory for eHealth (GOe) defined mHealth as the “medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices.”[3]

Wearables have become widely available commercial products and are increasingly being used to collect and process raw physiological parameters into salient digital health information (i.e., digital signatures, digital phenotypes[4], novel disease diagnostic criteria, etc.). A large collection of relatively low-cost technology has emerged which provides continuous or frequently-measured physiological parameters and provides the opportunity to detect changes in an individual patient’s health status or across populations of patients (Figure 1). Many consumer wearable devices collect physiological measurements such as heart rate (HR), skin temperature, and peripheral capillary oxygen saturation (SpO2) and can also collect information about physical activity, geolocation, and other ambient environmental variables.[5]

The data collected by wearables are currently being investigated across a broad set of clinical domains and patient populations. Early population-level research projects that collected physical activity data from wearables include the National Health and Nutrition Examination Survey (NHANES) studies (e.g., accelerometry data) and the UK Biobank Study cohorts (e.g., accelerometry and personal camera data).[6, 7] As of 2016, the National Institute of Health’s Precision Medicine Initiative has funded the “All of Us” research project intended to study a 1 million patient cohort over 10 years, with efforts to collect digital health data[8] and a multi-institutional research initiative called “Project Baseline” has been enrolling patients since June 2017 and seeking a patient sample size of ten thousand people.[9] Large research efforts such as these aim to integrate wearable and other remote sensor technology with longitudinal samples of traditional clinical data, biospecimen

data, lab and imaging data, and health survey information to improve understanding of disease trajectory, population and personalized characteristics of disease, and inform preventive health strategies.

Our aim is to describe an important selection of fitness and health indicators that have gained the widest use in commercial wearable technologies and have the potential for monitoring and influencing health at the population scale (Figure 2). These include fitness-related indicators (step count, energy expenditure [EE]) and other prominent health-related indicators (atrial fibrillation [AF] detection, autonomic nervous system [ANS] function, fall detection, sleep stage estimation, and serum glucose tracking) (Table 1). Other examples of emerging research exist in neurology (e.g., Parkinson's Disease)[10], psychiatric and mental health[11] (e.g., mood disorders[12–15], substance use disorders[16]), and infectious disease recognition, among many other areas of application. As of 2018, there have been 171 completed clinical trial studies in the U.S. involving wearables (Figure 3). [17]

Wearables Data: Algorithms and Analytic Methods

Data Pre-Processing

Given the array of sensing units in many wearables, data pre-processing often must occur prior to initiating the highest order analytical goals. These concerns broadly fall under the domain of signal processing and other possible data pre-processing steps such as labeling, filtering, segmentation, feature extraction and feature selection.[18] In a discussion of wearable biosensing, Celka et al. described an architecture for data fusion, which includes capturing information from multiple biosensors and transferring the data through a low-level sensing layer (e.g., local sensors, local signal conditioning and signal processing) followed by a higher-level processing layer divided into two global sub-layers: feature extraction/selection and classification.[19] Others have noted that while some devices accommodate data fusion across sensor modules, there are additional limitations which must be overcome before a device can be scaled and used effectively in clinical care.[20]

Machine Learning

Data generated by wearables can be often characterized as “Big Data”, which presents promise and challenges.[21] Various machine learning (ML) approaches for predicting health outcomes have become a common area of focus across industry, academia, and healthcare research. ML is defined by computational algorithms that are designed to extract desired information from data via different probabilistic learning paradigms. Classically, medical research has aimed to utilize ML algorithms with clinical data (e.g., age, gender, physical exam findings, symptoms, vital signs, lab values, imaging variables, test result values) in order to predict clinical outcomes or to identify relationships between predictor variables, known as features, and clinical outcomes. While wearables data can be integrated with these broader sets of patient data to provide additional context about a patient, the raw sensor data can also serve as a direct input into an ML algorithm to predict a clinical outcome (i.e., a physiological or pathophysiological state) or extract meaningful data features.

A general distinction between ML algorithms is made based on the type of training data they utilize (labeled or unlabeled) and whether the goal is to predict specific outcomes or to learn patterns in the data. These scenarios include supervised (e.g., logistic regression, naïve Bayes, decision tree, nearest neighbor, random forest, discriminant analysis, support vector machine, neural networks), unsupervised (e.g., clustering algorithms, principal component analysis, etc.), or semi-supervised learning paradigms.[22] An additional area of research and application has turned towards deep learning architectures[23], such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), for a wide variety health-related applications, including molecular drug activity[24], neural connectomics[25], genomics[26, 27], image recognition[28-30], and speech recognition.[31, 32] Another relevant distinction within ML algorithms include those which perform “ensemble” goals; they are algorithms that weight multiple individual classifiers and combine them to obtain a classifier that outperforms the individual starting classifiers (e.g., Naïve Bayes Optimal Classifier, Bootstrap Aggregating or “Bagging”, Boosting, Bayes Model/Parameter Averaging, Bucket of Models, Stacking, etc.).[33]

Research from across the clinical spectrum has utilized ML methods. Fall detection research, based on wearables-collected data, has produced optimized detection of falls using support vector machines (SVM)[34], naïve Bayes[35], and artificial neural network (ANN)[36] approaches, among others.[37-39] Others have explored detection of atrial AF using k Nearest Neighbors (k-NN)[40] and SVM.[41] Additional studies in the cardiology field have included detecting clinical status of heart failure patients using graph mining of electrocardiogram (ECG) and seismocardiogram sensing patch data. [42] In the neurology patient population, classification of movement patterns from patients with Parkinson’s Disease has been advanced with translational research utilizing ML algorithms.[43-45] Other studies have focused on detecting different phenotypes of characteristic glucose regulation or dysregulation in non- diabetics using wearable data.[46, 47] While algorithms have been studied in various settings and with different types of data, it is clear that ML algorithms offer an opportunity to improve feature development methods and prediction accuracy when applied with technical expertise and appropriate domain knowledge.

Other Relevant Approaches to Analysis

In addition to ML, foundational statistical modeling principles often apply to wearable sensor data. Both classical parametric and non-parametric summary statistics and regression techniques (e.g., linear regression, logistic regression, multivariate adaptive regression splines) are still commonly appropriate for generating salient, interpretable insights into wearable data and health, depending on the types of data, variable distributions, and relationships between variables. These techniques are commonly used to define how change in a predictor variable impacts an outcome variable, for example by modeling the change in steady-state plasma glucose (SSPG) with respect to HR, activity, and BMI.[5] Furthermore, given the ubiquity of temporally dense, longitudinal measures in wearable data streams, time series analysis is also a common approach to feature detection.[48] However, while concepts such as multivariate time series analysis, multiple comparisons of correlated tests, and dimension reduction for correlated covariates are frequently used in digital phenotyping projects, there are common errors that can arise if not applied correctly.[49]

Validation Studies of Commercial Wearable Device Algorithms

Fitness-related Indicators

A majority of people in the U.S. would have moderately to extremely significant health benefits from increasing average activity levels during each day or across the span of each week lived.[50] In the Physical Activity Guidelines for Americans, the US Department of Health and Human Services recommends regular performance of moderate-to-vigorous physical activity (MVPA) for reduced risk of many diseases and disorders.[51] Absolute activity intensity refers to the magnitude of effort exerted by a person during physical activity, varying between people depending on relative fitness level, and ranges between light, moderate, and vigorous intensities—measured by metabolic equivalents (METs) which are ratios of metabolic rate relative to a person's resting metabolic rate (Table 2).

Many research groups have also evaluated the accuracy of fitness device step count algorithms against gold standard pedometer-based algorithms under various conditions. In a systematic review of 22 studies performed before or during 2015, there was a high correlation between tracker-assessed steps and step counts reported by both Fitbit and Jawbone wearables. However, the same review indicated that Fitbit over-estimated distance at slower speeds and under-estimation at faster speeds. There was high inter-device reliability for steps and distance during walking and running trials for the Fitbit.[52]

Step Counts and Physical Activity Classification—General algorithmic methods for step detection implement windowed peak detection on the magnitude of the triaxial acceleration vector via low-pass filtering with a delay.[53] The periodicity of steps during walking also makes frequency- and temporal-based methods including fourier and wavelet transforms useful for step detection. Sensor orientation and the earth's gravitational acceleration are usually accounted for using high-pass filters on the raw signal prior to use. Adaptive thresholding for the filters, peak calling, and frequency domain features make the step count measurements more robust to different users and environments. Several correction methods have been used to eliminate inaccuracies, including customizing settings based on individual user metrics and detecting a minimum duration of walking before initiating step counting.[54] Pedometer devices worn on the hip or foot detect steps more accurately than smartphone or wrist-worn accelerometers.

There have been several comprehensive assessments of step count, posture, and physical activity classification algorithms of commercially available fitness trackers. Storm et al found that the Movemonitor, Fitbit One, ActivPAL, Nike+ Fuelband and Bodymedia Sensewear Armband Mini underestimated step count and the Kineteks Tractivity significantly overestimated step count. The Movemonitor performed best in walking recognition and step count but had difficulty discriminating between standing and sitting. [55] In a comparison of 14 step-counting methods using 8 step-counting devices (StepWatch, activPAL, Fitbit Zip, Yamax Digi-Walker SW-200, New Lifestyles NL-2000, and ActiGraph GT9X, Fitbit Charge and Fitbit AG), Toth et al found the StepWatch to most accurately record steps.[56]

The accuracy of step counters at low walking speeds is important to assess patients with decreased gait velocity and/or gait impairment. Results of studies evaluating step count accuracy under different walking speeds and gait conditions have varied. Beevi et al. found four commercial pedometers (Yamax, Digi-Walker SW-200, Omron HJ-720, and the Fitbit Zip) to be less accurate at lower walking speeds, while Madigan found no difference in accuracy across walking speeds using the Garmin Vivofit2, Fitbit Flex, Up3 and Microsoft Band but did find that the use of a walking cane decreased step counter accuracy.[54, 57, 58] Chandrasekar found commercial pedometers (Fitbit-Zip and ActiGraph) to be less accurate in patients with altered gait, including lower gait velocity, reduced stride length, longer double-limb support phase and greater self-reported functional impairment, than in healthy subjects.[59] Adding the ActiGraph low-frequency extension improved the accuracy of stair step counts. Gaglani et al. found the Fitbit Ultra step count to be more accurate in gait-compromised patients than the Omron Step Counter HJ-113, the New Lifestyles 2000, or the Nike Fuelband.[60]

Activity intensity estimates from commercial fitness wearable devices were found to be less accurate than those from research-grade accelerometers.[61] Schneider & Chau evaluated the Fitbit Zip daily step counts and minutes of moderate-to-vigorous physical activity (MVPA) against the ActiGraph and found that the Zip overestimated step counts, and this overestimation was higher prior to Fitbit's algorithm update calculating MPVA only from activity bouts >10 minutes.[62] Winfree & Dominick improved the Fitbit Flex classification of MVPA with a naive Bayes classifier using Fitbit-reported steps, metabolic equivalents, and intensity levels, using the ActiGraph GT3X/Freedson vector magnitude (VM3) algorithm as a ground truth.[61]

Physical activity other than walking can also be detected from wearables. Mooney et al. evaluated two swim stroke classification algorithms from the Finis Swimsense and the Garmin Swim. They found four swim strokes could be readily classified in advanced athletes, but laps at the beginning and end of an interval were not as accurately timed and stroke classification was more difficult in recreational swimmers due to individual variances in stroke technique.[63]

Energy Expenditure Estimation—Many wearables calculate estimated EE through algorithms that consider resting energy expenditure (REE) and activity energy expenditure (AEE). Gold standards for estimated EE include Doubly Labeled Water (DLW) technique and Indirect Calorimetry, based on measurements of oxygen consumption (VO_2) and carbon dioxide production. Estimates of REE (similar to basal metabolic rate) and EE often utilize the Mifflin-St. Joer[64], Katch- McArdle[65], or Cunningham[66] formulas to estimate based on sex, age, height, and weight inputs, though others have been proposed.[67] However, use of gold standard methods is often limited by instrument availability, logistical barriers, and inability to measure specific activity effects on EE.

