PLOS ONE
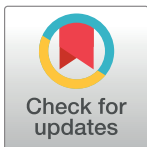
Rett syndrome severity estimation with the BioStamp nPoint using interactions between heart rate variability and body movement

Pradyumna Byappanahalli Suresha[ORCID][1,2]*, **Heather O'Leary**[3], **Daniel C. Tarquinio**[4], **Jana Von Hehn**[3], **Gari D. Clifford**[2,5]

1 School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, United States of America, **2** Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA, United States of America, **3** Rett Syndrome Research Trust, Trumbull, CT, United States of America, **4** Center for Rare Neurological Diseases, Norcross, GA, United States of America, **5** Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, United States of America

* pradyumna@gatech.edu

Abstract

Rett syndrome, a rare genetic neurodevelopmental disorder in humans, does not have an effective cure. However, multiple therapies and medications exist to treat symptoms and improve patients' quality of life. As research continues to discover and evaluate new medications for Rett syndrome patients, there remains a lack of objective physiological and motor activity-based (physio-motor) biomarkers that enable the measurement of the effect of these medications on the change in patients' Rett syndrome severity. In our work, using a commercially available wearable chest patch, we recorded simultaneous electrocardiogram and three-axis acceleration from 20 patients suffering from Rett syndrome along with the corresponding Clinical Global Impression—Severity score, which measures the overall disease severity on a 7-point Likert scale. We derived physio-motor features from these recordings that captured heart rate variability, activity metrics, and the interactions between heart rate and activity. Further, we developed machine learning (ML) models to classify high-severity Rett patients from low-severity Rett patients using the derived physio-motor features. For the best-trained model, we obtained a pooled area under the receiver operating curve equal to 0.92 via a leave-one-out-patient cross-validation approach. Finally, we computed the feature popularity scores for all the trained ML models and identified physio-motor biomarkers for Rett syndrome.

Introduction

Rett syndrome is a rare genetic neurodevelopmental disorder that occurs primarily in girls [1]. Though Rett syndrome is a clinical diagnosis, more than 95% of cases are caused by mutations in the gene encoding the methyl-CpG binding protein 2 (MECP2), a transcriptional regulator involved in chromatin remodeling and the modulation of RNA splicing [2]. Rett syndrome affects 1 in 10000 females and is characterized by a period of apparently normal postnatal development followed by developmental delay and loss of acquired skills resulting in

psychomotor regression, development of stereotypical hand movements, and dysautonomia [3]. It leads to the deterioration of the autonomous nervous system, impacting breathing regularity, heart rate (HR), gut motility. Further, it impairs motor planning and locomotion, resulting in significantly impaired mobility, no purposeful hand use, and largely absent verbal communication. There is no permanent cure for Rett syndrome in humans, and symptom management remains the standard of care [4]. When new drugs are discovered to alleviate specific Rett symptoms, clinical trials are conducted to learn about their efficacy, safety, and side effects. An essential step in measuring the efficacy of a drug or treatment method is to assess the associated benefits and risks through clinical trials. However, objective measures of symptom severity are not yet available for Rett syndrome or neurological conditions generally, and efforts to develop objective measures of autonomic symptoms could significantly enhance the ability to understand therapeutic efficacy. The key benefit we would like to see in Rett patients is improving their autonomous nervous system's function and locomotion. Amongst various indices measured in Rett clinical trials, the Clinical Global Impression—Severity (CGI-S) is used to measure overall disease severity in Rett subjects [5, 6]. The CGI-S is a 7−point Likert rating scale that reflects experts' clinical judgment of the patient based on the clinician's total experience with the Rett syndrome population. The CGI-S ranges from 1 to 7 and each score corresponds to the following patient states: (1) normal, not at all ill, (2) borderline ill, (3) mildly ill, (4) moderately ill, (5) markedly ill, (6) severely ill, (7) amongst the most extremely ill. It has been widely used as an outcome measure in Rett syndrome and other neurodevelopmental disorders such as Autism and Fragile X Syndrome [5]. In our work, we measured CGI-S in all our patients during all clinic visits included in our experiments to assess their global clinical state.

Dysautonomia or autonomic dysfunction is the abnormal function of the autonomous nervous system. It adversely affects involuntary body functions, including blood-pumping by the heart, maintaining proper blood pressure and respiration. Unfortunately, dysautonomia is a cardinal feature in Rett syndrome [7–12]. A principled approach to characterize dysautonomia is to utilize the electrocardiogram (ECG) [8, 13–15]. Researchers have used the ECG to study the variations in breathing and HR in Rett girls [16, 17]. In our work, we captured the ECG signal from girls suffering from Rett syndrome using the BioStamp® nPoint wearable biosensors to capture the severity of autonomic dysfunction and, in turn, develop Rett severity classification models using ECG. Specifically, we used the heart rate variability (HRV) metrics for this purpose. The HRV is a physiological phenomenon of the variation of the time interval between the heartbeats. Typically, it is measured using a set of metrics known as the HRV metrics. We describe the HRV metrics utilized in this work in detail in the section *Materials and methods*. Dystonia is a movement disorder where a subset of muscles contract uncontrollably. The contractions cause the affected body parts to twist involuntarily, resulting in repetitive movements. In Rett syndrome, dystonia, psychomotor regression, and stereotypical hand movements are fundamental concerns and cause significant stress on patient and caregiver quality of life. We refer the readers to the following works [18–21] for a detailed explanation of dystonia in Rett syndrome. The BioStamp® nPoint wearable biosensors measured the body movements of Rett patients by recording the three-axis acceleration signal via an accelerometer [22], an electromechanical sensor that senses static and dynamic forces of acceleration. To further characterize body movements, we derived rest activity metrics [23] and the cosinor rhytmometry features from the captured three-axis acceleration signals. Together, we call them actigraphy metrics. We describe these metrics in the section *Materials and methods*. Interactions between the HR and body movement were quantified using multiscale transfer entropy (MSTE) [24, 25] and multiscale network representation (MSNR) [26]. The MSTE metrics measured the information flow between two simultaneously sampled time series at multiple time scales. In MSNR, we constructed network representations of simultaneously sampled 3-dimensional time series at

multiple time scales and derived network characteristics at each time scale. These network representations revealed more nuanced characteristics of the time series being analyzed.

