Windows into human health through wearables data analytics
Daniel R. Witt\textsuperscript{1,4}, Ryan A. Kellogg\textsuperscript{2}, Michael P. Snyder\textsuperscript{2} and Jessilyn Dunn\textsuperscript{2,3,4}

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), electrocardiogram (ECG) algorithm (8287), and 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 and 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 articles 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; and diabetes mellitus type 2. All studies uncovered through this search on commercially available device algorithms and 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.

Addresses

1 Mayo Clinic Alix School of Medicine, Mayo Clinic, Rochester, MN, USA
2 Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA
3 Department of Biomedical Engineering, Duke University, Durham, NC, USA
4 Department of Biostatistics & Bioinformatics, Duke University, Durham, NC, USA

Corresponding author: Dunn, Jessilyn (jessilyn@stanford.edu)

Introduction: mapping the digitome

The use of wearable biosensors, referred to as ‘wearables’, has been encompassed in the broader interdisciplinary health services effort to use mHealth to enhance data collection, diagnoses, treatment monitoring, and health insights [1,2]. Although definitions vary, the World Health Organization’s Global Observatory for eHealth defined mHealth as the “medical and public health practice supported by mobile devices, such as mobile phones, patient-monitoring devices, personal digital assistants, 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 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 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 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, laboratory and imaging data, and health survey information to improve understanding of disease trajectory, population, and personalized characteristics of disease and to 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 United States involving wearables (Figure 3) [17].

Wearable data: algorithms and analytic methods

Data preprocessing

Given the array of sensing units in many wearables, data preprocessing often must occur before initiating the highest order analytical goals. These concerns broadly fall under the domain of signal processing and other possible data preprocessing steps such as labeling, filtering, segmentation, feature extraction, and feature selection [18]. In a discussion of wearable biosensing, Celka et al. [19] 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 sublayers: feature extraction/selection and classification. Others have noted that, although 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, 20a].