Other measurements of EE have emerged, using the concept of METs and use of wearables-derived activity data. One MET is defined as consumption of oxygen required at rest; actual energy cost will vary between individuals due to differences in body mass, adiposity, age, gender, and environmental conditions, and is thus calculated as $1 \text{ MET} = \sim 3.5$

mL/O₂/min/kg × min.[68] Various activities have been compared on relative oxygen uptake or intensity of activities as multiples of 1 MET. While METs are a common standard for approximating EE for researchers and clinicians, their use in estimation of AEE has been shown to be inaccurate across individuals of different body mass and body fat categories.[69] Wearables are often convenient, relatively inexpensive alternatives to estimating EE, through measurements of HR, accelerometry data, and step-counts in addition to baseline user information. HR data can help characterize frequency, intensity and duration of physical activity, using calculations of relationships between HR and VO₂. [70] Yet, current standards suggest that no single technique is able to perfectly quantify EE associated with physical activity under free-living conditions and multiple complementary methods are recommended including HR, accelerometry and pedometer-measured step-counts.[71]

Pande et al. used smartphone sensors (accelerometer and barometer sensor) and showed that using a bagged regression tree ML algorithm with the addition of barometer data yielded a 96% correlation with actual EE determined by gold standard calorimetry (COSMED K4b2). [72] Surprisingly, smartphone sensor-based calculations had higher accuracy for EE when compared to EE derived by Fitbit and Nike+FuelBand. Other research has compared three commercial sport watches (Suunto Ambit2, Garmin Forerunner920XT, and Polar V800) and identified that the accuracy of the EE estimations was intensity-dependent for all tested watches, relative to indirect calorimetry.[73] In this study, the wearables had mean absolute percentage error values from -25% to +38% for aerobic running (4–11 km/h), with the Polar V800 performing most accurately (-3.6 to -12.2% mean error); all three watches significantly underestimated EE by -21.62% to -49.30% during anaerobic running (14–17 km/h).

Yet, several studies have identified lack of accuracy in EE estimates by wearables. Xi et al. tested 6 devices (Apple Watch 2, Samsung Gear S3, Jawbone Up3, Fitbit Surge, Huawei Talk Band B3, and Xiaomi Mi Band 2) and 2 smartphone apps (Dongdong and Ledongli) in 44 healthy participants; the authors measured five major health indicators (HR, number of steps, distance, energy expenditure, and sleep duration) under various activity states (resting, walking, running, cycling, and sleeping) against gold standard measurements. The tested wearables had high measurement accuracy with respect to heart rate, number of steps, distance, and sleep duration, but EE measurements made by these wearables were associated with lower measurement accuracy.[74] Shcherbina et al. evaluated seven wrist-worn devices (Apple Watch, Basis Peak, Fitbit Surge, Microsoft Band, Mio Alpha 2, PulseOn, and Samsung Gear S2) in estimating HR and EE against continuous telemetry and indirect calorimetry while 60 volunteers engaged in sitting, walking, running, and cycling. All devices reported the lowest error for cycling and the highest error during walking. The authors observed higher error for males, people with greater body mass index, participants with darker skin tone, and during walking. Most wrist-worn devices adequately measure HR in laboratory-based activities, but poorly estimated EE.[75]

Health-related Indicators

In addition to fitness-related indicators, another broad set of clinical goals for wearables include using raw wearable data and relevant algorithms to accurately and precisely define

and detect pathophysiological phenomena. While a large portion of clinical care relies on the use of patient-specific health data (e.g., history and physical exam, lab and other test results, imaging tests, etc.) and human clinical decision making, much of this care occurs in the traditional brick-and-mortar health setting, under a multitude of systemic constraints.[76] Given that changes in health status often occur gradually outside of the hospital and clinic[77], there is a clear role for remote monitoring of various patient populations to collect and process longitudinal health data into diagnostic, prognostic, and treatment-related insights.

Atrial Fibrillation Classification

AF is the most common clinically significant cardiac arrhythmia in the US, with economic costs estimated to be \$16-26 billion in annual US healthcare expenses.[78] Because AF increases risk for stroke, it is critical to treat and monitor accurately.[79] Recent studies have shown that both mobile, internet-enabled electrocardiography (iECG) and photoplethysmography (PPG) based methods are able to detect AF.[80]

Wearable devices most commonly aim to measure cardiac function through the use of PPG, which assesses changes in blood vessel expansion and contraction during the cardiac cycle via changes in light absorption and reflection of light emitting diode (LED) and photodiodes. Raw data from PPG sensors often must be processed with beat-detection algorithms which often employ pre-processing steps (including signal derivatives, digital filters, wavelets, or filter banks) and subsequent extraction of pressure pulse morphology and HR occurs using spectral HR estimation (e.g., power spectral density), peak detection (e.g., decision logic using k-NN algorithms), and final classification of inter-beat-intervals (IBI) of the time series containing the extracted candidate peaks.[81]

PPG has been used to detect AF.[82] Bonomi et al. used an ECG validated first-order Markov model to assess the probability of irregular rhythm due to AF using PPG-derived IBIs and observed $97 \pm 2\%$ sensitivity and $99 \pm 3\%$ specificity for AF detection against a Holter ECG monitor standard.[83] Early studies of PPG data from smartphone built-in cameras indicated that high accuracy of AF detection could be accomplished using Root Mean Square of Successive Differences (RMSSD), Shannon entropy (ShE) and Sample entropy (SampE), with sensitivity of 91.5% (85.9-95.4) and specificity of 99.6% (97.8-100). [84] In a study of the Empatica E4 wristband, equipped with PPG sensors, classification of normal sinus rhythm (NSR), AF, and other arrhythmias (ARR) reached an accuracy of 0.9, 0.9, and 0.8, for respective diagnostic groups.[85]

Additional use of ECG devices (e.g., 24-hour Holter monitor) has been a gold standard for AF detection.[86] A large number of algorithms have been proposed and tested for automated AF detection. Xia et al used short-term Fourier transform (STFT) and stationary wavelet transform (SWT) with subsequent deep CNN analysis (sensitivity = 98.79%, specificity = 97.87% and accuracy = 98.63%). Using SWT and SVM algorithms, Asgari et al. (2015) achieved sensitivity and specificity of 97.0% and 97.1%, respectively, and Xu et al. tested modified frequency SWT and CNN methods and achieved sensitivity of 74.96%, specificity of 86.41%, and accuracy of 81.07%.[87, 88] Other methods have largely used different parameters of ECG with CNN classification.[89, 90]

Several consumer-grade wearables have been studied with respect to their ability to successfully detect AF, such as AliveCor's KardiaMobile and KardiaBand or the Fibrichick system.[91-94] Bumgarner et al. studied the Kardia Band (KB), a novel wristband add-on to the Apple smartwatch designed for taking user-initiated measurements of cardiac rhythm strips. In a sample of 100 patients presenting for cardioversion, out of 169 simultaneous ECG and KB recordings, 57 were non-interpretable by the KB and subsequent interpretation of these 57 ECG by electrophysiologists diagnosed AF with 100% sensitivity, 80% specificity, and a K coefficient of 0.74. In the ECG recordings that KB interpreted, the authors observed that KB interpreted AF with 93% sensitivity, 84% specificity, and a K coefficient of 0.77 compared to physician interpretation of KB recordings (99% sensitivity, 83% specificity, and a K coefficient = 0.83).[95]

Additional research within the AF population has considered the performance and trade-offs between the ECG patch, wrist-worn ECG wearables, and gold standard clinical tools. According to research by Hernandez-Silveira et al., a digital patch device compared to clinical monitoring of HR and respiratory rate, had statistically significant ($p < 0.0001$) correlation, with lower correlation between respiratory rate in patients with atrial fibrillation ($p = 0.02$).[96] In a small sample ($n = 10$) of patients with AF, correlations of up to 85% were found between 24-h Holter and ECG patch AF detection. RR-interval analysis of both systems resulted in very high correlations of 99% and higher.[97] Further studies demonstrated that in patients with high CVD risk profile, a randomized-controlled trial of home-based wearable ECG sensor patch compared with delayed monitoring had a higher rate of AF diagnosis in the group with the wearable ECG patch.[98]

Further research has been done using algorithms that integrate information from Holter ECG and smart watch PPG data (Samsung Simband). Researchers retrieved spectral features from ECG signal using wavelet analysis, extracted features with a CNN, and used a bidirectional recurrent neural network (BRNN) and soft attention mechanism on top of the BRNN to achieve accurate detection of paroxysmal AF with area under the curve (AUC) = 0.94.

Autonomic Nervous System Function and Dysfunction

While heart rate reflects the average number of cardiac cycles per minute, heart rate variability (HRV) measures beat-to-beat variation in time between successive heart beats. Heart rate is determined by the ANS, and the sinoatrial node of the heart receives sympathetic and parasympathetic inputs that lead to variations in heart rate characteristics. Parasympathetic input is mediated primarily by the vagus nerve. Respiration affects heart rate, with increasing heart rate upon inhalation and decreasing heart rate with exhalation.[99] Additional factors that impact heart rate variability include physical activity and stress.