The goal of our study was two-fold. First, we wanted to develop machine learning (ML) classification models to classify patients with low-severity Rett syndrome (CGI-S $\leq$4) from patients with high-severity Rett syndrome (CGI-S >4) based on the objective measures attained from a wearable biosensor. Second, through the classification experiment, we wanted the trained models to provide us with important features (physio-motor biomarkers) that could help us distinguish the two groups. Hence, we developed binary classifier models based on raw data recordings using metrics derived from the following feature sets: (1) HRV metrics, (2) Actigraphy metrics, (3) MSTE-features, and (4) MSNR-features. We used the least absolute shrinkage and selection operator (LASSO) for model training and developed logistic regression models with the L1-penalty. We developed separate models for each of the four feature sets listed above and for all possible two, three, and four combinations of these feature sets. In total, we developed 15 binary-classification models for Rett syndrome severity classification. Finally, we listed the models' features that were important for Rett syndrome severity classification. We illustrate the complete pipeline of our work in Fig 1.

Materials and methods

Data collection

The dataset for this work was sourced from two Institutional Review Board (IRB)-approved studies: (1) The Triheptanoin-clinical trial [27] (2) The Outcome measures and biomarkers
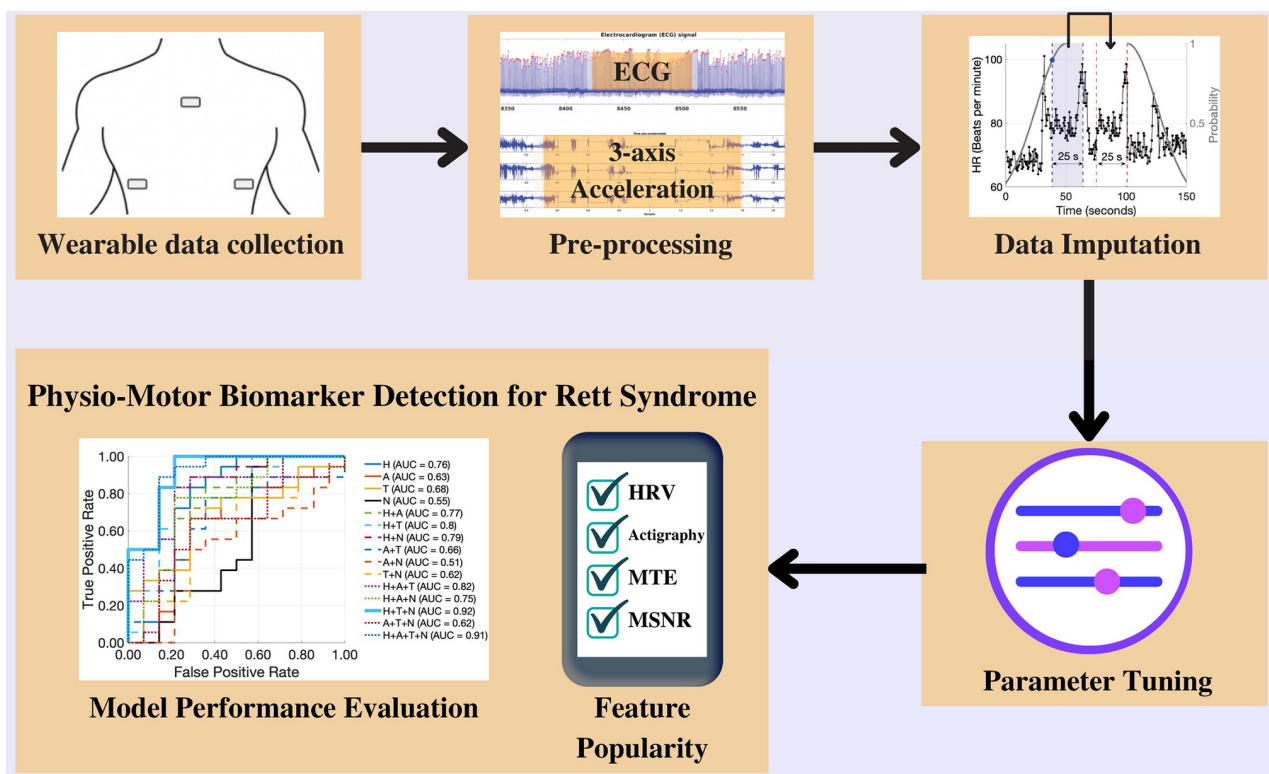
Fig 1. The project pipeline. We utilized the BioStamp® nPoint biosensor wearables to collect simultaneous ECG and three-axis acceleration data from multiple locations on the chest. We performed data pre-processing and implemented multiple data imputation techniques to improve data quality. We trained L1-regularized logistic regression classifier models and tuned the model weights using the imputed data. Finally, we visualized model performance and computed feature popularity scores.

