Smartphone- and Smartwatch-Based Remote Characterisation of Ambulation in Multiple Sclerosis during the Two-Minute Walk Test

A.P. Creagh*[1,2], C. Simillion[2], A.K. Bourke[2], A. Scotland[2], F. Lipsmeier[2], C. Bernasconi[2], J. van Beek[2], M. Baker[2], C. Gossens[2], M. Lindemann†[2] and M. De Vos†[1,3,4]

I. Supplementary Material

Smartphone and smartwatch devices contain 3-axis accelerometer $(\boldsymbol{a}_x, \boldsymbol{a}_y, \boldsymbol{a}_z)$ and gyroscope $(\boldsymbol{g}_x, \boldsymbol{g}_y, \boldsymbol{g}_z)$ sensors. The orientation invariant signal magnitude is defined for example by $\|\boldsymbol{a}\| = (\mathbf{x}^2 + \mathbf{y}^2 + \mathbf{z}^2)^{\frac{1}{2}}$, where $\mathbf{x} = (a_{x_1}, a_{x_2}, ..., a_{x_T})$, $\mathbf{y} = (a_{y_1}, a_{y_2}, ..., a_{y_T})$ and $\mathbf{z} = (a_{z_1}, a_{z_2}, ..., a_{z_T})$ for signals of length T.

Table I: Mathematical description of the features extracted from the remote Two-Minute Walk Test

Feature	Description	Parameterization
$steps(\|\boldsymbol{a}\|)$	The number of steps counted, as per [1].	See [1].
$mean(\boldsymbol{a}_x)$ $mean(\boldsymbol{a}_y)$ $mean(\boldsymbol{a}_y)$ $mean(\|\boldsymbol{a}\|)$	The mean value of vector $\boldsymbol{x} \in \mathbb{R}^{1\times N}$, made up of N scalar observations: $$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (1)$$	
$var(\boldsymbol{a}_x)$ $var(\boldsymbol{a}_y)$ $var(\boldsymbol{a}_z)$ $var(\|\boldsymbol{a}\|)$	The variance which characterises the spread of a set of (random) numbers from their average value. For a random variable vector $\boldsymbol{x} \in \mathbb{R}^{1\times N}$ made up of N scalar observations, the variance is defined as: $$var = \frac{1}{N-1}\sum_{i=1}^{N} \|x_i - \overline{x}\|^2 \quad (2)$$ where $\overline{x}$ is the mean of $\boldsymbol{x}$ (1).	
$std(\boldsymbol{a}_x)$ $std(\boldsymbol{a}_y)$ $std(\boldsymbol{a}_z)$ $std(\|\boldsymbol{a}\|)$	For a random variable vector $\boldsymbol{x}$ made up of N scalar observations, the standard deviation is defined as: $$std = \left(\frac{1}{N-1}\sum_{i=1}^{N} \|x_i - \overline{x}\|^2\right)^{1/2} \quad (3)$$ where $\overline{x}$ is the mean of $\boldsymbol{x}$ (1).	

[1]Department of Biomedical Engineering, University of Oxford, UK;
[2]F. Hoffmann-La Roche Ltd, Basel, CH;
[3]Department of Electrical Engineering, KU Leuven, BE;
[4]Department of Development and Regeneration, KU Leuven, BE;
*Corresponding author e-mail: (andrew.creagh@eng.ox.ac.uk);
†Shared last authorship.

Table I: Mathematical description of the features extracted from the remote Two-Minute Walk Test