Decreased parasympathetic or increased sympathetic activity leads to reduced HRV. Higher HRV is associated with better aerobic fitness and also correlates inversely with age.[100] Decreased HRV reflects lowered autonomic tone and correlates with a variety of negative health states. Reduced HRV has been consistently shown to predict autonomic neuropathy in diabetic patients before the onset of symptoms.[101] Similarly, decreased HRV may have associations with congestive heart failure[102], mortality risk after myocardial infarction[103], sudden cardiac death[104], liver cirrhosis[105], and sepsis.[106] HRV is

moreover diminished in fitness overtraining and is used as a metric in managing recovery. [107]

The use of HRV derived from wearables has been implemented in studies of ANS pathophysiology across multiple patient populations. In a study by Cichosz et al., 21 patients with type 1 diabetes mellitus were monitored with continuous glucose monitors and Holter devices while they performed normal daily activities and were assessed for evidence of cardiovascular autonomic neuropathy during periods of hypoglycemia. The study indicated that the low frequency (LF) HRV parameter was significantly ($P = 0.029$) reduced during hypoglycemia for all patients, but surprisingly LF during hypoglycemia did not differ between patients with and without cardiac autonomic neuropathy as diagnosed by clinical criteria ($P = 0.74$). [108] Others have measured HRV and electrodermal response in older adults during simulated falls and during standing-lying transitions. Using SVM, positional changes were detectable at 70.37% sensitivity and 80% specificity with a polynomial kernel ($p = 5$). [109] In a separate study, researchers measured HR and locomotor activity via wearable patches in 16 patients with schizophrenia and 19 healthy controls; various signal processing techniques were applied and SVM trained on these features could discriminate patients with schizophrenia from controls with an AUC of 1.00 for test data, though reduced AUC was observed in distinguishing non-psychiatric patients from controls. [110]

Fall Detection

Every year in the United States, an estimated one third of adults 65 years and older fall at least one time during the year, and up to 30% of these falls can result in moderate to severe debilitation, morbidity, and mortality. [111] Identifying falls and initiating a time-sensitive response can reduce morbidity and mortality. [112] Research in automated fall detection systems is often classified by the type of sensors used, including video, acoustic, smartphone and wearable sensor based methods. [113]

Approaching fall detection using wearable sensors has often proven to be most practical for real-world applications and less prone to signal error. [34] Özdemir et al. studied falls with six wearable MTw sensor units (tri-axial devices with accelerometer, gyroscope, and magnetometer/compass) and compared six machine learning classifier techniques (k-NN, least squares method (LSM), SVM, Bayesian decision making, dynamic time warping, and artificial neural networks); considering performance and computational complexity, the k-NN classifier and LSM were superior with sensitivity, specificity, and accuracy all greater than 99%. [39] Gao et al. (2014) conducted a comparison of single versus multi-sensor systems and observed overall recognition accuracy of 96.4% using a decision tree classifier on mean and variance features, and observed decreased computational complexity of the classifier. [114] In a study by Aicha et al. (2018), the performance of three deep learning model architectures had strong recognition of identity of the subject, but these models only slightly outperformed other baseline methods on fall risk assessment. [115]

In addition to evaluating the number and type of wearable sensors needed for fall detection, several studies have assessed sensor placement on fall detection outcomes. One study identified that sensor placement on the waist region had 99.96% fall detection sensitivity (k-NN classifier) compared to 97.37% sensitivity for sensor placement on the wrist. [116] Attal

et al. (2015) tested multiple sensors on the chest, right thigh, and left ankle of participants; after testing both supervised and unsupervised ML algorithms in detecting fall events, the k-NN classifier provided the best performance compared to other supervised classification algorithms. A Hidden Markov Model classifier performed best among unsupervised classification algorithms.[117] Other studies placed sensors on the insole of a shoe, monitored participants during activities of daily living, and used SVM model and Leave-One-Out cross-validation to reach very accurate fall detection outcomes (sensitivity=0.996, specificity=1.000, and accuracy=0.999).[118]

While these and other studies have described devices and data analysis models for fall assessment, some research on fall detection devices has questioned their accuracy in practical settings. Chaudhuri et al. (2015) observed 25% sensitivity, 92% specificity, 2% positive predictive value, and greater than 99% negative predictive value in a fall detection device (with unknown detection algorithms used) when piloted with 7 older adults over a 4-month period of time.[119]

Detection of Behavior Change

Detecting changes in human behavioral patterns is an ongoing area of study, with research spanning different patient populations, sensor types, and data analysis methods. Whether behavioral endpoints are focused on diet, physical activity, or psychosocial markers, there is continued research to better understand the value of wearable technology and behavioral-indicator feedback for users and members of a healthcare team.

A large collection of research has been conducted on wearables-based, automated, user-independent methods of measuring diet (i.e., estimation of calorie consumption, food type identification, behavioral eating patterns, etc.).[120] Amft et al. developed a wearable ear pad sensor system for food classification, which recognized 19 types of foods with a performance accuracy of 86.6%.[121] The analysis methods utilized spectral signal analysis of acoustic data, Linear Predictive Coefficients, feature extraction (e.g., Mel-frequency cepstral coefficients, auto-regressive coefficients) and subsequent hierarchical agglomerative clustering approach (i.e., unsupervised ML). In addition to identifying type of food or quantity, eating behavior is affected by habit (i.e., night eating, evening overeating, weekend overeating, social factors, stress-correlated eating patterns). Sazanov et al. utilized mel-scale Fourier spectrum, wavelet packets and SVM algorithms to differentiate swallowing events from respiration, intrinsic speech, head movements, food ingestion and ambient noise, thus attempting to characterize food intake behaviors such as number and frequency of chews and swallows.[122]

In addition to diet, a general category of interest in behavioral change is detecting overall physical activity and sedentary behavior patterns. Some methods have looked at classical statistical changes of behavior frequency over time[123], using activity density map visualization[124], algorithm-derived activity curves[125], change-point detection in time series activity data[126], and novel supervised learning applications for change analysis. [127]

Sprint et al. (2016) developed a method for detecting physical activity change over varying time windows using Fitbit data collected from older adults who participated in a health intervention study. Other research focused on Fitbit Zip wearable data to assess daily step counts in overweight participants; researchers assigned daily step goals using a pseudo-random multisine algorithm that was a function of each participant's median baseline steps and used linear mixed effects model to detect an average increase of 2650 daily steps ($p < 0.01$) between baseline and completion of the intervention.[128] Other sensor methods assessed how longitudinal changes in actigraphy correlated with clinical and functional assessment scores for elderly patients in nursing homes.[129]

Beyond diet and physical activity behavior change, work has been done in automated psycho-physiological stress detection and changes over time. Gjoreski et al. provided a good review of research on wearables-based stress detection and presented a method for automatically measuring subjective stress using Empatica E3 and E4 wrist devices.[130] Their method utilized features extracted from PPG data (HR and IBI features), electrodermal activity (EDA), skin temperature data, and acceleration data (3-axis accelerometer data). Feature selection methods (e.g., correlation coefficients between features, information-gain rankings using the leave-one-subject-out technique) were implemented followed by classifier algorithm comparisons (Decision Tree, Naïve Bayes, k-NN, SVM) and ensemble algorithm comparisons (Bagging, Boosting, Random Forest, Ensemble Selection). Using 55 days of real-life data, the described methods detected 70% of the stress events with a precision of 95%.[130] Future work in this area would involve analyzing changes in frequency and correlates of stressful events to identify changes in stress response patterns.

Sleep Stage Estimation

Quantification of sleep stages can provide important clinical information for diagnosing patients with sleep disorders and identifying lifestyle factors involved in complex disease. Under the visual scoring system for sleep stages, sleep medicine has traditionally used polysomnography (PSG) approaches, which can include various combinations of data: EEG, airflow through nose and mouth, respiratory rate, BP changes, blood oxygen level, electrooculography (EOG) signals, and chin and legs surface electromyography (EMG). [131] Traditional visual scoring of the entire sleep session is highly time-consuming and observer-dependent; thus, there have been research efforts to automate the measurement of sleep stages by using novel preprocessing, feature extraction, feature selection and classifier methods. While some research avenues have focused on establishing clinical gold standard electroencephalogram (EEG)-based ground-truth criteria for sleep stage estimation[132], there have also been efforts to categorize sleep stages using wearable data.

In qualitative analysis, PSGs of an entire sleep session were traditionally graded according to traditional Rechtschaffen and Kales (R&K) criteria[133], and more recently, American Academy of Sleep Medicine (AASM) standards.[134] EEG has had the most clinical utility and sleep epochs are categorized according to Delta, Theta, Alpha, Beta bands and the presence or absence of sleep spindles and K-complexes.[132] The AASM manual defines the following sleep stages: Awake state (stage W); Stage N1; Stage N2; Stage N3; and Rapid Eye Movement (REM).[134] Due to limitations and constraints associated with qualitative

assessment, there has been a large increase in proposed automated sleep stage classification methods and increasing interest in exploring wearable detection methodologies.[135, 136]

Beattie et al. (2017) reported that the use of accelerometer and PPG data from a Fitbit Surge wearable to assess movement, breathing, and HR variability had a 69% classification accuracy for sleep stage when compared against EEG recordings and classification.[137] Early tests of Fitbit (unknown model) wearables compared to actigraphy and PSG gold standard by Montgomery-Downs et al. indicated overestimation of both sleep time and quality, and observed high sensitivity of both Fitbit and actigraphy for accurately identifying sleep within all sleep stages and during arousals, but poor specificity of both for accurately identifying Awake Stage.[138] Along similar lines, Castner et al. (2018) observed that the Fitbit Charge overestimated sleep efficiency and underestimated wake counts compared to actigraphy in a sample of women with poorly-controlled asthma.[139] Other research in healthy subjects raised concerns that Fitbit Flex wearables failed to capture a significant amount of sleep data at baseline: only 14% of Fitbit Flexes recorded sleep after 5 nights into the trial and nearly 35% failed to record any nights of sleep in a cohort of 107 students with an attempt to control for user error and battery charge limitations.[140]

In addition to estimating sleep stages against gold standards, future work is likely to focus on distinguishing parameters of sleep in populations with known sleep disorders. Research should focus on improving the accuracy and precision of distinguishing awake resting versus sleep states, and comparisons of wearable versus non-wearable (e.g., remote monitoring hardware under a mattress, etc.) approaches to sleep quality detection.

Serum Glucose Control and Diabetes Management

Diabetes mellitus (DM, types 1 and 2) was the seventh leading cause of death in the U.S. and has an associated direct and indirect cost \$245 billion in the U.S.[141, 142] Given the need for diagnosis and effective management of serum glucose levels, continuous glucose monitor (CGM) devices have emerged as a promising clinical tools for management of both DM type 1, DM type 2, and pre-diabetic phenotypes. Industry leaders in the field include Dexcom, Abbott, and Medtronic whose CGM devices all employ subdermal enzymatic sensors with sensor life of 1-2 weeks.[143] A newer device by Senseonics Inc. incorporates an implanted fluorescence-based sensor with a longer lifetime of 3 months.[144] Broadly, CGMs output interstitial glucose values that are proportional to actual serum glucose levels. Calibration algorithms are used to convert the raw sensor signal (often in nanoamps) to a blood glucose estimate (in the U.S., milligrams per deciliter) using linear modeling and linear regression, followed by filtering algorithms (e.g., median filters, finite and infinite impulse response filters, Kalman filtering[145]), hypo- and hyper-glycemic threshold alarm systems, blood-interstitial fluid glucose dynamics and lag, and signal dropouts and artifact analysis.[146]

CGM provides input to models for predicting fluctuations in blood glucose concentration over time, allowing monitoring and predictions of serum glucose for both DM type 1 and type 2 patient populations. In DM type 1, there is value in predicting life-threatening hypoglycemic events. Typically such predictive algorithms have depended on inputting meal intake, activity, infection, and emotional factors that influence glucose metabolism. Recent

advances in glucose prediction based on CGM readings include methods based on deep learning and SVM that do not require these additional inputs and predict 60 minutes of glucose changes with mean squared error 12.95mg/dl.[147, 148] In addition, there has been extensive research and development of fully closed-loop Artificial Pancreas Device Systems (i.e., closed loop glucose control system).[149, 150] These devices aim to combine CGM with responsive insulin and glucagon delivery as functions of a series of robust glycemic control algorithms, to improve safe treatment of patients with DM type 1. In patients with DM type 2, there is potential value in having non-invasive CGM devices that reduce the need for finger-stick glucose monitoring and increase the clinical insight into changing serum glucose levels in the complex milieu of individual patient health characteristics and more fine-grained analysis of treatment efficacy and effectiveness.