development study [28]. The data were collected between January 2016 and December 2018—a three-year period. We used the body-worn patch BioStamp® (MC10 Inc., Cambridge, MA, USA) [29] to record ECG and three-axis acceleration from all the participants. While some ECG records were captured at a sampling rate of 125Hz, others were captured at a sampling rate of 250Hz. Concurrently, the three-axis acceleration records were captured at the sampling rates of 31.25Hz and 62.5Hz, respectively. These differences did not meaningfully influence the HRV and activity metrics we extracted [30]. We captured the ECG signal and the three-axis acceleration from the following four locations on the body: (1) Medial chest, (2) Left Hypochondrium, (3) Right Hypochondrium, and (4) Left Pectoralis. Per the protocols, all four locations were not used for all the participants, and only a subset of these locations was used for each participant. In conjunction to the signal data obtained from the biosensors, caretaker and physician surveys were conducted to obtain symptom severity for all 20 patients enrolled in the study. Specifically, the CGI-S scores were obtained through physician surveys to assign a binary label (low-severity vs. high-severity) for each patient-visit. A patient-visit was assigned to the low-severity category if the CGI-S ≤4 and was assigned to the high-severity category if the CGI-S >4. For each patient-visit we needed two consecutive days of signal data for the feature extraction. By applying this filter, we obtained a total of 32 patient-visits with two consecutive days of signal data and the associated CGI-S label. Among the 32 patient-visits, we had 18 high-severity visits corresponding to 10 unique patients and 14 low-severity visits corresponding to 11 unique patients. One patient had both low-severity and high-severity visits. We considered each patient-visit a data point and had a total of 32 data points with an associated binary label for model development and analysis. The racial, ethnic and age breakdown of the study subjects are illustrated in Fig 2A–2C respectively. The low and the high-severity groups had a similar split in patients with respect to the race, ethnicity and age. We had 8 patients each in low and high-severity groups whose race was "Caucasian" and 2 patients each in the two groups who's race was "Asian". One patient's race was unknown. Further, we had 8 patients each in low and high-severity groups whose ethnicity was "Non-Hispanic" and 2 patients each in the two groups whose race was "Hispanic". One patient's race was unknown. Finally, the median age at enrollment in the low and high-severity groups were 8.5 and 8 years respectively.

Study approval. This study was approved by the Emory Institutional Review Board (IRB00088492: Outcome Measures and Biomarkers Development for Rett Syndrome and Multisite development of standardized assessments for use in clinical trials). A written informed consent was received prior to the participation from the parents of the patients.
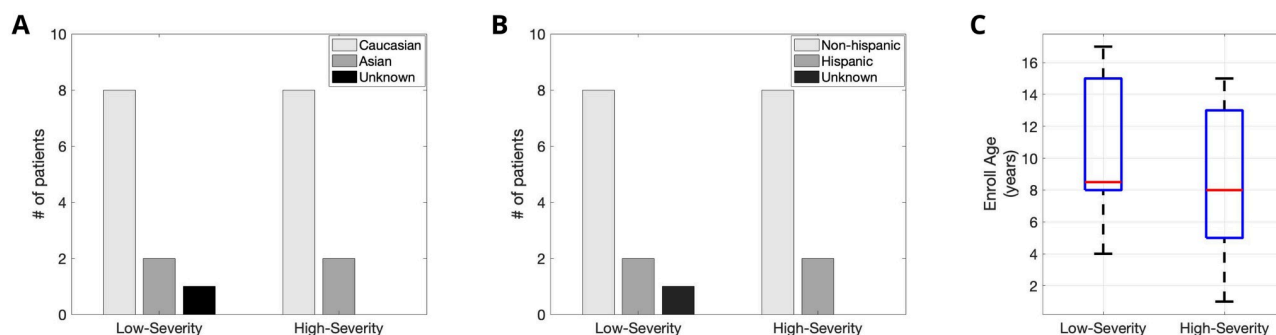
Fig 2. The race, ethnic and age split in the cohort. (A) We illustrate the number of patients in each of the caucasian, asian and unknown races in both Rett severity groups. (B) We illustrate the number of patients in each of the non-hispanic, hispanic and unknown ethnicities in both Rett severity groups. (C) We depict a box plot created with the ages of the patients in the two Rett severity groups, showing the median, 25th percentile and the 75th percentile.

Missing data

Considerable amounts of missing data were present in the dataset due to the following reasons: (1) Device charging and data upload, (2) Motion artifacts, and (3) In some cases, low compliance by the caretakers. The caretakers were trained in-clinic for data collection. Per protocol, caretakers collected approximately an hour of in-clinic data from the patients when learning how to use the devices. The caretakers were then asked to record data at home for up to seven days in intervals up to 24 hours at a time. The Biostamp sensor needed removal during bathing. Further, the device needed to be fully charged every 20 to 24 hours depending on the device age and a full charging of the device took up to two hours. To minimize data loss, the caretakers were advised to charge the devices while the patients were bathing. However, due to various factors such as forgetfulness and technological difficulties caretakers could have exhibited lower compliance. To tackle the missing data issue we implemented three signal imputation techniques to improve data quality and increase the amount of available data for analysis. Namely,

1. Signal quality index-based ECG data imputation.

2. Data imputation for activity counts.

3. Stochastic surrogate data imputation.

Signal quality index-based ECG data imputation. The ECG signal recorded using the wearable patches contained sections of data corrupted by motion artifacts. To improve data quality, we sourced data from multiple sensors. As discussed earlier, we simultaneously captured ECG from up to four unique locations on the body. For a time-window t, say we had N (N $\leq$ 4) ECG signal snippets recorded from N locations, we chose the one signal snippet with the highest Signal Quality Index (SQI). The SQI for ECG provides the percentage of beats that match when detected by multiple annotation generators with highly differing noise responses [31]. We refer the readers to Li et. al [32] for a detailed explanation of the SQI computation algorithm. If this value was greater than 0.75, we used it in our analysis; otherwise, we discarded all ECG data for the time-window t. By switching between N signals for each time-window t to form a single 1−D ECG signal, we maximized the amount of good data available for the analysis.