**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 health-care 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 use ML algorithms with clinical data (e.g., age, gender, physical examination findings, symptoms, vital signs, laboratory values, imaging variables, test result values) to predict clinical outcomes or to identify relationships between predictor variables, known as features, and clinical outcomes. Although 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 use (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 [SVM], 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, for a wide variety of 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 used ML methods. Fall detection research, based on wearable-collected data, has produced optimized detection of falls using SVMs [34], Naïve Bayes [35], and artificial neural network [36] approaches, among others [37–39]. Others have explored detection of 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 using ML algorithms [43–45]. Other studies have focused on detecting different phenotypes of characteristic glucose regulation or dysregulation in nondiabetics 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 nonparametric 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 with respect to HR, activity, and body mass index [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, although 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 United States 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
<table>
<thead>
<tr>
<th>Health domain</th>
<th>Sensors/devices</th>
<th>Analytical methods</th>
<th>Cohort</th>
<th>Accuracy</th>
<th>Ref(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step counts</td>
<td>Triaxial accelerometers (different devices and locations on body)</td>
<td>Spatiotemporal gait parameter detection</td>
<td>n = steps: 986±127 (slow speed); 1127 ± 103 (self-select speed); 1289 ± 115 (fast speed)</td>
<td>Lowest MAPE: 1.98 ± 1.50–0.93 ± 0.79% (Movemonitor)</td>
<td>[55]</td>
</tr>
<tr>
<td></td>
<td>StepWatch (SW), activPAL, Fitbit Zip, Yamax Digi-Walker SW-200, New Lifestyles NL-2000, ActiGraph GT9X (AG), Fitbit Charge</td>
<td>Moving average vector, low frequency extension</td>
<td>n = 12 adults (age: 35 ± 13 yr)</td>
<td>Highest accuracy: SW (95.3%–102.8% of actual steps taken throughout the day (P &gt; 0.05))</td>
<td>[56]</td>
</tr>
<tr>
<td></td>
<td>Yamax, Digi-Walker SW-200, Omron HJ-720, Fitbit Zip</td>
<td>Unknown device algorithm; chi-square test</td>
<td>n = 14 participants (age: 29.93 ± 4.93 yr)</td>
<td>Error rates = 22.46–87.11% (Yamax); −0.63–95.96% (Omron), −0.70–95.98% (Fitbit)</td>
<td>[57]</td>
</tr>
<tr>
<td></td>
<td>Garmin Vivofit2, Fitbit Flex, Upch, Microsoft Band</td>
<td>Unknown device algorithm; t-test, Bland-Altman concordance, linear regression</td>
<td>n = 30 older adults from a continuing care retirement community w/assistive ambulatory devices</td>
<td>Accuracy (of &lt;20% error in actual steps) = 40% (Garmin), 40% (Fitbit), 40% (Microsoft band), 37% (Upch)</td>
<td>[58]</td>
</tr>
<tr>
<td></td>
<td>Fitbit Ultra, Omron Step Counter HJ-113, New Lifestyles 2000, Nike Fuelband</td>
<td>Unknown device algorithm; Bland–Altman concordance</td>
<td>n = 17 patients with confirmed idiopathic normal pressure hydrocephalus</td>
<td>MAPE = 6 ± 17% (walking) and 12 ± 20% (stairs) (Fitbit, waist); 10 ± 16% (walking) and 8 ± 9% (stairs) (Actigraph) 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); p = −0.08 (CI: −0.29, 0.12) (Nike)</td>
<td>[59]</td>
</tr>
<tr>
<td>Activity classification</td>
<td>Fitbit Ultra, Fitbit Zip, ActiGraph</td>
<td>Unknown device algorithm; Bland–Altman concordance</td>
<td>n = 27 people with polymyalgia rheumatica</td>
<td>Accuracy of &lt;20% error in actual steps) = 40% (Garmin), 40% (Fitbit), 40% (Microsoft band), 37% (Upch)</td>
<td>[60]</td>
</tr>
<tr>
<td></td>
<td>Finis Swimsense, Garmin Swim</td>
<td>Unknown device algorithm; chi-square test, linear mixed effects model</td>
<td>n = 10 national level swimmers (5 men, 5 women; age: 15.3 ± 1.3y ears)</td>
<td>Identify swim strokes: Garmin: X²(3) = 31.292 (p &lt; 0.05); Finis: X²(3) = 33.004 (p &lt; 0.05)</td>
<td>[63]</td>
</tr>
<tr>
<td></td>
<td>ActiGraph GT3X, Fitbit Flex</td>
<td>Freedson vector magnitude (VM3); Pearson’s correlation, Bland–Altman concordance, linear regression classifier, quadratic discriminant analysis classifier, Naïve Bayesian classifier and a Mahalanobis distance–based classifier</td>