CGMs also enable discovery of individual-specific dynamic glucose patterns and may also be informative for pre-diabetics or non-diabetics for identifying risk factors and trajectories of developing diabetes. Individuals classified as non-diabetic by clinical metrics including HbA1C, fasting blood glucose, and oral glucose tolerance can still have frequent spikes in blood glucose that reach the diabetic range (> 200 mg/dl).[47] The authors used Dexcom G4 CGM devices and characterized glucose variability by dynamic time warping, spectral clustering using von Luxburg's methods, subsequent k-NN unsupervised training techniques and resultant analysis of spectral clustering to detect unique glycemic signatures as low, intermediate, and severe glucotypes.

As CGM technology continues to become more accessible, it will likely be a useful tool for both diabetics and non-diabetics for dietary and lifestyle management, in addition to more fine-grained clinical understanding of intra- and inter-patient changes in glycemic levels and response to treatment.

Conclusion

Wearable health technology is rapidly growing in both commercial markets and biomedical fields. These devices aim to collect and process raw physiological or environmental parameters into salient digital health information. Here we sought to highlight several of the emerging applications of wearables algorithms that hold promise for understanding both normal baseline physiology and pathophysiology. Much of the current and future utility of wearables lies in the signal processing steps and algorithms used for rapidly analyzing large volumes of data. Continued algorithmic development and advances in machine learning techniques will likely lead to further improvements in analytic capabilities. In this context, our review aims to detail select fitness- and health-related indicators currently available through commercial wearable technologies.

REFERENCES WITH SELECTED RESEARCH OF INTEREST:

•• = outstanding interest

• = special interest

1. Eapen ZJ, et al., Defining a Mobile Health Roadmap for Cardiovascular Health and Disease. J Am Heart Assoc, 2016 5(7).

2. Neubeck L, et al., The mobile revolution--using smartphone apps to prevent cardiovascular disease. *Nat Rev Cardiol*, 2015 12(6): p. 350–60. [PubMed: 25801714]
3. eHealth, W.G.O.f., MHealth: new horizons for health through mobile technologies. 2011, Geneva: World Health Organization.
4. Jain SH, et al., The digital phenotype. *Nat Biotechnol*, 2015 33(5): p. 462–3. [PubMed: 25965751]
- 5. Li X, et al., Digital Health: Tracking Physiomes and Activity Using Wearable Biosensors Reveals Useful Health-Related Information. *PLoS Biol*, 2017 15(1): p. e2001402. [PubMed: 28081144]
An important study incorporating over 250,000 daily measurements of 43 people to analyze baseline physiological health parameters, changes in these parameters in different ambient environments, and indicators of pathological changes in these baseline characteristics.
- 6. Doherty A, et al., Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. *PLoS One*, 2017 12(2): p. e0169649. [PubMed: 28146576] The study is notable for use of accelerometer measured physical activity in over 100,000 participants of the UK Biobank study, and a good feasibility study for collection and analysis of objective physical activity data in large studies.
7. Wolff-Hughes DL, Bassett DR, and Fitzhugh EC, Population-referenced percentiles for waist-worn accelerometer-derived total activity counts in U.S. youth: 2003 – 2006 NHANES. *PLoS One*, 2014 9(12): p. e115915. [PubMed: 25531290]
8. All of Us Research Program Protocol | National Institutes of Health (NIH) — All of Us. [cited 2018 11/9]; Available from: <https://allofus.nih.gov/about/all-us-research-program-protocol>.
9. Maxmen A, Google spin-off deploys wearable electronics for huge health study. *Nature*, 2017 547(7661): p. 13–14. [PubMed: 28682357]
10. Pasluosta CF, et al., An Emerging Era in the Management of Parkinson's Disease: Wearable Technologies and the Internet of Things. *IEEE J Biomed Health Inform*, 2015 19(6): p. 1873–81. [PubMed: 26241979]
11. Dunn J, Runge R, and Snyder M, Wearables and the medical revolution. *Per Med*, 2018 15(5): p. 429–448. [PubMed: 30259801]
12. Faedda GL, et al., Actigraph measures discriminate pediatric bipolar disorder from attention-deficit/hyperactivity disorder and typically developing controls. *J Child Psychol Psychiatry*, 2016 57(6): p. 706–16. [PubMed: 26799153]
13. Gentili C, et al., Longitudinal monitoring of heartbeat dynamics predicts mood changes in bipolar patients: A pilot study. *J Affect Disord*, 2017 209: p. 30–38. [PubMed: 27870943]
14. Kim J, et al., Co-variation of depressive mood and locomotor dynamics evaluated by ecological momentary assessment in healthy humans. *PLoS One*, 2013 8(9): p. e74979. [PubMed: 24058642]
15. Roh T, Hong S, and Yoo HJ, Wearable depression monitoring system with heart-rate variability. *Conf Proc IEEE Eng Med Biol Soc*, 2014 2014: p. 562–5.
16. Carreiro S, et al., Real-time mobile detection of drug use with wearable biosensors: a pilot study. *J Med Toxicol*, 2015 11(1): p. 73–9. [PubMed: 25330747]
17. Search of: wearable | Completed Studies - Results by Topic - [ClinicalTrials.gov](https://clinicaltrials.gov/ct2/results/browse?term=wearable&recrs=e&brwse=cond_alpha_all). [cited 2018 11/9]; Available from: https://clinicaltrials.gov/ct2/results/browse?term=wearable&recrs=e&brwse=cond_alpha_all.
18. Janidarmian M, et al., A Comprehensive Analysis on Wearable Acceleration Sensors in Human Activity Recognition. *Sensors (Basel)*, 2017 17(3).
19. Celka PV, R.; Renevey P; Verjus C; Neuman V, Wearable biosensing: signal processing and communication architectures issues. *Journal of Telecommunications and Information Technology*, 2005: p. 15.
- 20. King RC, et al., Application of data fusion techniques and technologies for wearable health monitoring. *Med Eng Phys*, 2017 42: p. 1–12. [PubMed: 28237714] King et al. provide a short overview of data fusion techniques and algorithms that can be used to interpret wearable sensor data in the context of health monitoring applications.
20. Luo J, et al., Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomed Inform Insights*, 2016 8: p. 1–10.
21. Jiang F, et al., Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*, 2017 2(4): p. 230–243. [PubMed: 29507784]

22. LeCun Y, Bengio Y, and Hinton G, Deep learning. *Nature*, 2015 521(7553): p. 436–44. [PubMed: 26017442]
23. Ma J, Sheridan RP, Liaw A, Dahl GE & Svetnik V, Deep neural nets as a method for quantitative structure-activity relationships. *J. Chem. Inf. Model*, 2015 55: p. 263–274. [PubMed: 25635324]
24. Helmstaedter M, et al., Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 2013 500(7461): p. 168–74. [PubMed: 23925239]
25. Leung MK, et al., Deep learning of the tissue-regulated splicing code. *Bioinformatics*, 2014 30(12): p. i121–9. [PubMed: 24931975]
26. Xiong HY, et al., RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science*, 2015 347(6218): p. 1254806. [PubMed: 25525159]
27. Krizhevsky A, Sutskever I & Hinton G, ImageNet classification with deep convolutional neural networks. *Proc. Advances in Neural Information Processing Systems*, 2012 25: p. 1090–1098.
28. Farabet C, et al., Learning hierarchical features for scene labeling. *IEEE Trans Pattern Anal Mach Intell*, 2013 35(8): p. 1915–29. [PubMed: 23787344]
29. Tompson J, Jain A, LeCun Y & Bregler C, Joint training of a convolutional network and a graphical model for human pose estimation. *Proc. Advances in Neural Information Processing Systems*, 2014 27: p. 1799–1807.
30. Mikolov T, Deoras A, Povey D, Burget L & Cernocky J, Strategies for training large scale neural network language models. *Proc. Automatic Speech Recognition and Understanding*, 2011: p. 196–201
31. Hinton G.e.a., Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, 2012 29: p. 82–97
32. Rokach L, Ensemble-based classifiers. *Artificial Intelligence Review*, 2010 33(1-2): p. 1–39.
- 33. Aziz O, et al., Validation of accuracy of SVM-based fall detection system using real-world fall and non-fall datasets. *PLoS One*, 2017 12(7): p. e0180318. [PubMed: 28678808] Aziz et al. use data from falls in the real-world setting (instead of only laboratory-derived data) and use Support Vector Machine classification algorithms to achieve relatively high accuracy of a fall detection system.
34. Dinh A, et al., A fall and near-fall assessment and evaluation system. *Open Biomed Eng J*, 2009 3: p. 1–7. [PubMed: 19662151]
35. Nukala BT, et al., Real-Time Classification of Patients with Balance Disorders vs. Normal Subjects Using a Low-Cost Small Wireless Wearable Gait Sensor. *Biosensors (Basel)*, 2016 6(4).
36. Mirmahboub B, et al., Automatic monocular system for human fall detection based on variations in silhouette area. *IEEE Trans Biomed Eng*, 2013 60(2): p. 427–36. [PubMed: 23192468]
37. Xi X, et al., Evaluation of Feature Extraction and Recognition for Activity Monitoring and Fall Detection Based on Wearable sEMG Sensors. *Sensors (Basel)*, 2017 17(6).
38. Ozdemir AT and Barshan B, Detecting falls with wearable sensors using machine learning techniques. *Sensors (Basel)*, 2014 14(6): p. 10691–708. [PubMed: 24945676]
39. Gilani M, Eklund JM, and Makrehchi M, Automated detection of atrial fibrillation episode using novel heart rate variability features. *Conf Proc IEEE Eng Med Biol Soc*, 2016 2016: p. 3461–3464. [PubMed: 28269045]
40. Smisek R, et al., Multi-stage SVM approach for cardiac arrhythmias detection in short single-lead ECG recorded by a wearable device. *Physiol Meas*, 2018 39(9): p. 094003. [PubMed: 30102239]
41. Nemati S, et al., Monitoring and detecting atrial fibrillation using wearable technology. *Conf Proc IEEE Eng Med Biol Soc*, 2016 2016: p. 3394–3397. [PubMed: 28269032]
- 42. I Inan OT, et al., Novel Wearable Seismocardiography and Machine Learning Algorithms Can Assess Clinical Status of Heart Failure Patients. *Circ Heart Fail*, 2018 11(1): p. e004313. [PubMed: 29330154] Patients with compensated and decompensated HF were fitted with a wearable ECG and seismocardiogram sensing patch and machine learning algorithms (including graph score analysis) detected compensated and decompensated HF states by analyzing cardiac response to submaximal exercise.
43. Ireland D, et al., Classification of Movement of People with Parkinsons Disease Using Wearable Inertial Movement Units and Machine Learning. *Stud Health Technol Inform*, 2016 227: p. 61–6. [PubMed: 27440290]