Feature	Description	Parameterization
$skew(\boldsymbol{a}_x)$ $skew(\boldsymbol{a}_y)$ $skew(\boldsymbol{a}_y)$ $skew(\|\boldsymbol{a}\|)$	Skewness as a measure of the asymmetry of the probability distribution of the signal values $\boldsymbol{x} \in \mathbb{R}^{1\times N}$. The relative magnitude of how far a distribution deviates from the normal which used as a proxy for smooth stable ambulatory movements. Positive values indicate that the mass of the distribution is predominantly skewed to the left, while the inverse is true for negative values. Skewness is defined as [2]: $$skew = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}{\left[\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2\right]^{3/2}} \quad (4)$$ where x_i denotes each i^{th} sample in $\boldsymbol{x}$ and $\overline{x}$ denotes the mean of the samples in $\boldsymbol{x}$.	
$kurt(\boldsymbol{a}_x)$ $kurt(\boldsymbol{a}_y)$ $kurt(\boldsymbol{a}_y)$ $kurt(\|\boldsymbol{a}\|)$	Kurtosis measures the level of 'taildness' in the probability distribution of the magnitude of $\boldsymbol{x} \in \mathbb{R}^{1\times N}$ signal values, compared to the normal distribution, and is used as a measure of how outlier prone a distribution is. Values > 3 indicate tails that exponentially decay at a slower rate to the normal distribution (more outliers present) and vice-versa. Kurtosis is defined by [2]: $$kurt = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^4}{\left[\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2\right]^{2}} \quad (5)$$ where x_i denotes each i^{th} sample in $\boldsymbol{x}$ and $\overline{x}$ denotes the mean of the samples in $\boldsymbol{x}$.	
$H(\boldsymbol{a}_x)$ $H(\boldsymbol{a}_y)$ $H(\boldsymbol{a}_z)$ $H(\|\boldsymbol{a}\|)$	For a random variable vector $\boldsymbol{x}$ made up of N scalar observations, the "Shannon" (non-normalised) Entropy (H) is defined as [3]: $$H(x) = -\sum_{i=1}^{N} x_i^2 \log(x_i^2) \quad (6a)$$ whereas the "Log Energy" Entropy (H) is defined as: $$H(x) = \sum_{i=1}^{N} \log(x_i^2) \quad (6b)$$	
$p_{25}(\|\boldsymbol{a}\|)$	The 25^{th} percentile value of $\boldsymbol{x} \in \mathbb{R}^{1\times N}$.	
$p_{75}(\|\boldsymbol{a}\|)$	The 75^{th} percentile value of $\boldsymbol{x} \in \mathbb{R}^{1\times N}$.	
$iqr(\|\boldsymbol{a}\|)$	The interquartile range is the difference between 75^{th} and 25^{th} percentiles of $\boldsymbol{x} \in \mathbb{R}^{1\times N}$.	
$med(\|\boldsymbol{a}\|)$	The median value of $\boldsymbol{x} \in \mathbb{R}^{1\times N}$.	
$mode(\|\boldsymbol{a}\|)$	The mode of $\boldsymbol{x} \in \mathbb{R}^{1\times N}$.	
$range(\|\boldsymbol{a}\|)$	The range, $\|\max(x) - \min(x)\|$, of $\boldsymbol{x} \in \mathbb{R}^{1\times N}$.	

Table I: Mathematical description of the features extracted from the remote Two-Minute Walk Test

Feature	Description	Parameterization
$ApEn(\|\boldsymbol{a}\|)$	Approximate entropy characterises the irregularity and subtle fluctuations in a time series $\boldsymbol{x}$ of length N. The sequence is first divided into m length consecutive segments (templates), which are recursively compared to all others. The total number of similar templates are determined $C_i^{(m)}$ where pairs i of templates are defined as similar by calculating the Chebyshev distance, the maximum absolute difference between each pair of respective template values $< r$, a threshold value. ApEn is then defined as: $ApEn = \Phi^m - \Phi^{m+1}, \quad (7a)$ where $\Phi^m(r) = \frac{\sum_{i=1}^{N-m+1} \ln C_i^m(r)}{N-m+1} \quad (7b)$ $ApEn \rightarrow 0$ denotes similar signal behaviour after $m = m+1$, implying greater regularity within $\boldsymbol{x}$. See Pincus (1991) for more detailed description of $ApEn$ [4].	
$RPDE(\|\boldsymbol{a}\|)$	Recurrence period density entropy ($RPDE$) is a method used to characterise the deviations from exact periodicity and stochasticity within a signal [5], and was proposed here to capture the ability to maintain consistent gait rhythm. The time series is first embedded in phase space such that: $\mathbf{X}_n = [x_n, x_{n+\tau}, x_{n+2\tau}, \ldots, x_{n+(M-1)\tau}]$ for each value of x_n, where M is the embedding dimension, and τ is the embedding delay. Then, around each point $\mathbf{X}_n$ in the embedded phase space, a recurrence neighbourhood of radius ε is formed[6]. Each recurrence of the time series through this neighbourhood is recorded, and the time difference T between successive returns is recorded in a histogram. The normalised entropy of this recurrence period density function $P(T)$ is then computed as: $H_{norm} = -(\ln T_{\max})^{-1} \sum_{t=1}^{T_{\max}} P(t) \ln P(t) \quad (8)$ is the RPDE value, where $T_{\max}$ is the largest recurrence value. For more information we refer readers to [5].	
$H(\boldsymbol{a}_y\|\boldsymbol{a}_z)$ $H(\boldsymbol{a}_x\|\boldsymbol{a}_z)$ $H(\boldsymbol{a}_y\|\boldsymbol{a}_x)$	The conditional entropy $H(Y\|X)$ of the medio-lateral and vertical, anterior-posterior and vertical, and medio-lateral and anterior-posterior axis. $H(Y\|X)$ quantifies the amount of information needed to describe the outcome of a random variable Y given that the value of another random variable X is known. The conditional entropy of Y given X is defined as [3]: $H(Y\|X) = -\sum_{x\in\mathcal{X}, y\in\mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)} \quad (9)$ where $\mathcal{X}$ and $\mathcal{Y}$ denote the support sets of X and Y.	