<td>n = 19 subjects, average of 1.8 weeks</td>
<td>Fitbit Flex prediction success: 83.68%</td>
<td>[61]</td>
</tr>
<tr>
<td>Energy expenditure</td>
<td>ActiGraph GT3X, Fitbit Zip</td>
<td>Freedson vector magnitude (VM3); Pearson’s correlations, Bland–Altman concordance, linear regression</td>
<td>Three cohorts (n’s = 25, 35, and 27) of middle-school students</td>
<td>MVPA: r = 0.67 (cohort 1); r = 0.79 (cohort 2), r = 0.94 (cohort 3)</td>
<td>[62]</td>
</tr>
<tr>
<td></td>
<td>Smartphone-based triaxial accelerometry and barometer, Nike + fuelband, Fitbit</td>
<td>Linear regression, Bagged regression trees, Artificial Neural Network (ANN)</td>
<td>n = 12 healthy subjects</td>
<td>Correlation (ρ) = 96% with gold standard and MSE = 0.70</td>
<td>[72]</td>
</tr>
<tr>
<td></td>
<td>Suunto Ambit2, Garmin Forerunner920XT, Polar V800</td>
<td>Unknown device algorithms; ANOVA, Pearson’s correlation, MAPE calculation</td>
<td>n = 20 subjects</td>
<td>Polar V800 highest accuracy, MAPE = −12.20% (stage 1), −3.61% (stage 2), and −4.29% (stage 3)</td>
<td>[73]</td>
</tr>
<tr>
<td>Sleep stage</td>
<td>Method</td>
<td>n</td>
<td>Data Description</td>
<td>Sensitivity (Sleep)</td>
<td>Specificity (Sleep)</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>----</td>
<td>----------------------------------------------------------------------------------</td>
<td>---------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td></td>
<td>Fitbit Surge (optical LED PPG and 3D accelerometer)</td>
<td>60</td>
<td>Adults (36 M, 24 F, age = 34 ± 10, BMI = 28 ± 6)</td>
<td>94.6%</td>
<td>69.3%</td>
</tr>
<tr>
<td></td>
<td>Linear discriminant classifiers, quadratic discriminant classifiers,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>random forests, SVM, Bland–Altman concordance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fitbit, Actiwatch 64 (AW-64)</td>
<td>24</td>
<td>Healthy adults</td>
<td>97.8%</td>
<td>69.6%</td>
</tr>
<tr>
<td></td>
<td>Bland–Altman concordance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown device algorithm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ANOVA, Bland–Altman concordance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fitbit Flex</td>
<td>107</td>
<td>Healthy college students</td>
<td>97%</td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td>Unknown device algorithm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation detection</td>
<td>Zio wearable patch-based device</td>
<td>75</td>
<td>Subjects with cardiovascular risk factors (all male, age 69 ± 8.0 years; ejection fraction 57% ± 8.7%)</td>
<td>5.3%; 28% ± 48%</td>
<td>44%</td>
</tr>
<tr>
<td></td>
<td>Spectral analysis (AR model), variability and irregularity analysis,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>shape analysis, k-NN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kardia Band (AliveCor) with Apple smartwatch</td>
<td>100</td>
<td>Patients (age 68 ± 11 yrs)</td>
<td>93%</td>
<td>84%</td>
</tr>
<tr>
<td></td>
<td>Empatica E4 wristband (electrodermal activity sensor, an infrared</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>thermopile, a 3-axis accelerometer and a PPG sensor which measures the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BVP signal)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spectral analysis (AR model), variability and irregularity analysis,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>shape analysis, k-NN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Empatica E4 wristband (electrodermal activity sensor, an infrared</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>thermopile, a 3-axis accelerometer and a PPG sensor which measures the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BVP signal)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spectral analysis (AR model), variability and irregularity analysis,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>shape analysis, k-NN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kardia Band (AliveCor) with Apple smartwatch</td>
<td>100</td>
<td>Patients (age 68 ± 11 yrs)</td>
<td>93%</td>
<td>84%</td>
</tr>
<tr>
<td></td>