- 44. Kubota KJ, Chen JA, and Little MA, Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures. *Mov Disord*, 2016 31(9): p. 1314–26. [PubMed: 27501026] A thorough, non-technical overview of machine learning algorithms and their potential applications to detecting, categorizing, and predicting changes in patients with motor signs/symptoms of Parkinson's Dementia.
- 45. Rovini E, et al., Comparative Motor Pre-clinical Assessment in Parkinson's Disease Using Supervised Machine Learning Approaches. *Ann Biomed Eng*, 2018 46(12): p. 2057–2068. [PubMed: 30030773]
- 46. Zeevi D, et al., Personalized Nutrition by Prediction of Glycemic Responses. *Cell*, 2015 163(5): p. 1079–1094. [PubMed: 26590418]
- 47. Hall H, et al., Glucotypes reveal new patterns of glucose dysregulation. *PLoS Biol*, 2018 16(7): p. e2005143. [PubMed: 30040822] Use of Dynamic Time Warping (i.e., machine learning approach) and other statistical methods to separate individuals into “glucotypes” based on response to glucose challenges using continuous glucose monitoring data. The authors found that even individuals considered normoglycemic by standard measures exhibit high glucose variability using CGM, with glucose levels reaching prediabetic and diabetic ranges 15% and 2% of the time.
- 48. Woodbridge J, et al., Aggregated Indexing of Biomedical Time Series Data. *Proc IEEE Int Conf Healthc Inform Imaging Syst Biol*, 2012 2012: p. 23–30. [PubMed: 27617298]
- 49. Barnett I, et al., Beyond smartphones and sensors: choosing appropriate statistical methods for the analysis of longitudinal data. *J Am Med Inform Assoc*, 2018.
- 50. Warburton DE, Nicol CW, and Bredin SS, Prescribing exercise as preventive therapy. *CMAJ*, 2006 174(7): p. 961–74. [PubMed: 16567757]
- 51. Piercy KL, et al., The Physical Activity Guidelines for Americans. *JAMA*, 2018 320(19): p. 2020–2028. [PubMed: 30418471]
- 52. Evenson KR, Goto MM, and Furberg RD, Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act*, 2015 12: p. 159. [PubMed: 26684758]
- 53. Ahola T, Pedometer for Running Activity Using Accelerometer Sensors on the Wrist. *Medical Equipment Insights*, 2010 3.
- 54. Bassett DR Jr., et al., Step Counting: A Review of Measurement Considerations and Health-Related Applications. *Sports Med*, 2017 47(7): p. 1303–1315. [PubMed: 28005190]
- 55. Storm FA, Heller BW, and Mazza C, Step detection and activity recognition accuracy of seven physical activity monitors. *PLoS One*, 2015 10(3): p. e0118723. [PubMed: 25789630] A comprehensive study comparing the seven commercially available activity (Movemonitor (Mc Roberts), Up (Jawbone), One (Fitbit), ActivPAL (PAL Technologies Ltd.), Nike+ Fuelband (Nike Inc.), Tractivity (Kineteks Corp.) and Sensewear Armband Mini (Bodymedia)) and measuring accuracy of step counts and activity detection (e.g., sitting vs. standing).
- 56. Toth LP, et al., Video-Recorded Validation of Wearable Step Counters under Free- living Conditions. *Med Sci Sports Exerc*, 2018 50(6): p. 1315–1322. [PubMed: 29381649]
- 57. Beevi FH, et al., An Evaluation of Commercial Pedometers for Monitoring Slow Walking Speed Populations. *Telemed J E Health*, 2016 22(5): p. 441–9. [PubMed: 26451900]
- 58. Madigan EA, Fitness band accuracy in older community dwelling adults. *Health Informatics J*, 2017: p. 1460458217720399.
- 59. Chandrasekar A, et al., Preliminary concurrent validity of the Fitbit-Zip and ActiGraph activity monitors for measuring steps in people with polymyalgia rheumatica. *Gait Posture*, 2018 61: p. 339–345. [PubMed: 29427859]
- 60. Gaglani SM, Moore J; Haynes R; Hoffberger J; Rigamonti D, Using Commercial Activity Monitors to Measure Gait in Patients with Suspected iNPH: Implications for Ambulatory Monitoring. *Cureus*, 2015 7(11).
- 61. Winfree KN and Dominick G, Modeling Clinically Validated Physical Activity Assessments Using Commodity Hardware. *IEEE J Biomed Health Inform*, 2018 22(2): p. 335–345. [PubMed: 29505401]

62. Schneider M and Chau L, Validation of the Fitbit Zip for monitoring physical activity among free-living adolescents. *BMC Res Notes*, 2016 9(1): p. 448. [PubMed: 27655477]
- 63. Mooney R, et al., Evaluation of the Finis Swimsense(R) and the Garmin Swim activity monitors for swimming performance and stroke kinematics analysis. *PLoS One*, 2017 12(2): p. e0170902. [PubMed: 28178301] Mooney et al. measure accuracy of two wearables in their quantification of temporal and kinematic swimming variables, identifying limitations in accuracy of stroke rate, stroke length and average speed scores but reasonable measurement of individual lap times and stroke counts.
64. Mifflin MD, et al., A new predictive equation for resting energy expenditure in healthy individuals. *Am J Clin Nutr*, 1990 51(2): p. 241–7. [PubMed: 2305711]
65. McArdle WK, F., *Essentials of Exercise Physiology*. 5th, North American edition ed. 2015, Philadelphia: LWW.
66. Cunningham JR and Johns HE, Calculation of the average energy absorbed in photon interactions. *Med Phys*, 1980 7(1): p. 51–4. [PubMed: 7366541]
67. Harris JA and Benedict FG, A biometric study of basal metabolism in man Carnegie Institution of Washington publication. 1919, Washington,: Carnegie Institution of Washington vi, 266 p. incl. tables.
68. Jette M, Sidney K, and Blumchen G, Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. *Clin Cardiol*, 1990 13(8): p. 555–65. [PubMed: 2204507]
69. Byrne NM, et al., Metabolic equivalent: one size does not fit all. *J Appl Physiol* (1985), 2005 99(3): p. 1112–9. [PubMed: 15831804]
70. Ceesay SM, et al., The use of heart rate monitoring in the estimation of energy expenditure: a validation study using indirect whole-body calorimetry. *Br J Nutr*, 1989 61(2): p. 175–86. [PubMed: 2706223]
71. Schutz Y, Weinsier RL, and Hunter GR, Assessment of free-living physical activity in humans: an overview of currently available and proposed new measures. *Obes Res*, 2001 9(6): p. 368–79. [PubMed: 11399784]
72. Pande A, et al., Using Smartphone Sensors for Improving Energy Expenditure Estimation. *IEEE J Transl Eng Health Med*, 2015 3: p. 2700212. [PubMed: 27170901]
73. Roos L, et al., Validity of sports watches when estimating energy expenditure during running. *BMC Sports Sci Med Rehabil*, 2017 9: p. 22. [PubMed: 29296281]
74. Xie J, et al., Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study. *JMIR Mhealth Uhealth*, 2018 6(4): p. e94. [PubMed: 29650506]
- 75. Shcherbina A, et al., Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy Expenditure in a Diverse Cohort. *J Pers Med*, 2017 7(2).A thorough assessment of the accuracy of seven commercially available wrist-worn devices in estimating heart rate and energy expenditure and a proposed wearable sensor evaluation framework.
76. Reid PP, Compton WD; Grossman J; Fanjiang G, *A New Partnership between Systems Engineering and Health Care*, ed. N.A.o.E.U.a.I.o.M.U.C.o.E.a.t.H. Care. 2005: National Academies Press (US).
77. Ashman JJ, Rui P, and Okeyode T, Characteristics of Office-based Physician Visits, 2014. *NCHS Data Brief*, 2017(292): p. 1–8.
78. Chugh SS, et al., Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation*, 2014 129(8): p. 837–47. [PubMed: 24345399]
79. Wolf PA, Abbott RD, and Kannel WB, Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke*, 1991 22(8): p. 983–8. [PubMed: 1866765]
80. Turakhia MP, et al., Feasibility of extended ambulatory electrocardiogram monitoring to identify silent atrial fibrillation in high-risk patients: the Screening Study for Undiagnosed Atrial Fibrillation (STUDY-AF). *Clin Cardiol*, 2015 38(5): p. 285–92. [PubMed: 25873476]
81. Aboy M, et al., An automatic beat detection algorithm for pressure signals. *IEEE Trans Biomed Eng*, 2005 52(10): p. 1662–70. [PubMed: 16235652]