Table I: Mathematical description of the features extracted from the remote Two-Minute Walk Test

Feature	Description	Parameterization
$\max\left(jerk(\|\boldsymbol{a}\|)\right)$ $\min\left(jerk(\|\boldsymbol{a}\|)\right)$ $mean(jerk(\|\boldsymbol{a}\|))$ $std(jerk(\|\boldsymbol{a}\|))$ $skew(jerk(\|\boldsymbol{a}\|))$ $kurt(jerk(\|\boldsymbol{a}\|))$	The maximum, minimum, mean, variability, skewness and kurtosis of the jerk values. Jerk is the third derivative, $\frac{d^3x}{dt^3}$ of $\boldsymbol{x}$ and helps characterises sharp sensor movements.	
$zcr(\boldsymbol{a}_x)$ $zcr(\boldsymbol{a}_y)$ $zcr(\boldsymbol{a}_z)$ $zcr(\|\boldsymbol{a}\|)$	The zero-crossing rate calculates the rate of sign-changes in a signal, roughly capturing the static-to-dynamic transitions within gait. Assuming a zero-mean signal $\boldsymbol{x}$, zcr can be defined as [7]: $$zcr = \frac{1}{T-1}\sum_{t=1}^{T-1} \mathbb{1}_{\mathbb{R}_{<0}}(x_t, x_{t-1}) \quad (10)$$ where $\mathbb{1}_{\mathbb{R}_{<0}}$ is an indicator function.	
$\max(\widehat{P}(f))$	The maximum power spectral density (PSD) of the sequence $\boldsymbol{x}$, in this case given as $\|\boldsymbol{a}\|$. The PSD estimate of $\boldsymbol{x}$ is performed using the modified periodogram, calculated using Welch's overlapped segment averaging estimator, with a Hamming window function h_n. The modified periodogram PSD is defined by: $$\widehat{P}(f) = \frac{\Delta t}{N}\left\|\sum_{n=0}^{N-1} h_n x_n e^{-j2\pi f n}\right\|^2 \quad (11)$$ where the periodogram is the Fourier transform of the biased estimate of the autocorrelation sequence for a signal x_n, sampled at f_s samples per unit time. Δt is the sampling interval and $-\frac{1}{2\Delta t} < f \leq \frac{1}{2\Delta t}$	
$\arg\max\limits_{f}(\widehat{P}(f))$	The dominant frequency, f (Hz), in the sequence $\boldsymbol{x}$, in this case given as $\|\boldsymbol{a}\|$.	
$\max(\widehat{P}(f_g))$	The maximum power spectral density (PSD) of the sequence $\boldsymbol{x}$ within the gait domain such that $f_g \in [0.5 - 3.5]$ Hz. $\boldsymbol{x}$ is given as $\|\boldsymbol{a}\|$.	
$\arg\max\limits_{f_g}(\widehat{P}(f_g))$	The dominant frequency, f_g (Hz), in the sequence $\boldsymbol{x}$, within the gait domain such that $f_g \in [0.5 - 3.5]$ Hz.	
$DFA(\|\boldsymbol{a}\|)$	Detrended Fluctuation Analysis is a method for determining the statistical rate of decay of a series' autocorrelation [8] and has been used to compare alterations in gait dynamics in healthy and diseased gait [9], [10], [11].	Parameters given by [10].