Data imputation for activity counts. The activity count signal was extracted from the three-axis acceleration in order to further extract actigraphy metrics using the Oakley's method [33]. When analyzing activity count signal in isolation, as per the scripts provided in the actigraphy toolbox [34], we combined the data to obtain one value per hour. When there was no data in a given hour, we imputed those samples using the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) interpolation as recommended by Cakmak et al. [35].

Stochastic surrogate data imputation. The computation of transfer entropy and multi-scale network representation required us to impute the missing data in the 48-hour HRV features and the activity count signals. For this, we developed the surrogate data imputation method, a stochastic technique developed to impute missing data in a timeseries using data in the vicinity of the missing sections (or *gaps*). The data imputation algorithm works as follows. Given a time series ($\mathcal{S}$) and its timestamps (t), we find all the N *gaps* $g[i]$ $\forall i \in \{1, 2, 3, ..., N\}$ in $\mathcal{S}$ which are greater than a fixed threshold th_g. The gaps are then sorted in increasing order. We impute the gaps with surrogate signal snippets in increasing order of the gap length as follows. We denote the gap length in seconds as g_l, while t_b and t_e denote the time stamp where the gap begins and ends, respectively. Next, for each gap $g[i]$ we draw a sample x_r from the
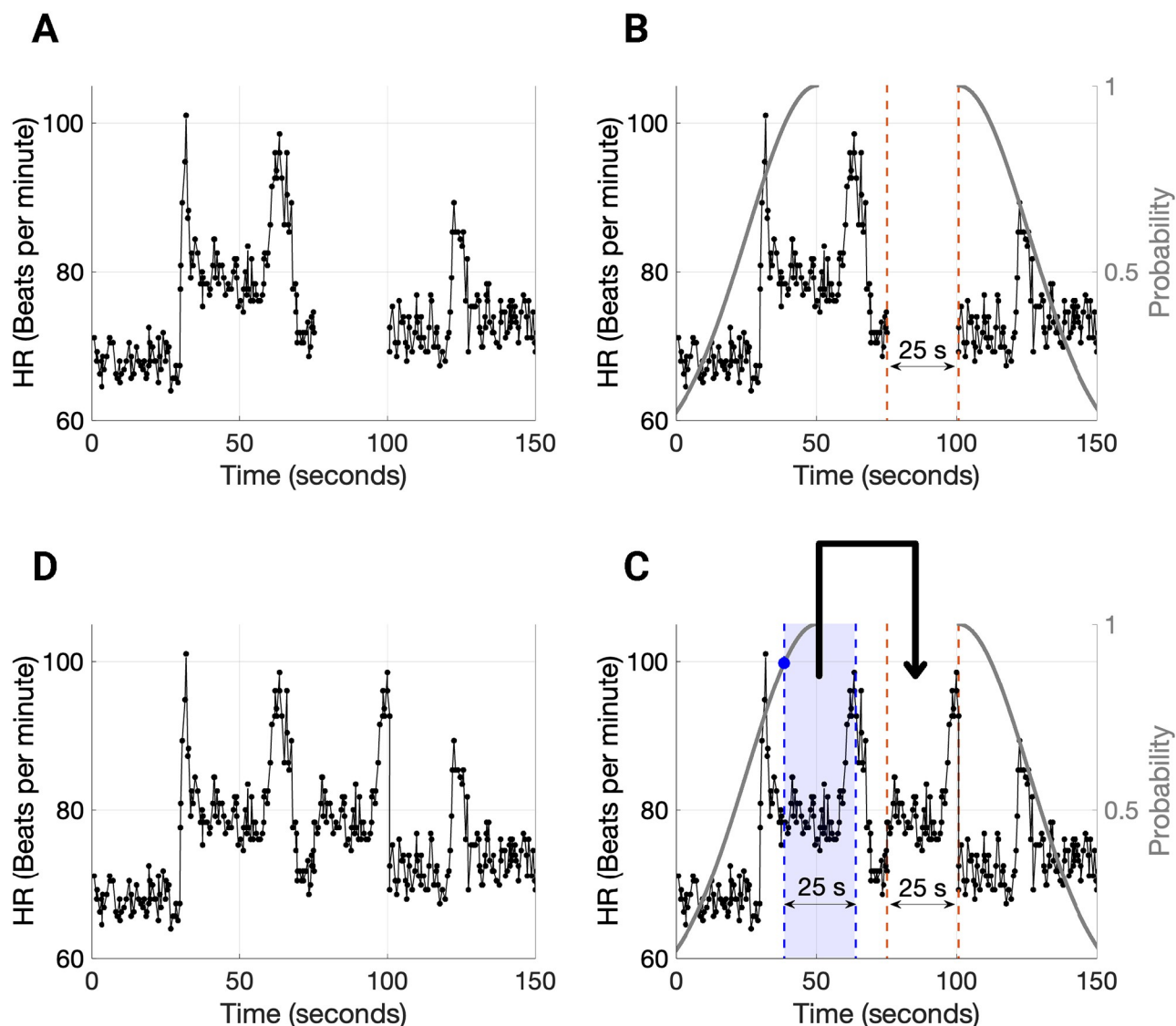
Fig 3. Stochastic surrogate data imputation technique fills gaps in heart rate (HR) and activity count signals by choosing a contiguous HR segment in the gap's neighborhood of the same length as the gap. (A) We illustrate an example HR signal with missing data. (B) The HR signal contains a 25-second-long gap that has been identified. Based on the boundaries and length of the gap, two folded normal distributions are constructed for stochastically choosing a 25-second-long HR segment. (C) We perform a coin toss experiment, and based on the outcome, we sample from one of the two folded normal distributions and accordingly select a contiguous 25-second-long HR segment and copy it over to the gap. Further, we add a small noise to this imputed signal (we do not show it in the figure for convenience). (D) The imputed HR signal with no gaps.