82. Lewis M, et al., Screening for atrial fibrillation: sensitivity and specificity of a new methodology. *Br J Gen Pract*, 2011 61(582): p. 38–9. [PubMed: 21401988]
- 83. Bonomi AG, et al., Atrial Fibrillation Detection Using a Novel Cardiac Ambulatory Monitor Based on Photo-Plethysmography at the Wrist. *J Am Heart Assoc*, 2018 7(15): p. e009351. [PubMed: 30371247] The authors measured wearable photoplethysmography data and detected AF with >96% accuracy and compared cohorts undergoing cardioversion and in outpatient setting. They developed an algorithm with <0.2% false positive AF detection rate, but noted that the proposed algorithm did not provide AF classification for 44+/-16% of the monitoring period in the outpatient cohort due to motion artifacts. This represents high accuracy of AF detection methods and room for improvement to accommodate realistic artifacts.
84. Lee J RB, McManus DD, Mathias O, Chon KH, Atrial fibrillation detection using a smart phone. *Conf Proc IEEE Eng Med Biol Soc*, 2012(2012): p. 1177–80. [PubMed: 23366107]
85. Corino VDA, et al., Detection of atrial fibrillation episodes using a wristband device. *Physiol Meas*, 2017 38(5): p. 787–799. [PubMed: 28151434]
86. January CT, et al., 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the Heart Rhythm Society. *Circulation*, 2014 130(23): p. e199–267. [PubMed: 24682347]
87. Asgari S, Mehrnia A, and Moussavi M, Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine. *Comput Biol Med*, 2015 60: p. 132–42. [PubMed: 25817534]
88. Xu X, et al., Atrial Fibrillation Beat Identification Using the Combination of Modified Frequency Slice Wavelet Transform and Convolutional Neural Networks. *J Healthc Eng*, 2018 2018: p. 2102918. [PubMed: 30057730]
89. Fan X, et al., Multiscaled Fusion of Deep Convolutional Neural Networks for Screening Atrial Fibrillation From Single Lead Short ECG Recordings. *IEEE J Biomed Health Inform*, 2018 22(6): p. 1744–1753. [PubMed: 30106699]
90. He R, et al., Automatic Detection of Atrial Fibrillation Based on Continuous Wavelet Transform and 2D Convolutional Neural Networks. *Front Physiol*, 2018 9: p. 1206. [PubMed: 30214416]
91. Vandenberk T, et al., Clinical Validation of Heart Rate Apps: Mixed-Methods Evaluation Study. *JMIR Mhealth Uhealth*, 2017 5(8): p. e129. [PubMed: 28842392]
- 92. William AD, et al., Assessing the accuracy of an automated atrial fibrillation detection algorithm using smartphone technology: The iREAD Study. *Heart Rhythm*, 2018 15(10): p. 1561–1565. [PubMed: 30143448] William et al. conducted a single-center, adjudicator-blinded case series of 52 consecutive patients with AF and tested the Kardia Mobile Cardiac Monitor to detect atrial fibrillation with automated algorithm interpretation achieving 96.6% sensitivity and 94.1% specificity for AF detection. The authors also observed that some patient recordings were uninterpretable by the Kardia device, indicating further algorithm development to accommodate clinical data.
- 93. Brasier N, et al., Detection of atrial fibrillation with a smartphone camera: first prospective, international, two-centre, clinical validation study (DETECT AF PRO). *Europace*, 2018. This study was a prospective, two-center, clinical validation study of 592 patients with presumed atrial fibrillation or controls, using photoplethysmography data collected by a smartphone camera with subsequent rhythm analysis. The authors observed detection of atrial fibrillation with a sensitivity of 91.5% (85.9-95.4) and specificity of 99.6% (97.8-100), indicating the potential validity of using smartphone-camera-based methods of collecting PPG data and making subsequent classification of atrial fibrillation.
94. Lowres N, et al., Feasibility and cost-effectiveness of stroke prevention through community screening for atrial fibrillation using iPhone ECG in pharmacies. The SEARCH-AF study. *Thromb Haemost*, 2014 111(6): p. 1167–76. [PubMed: 24687081]
95. Bumgarner JM, et al., Smartwatch Algorithm for Automated Detection of Atrial Fibrillation. *J Am Coll Cardiol*, 2018 71(21): p. 2381–2388. [PubMed: 29535065]
96. Hernandez-Silveira M, et al., Preliminary assessment of the SensiumVitals(R): A low-cost wireless solution for patient surveillance in the general wards. *Conf Proc IEEE Eng Med Biol Soc*, 2015 2015: p. 4931–7.

97. Torfs T, et al., Clinical validation of a low-power and wearable ECG patch for long term full-disclosure monitoring. *J Electrocardiol*, 2014 47(6): p. 881–9. [PubMed: 25201418]
98. Steinhubl SR, et al., Effect of a Home-Based Wearable Continuous ECG Monitoring Patch on Detection of Undiagnosed Atrial Fibrillation: The mSToPS Randomized Clinical Trial. *JAMA*, 2018 320(2): p. 146–155. [PubMed: 29998336]
99. Rajendra Acharya U, et al., Heart rate variability: a review. *Med Biol Eng Comput*, 2006 44(12): p. 1031–51. [PubMed: 17111118]
100. De Meersman RE, Heart rate variability and aerobic fitness. *Am Heart J*, 1993 125(3): p. 726–31. [PubMed: 8438702]
101. Vinik AI, et al., Diabetic autonomic neuropathy. *Diabetes Care*, 2003 26(5): p. 1553–79. [PubMed: 12716821]
102. Nolan J, et al., Prospective study of heart rate variability and mortality in chronic heart failure: results of the United Kingdom heart failure evaluation and assessment of risk trial (UK-heart). *Circulation*, 1998 98(15): p. 1510–6. [PubMed: 9769304]
103. Kleiger RE, et al., Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *Am J Cardiol*, 1987 59(4): p. 256–62. [PubMed: 3812275]
104. Molgaard H, Sorensen KE, and Bjerregaard P, Circadian variation and influence of risk factors on heart rate variability in healthy subjects. *Am J Cardiol*, 1991 68(8): p. 777–84. [PubMed: 1892086]
105. Mani AR, et al., Decreased heart rate variability in patients with cirrhosis relates to the presence and degree of hepatic encephalopathy. *Am J Physiol Gastrointest Liver Physiol*, 2009 296(2): p. G330–8. [PubMed: 19023029]
106. Griffin MP and Moorman JR, Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis. *Pediatrics*, 2001 107(1): p. 97–104. [PubMed: 11134441]
107. Lucini D, et al., Heart rate variability to monitor performance in elite athletes: Criticalities and avoidable pitfalls. *Int J Cardiol*, 2017 240: p. 307–312. [PubMed: 28495249]
108. Cichosz SL, et al., Are Changes in Heart Rate Variability During Hypoglycemia Confounded by the Presence of Cardiovascular Autonomic Neuropathy in Patients with Diabetes? *Diabetes Technol Ther*, 2017 19(2): p. 91–95. [PubMed: 28118049]
109. Nocua R, et al., Evaluation of the autonomic nervous system for fall detection. *Conf Proc IEEE Eng Med Biol Soc*, 2009 2009: p. 3225–8.
- 110. Reinertsen E, et al., Multiscale network dynamics between heart rate and locomotor activity are altered in schizophrenia. *Physiol Meas*, 2018 39(11): p. 115001. [PubMed: 30222594] Reinertsen et al. used wearable-derived data to quantify and measure interactions between heart rate and locomotor activity in patient with diagnosed schizophrenia. The presented algorithm achieved perfect discrimination of subjects with schizophrenia from controls, raising the possibility of using wearables to monitor other physiological and pathophysiological correlates of patients with psychiatric disorders.
111. Axer H, et al., Falls and gait disorders in geriatric neurology. *Clin Neurol Neurosurg*, 2010 112(4): p. 265–74. [PubMed: 20089351]
112. Gurley RJ, et al., Persons found in their homes helpless or dead. *N Engl J Med*, 1996 334(26): p. 1710–6. [PubMed: 8637517]
113. Khan SS and Hoey J, Review of fall detection techniques: A data availability perspective. *Med Eng Phys*, 2017 39: p. 12–22. [PubMed: 27889391]
114. Gao L, Bourke AK, and Nelson J, Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Med Eng Phys*, 2014 36(6): p. 779–85. [PubMed: 24636448]
115. Nait Aicha A, et al., Deep Learning to Predict Falls in Older Adults Based on Daily-Life Trunk Accelerometry. *Sensors (Basel)*, 2018 18(5).
116. Ozdemir AT, An Analysis on Sensor Locations of the Human Body for Wearable Fall Detection Devices: Principles and Practice. *Sensors (Basel)*, 2016 16(8).
117. Attal F, et al., Physical Human Activity Recognition Using Wearable Sensors. *Sensors (Basel)*, 2015 15(12): p. 31314–38. [PubMed: 26690450]

118. Cates B, et al., A Novel Detection Model and Its Optimal Features to Classify Falls from Low- and High-Acceleration Activities of Daily Life Using an Insole Sensor System. *Sensors (Basel)*, 2018 18(4).
119. Chaudhuri S, et al., Real-World Accuracy and Use of a Wearable Fall Detection Device by Older Adults. *J Am Geriatr Soc*, 2015 63(11): p. 2415–6. [PubMed: 26603067]
120. Tri Vu FL, Nabil Alshurafa, Wenyao Xu, Wearable Food Intake Monitoring Technologies: A Comprehensive Review. *Computers* 2017 6(4): p. 1–28.
121. Amft O, A wearable earpad sensor for chewing monitoring *Sensors*, 2010 IEEE, 2010.
122. Sazonov EMO; Schuckers S; Lopez-Meyer P; Melanson EL; Neuman MR, Automatic Detection of Swallowing Events by Acoustical Means for Applications of Monitoring of Ingestive Behavior. *IEEE Trans. Biomed. Eng.*, 2010 57: p. 626–633. [PubMed: 19789095]
123. Paavilainen P, et al., Circadian activity rhythm in demented and non-demented nursing-home residents measured by telemetric actigraphy. *J Sleep Res*, 2005 14(1): p. 61–8. [PubMed: 15743335]
124. Wang C, et al., A Low-Power Fall Detector Balancing Sensitivity and False Alarm Rate. *IEEE J Biomed Health Inform*, 2018 22(6): p. 1929–1937. [PubMed: 29990072]
125. Dawadi PN, Cook DJ, and Schmitter-Edgecombe M, Modeling Patterns of Activities using Activity Curves. *Pervasive Mob Comput*, 2016 28: p. 51–68. [PubMed: 27346990]
126. Liu S, et al., Change-point detection in time-series data by relative density-ratio estimation. *Neural Netw*, 2013 43: p. 72–83. [PubMed: 23500502]
127. Hido SI, Ide T; Kashima H; Kubo H; Matsuzawa H, Unsupervised Change Analysis Using Supervised Learning, in *Advances in Knowledge Discovery and Data Mining*. 2008, Springer Berlin Heidelberg p. 148–159.
- 128. Korinek EV, et al., Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. *J Behav Med*, 2018 41(1): p. 74–86. [PubMed: 28918547] Korinek et al. present a feasibility study on using adaptive interventions as a behavioral intervention strategy to monitor walking in adults. They observed that tailoring of intervention components over time to a person's evolving needs with rewards grounded in Social Cognitive Theory, in addition to collecting and analyzing wearable step data, showed statistically significant increase of walking behavior in overweight adults.
129. Merilahti J, Viramo P, and Korhonen I, Wearable Monitoring of Physical Functioning and Disability Changes, Circadian Rhythms and Sleep Patterns in Nursing Home Residents. *IEEE J Biomed Health Inform*, 2016 20(3): p. 856–864. [PubMed: 25861091]
- 130. Gjoreski M, et al., Monitoring stress with a wrist device using context. *J Biomed Inform*, 2017 73: p. 159–170. [PubMed: 28803947] Gjoreski et al. describe a novel context-based stress-detection method consisting of three machine-learning components with 70% accuracy for detection of stress events with a precision of 95%.
131. Penzel T and Conradt R, Computer based sleep recording and analysis. *Sleep Med Rev*, 2000 4(2): p. 131–148. [PubMed: 12531163]
132. Boostani R, Karimzadeh F, and Nami M, A comparative review on sleep stage classification methods in patients and healthy individuals. *Comput Methods Programs Biomed*, 2017 140: p. 77–91. [PubMed: 28254093]
133. Kales ARA, A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. 1968, Bethesda, Md.: U. S. National Institute of Neurological Diseases and Blindness, Neurological Information Network.
134. Berry RB, et al., Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine. *J Clin Sleep Med*, 2012 8(5): p. 597–619. [PubMed: 23066376]
135. Fraiwan L, et al., Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier. *Comput Methods Programs Biomed*, 2012 108(1): p. 10–9. [PubMed: 22178068]