Table I: Mathematical description of the features extracted from the remote Two-Minute Walk Test

Feature	Description	Parameterization
$\max(r(\Vert\boldsymbol{a}\Vert))$	The maximum autocorrelation within the time-series $\boldsymbol{x} \in \mathbb{R}^{1\times N}$ [12]. The autocorrelation function measures the correlation between x_t and x_{t+k}, where $k = 0, \ldots, K$. The autocorrelation for lag k is: $$r_k = \frac{c_k}{c_0} \quad (12a)$$ where $$c_k = \frac{1}{T}\sum_{t=1}^{T-k}(x_t - \overline{x})(x_{t+k} - \overline{x}) \quad (12b)$$ and c_0 is the sample variance (2) of the time-series.	
$r_1(\Vert\boldsymbol{a}\Vert))$	The autocorrelation coefficient, r at time lag 1 within the time-series [12], [13].	
$MI(\boldsymbol{a}_z, \boldsymbol{a}_y)$	The mutual information [14] between the anterio-posterior and medio-lateral sensor axis.	
$\rho(\boldsymbol{a}_x, \boldsymbol{a}_y)$	Spearman's correlation between the anterio-posterior and vertical sensor axis.	
$\rho(\boldsymbol{a}_y, \boldsymbol{a}_z)$	Spearman's correlation between the vertical and medio-lateral sensor axis.	
$\rho(\boldsymbol{a}_x, \boldsymbol{a}_z)$	Spearman's correlation between the anterio-posterior and medio-lateral sensor axis.	
$\max(\widehat{f}_h(\Vert\boldsymbol{a}\Vert))$	The maximum Kernel density estimation (KDE) value [15], [16]. For any values of $\boldsymbol{x} \in \mathbb{R}^{1\times N}$, the KDE is given by: $$\widehat{f}_h(x) = \frac{1}{hN}\sum_{i=1}^{N} K\Big(\frac{x - x_i}{h}\Big) \quad (13)$$ where $[x_1, x_2, \ldots, x_N]$ are random samples from an unknown distribution, N is the sample size, $K(\cdot)$ is the kernel smoothing function, and h is the bandwidth.	
$\arg\max\limits_{\Vert\boldsymbol{a}\Vert}(\widehat{f}_h(\Vert\boldsymbol{a}\Vert)$	The value of x that maximises the KDE (13).	

Table I: Mathematical description of the features extracted from the remote Two-Minute Walk Test