normal distribution with mean 0 and variance equal to 1, i.e., $x_r \sim \mathcal{N}(0, 1)$. The normal distribution from which we picked a sample is then mapped to the timeseries $\mathcal{S}$ in the following way. We map the left half of the distribution ($0.5 \times$ folded normal distribution) to $\mathcal{S}$ where $t < t_b - g_l$ and the right half of the distribution ($0.5 \times$ folded normal distribution) to $\mathcal{S}$ where $t > t_e$. We illustrate this mapping of the Gaussian distribution onto the timeseries in Fig 3. Accordingly, we copy the signal snippet of length g_l starting from the point in time that corresponds to x_r on the timestamp signal t. We insert this copied signal in the gap $g[i]$ and add a noise signal which is 5% of the sample sampled from a Gaussian distribution with mean ($\mu_{\mathcal{S}}$) and variance ($\sigma_{\mathcal{S}}^2$) equal to the mean and variance of the signal $\mathcal{S}$. Further, we update both the

timeseries $\mathcal{S}$ and the corresponding timestamps t. This procedure is repeated iteratively until all gaps $\{g[i]\}$ in $\mathcal{S}$ greater than th_g are imputed.

Feature extraction

We extracted HRV metrics from the ECG signals and actigraphy features from the three-axis accelerometer signals in the dataset. The HRV metrics were extracted using the open-source PhysioNet Cardiovascular Signal Toolbox provided by Vest et al. [31]. We extracted 24 distinct HRV metrics, including time-domain measures, frequency-domain measures, entropy measures, phase rectified signal averaging (PRSA) measures, and other non-linear metrics. We used the default window length settings provided in the toolbox and thus used a 300-second-long feature extraction window with a 30-second shift. We used the SQI based ECG data imputation to maximize the amount of good ECG data. For a given patient-visit, we computed the mean and variance of each HRV metric between the times 10 PM and 10 AM. We chose this period to include ECG data during sleep and discard the noisier signal recorded during daytime and evenings. Hence, each HRV metric provided two features resulting in 48 features from 24 HRV metrics.

We extracted the actigraphy features from the z-axis of the acceleration signal from the right hypochondrium using the open-source Actigraphy Toolbox [34]. We converted the acceleration signal to activity counts using the Oakley method described by Borazio et al. [33]. First, we used Oakley's method for converting accelerometer signals to activity count. We then extracted the following eight features using the toolbox: (1) Interday stability, (2) Intraday variability, (3) Least active 5 hours, (4) Most active 10 hours, (5) Rest activity, (6) Mesor, (7) Amplitude, and (8) Acrophase. The last three features were based on Cosinor Rhythmometry. The Actigraphy features needed two consecutive 24-hour periods (midnight to midnight) of data for feature computation. Thus, we identified the best two consecutive 24-hour periods with the least missing data for each patient-visit. If both days did not have at least 12-hours of acceleration data per day, those patient-visits were discarded. To impute missing data, we used the PCHIP interpolation.

Finally, we computed MTE and MSNR features using 2-day consecutive HR and activity count signals. We utilized the *Stochastic surrogate data imputation* technique to impute missing data. We processed the HR signal, deceleration capacity (DC) of the RR-interval [36] signal, and the activity count (Act) signal for computing the features. Transfer entropy depicted as $TE_{X \to Y}$ is a measure of directional coupling between two concurrently sampled timeseries $X = \{x_1, x_2, \ldots, x_N\}$ and $Y = \{y_1, y_2, \ldots, y_N\}$. Formally, $TE_{X \to Y}$ is a reduction in uncertainty, given by the conditional entropy of y_i given its past values minus the conditional entropy of y_i given both its past values and past values of the other variable $y_{i-w}^{(l)}$:

$$TE_{X \to Y} = H(y_i | y_{i-w}^{(l)}) - H(y_i | y_{i-w}^{(l)}, x_{i-t}^{(k)}) \tag{1}$$

where i indicates a given point in time, t and w are the time lags in X and Y respectively, and k and l are the block lengths of past values in X and Y respectively. The k and l were both set to 1 to improve computational speed, and t and w were both set to 1 under the assumption that the maximum auto-transfer of information occurs from the data point in X immediately before the target value in Y, and vice versa. These choices of $k = l = t = w = 1$ are appropriate in biomedical experiments as the absolute values of auto-correlation functions tend to decrease monotonically as time lag increases [25]. In our experiments, we computed the TE between the following signals: (1) HR-DC, (2) HR-Act, (3) DC-Act, (4) DC-HR, (5) Act-HR, and (6) Act-DC. We computed these TE values at scales 1 to 10 using the coarse-graining algorithm [37] to obtain Multiscale Transfer Entropy (MSTE) features. The probability densities for the