136. Lajnef T, et al., Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines. *J Neurosci Methods*, 2015 250: p. 94–105. [PubMed: 25629798]
137. Beattie Z, et al., Estimation of sleep stages in a healthy adult population from optical plethysmography and accelerometer signals. *Physiol Meas*, 2017 38(11): p. 1968–1979. [PubMed: 29087960]
138. Montgomery-Downs HE, Insana SP, and Bond JA, Movement toward a novel activity monitoring device. *Sleep Breath*, 2012 16(3): p. 913–7. [PubMed: 21971963]
139. Castner J, et al., Validation of fitness tracker for sleep measures in women with asthma. *J Asthma*, 2018: p. 1–12.
140. Baroni A, et al., Fitbit Flex: an unreliable device for longitudinal sleep measures in a non-clinical population. *Sleep Breath*, 2016 20(2): p. 853–4. [PubMed: 26449552]
141. Prevention, C.f.D.C.a. About Underlying Cause of Death 1999–2015. CDC WONDER Database 2016 [cited 2018 11 15]; Available from: <http://wonder.cdc.gov/ucd-icd10.html>.
142. American Diabetes, A., Economic costs of diabetes in the U.S. in 2012. *Diabetes Care*, 2013 36(4): p. 1033–46. [PubMed: 23468086]
143. Chen C, et al., Current and Emerging Technology for Continuous Glucose Monitoring. *Sensors (Basel)*, 2017 17(1).
- 144. Kropff J, et al., Accuracy and Longevity of an Implantable Continuous Glucose Sensor in the PRECISE Study: A 180-Day, Prospective, Multicenter, Pivotal Trial. *Diabetes Care*, 2017 40(1): p. 63–68. [PubMed: 27815290] Kropff et al. studied the Eversense (Senseonics Inc.) implantable CGM sensor in 71 participants aged 18 years and older with type 1 and type 2 diabetes in a 180-day multinational, multicenter trial. The authors observed 81% percent of hypoglycemic events were detected by the CGM system within 30 min and no serious adverse events occurred during the study.
145. Knobbe EJ and Buckingham B, The extended Kalman filter for continuous glucose monitoring. *Diabetes Technol Ther*, 2005 7(1): p. 15–27. [PubMed: 15738701]
146. Bequette BW, Continuous glucose monitoring: real-time algorithms for calibration, filtering, and alarms. *J Diabetes Sci Technol*, 2010 4(2): p. 404–18. [PubMed: 20307402]
147. Hamdi TBAJ; Di Costanzo V; Fnaiecha F; Moreau E; Ginoux J, Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm. *Biocybernetics and Biomedical Engineering*, 2018 38(2): p. 362–372.
- 148. Mhaskar HN, Pereverzyev SV; van der Walt M, A Deep Learning Approach to Diabetic Blood Glucose Prediction. *Front. Appl. Math. Stat*, 2017 3(14). A deep learning algorithm is presented for 30-min prediction of blood glucose levels measured by continuous glucose monitoring devices with prediction accuracy of 96.43% in the hypoglycemic range, 97.96% in the euglycemic range, and 85.29% in the hyperglycemic range.
149. Cobelli C, Renard E, and Kovatchev B, Artificial pancreas: past, present, future. *Diabetes*, 2011 60(11): p. 2672–82. [PubMed: 22025773]
- 150. Bekiari E, et al., Artificial pancreas treatment for outpatients with type 1 diabetes: systematic review and meta-analysis. *BMJ*, 2018 361: p. k1310. [PubMed: 29669716] A comprehensive systematic review and meta-analysis of artificial pancreas treatment for patients with type 1 diabetes mellitus. The authors identified that artificial pancreas systems are an efficacious and safe approach for treating outpatients with type 1 diabetes, limitations in the research include inconsistency in outcome reporting, small sample size, and short follow-up duration of individual trials.
151. Sprint G, Cook DJ, and Schmitter-Edgecombe M, Unsupervised detection and analysis of changes in everyday physical activity data. *J Biomed Inform*, 2016 63: p. 54–65. [PubMed: 27471222]

Figure 1: Visual schema of wearable data, data analysis, and health domains.

Adapted from: Kamisalic, A., et al. (2018). "Sensors and Functionalities of Non-Invasive Wrist-Wearable Devices: A Review." *Sensors (Basel)* 18(6).

Abbreviations: Musculoskeletal (MSK); Heart Rate (HR); Blood Pressure (BP); Peripheral Capillary Oxygen Saturation (SpO2); Electrodermal Activity (EDA); Inertial Measurement Unit (IMU); Global Positioning System (GPS); Family History (Family Hx); Patient History (Patient Hx)

Figure 2: Visual Abstract of Concepts in Review.

Figure 3. The number of completed clinical trials on [ClinicalTrials.gov](https://clinicaltrials.gov) that have incorporated “wearables” into methods or research aims (total = 171), through 2018 (note: categories not mutually exclusive and some studies occupy multiple categories above).

Abbreviations: Ear, Nose, and Throat (ENT)

Table 1.

Accuracy of wearables and associated algorithms, grouped by fitness- and health-related indicators.

Health Domain	Sensors/ Devices	Analytical Methods	Cohort	Accuracy	Ref(s)
Step Counts	Triaxial Accelerometers (different devices and locations on body)	Spatiotemporal gait parameter detection	n = steps: 986 ± 127 (slow speed); 1127 ± 103 (selfselect speed); 1289 ± 115 (fast speed)	Lowest MAPE: 1.98 ± 1.50 - 0.93 ± 0.79 % (Movemonitor) Highest MAPE: 12.22 ± 7.04 - 35.39 ± 21.17 % (Nike + Fuelband)	[55]
	StepWatch (SW), activPAL, Fitbit Zip, Yamax Digi-Walker SW-200, New Lifestyles NL-2000, ActiGraph GT9X (AG), Fitbit Charge	Moving Average Vector, Low Frequency Extension	n= 12 adults (age: 35 ± 13yr)	Highest Accuracy: SW (95.3% - 102.8% of actual steps taken throughout the day (P>0.05))	[56]
	Yamax, Digi-Walker SW-200, Omron HJ-720, Fitbit Zip	Unknown device algorithm; chi-square test	n=14 participants (age: 29.93 ± 4.93 yr)	Error rates = 22.46-87.11% (Yamax); -0.63 - 95.96% (Omron); -0.70 - 95.98% (Fitbit)	[57]
	Garmin Vivofit2, Fitbit Flex, Up3, Microsoft Band	Unknown device algorithm; t-test, Bland-Altman concordance, linear regression	n= 30 older adults from a continuing care retirement community w/ assistive ambulatory devices	Accuracy (of < 20% error in actual steps) = 40% (Garmin), 40% (Fitbit), 40% (Microsoft band), 37% (Up3)	[58]
	Fitbit-Zip, ActiGraph	Unknown device algorithm; Bland-Altman concordance	n= 27 people with Polymyalgia Rheumatica	MAPE = 6 +/- 17% (walking) and 12 +/- 20% (stairs) (Fitbit, waist); 10 +/- 16% (walking) and 8 +/- 9% (stairs) (Actigraph)	[59]
	Fitbit Ultra, Omron Step Counter HJ-113, New Lifestyles 2000, Nike Fuelband	Unknown device algorithm; bivariate correlation analysis	n= 17 patients with confirmed idiopathic Normal Pressure Hydrocephalus	Correlation (ρ) = 0.72 (CI: 0.56, 0.89) (Fitbit); ρ = 0.19 (CI: - 0.14, 0.53) (Omron); ρ = -0.27 (CI: -0.56, 0.02) (New Lifestyles); ρ = -0.08 (CI: -0.29, 0.12) (Nike)	[60]
Activity Classification	Finis <i>Swimsense</i> , Garmin <i>Swim</i>	Unknown device algorithm; chi-square test, linear mixed effects model	n= 10 national level swimmers (5 male, 5 female; age: 15.3 ± 1.3y ears)	Identify swim strokes: Garmin: X ² (3) = 31.292 (p<0.05); Finis: X ² (3) = 33.004 (p<0.05)	[63]
	ActiGraph GT3X, Fitbit Flex	Freedson vector magnitude (VM3); linear discriminant analysis classifier, quadratic discriminant analysis classifier, naïve Bayesian classifier and a Mahalanobis distance based classifier	n= 19 subjects, average of 1.8 weeks	Fitbit Flex prediction success: 83.68%	[61]
	ActiGraph GT3X, Fitbit Zip	Freedson vector magnitude (VM3); Pearson's correlations, Bland-Altman concordance, linear regression	Three cohorts (n's = 25, 35, and 27) of middle-school students	MVPA: r = 0.67 (cohort 1); r = 0.79 (cohort 2), r = 0.94 (cohort 3)	[62]
Energy Expenditure	Smartphone-based triaxial accelerometry and barometer, Nike + fuelband, Fitbit (ANN)	Linear regression, Bagged Regression Trees, Artificial Neural Network	n=12 healthy subjects	Correlation (ρ) = 96% with gold-standard and RMSE = 0.70	[72]
	Suunto Ambit2, Garmin Forerunner920XT, Polar V800	Unknown device algorithms; ANOVA, Pearson's correlation, MAPE calculation	n=20 subjects	Polar V800 highest accuracy, MAPE = -12.20% (stage 1), -3.61% (stage 2), and -4.29% (stage 3)	[73]