<table>
<tr><th>Feature</th><th>Description</th><th>Parameterization</th></tr>
<tr><td>$\overline{cA_j}(\|\boldsymbol{a}\|)$
$\overline{cD_j}(\|\boldsymbol{a}\|)$
$std(cA_j(\|\boldsymbol{a}\|))$
$std(cD_j(\|\boldsymbol{a}\|))$
$\max(cA_j(\|\boldsymbol{a}\|))$
$\max(cD_j(\|\boldsymbol{a}\|))$
$\min(cA_j(\|\boldsymbol{a}\|))$
$\min(cD_j(\|\boldsymbol{a}\|))$</td><td>The average, standard deviation, maximum and minimum values of the j^{th} DWT approximation cA_j and detail cD_j coefficients. A sparse representation of gait signals were obtained using the Discrete Wavelet Transform (DWT), where the signal was decomposed into a number of different bandwidths expressed by approximation cA_j and detail cD_j coefficients at level $j = 1, 2, 3, ...L$. This is achieved by convolution between the signal $\boldsymbol{x} \in \mathbb{R}^{1\times N}$ (of any given length, N) and the high- or low-pass filter, $h \in \mathbb{R}^{1\times M}$ (of length $M < N$), and each i^{th} sample of the resulting signal is calculated as:
$$y_i = \sum_{k=-\infty}^{\infty} x_k h_{i-k} \tag{14}$$
where the detail coefficients for the j^{th} level are obtained from the high-pass filter and approximation coefficients from the low-pass filter. By then down-sampling the cA_j by a factor of two, the above filtering operations are then repeated on the resulting signal to produce cD_{j+1} and cA_{j+1} and so on. The result is the decomposition of $\boldsymbol{x}$ into multiple frequency bands, where after each level, the frequency resolution is doubled while the time resolution is halved. Therefore the decomposed signal y at the j^{th} level, of length 2^N, roughly corresponds to a signal in the pass-band $\left[\frac{\pi}{2^j}, \frac{\pi}{2^{j-1}}\right]$, and coefficients in the range $\left[2^{N-j}, 2^{N-j+1}\right]$.</td><td>We extracted the wavelet coefficients experimenting with three wavelet families [17]:
• Daubechies
• Symlets
• Coiflets</td></tr>
<tr><td>$E(cA_j(\|\boldsymbol{a}\|))$
$E(cD_j(\|\boldsymbol{a}\|))$</td><td>Energy E of the j^{th} DWT approximation cA_j and detail cD_j coefficient. The energy is defined as:
$$E(y) = \sum_{i=1}^{N} \lvert y_i \rvert^2 \tag{15}$$
where $y = cD_j$; $y = cA_j$; are the detail and approximation coefficients at level $j = 1, 2, 3, ...L$.</td><td></td></tr>
<tr><td>$H(cA_j(\|\boldsymbol{a}\|))$
$H(cD_j(\|\boldsymbol{a}\|))$</td><td>Entropy H of the j^{th} DWT approximation cA_j and detail cD_j coefficients. This quantifies the predictability of the decomposed gait signal at the j^{th} level, of length 2^N, roughly corresponding to the pass-band $\left[\frac{\pi}{2^j}, \frac{\pi}{2^{j-1}}\right]$, and coefficients in the range $\left[2^{N-j}, 2^{N-j+1}\right]$. Wavelet "Shannon" (non-normalised) Entropy H is defined as:
$$H(y) = -\sum_{i=1}^{N} y_i^2 \log(y_i^2) \tag{16a}$$
whereas the "Log Energy" Entropy is defined as:
$$H(y) = \sum_{i=1}^{N} \log(y_i^2) \tag{16b}$$
where $y = cD_j$; $y = cA_j$; are the detail and approximation coefficients at level $j = 1, 2, 3, ...L$.</td><td>Both log entropy and Shannon's entropy were computed.</td></tr>
</table>

Table I: Mathematical description of the features extracted from the remote Two-Minute Walk Test

Feature	Description	Parameterization
$\overline{\Phi}(cA_j(\|\boldsymbol{a}\|))$ $\overline{\Phi}(cD_j(\|\boldsymbol{a}\|))$ $std(\Phi(cA_j(\|\boldsymbol{a}\|)))$ $std(\Phi(cD_j(\|\boldsymbol{a}\|)))$ $\max(\Phi(cA_j(\|\boldsymbol{a}\|)))$ $\max(\Phi(cD_j(\|\boldsymbol{a}\|)))$	The average, standard deviation and maximum value of the Teager-Kaiser Energy Operator (TKEO, Φ) of the j^{th} DWT approximation cA_j and detail cD_j coefficients. TKEO is calculated directly in the time-domain, however it incorporates both amplitude- and frequency-specific information by measuring local differences between adjacent signal samples and is defined as [18]: $$\Phi(x_i) = x_i^2 - x_{i+1} \cdot x_{i-1} \quad (17)$$ where i denotes the index of the input vector.	
E_{pk}	The maximum scale-dependent energy density E_s of the CWT. $$E_{pk} = \max([E_1, ..., E_S]) \quad (18a)$$ where S are the number of scales. The CWT of a discrete time signal x_n with fixed sampling period δ_t, is defined as the convolution of x_n with a scaled and translated mother wavelet $\psi_0(\eta)$ [19]: $$W_n(s) = \sum_{n'=0}^{N-1} x_{n'} \psi^* \left[\frac{(n' - n)\delta_t}{s} \right] \quad (18b)$$ where (*) denotes the complex conjugate, s is the wavelet scaling factor and n is the localised time index. The subscript $_0$ on ψ has been dropped to indicate that this ψ_0 has been multiplied by $\left(\frac{\delta_t}{s}\right)^{1/2}$, in order to normalise ψ to have unit energy. This ensures that the wavelet transforms $W_n(s)$ at each scale s are directly comparable to each other and to the transforms of other time series; see [19]. A Morlet wavelet was used, which consists of a plane wave modulated by a Gaussian: $$\psi_0(\eta) = \frac{1}{\sqrt[4]{\pi}} e^{i \cdot w_0 \cdot \eta} \cdot e^{\frac{-\eta^2}{2}} \quad (18c)$$ where and η is the a non-dimensional time parameter and w_0 is the non-dimensional frequency, here taken to be 6, as per [19], to satisfy the admissibility condition. The total signal energy at a specific scale can be measured by the scale-dependent energy density spectrum E_s: $$E_s = \sum_{n=0}^{N-1} \|W_n(s)\|^2 \quad (18d)$$ where $s \in [1, S]$ and $\|W_n(s)\|^2$ is the 2-D wavelet energy density (scalogram) that measures the total energy distribution of the signal.	