estimation of MTE were estimated using the Darbellay-Vajda (D-V) adaptive partitioning algorithm [25, 26, 38]. Further, we computed 3D D-V partitioning using the HR-DC-Act signals and computed multiscale network representation (MSNR) features. The network representation features included the following 11 metrics: (1) Number of nodes (total number of nodes in the network), (2) Average degree (the average value of the degree of all nodes in the network, where the degree of a node is defined as the total number of its neighboring edges), (3) Number of loops (the total number of independent loops in the network, also known as the "cyclomatic number" or the number of edges that need to be removed so that the network cannot have cycles), (4) LOOP3 (the total number of loops of size 3 in the network), (5) LOOP4 (the total number of loops of size 4 in the network), (6) Average clustering coefficient—1 (number of LOOP3s divided by the number of connected triples in the network), (7) Average clustering coefficient—2 (the clustering coefficient $c(u)$ for node u can be defined as the ratio of the number of actual edges between the neighbors of u to the number of possible edges between them, and the average clustering coefficient $C(G)$ of a network is the average of $c(u)$ taken over all the nodes in the network), (8) Graph radius (the eccentricity of a node u is defined as $e(u) = max\{d(u, v):v \in V\}$, where the distance $d(u, v)$ is the length of the shortest path from u to v, and V is the set of all nodes; the graph radius is the minimum of eccentricities over all nodes in the network), (9) Spectral radius (the largest magnitude eigenvalue of the adjacency matrix of the network), (10) Trace (sum of the eigenvalues of the adjacency matrix, i.e., $\sum_i^n \lambda$), and (11) Energy (squared sum of the eigenvalues of the adjacency matrix A. More formally, the energy of a network G is: $E(G) = \sum_i^n \lambda_i^2$). We computed the above-described MSNR features at scales 1 to 10 using the coarse-graining technique [37]. Using the surrogate data imputation technique, we obtained 100 imputations for each patient-visit. Thus, we generated 3200 datapoints. We computed the MTE and MSNR features for all the 3200 datapoints and then computed the mean and variance over 100 imputations for each patient-visit. In the end, we obtained 32 vectors of length 120 (60 mean values and 60 variance values) as the MTE features and obtained 32 vectors of length 220 (110 mean values and 110 variance values) as the MSNR features. Consequently, we obtained 48 HRV features, 8 Actigraphy features, 120 MTE features, and 220 MSNR features for each of the 32 patient visits.

Rett syndrome severity classification experiments

We developed separate models for each feature sub-group and combinations of feature sub-groups to obtain 15 classifiers corresponding to the following feature combinations: (1) HRV, (2) Actigraphy, (3) MSTE, (4) MSNR, (5) HRV + Actigraphy, (6) HRV + MSTE, (7) HRV + MSNR, (8) Actigraphy + MSTE, (9) Actigraphy + MSNR, (10) MSTE + MSNR, (11) HRV + Actigraphy + MSTE, (12) HRV + Actigraphy + MSNR, (13) HRV + MTE + MSNR, (14) Actigraphy + MTE + MSNR, and (15) HRV + Actigraphy + MTE + MSNR. We used the LASSO based logistic regression classifier to assess the performance of different feature combinations. The models were assessed via a leave-one-patient out cross-validation (LOOCV) experiment. Given that the dataset was small (n = 20), we chose the LOOCV approach against k-fold validation as with LOOCV we had a larger number of datapoints to learn the model from. The LOOCV method is approximately unbiased, because the difference in sample size between a training fold and the entire dataset is one datapoint [39, 40]. However, we note that the LOOCV tends to have a higher variance. This issue can potentially be resolved with a larger dataset and we expand on this in the *Discussion* section of the paper. The within-training-fold cross-validation (CV) is a standard approach for variable selection and hyper parameter tuning [41, 42]. Therefore, the hyperparameter tuning was performed using a within-training-three-fold CV. We chose a three-fold CV instead of a larger k-fold CV to keep the number

computations relatively smaller. We measured the classification performance using the metric —area under the receiver operating curve (AUC). Using this metric, we compared the 15 different classifiers and assessed their classification performance. For each feature combination, the entire classification experiment was repeated five times with different random seeds to account for variability in model coefficients that arose during hyperparameter tuning (the within-training-three-fold data split changed each time). The model's outputs for the five repetitions of the experiment were combined by computing the median value for the classification probability output. The final AUC for each feature combination was determined by comparing this median output against the ground truth labels.

Feature popularity score

Apart from measuring classification performance, we computed a novel feature popularity score for the features used in classification, which allowed us to measure feature importance and compare features. Since all classification experiments comprised 20 patients, the leave-one-patient out cross-validation approach produced 20 models. As described previously we repeated the classification experiment five times for a given feature combination, resulting in $5 \times 20 = 100$ classifier models per feature combination. In these 100 models, for each feature f, we counted the number of models in which it had a non-zero LASSO coefficient. For each model, most model weights were zero and only few features were picked. Thus, the normalized count of the number of times a feature was picked was a natural choice for us to express the feature popularity. Then, the feature popularity score (ρ) for the feature f was computed as:

$$\rho = \frac{\# \text{ of nonzero coefficients for } f}{100} \tag{2}$$

This metric measured the popularity of the feature; the greater number of times the feature was picked by the LASSO-based classifiers, the more popular the feature was. Please note that we do not use the actual model weight value associated with a given feature in computing feature popularity. Further, we refer the readers to additional reading [43–45] to get a comprehensive understanding of other feature selection methods.

Results

The HRV, breathing, and physical activity are thought to influence Rett syndrome severity, but the underlying correlations are yet to be measured. Our experiments investigated the effect of HRV-metrics, actigraphy, MSTE, and MSNR features on Rett syndrome severity by assessing 15 binary classification models (Fig 4A) described in the section *Rett syndrome severity classification experiments*. For this, as described in the section *Data collection*, we utilized a cohort of 20 patients with Rett syndrome, of which 10 patients had low-severity patient-visits, nine patients had high-severity patient-visits and one individual had both low-severity and high-severity patient-visits. We illustrate this in Fig 4B.