<td>Empatica E4 wristband (electrodermal activity sensor, an infrared</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>thermopile, a 3-axis accelerometer and a PPG sensor which measures the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BVP signal)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spectral analysis (AR model), variability and irregularity analysis,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>shape analysis, k-NN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Empatica E4 wristband (electrodermal activity sensor, an infrared</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>thermopile, a 3-axis accelerometer and a PPG sensor which measures the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BVP signal)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spectral analysis (AR model), variability and irregularity analysis,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>shape analysis, k-NN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Empatica E4 wristband (electrodermal activity sensor, an infrared</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>thermopile, a 3-axis accelerometer and a PPG sensor which measures the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autonomic nervous system pathology</td>
<td>Digital Holter monitor (SpiderView Plus), Guardian Real-Time continuous</td>
<td>21</td>
<td>Adults with T1D</td>
<td>90.2%</td>
<td>3.9%</td>
</tr>
<tr>
<td></td>
<td>QRIS detection (bandpass filter, differentiating, squaring, moving-window)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued on next page)
<table>
<thead>
<tr>
<th>Health domain</th>
<th>Sensors/devices</th>
<th>Analytical methods</th>
<th>Cohort</th>
<th>Accuracy</th>
<th>Ref(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood glucose monitoring system (Medtronic MiniMed)</td>
<td>Thermistor (skin temperature sensor), skin resistance sensor, and ECG</td>
<td>ECG (bandpass Butterworth filter, Pan–Thomkins' algorithm); HRV (Fast Fourier transform, Poincare Plot); SVM with Gaussian vs. polynomial kernel</td>
<td>n = 7 adult subjects (28 ± 7 years)</td>
<td>Using polynomial kernal (p = 5 and C = 10): sensitivity = 70.3%, specificity = 80%, positive predictor = 73%</td>
<td>[109]</td>
</tr>
<tr>
<td>Adhesive patch sensor (Proteus Biomedical), wrist-worn Simband device (Samsung)</td>
<td>SVM, time series analysis (mutual information, transfer entropy), multiscale network representation</td>
<td>Cohort 1: n = 16 patients with schizophrenia, n = 19 healthy controls Cohort 2: n = 41 patients with AF and 53 controls</td>
<td>Schizophrenia cohort: AUC = 0.81–1.0 (depending on feature groups) AF cohort: AUC = 0.77–0.93 (depending on feature groups)</td>
<td></td>
<td>[110]</td>
</tr>
<tr>
<td>Fall detection</td>
<td>Triaxial devices (accelerometer, gyroscope, and magnetometer/compass)</td>
<td>k-NN, LSM, SVM, BDM, DTW, ANNs</td>
<td>n = 14 volunteers</td>
<td>k-NN classifier accuracy = 99.91%</td>
<td>[39]</td>
</tr>
<tr>
<td></td>
<td>Accelerometer, magnetometer, and gyroscope (unknown device)</td>
<td>Unknown device algorithm</td>
<td>n = 18 participants (only 8 completed)</td>
<td>Sensitivity = 0.25, specificity = 0.92</td>
<td>[119]</td>
</tr>
<tr>
<td></td>
<td>Shimmer wireless sensor (triaxial accelerometer)</td>
<td>ANN, decision tree, k-NN, Naive Bayes, SVM k-NN, BDM, SVM, LSM, DTW, ANNs</td>
<td>n = 8 older adults (76.50 ± 4.41 years)</td>
<td>Accuracy = 96.8% (ANN), 96.4% (decision tree), 96.2% (k-NN), 89.5% (Naive Bayes), 92.7% (SVM) Accuracy = 99.87% with (sensor on waist and k-NN algorithm)</td>
<td>[114]</td>
</tr>
<tr>
<td></td>
<td>Triaxial gyroscope, triaxial accelerometer, triaxial magnetometer, and pressure sensors</td>
<td>k-NN, BDM, SVM, DTW, ANNs</td>
<td>n = 14 adults</td>
<td></td>
<td>[116]</td>
</tr>
<tr>
<td></td>
<td>Accelerometer from a single IMU (MPU-6050, InvenSense), FSR sensors (FSR 402 Short, Interlink Electronics)</td>
<td>SVM (leave-one-out cross-validation)</td>
<td>n = 20 male subjects</td>
<td>Sensitivity = 0.996, Specificity = 1.000, and Accuracy = 0.999</td>
<td>[118]</td>
</tr>
<tr>
<td>Detection of behavior change</td>
<td>Triaxial accelerometer from Dynaport MoveMonitor</td>
<td>CNN, RNN-LSTM, ConvLSTM</td>
<td>n = 296 participants</td>
<td>AUC: 65 ± 9% (subject-level partitioning), 94 ± 7% (sample level partitioning) Largest number of changes detected: virtual classifier = 51</td>
<td>[115]</td>
</tr>
<tr>
<td></td>
<td>Fitbit Charge, Fitbit Flex</td>
<td>Physical activity change detection algorithms, relative unconstrained least-squares importance fitting, texture-based dissimilarity, permutation-based change detection in activity routine, virtual binary classifiers</td>