Table I: Mathematical description of the features extracted from the remote Two-Minute Walk Test

Feature	Description	Parameterization
$\arg\max_{f}(E_s)$	The frequency (f_{max}), in Hz, which maximises E_s over all scales s (18d). f, in Hz, can be approximated from the wavelet scaling factor s such that [20]: $$f = \frac{f_c}{s} \quad (19a)$$ where the center frequency in Hz can be defined by [19]: $$f_c = \frac{w_0 + \sqrt{2 + w_0^2}}{4\pi} \quad (19b)$$	
$\sum E_s$	The total scale dependent energy E over all scales s (18d).	
$\overline{E}_s$	The average E value over all scales s (18d).	
$std(E_s)$	The standard deviation in the values of E over all scales s (18d).	
$skew(E_s)$	Skewness (4) as a measure of the asymmetry of the probability distribution of the scale dependent energy distribution E over all scales s (18d).	
$kurt(E_s)$	The kurtosis in the values of E over all scales s (18d).	
$AUC(E_s)$	The area under the curve (AUC) approximate integral of E over all scales s (18d), with with respect to the frequencies f, using the trapezoidal method as defined by: $$AUC(E_s) = \int_f E_s dS \approx \frac{\Delta f}{2N} \sum_{s=1}^{S} (E_{s-1} + E_s) \quad (20)$$ where s represents the scale factor, S are the number of scales and the spacing between each point is equal to the scalar value $\frac{\Delta f}{N}$.	
$prom(E_{pk})$	The prominence of the maximum scale-dependent energy density peak E_{pk} (18a). Prominence measures how much a peak stands out due to its intrinsic height and its location relative to other peaks.	
$\frac{E_{pk1}}{E_{pk2}}$	The ratio of the maximum scale-dependent energy peak density to next highest peak. $$\frac{E_{pk1}}{E_{pk2}} = \frac{max_1(E_s(\lVert\boldsymbol{a}\rVert))}{max_2(E_s(\lVert\boldsymbol{a}\rVert))} \quad (21)$$	
$width(E_{pk})$	The width E_w of the maximum scale-dependent energy density E_{pk}. The width of the peak is estimated as the distance between the points where the descending edges intercept a horizontal reference line beneath the peak at a vertical distance equal to half the peak prominence ($prom$).	
$\frac{E_{w1}}{E_{w2}}$	The ratio of the width of the maximum scale-dependent energy peak density to next highest peak. $$\frac{E_{w1}}{E_{w2}} = \frac{width\,(max_1(E_s(\lVert\boldsymbol{a}\rVert)))}{width\,(max_2(E_s(\lVert\boldsymbol{a}\rVert)))} \quad (22)$$	

Table I: Mathematical description of the features extracted from the remote Two-Minute Walk Test