Performance comparison of different feature combinations

The best binary Rett severity classifier used the feature combination of HRV-metrics, MSTE-features, and MSNR-features, and obtained a pooled-AUC equal to 0.92. When we used the four feature sub-groups separately for classification, HRV-metrics performed the best with a pooled-AUC equal to 0.76. This was followed by MSTE-features (AUC = 0.68), Actigraphy-metrics (AUC = 0.63) and MSNR-features (AUC = 0.55) respectively. When we used 2-combinations of feature sets, the feature combination of HRV and MSTE performed the best with an AUC = 0.80. This was followed by the following 2-combinations of feature sets: (1) HRV
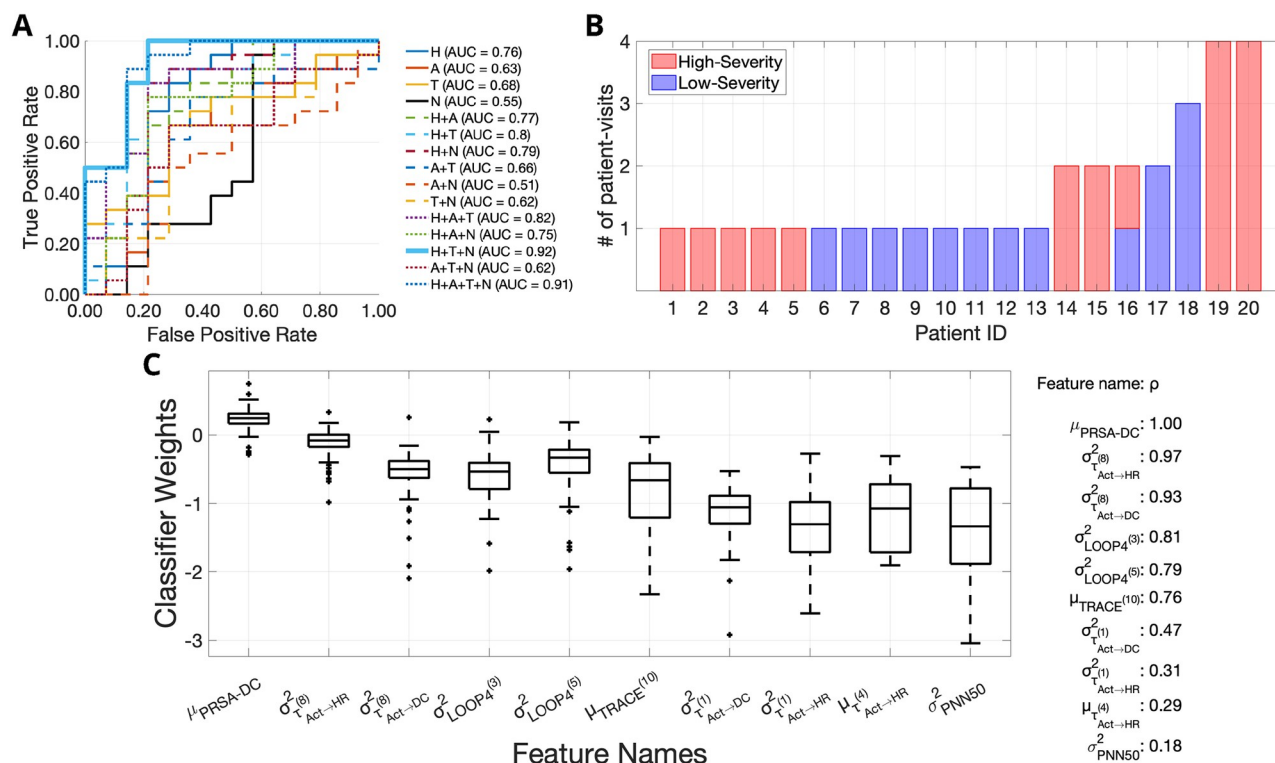
Fig 4. The feature combination of heart rate variability (HRV) metrics, multiscale transfer entropy, and multiscale network representation achieves the highest area under the receiver operating curve (AUC), equal to 0.92. (A) The ROCs for the 15 different Rett severity classifiers are provided. Each classifier uses a different subset-combination of the four feature sets, namely: (1) HRV metrics (H), (2) Actigraphy metrics (A), (3) Multiscale transfer entropy features (T), (4) Multiscale network representation features (N). The combination of H+T+N performed the best with a leave one-patient out cross-validation pooled-AUC equal to 0.92. The individual ROCs corresponding to the individual classifiers are shown using a combination of line styles and colors. The figure legend shows the pooled-AUC values for each feature combination. (B) A depiction of the number of patient-visits for each of the 20 patients showing the low-severity and high-severity patient-visits in different colors. (C) The top 10 most popular features used by the best classifier (H+T+N) for Rett severity classification are shown here, along with the corresponding feature coefficients. The mean deceleration capacity ($\mu_{PRSA-DC}$) is the most popular feature with a feature popularity score of 1 followed by $\sigma^2_{\tau^{(8)}_{Act\to HR}}$ and $\sigma^2_{\tau^{(8)}_{Act\to DC}}$ with feature popularity scores 0.97 and 0.93, respectively. In the top-10 most popular features for the H+T+N feature combination, two features were HRV-metrics ($\mu_{PRSA-DC}$ and σ^2_{PNN50}), three features were MSNR-features ($\sigma^2_{LOOP4^{(3)}}$, $\sigma^2_{LOOP4^{(5)}}$ and $\mu_{TRACE^{(10)}}$) and the rest five were MSTE-features.