Health Domain	Sensors/ Devices	Analytical Methods	Cohort	Accuracy	Ref(s)
Sleep Stage	Fitbit Surge (optical LED PPG and 3D accelerometer)	Linear discriminant classifiers, quadratic discriminant classifiers, random forests, SVM, Bland-Altman concordance	n= 60 adults (36 M, 24 F, age = 34 ± 10 , BMI = 28 ± 6)	Sensitivity (classifying sleep) = 94.6%; specificity (classifying wake state) = 69.3%; light sleep (69.2%); deep sleep (62.4%); REM sleep (71.6%)	[137]
	Fitbit, Actiwatch 64 (AW-64)	Unknown device algorithm; ANOVA, Bland-Altman concordance	n= 24 healthy adults	Fitbit: Sensitivity (Sleep) = 97.8% (CI: 92-100%); (stage N1) = 91.4% (CI: 63.2-100%); (stage N2) = 97.7% (92.8-100%); (stage N3) = 98.3% (77.9-100%); (REM) = 98.8% (85.4-100%); (Arousal epochs) = 97.5% (91.4-100%); Specificity (wake) = 19.8% (2.1-78.0%); (wake before sleep onset) = 42.9% (0-100%); (wake after sleep onset) = 11.8% (0-43.4%) Actiwatch: Sensitivity (Sleep) = 95.7% (CI: 90.2-99.6%); (stage N1) = 86.5% (CI: 52.9-100%); (stage N2) = 96.1% (88.7-100%); (stage N3) = 96.8% (81.7-100%); (REM) = 95.9% (85.4-100%); (Arousal epochs) = 86.1% (69.7-97.3%); Specificity (wake) = 38.9% (4.8-74.8%); (wake before sleep onset) = 35.8% (0-100%); (wake after sleep onset) = 37.5% (1.2-77.8%)	[138]
	Fitbit Flex	Unknown device algorithm	n= 107 healthy college students	Out of 7 nights of recording, 14 % of participants had successful recording (nights 6 and 7) and 35 % had no successful recordings at night.	[140]
	Fitbit Charge, Actigraph wGT3X+	Unknown device algorithm	n= 47 women with poorly controlled asthma (424,938 min utes, 738 nights, and 833 unique sleep segments)	Sensitivity = 97%, Specificity = 40% to identify sleep	[139]
Atrial Fibrillation Detection	Zio wearable patch-based device	Unknown device algorithm	n= 75 subjects with cardiovascular risk factors (all male, age 69 ± 8.0 years; ejection fraction 57% $\pm 8.7\%$)	Detected AF in 4 subjects (5.3%; AF burden 28% $\pm 48\%$); AT present in 67% (4 beats), 44% (8 beats), and 6.7% (60 seconds) of subjects.	[80]
	Kardia Band (AliveCor) with Apple smartwatch	Unknown device algorithm	n= 100 patients (age 68 $\pm$ 11 yrs)	93% sensitivity, 84% specificity, Kappa = 0.77; 57 (34%) patient-recordings were non-interpretable by the KB.	[95]
	Empatica E4 wristband (electrodermal activity sensor, an infrared thermopile, a 3-axis accelerometer and a PPG sensor which measures the BVP signal)	Spectral analysis (AR model), variability and irregularity analysis, shape analysis, k-NN	n= 70 subjects (n=30 AF, n=9 ARR, n=31 NSR)	NSR: Sensitivity= 0.773, Specificity = 0.928 AF: Sensitivity = 0.754, Specificity = 0.963 ARR: Sensitivity = 0.758, Specificity = 0.768	[85]
	iRhythm Zio skin patch	Unknown device algorithm	n= 2659 patients	Incidence of new AF cases = 3.9% (53/1366) in immediate	[98]

Health Domain	Sensors/ Devices	Analytical Methods	Cohort	Accuracy	Ref(s)
				monitoring group vs 0.9% (12/1293) in the delayed monitoring group (absolute difference, 3.0% [95% CI, 1.8%-4.1%])	
Autonomic Nervous System Pathology	Digital Holter monitor (SpiderView Plus), Guardian Real-Time Continuous Glucose Monitoring System (Medtronic MiniMed)	Unknown CGS algorithm; QRS detection (bandpass filter, differentiating, squaring, moving-window integration), t-test, linear regression	n= 21 adults with T1D	LF p = 0.029, no significant changes in other measured parameters (p>0.05)	[108]
	Thermistor (skin temperature sensor), skin resistance sensor, and ECG	ECG (bandpass Butterworth filter, Pan-Thomkins' algorithm); HRV (Fast Fourier Transform, Poincare Plot); SVM with gaussian vs. polynomial kernel	n= 7 adult subjects (28±7 years)	Using polynomial kernel (p = 5 and C=10): sensitivity=70.3%, specificity=80%, positive predictor=73%	[109]
	Adhesive patch sensor (Proteus Biomedical), wrist-worn Simband device (Samsung)	SVM, time series analysis (mutual information, transfer entropy), multiscale network representations	Cohort 1: n=16 patients with schizophrenia, n= 19 healthy controls Cohort 2: n=41 patients with AF and 53 controls	Schizophrenia cohort: AUC= 0.81 - 1.0 (depending on feature groups) AF cohort: AUC= 0.77-0.93 (depending on feature groups)	[110]
Fall Detection	Tri-axial devices (accelerometer, gyroscope, and magnetometer/compass)	k-NN, LSM, SVM, BDM, DTW, ANNs	n= 14 volunteers	k-NN classifier accuracy = 99.91%	[39]
	Accelerometer, magnetometer and gyroscope (unknown device)	Unknown device algorithm	n= 18 participants (only 8 completed)	Sensitivity =0.25, specificity =0.92	[119]
	Shimmer wireless sensor (tri-axial accelerometer)	ANN, Decision Tree, k-NN, Naïve Bayes, SVM	n= 8 older adults (76.50 ± 4.41 years)	Accuracy = 96.8% (ANN), 96.4% (Decision Tree), 96.2% (k-NN), 89.5% (Naive Bayes), 92.7% (SVM)	[114]
	Triaxial gyroscope, triaxial accelerometer, triaxial magnetometer, and pressure sensors	k-NN, BDM, SVM, LSM, DTW, ANNs	n= 14 adults	Accuracy = 99.87% with (sensor on waist and k-NN algorithm)	[116]
	Accelerometer from a single IMU (MPU-6050, InvenSense), FSR sensors (FSR 402 Short, Interlink Electronics)	SVM (Leave-One-Out cross-validation)	n= 20 male subjects	Sensitivity= 0.996, Specificity= 1.000, and Accuracy= 0.999	[118]
	Triaxial accelerometer (Dynaport MoveMonitor)	CNN, RNN-LSTM, ConvLSTM	n= 296 participants	AUC: 65 +/- 9% (subject level partitioning), 94 +/- 7% (sample level partitioning)	[115]
Detection of Behavior Change	Fitbit Charge, Fitbit Flex	Physical Activity Change Detection algorithms, Relative Unconstrained Least-Squares Importance Fitting, texture-based dissimilarity, Permutation-based Change Detection in Activity Routine, virtual binary classifiers	n= 11 older adults (Male = 3, Female = 8; age 57.09 ± 8.79 years) participating in a 10-week health intervention	Largest number of changes detected: Virtual Classifier = 51	[151]
	Fitbit Zip, Just Walk smartphone application	Linear mixed effects model	n= 20 overweight (mean BMI = 33.8 ± 6.82 kg/m ²), sedentary adults (90% female)	Mean increase in steps per day= 2650 (t = 8.25, p < 0.01)	[128]

Health Domain	Sensors/ Devices	Analytical Methods	Cohort	Accuracy	Ref(s)
	Wrist-worn device (IST Vivago WristCare) with an integrated activity sensor	Weighted rank order correlation, time series feature extraction (interdaily stability, interdaily variability, relative amplitude)	n=16 nursing home residents (average age of 90.7 years, seven demented subjects, one female) in their daily life over several months (12–18 months)	Significant correlation between functional status and actigraphic parameters: interdaily stability ($p < 0.05$) and circadian rhythm strength ($p < 0.05$)	[129]
	Wrist-worn device (IST Vivago WristCare) with an integrated activity sensor	Pearson's chi-squared test, Mann-Whitney's nonparametric U-test, Partial Spearman's rank order correlation, Poincare (return) plot analysis	n= 42 volunteers (aged 56–97 years; 23 demented and 19 non-demented)	Significant ($p < 0.01$) partial correlations were found between functional scores and night/day activity ratios ($r = 0.55$), normalized day activity ($r = 0.53$), normalized night activity ($r = 0.52$), sine amplitude ($r = 0.46$), Poincare differences between delays of 24 h and 12 h ($r = 0.42$), Poincare 1-min ($r = 0.39$), and 24-h activity ($r = 0.39$)	[123]
	Passive infrared (PIR) motion sensors	Co-occurrence matrix-based dissimilarity, fuzzy logic system, feature space distance measurement	n= 9 apartments inhabited by one older adult (each) in assisted living	Dissimilarity results (weighted normalized Euclidean distance) with respect to case #3 individual): 0.30–0.52.	[124]
	Apartments equipped with discrete (on/off) combination motion/light sensors on the ceilings and combination door/temperature sensors on cabinets and doors	Activity-curve modeling, Kullback-Leibler divergence measure, DTW, Permutation-based Change Detection in Activity Routine, Pearson rank correlation, Spearman rank correlation	n= 18 smart homes with older adult residents	Correlation: physical motor function ($p = 0.43$, $p < 0.005$); cognitive status (no significant correlation)	[125]
Serum Glucose and "Glucotype" Detection	Freestyle Navigator CGM System (Abbott Laboratories)	SVM, differential evolution algorithm,	n = 12 patients with type 1 diabetes	Average of RMSE = 9.44, 10.78, 11.82 and 12.95 (mg/dL) for prediction horizons, respectively, of 15, 30, 45 and 60 mins	[147]
	CGM (unknown device)	CNN	n= 25 patients with type 1 diabetes (all ages <18 years)	or 30% training data, the Accuracy = 93.21% (hypoglycemic), 97.68% (euglycemic), 86.78% (hyperglycemic)	[148]
	Dexcom G4 CGM device	Complexity-invariant DTW, DTW, ANOVA, Spearman's correlation	n= 57 healthy participants (32 female; 25 male; median age 51 yr)	Low variability (glucotype L), moderate variability (glucotype M), and severe variability (glucotype S); Significant correlation: ($P < 0.05$) between the time spent in low glycemic signatures and lower values for fasting glucose, HbA1c, OGTT, SSPG, BMI, and age	[47]

Health Term Abbreviations—MVPA: moderate to vigorous physical activity; REM: rapid eye movement; BMI: body mass index; NSR: normal sinus rhythm; AF: atrial fibrillation; ARR: arrhythmia; ECG: electrocardiogram; SSPG: steady-state plasma glucose; OGTT: Oral Glucose Tolerance Test; HbA1c: Hemoglobin A1c

Methods/M Measurement Term Abbreviations—MAPE: Mean Absolute Percentage Error; RMSE: Root Mean Square Error; AUC: Area Under (receiver operating characteristic) Curve; ANOVA: Analysis of Variance; SVM: support vector machine; k-NN: k-Nearest Neighbor; LSM: Least Squares Method; BDM: Bayesian Decision Making; DTW: Dynamic Time Warping; ANN: Artificial Neural Network; CNN: Convolutional Neural Network; RNN-LSTM: Recurrent Neural Network (long short-term memory); Conv-LSTM: Mixed CNN-RNN (long short-term memory)

Table 2:

Metabolic equivalent (MET) chart with example activities. Adapted from (<https://www.hsph.harvard.edu/nutritionsource/mets-activity-table/>)

Light <3.0 METs	Moderate 3.0–6.0 METs	Vigorous >6.0 METs
Sitting Walking—slowly Standing—light work (cooking, washing dishes)	Walking—very brisk (4 mph) Cleaning—heavy (washing windows, vacuuming, mopping) Mowing lawn (walking power mower) Bicycling—light effort (10–12 mph)	Walking/hiking Jogging at 6 mph Shoveling snow Carrying heavy loads Bicycling fast (14–16 mph) Playing basketball, soccer, tennis games