Feature	Description	Parameterization
$\frac{f_{max1}}{f_{max2}}$	The ratio of the frequency that maximises the scale-dependent energy peak density, f_{max}, to the frequency that maximises the next highest peak.	
$\sum\left\lvert\frac{E(IMF_{(4:)}(\lVert\boldsymbol{a}\rVert))}{E(IMF_{(:3)}(\lVert\boldsymbol{a}\rVert))}\right\rvert$ $\sum\left\lvert\frac{\Phi(IMF_{(4:)}(\lVert\boldsymbol{a}\rVert))}{\Phi(IMF_{(:3)}(\lVert\boldsymbol{a}\rVert))}\right\rvert$ $\sum\left\lvert\frac{H(IMF_{(4:)}(\lVert\boldsymbol{a}\rVert))}{H(IMF_{(:3)}(\lVert\boldsymbol{a}\rVert))}\right\rvert$	Classical EMD decomposes a signal into a small finite number of intrinsic mode functions (IMFs) using the Hilbert-Huang transform (HHT) to encode instantaneous frequency and amplitude information[21]. For more detailed information on the calculation of EMD, IMFs and the HHT we refer the reader to [21]. The first 3 IMFs represents the "high-frequency (noise)" components with the latter IMFs capturing the relatively "low-frequency (signal)" components of gait rhythm. The sum of the signal-to-noise (SNR) energy, E (15), entropy, H (6), and TKEO, Φ (17), are then computed using IMFs. This is analogous for the variability in the ratio (amount) of gait to higher- frequency perturbations over the 2MWT.	Computed with the signal sampled every: $\Delta t = \{0.01, 0.5, 1\}$ seconds [s].
$MsEn_\tau(\lVert\boldsymbol{a}\rVert)$	Multiscale entropy ($MsEn$) calculates the sample entropy (*SampEn*) of a signal at increasingly coarser grains (scales) [22]. $MsEn$ characterises dynamic complexity of gait within a signal [23], [24]. Signals are first segmented by taking the mean of data points in non-overlapping windows of increasing length, τ. Given a time series, $[x_1, x_2, ..., x_N]$, we construct consecutive coarse-grained time series by averaging a successively increasing number of data points in non-overlapping windows. Each element of the coarse-grained time series $y_k^{(\tau)}$ is computed as: $$y_j^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i \quad (23)$$ where τ represented the scale factor and $1 \leq j \leq N/\tau$. The length of each coarse-grained time series is N/τ. $MsEn_\tau(x)$ therefore denotes the sample entropy (*SampEn*) [25] calculated for each coarse-grained time series, $y^{(\tau)}$, defined by τ. Lower values of *SampEn* indicate more self-similarity in the time series.	As per [23]: • Maximum scale factor $\tau = 20$ • Embedding dimension $m = 2$ • Tolerance $r = 0.2$
$\overline{MsEn_{(:)}}(\lVert\boldsymbol{a}\rVert)$	Average multiscale mntropy ($MsEn$) over all timescales.	
$std(MsEn_{(:)}(\lVert\boldsymbol{a}\rVert))$	Standard deviation in multiscale entropy ($MsEn$) over all scales.	
$AUC(MsEn_{(:)}(\lVert\boldsymbol{a}\rVert))$	The area under the curve (AUC) of $MsEn$ over all timescales. AUC is computed using numerical integration via the trapezoidal method such that $$AUC = \frac{b-a}{2N}\sum_{\tau=1}^{S}(MsEn_{(\tau-1)} + MsEn_{(\tau)}) \quad (24)$$ where τ represents the scale factor, S are the number of scales and the spacing between each point is equal to the scalar value $\frac{b-a}{N}$, in this case 1.	
$p_{25}(MsEn_{(:)}(\lVert\boldsymbol{a}\rVert))$	25^{th} percentile of $MsEn$ values over all scales.	
$p_{75}(MsEn_{(:)}(\lVert\boldsymbol{a}\rVert))$	75^{th} percentile of $MsEn$ values over all scales.	
$p_{25}(...$ $AUC(MsEn_{(:)}(\lVert\boldsymbol{a}\rVert)))$	25^{th} percentile in AUC values computed from $MsEn$ over all scales (24).	