<td>n = 11 older adults (Male = 3, Female = 8; age 57.09 ± 8.79 years) participating in a 10-week health intervention</td>
<td></td>
<td>[151]</td>
</tr>
<tr>
<td></td>
<td>Fitbit Zip, Just Walk smartphone application</td>
<td>Linear mixed effects model</td>
<td>n = 20 overweight (mean BMI = 33.8 ± 6.82 kg/m2), sedentary adults (90% female)</td>
<td>Mean increase in steps per day = 2650 (t = 8.25, p &lt; 0.01)</td>
<td>[128]</td>
</tr>
<tr>
<td>Device and Methodology</td>
<td>Data Collection and Analysis</td>
<td>Sample Size</td>
<td>Significant Findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wrist-worn device (IST Vivago WristCare) with an integrated activity sensor</td>
<td>Weighted rank order correlation, time series feature extraction (interdaily stability, interdaily variability, relative amplitude)</td>
<td>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)</td>
<td>Significant correlation between functional status and actigraphic parameters: interdaily stability ($p &lt; 0.05$) and circadian rhythm strength ($p &lt; 0.05$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wrist-worn device (IST Vivago WristCare) with an integrated activity sensor</td>
<td>Pearson’s chi-squared test, Mann–Whitney’s nonparametric U-test, Partial Spearman’s rank order correlation, Poincare (return) plot analysis</td>
<td>n = 42 volunteers (aged 56–97 years; 23 demented and 19 non demented)</td>
<td>Significant ($p &lt; 0.01$) 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 ($r = 0.39$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passive infrared (PIR) motion sensors</td>
<td>Co-occurrence matrix–based dissimilarity, fuzzy logic system, feature space distance measurement</td>
<td>n = 9 apartments inhabited by one older adult (each) in assisted living</td>
<td>Dissimilarity results (weighted normalized Euclidean distance) with respect to case #3 individual: 0.30–0.52.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apartments equipped with discrete (on/off) combination motion/light sensors on the ceilings and combination door/temperature sensors on cabinets and doors</td>
<td>Activity-curve modeling, Kullback–Leibler divergence measure, DTW, permutation-based change detection in activity routine, Pearson rank correlation, Spearman rank correlation</td>
<td>n = 18 smart homes with older adult residents</td>
<td>Correlation: physical motor function ($p = 0.43, p &lt; 0.005$); cognitive status (no significant correlation)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freestyle Navigator CGM System (Abbott Laboratories)</td>
<td>SVM, differential evolution algorithm,</td>
<td>n = 12 patients with type 1 diabetes</td>
<td>Average of RMSE = 9.44, 10.78, 11.82 and 12.95 (mg/dL) for prediction horizons, respectively, of 15, 30, 45 and 60 min</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGM (unknown device)</td>
<td>CNN</td>
<td>n = 25 patients with type 1 diabetes (all ages &lt;18 years)</td>
<td>or 30% training data, the accuracy = 93.21% (hypoglycemic), 97.68% (euglycemic), 86.78% (hyperglycemic)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dexcom G4 CGM device</td>
<td>Complexity- invariant DTW, DTW, ANOVA, Spearman’s correlation</td>
<td>n = 57 healthy participants (32 female; 25 male; median age 51 yr)</td>
<td>Low variability (glucotype L), moderate variability (glucotype M), and severe variability (glucotype S); significant correlation ($p &lt; 0.05$) between the time spent in low glycemic signatures and lower values for fasting glucose, HbA1c, OGTT, SSPG, BMI, and age</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CGM, continuous glucose monitor; 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; 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: Convolved Neural Network; RNN-LSTM: Recurrent Neural Network (long short-term memory); Conv-LSTM: Mixed CNN-RNN (long short-term memory).
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 overestimated distance at slower speeds and underestimated 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 before 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, 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, whereas 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. Gagliani 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 and Chau evaluated the Fitbit Zip daily step counts and minutes of MVPA.