Table I: Mathematical description of the features extracted from the remote Two-Minute Walk Test

Feature	Description	Parameterization
$p_{75}(...$ $AUC(MsEn_{(:)}(\|\boldsymbol{a}\|)))$	75^{th} percentile in AUC values computed from $MsEn$ over all scales (24).	
$\frac{MsEn(\|\boldsymbol{a}\|)_{(:10)}}{MsEn(\|\boldsymbol{a}\|)_{(11:)}}$	The ratio in $MsEn$ over first 10 to last 10 scales captures the dynamic complexity of gait versus that of random fluctuations in a signal [23].	
$\frac{std(MsEn(\|\boldsymbol{a}\|)_{(:10)})}{std(MsEn(\|\boldsymbol{a}\|)_{(11:)})}$	The ratio in the standard deviation in $MsEn$ over first 10 to last 10 scales captures the variance in the dynamic complexity of gait versus the variance in that of random fluctuations in a signal [23].	

REFERENCES

[1] H.-h. Lee, S. Choi, and M.-j. Lee, "Step detection against the dynamics of smartphones," *Sensors*, vol. 15, no. 10, pp. 27 230–27 250, 2015.

[2] D. Joanes and C. Gill, "Comparing measures of sample skewness and kurtosis," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 47, no. 1, pp. 183–189, 1998.

[3] R. M. Gray, *Entropy and information theory*. Springer Science Business Media, 2011.

[4] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.

[5] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *Biomedical engineering online*, vol. 6, no. 1, p. 23, 2007.

[6] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics reports*, vol. 438, no. 5-6, pp. 237–329, 2007.

[7] L. R. Rabiner, "Digital processing of speech signal," *Digital Processing of Speech Signal*, 1978.

[8] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, "Multifractal detrended fluctuation analysis of nonstationary time series," *Physica A: Statistical Mechanics and its Applications*, vol. 316, no. 1-4, pp. 87–114, 2002.

[9] J. M. Hausdorff, Y. Ashkenazy, C. K. Peng, P. C. Ivanov, H. E. Stanley, and A. L. Goldberger, "When human walking becomes random walking: fractal analysis and modeling of gait rhythm fluctuations," *Physica A*, vol. 302, no. 1-4, pp. 138–47, 2001.

[10] S. Damouras, M. D. Chang, E. Sejdić, and T. Chau, "An empirical examination of detrended fluctuation analysis for gait data," *Gait posture*, vol. 31, no. 3, pp. 336–340, 2010.

[11] S. Dutta, D. Ghosh, and S. Chatterjee, "Multifractal detrended fluctuation analysis of human gait diseases," *Frontiers in physiology*, vol. 4, p. 274, 2013.

[12] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley Sons, 2015.

[13] A. Zhan, M. A. Little, D. A. Harris, S. O. Abiola, E. Dorsey, S. Saria, and A. Terzis, "High frequency remote monitoring of parkinson's disease via smartphone: Platform overview and medication response detection," *arXiv preprint arXiv:1601.00960*, 2016.

[14] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley Sons, 2012.

[15] A. Qureshi, M. Brandt-Pearce, M. M. Engelhard, and M. D. Goldman, "Relationship between kernel density function estimates of gait time series and clinical data," in *2017 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*. IEEE, Conference Proceedings, pp. 329–332.

[16] S. R. Dandu, M. M. Engelhard, A. Qureshi, J. Gong, J. C. Lach, M. Brandt-Pearce, and M. D. Goldman, "Understanding the physiological significance of four inertial gait features in multiple sclerosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 40–46, 2018.

[17] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.

[18] J. F. Kaiser, "On a simple algorithm to calculate the'energy'of a signal," in *International conference on acoustics, speech, and signal processing*. IEEE, Conference Proceedings, pp. 381–384.

[19] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.

[20] P. Abry, *Ondelettes et turbulences: multirésolutions, algorithmes de décomposition, invariance d'échelle et signaux de pression*. Diderot éd., 1997.

[21] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," in *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, vol. 454. The Royal Society, Conference Proceedings, pp. 903–995.

[22] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical review letters*, vol. 89, no. 6, p. 068102, 2002.

[23] M. Costa, C. K. Peng, A. L. Goldberger, and J. M. Hausdorff, "Multiscale entropy analysis of human gait dynamics," *Physica A: Statistical Mechanics and its Applications*, vol. 330, no. 1, pp. 53–60, 2003.

[24] J. Yu, J. Cao, W.-H. Liao, Y. Chen, J. Lin, and R. Liu, "Multivariate multiscale symbolic entropy analysis of human gait signals," *Entropy*, vol. 19, no. 10, p. 557, 2017.

[25] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.