Against the ActiGraph and found that the Zip overestimated step counts, and this overestimation was higher before Fitbit’s algorithm update calculating MPVA only from activity bouts >10 min [62]. Winfree and Dominick improved the Fitbit Flex classification of MVPA with a naïve 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 because of individual variances in stroke technique [63].

Energy expenditure estimation

Many wearables calculate estimated EE through algorithms that consider resting energy expenditure and activity energy expenditure. Gold standards for estimated EE include doubly labeled water technique and indirect calorimetry, based on measurements of oxygen consumption (\(\text{VO}_2\)) and carbon dioxide production. Estimates of resting energy expenditure (similar to basal metabolic rate) and EE often use the Mifflin [64], McArdle [65], or Cunningham [66] formulas to estimate based on sex, age, height, and weight inputs, although 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 wearable-derived activity data. One MET is defined as consumption of oxygen required at rest; actual energy cost will vary between individuals because of differences in body mass, adiposity, age, gender, and environmental conditions, and is thus calculated as 1 MET = \(\sim 3.5 \text{ mL/O}_2/\text{min/ kg} \times \text{min} \) [68]. Various activities have been compared on relative oxygen uptake or intensity of activities as multiples of one MET. Although METs are a common standard for approximating EE for researchers and clinicians, their use in estimation of activity energy expenditure has been shown to be inaccurate across individuals of different body mass and body fat categories [69]. Wearables are often convenient, relatively

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Metabolic equivalent chart with example activities.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light</td>
<td>Moderate</td>
</tr>
<tr>
<td>&lt;3.0 METs</td>
<td>3.0–6.0 METs</td>
</tr>
<tr>
<td>Sitting Walking—slowly</td>
<td>Walking—very brisk (4 mph)</td>
</tr>
<tr>
<td>Standing—light work (cooking, washing dishes)</td>
<td>Cleaning—heavy (washing windows, vacuuming, mopping)</td>
</tr>
<tr>
<td>Mowing lawn (walking power mower)</td>
<td>Bicycling—light effort (10–12 mph)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MET: metabolic equivalent. Adapted from (https://www.hsph.harvard.edu/nutritionsource/mets-activity-table/).
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 with 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 six devices (Apple Watch 2, Samsung Gear S3, Jawbone Up3, Fitbit Surge, Huawei Talk Band B3, and Xiaomi Mi Band 2) and two smartphone apps (Dongdong and Ledongli) in 44 healthy participants; the authors measured five major 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, 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 men, 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 pathophysiologica phenomena. While a large portion of clinical care relies on the use of patient-specific health data (e.g., history and physical examination, laboratory 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.

AF classification
AF is the most common clinically significant cardiac arrhythmia in the United States, with economic costs estimated to be $16–26 billion in annual US health-care 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, 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 and photodiodes. Raw data from PPG sensors often must be processed with beat-detection algorithms which often use preprocessing 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 interbeat intervals (IBI) of the time series containing the extracted candidate peaks [81].

PPG has been used to detect AF [82]. Bonomi et al. [83] 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 ± 2% sensitivity and 99 ± 3% specificity for AF detection against a Holter ECG monitor standard. 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, Shannon entropy, and Sample entropy, 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, AF, and other arrhythmias reached an accuracy of 0.9, 0.9, and 0.8, for respective diagnostic groups [85].

Additional use of ECG devices (e.g., 24-h 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 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. 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 Fibricheck 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, of 169 simultaneous ECG and KB recordings, 57 were noninterpretable 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 the KB interpreted, the authors observed that the 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. As per the research by Hernandez-Silveira et al., a digital patch device compared with clinical monitoring of HR and respiratory rate had statistically significant (p < 0.0001) correlation, with lower correlation between respiratory rate in patients with AF (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 cardiovascular disease (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 carried out 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 and soft attention mechanism on top of the bidirectional recurrent neural network to achieve accurate detection of paroxysmal AF with area under the curve (AUC) = 0.94. **Autonomic nervous system function and dysfunction**

Heart rate reflects the average number of cardiac cycles per minute, whereas 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 HR characteristics. Parasympathetic input is mediated primarily by the vagus nerve. Respiration affects HR, with increasing HR upon inhalation and decreasing HR with exhalation [99]. Additional factors that impact HR 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 (DM) were monitored with continuous glucose monitors (CGMs) 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 HRV parameter was significantly (P = 0,029) reduced during hypoglycemia for all patients, but surprisingly low frequency 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

---

**Windows into human health through wearables data analytics** Witt et al. 39
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, although reduced AUC was observed in distinguishing nonpsychiatric patients from controls [110].

**Fall detection**

Every year in the United States, an estimated one-third of adults aged 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 wireless motion tracker (MTw) sensor units (triaxial devices with accelerometer, gyroscope, and magnetometer/compass) and compared six ML classifier techniques (k-NN, least squares method, SVM, Bayesian decision making, dynamic time warping, and artificial neural networks); considering performance and computational complexity, the k-NN classifier and least squares method were superior with sensitivity, specificity, and accuracy all greater than 99% [39]. Gao et al. conducted a comparison of single versus multisensor 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., 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 with 97.37% sensitivity for sensor placement on the wrist [116]. Attal et al. 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 with 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. 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 seven 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 health-care team.

A large collection of research has been conducted on wearable-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 used spectral signal analysis of acoustic data, linear predictive coefficients, feature extraction (e.g., Mel-frequency cepstral coefficients, autoregressive 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. used 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. 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 pseudorandom 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 the 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 carried out in automated psychophysiological stress detection and changes over time. Gjoreski et al. provided a good review of research on wearable-based stress detection and presented a method for automatically measuring subjective stress using Empatica E3 and E4 wrist devices [130]. Their method used features extracted from PPG data (HR and IBI features), electrodermal activity, 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: electroencephalogram (EEG), airflow through the nose and mouth, respiratory rate, BP changes, blood oxygen level, electrooculography signals, and chin and legs surface electromyography [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 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 as per traditional Rechtschaffen and Kales criteria [133], and more recently, American Academy of Sleep Medicine standards [134]. EEG has had the most clinical utility, and sleep epochs are categorized as delta, theta, alpha, beta bands and the presence or absence of sleep spindles and K-complexes [132]. The American Academy of Sleep Medicine manual defines the following sleep stages: awake state (stage W); stage N1; stage N2; stage N3; and rapid eye movement [134]. Owing 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. 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 with 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. observed that the Fitbit Charge overestimated sleep efficiency and underestimated wake counts compared with 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 the baseline; only 14% of Fitbit Flexes recorded sleep after five 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 nonwearable (e.g., remote monitoring hardware under a mattress, etc.) approaches to sleep quality detection.
Serum glucose control and diabetes management

DM (types 1 and 2) was the seventh leading cause of death in the United States and has an associated direct and indirect cost $245 billion in the United States [141,142]. Given the need for diagnosis and effective management of serum glucose levels, CGM devices have emerged as a promising clinical tools for management of both DM type 1, DM type 2, and prediabetic phenotypes. Industry leaders in the field include Dexcom, Abbott, and Medtronic whose CGM devices all use 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 nanomols) to a blood glucose estimate (in the United States, 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]), hypoglycemic and hyperglycemic 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 min of glucose changes with mean squared error 12.95 mg/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 noninvasive 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 prediabetics or nondiabetics for identifying risk factors and trajectories of developing diabetes. Individuals classified as nondiabetic 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 nondiabetics for dietary and lifestyle management, in addition to more fine-grained clinical understanding of intrapatient and interpatient 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 ML 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.

Acknowledgements

This work was funded by NIH U54DK10255603 and R01DK11018602.

Conflict of interest statement

Nothing declared.

References

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

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.


20. 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.


A thorough, 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.


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.


A comprehensive study comparing the seven commercially available activity monitors (Fitbit Charge, Fitbit Zip and ActiGraph) and measuring accuracy of seven physical activity monitors. Cureus 2018, 10:e118723.


A comprehensive study comparing the seven commercially available activity monitors using CGM, with glucose levels reaching prediabetic and diabetic ranges 15% and 2% of the time.

Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. JMIR Mhealth Uhealth 2018, 6:e94.


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.


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 interpretable by the Kardia device, indicating further algorithm development to accommodate clinical data.


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.


Reinertsen et al. used wearable-derived data to quantify and measure interactions between heart rate and locomotor activity in patients 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.


Gjoreski et al. describe a novel context-based stress-detection method.

Sensors (Basel) 2016, 16.


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.


Kropff et al. studied the Eversense (Sensornics 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.


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.


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.