+ MSNR (AUC = 0.79), (2) HRV + Actigraphy (AUC = 0.77), (3) Actigraphy + MSTE (AUC = 0.66), (4) MSTE + MSNR (AUC = 0.62), and (5) Actigraphy + MSNR (AUC = 0.51). When we used 3−combinations of feature sets, we obtained the following descending order of classification performances: (1) HRV + MSTE + MSNR (AUC = 0.92), (2) HRV + Actigraphy + MSTE (AUC = 0.82), (3) HRV + Actigraphy + MSNR (AUC = 0.75), (4) Actigraphy + MSTE + MSNR (AUC = 0.62). Finally, when all feature subsets were used together, we obtained an AUC = 0.91, only second to the combination of HRV-metrics, MSTE-features and MTNR-features (AUC = 0.92).

Feature popularity

The mean deceleration capacity ($\mu_{PRSA-DC}$) was the most popular feature for Rett syndrome severity classification. Using the novel formula described in the section *Feature popularity score*, we extracted the top-10 most popular features utilized by the best classifier (HRV, MSTE, and MSNR) and the distribution of their corresponding weights (Fig 4C). The feature $\mu_{PRSA-DC}$ came out on top with a feature popularity score $\rho = 1.00$. It was followed by the

following 9 features: (1) Variance (across all surrogate representations) of transfer entropy from the activity counts signal to the HR signal at the 8th coarse-graining scale ($\sigma^2_{\tau^{(8)}_{Act \to HR}}$) ($\rho =$ 0.97), (2) Variance (across all surrogate representations) of transfer entropy from the activity counts signal to the deceleration capacity signal at the 8th coarse-graining scale ($\sigma^2_{\tau^{(8)}_{Act \to DC}}$) ($\rho =$ 0.93), (3) Variance (across all surrogate representations) of the number of 4−loops in the network representation of the tuple—(HR signal, activity count signal, DC signal) at the 3rd coarse-graining scale ($\sigma^2_{LOOP4^{(3)}}$) ($\rho = 0.81$), (4) Variance (across all surrogate representations) of the number of 4−loops in the network representation of the tuple—(HR signal, activity count signal, DC signal) at the 5th coarse-graining scale ($\sigma^2_{LOOP4^{(5)}}$) ($\rho = 0.79$), (5) Mean (across all surrogate representations) of the trace of the adjacency matrix of the network representation of the tuple—(HR signal, activity count signal, DC signal) at the 10th coarse-graining scale ($\mu_{TRACE^{(10)}}$) ($\rho = 0.76$), (6) Variance (across all surrogate representations) of transfer entropy from the activity counts signal to the deceleration capacity signal at the 1st coarse-graining scale ($\sigma^2_{\tau^{(1)}_{Act \to DC}}$) ($\rho = 0.47$), (7) Variance (across all surrogate representations) of transfer entropy from the activity counts signal to the HR signal at the 1st coarse-graining scale ($\sigma^2_{\tau^{(1)}_{Act \to HR}}$) ($\rho =$ 0.31), (8) Mean (across all surrogate representations) of transfer entropy from the activity counts signal to the HR signal at the 4th coarse-graining scale ($\mu_{\tau^{(4)}_{Act \to HR}}$) ($\rho = 0.29$), (9) Variance of the 5-minute PNN50 measure i.e., the average number of pairs of adjacent beat-to-beat intervals differing by more than 50 ms, between the times 10 PM and 10 AM (σ^2_{PNN50}) ($\rho =$ 0.18). The feature weight for ($\mu_{PRSA-DC}$) was greater than zero for 93% of the time, suggesting an inverse relationship (due to the negative-log relationship between features and odds) between the feature values and the Rett disease severity (i.e., a higher values of $\mu_{PRSA-DC}$ corresponded to a higher chance of low-severity Rett syndrome). It conformed with our feature values as shown in Fig 5B. In S1 Fig we illustrate the top 5 most popular features for each of the 15 classifiers. Whenever HRV-metrics were used for classification (8 of the 15 instances), $\mu_{PRSA-DC}$ was the most important feature with a consistent feature popularity score of 1. When only Actigraphy-metrics were used for classification, the amplitude of the circadian rhythm measured using the cosinor rhytmometry was the most popular feature with a feature importance score of 0.94. When we used MSTE-features alone, $\sigma^2_{\tau^{(8)}_{Act \to DC}}$ was the most popular feature with a feature popularity score of 0.94. Finally, when we used the MSNR features alone, ($\sigma^2_{LOOP4^{(3)}}$) was the most popular feature with a feature popularity score of 0.90. Interestingly, the top features from the individual feature set models for HRV-metrics ($\mu_{PRSA-DC}$), MSTE-features ($\sigma^2_{\tau^{(8)}_{Act \to DC}}$), and MSNR-features ($\sigma^2_{LOOP4^{(3)}}$) were all featured as one among the top−5 most popular features in the best model (AUC = 0.92) that used the feature combination of HRV-metrics, MSTE-features, and MSNR-features. For the classifier that used all available features (HRV + Actigraphy + MSTE + MSNR), the following metrics featured as the top-5 most popular features: (1) $\mu_{PRSA-DC}$ ($\rho = 1.00$), (2) $\sigma^2_{\tau^{(8)}_{Act \to HR}}$ ($\rho = 0.98$), (3) $\sigma^2_{\tau^{(8)}_{Act \to DC}}$ ($\rho = 0.97$), (4) Inter-daily Stability ($\rho = 0.82$); (5) $\sigma^2_{LOOP4^{(3)}}$ ($\rho = 0.81$).

Discussion

As of 2022, no clinically meaningful disease-modifying treatments exist for patients with Rett syndrome. We instead rely on multiple therapeutics and symptomatic treatment strategies geared towards managing respiratory ailments, treating seizures, improving gastrointestinal function, and improving motor skills [46]. As new drugs and therapeutics are discovered, the
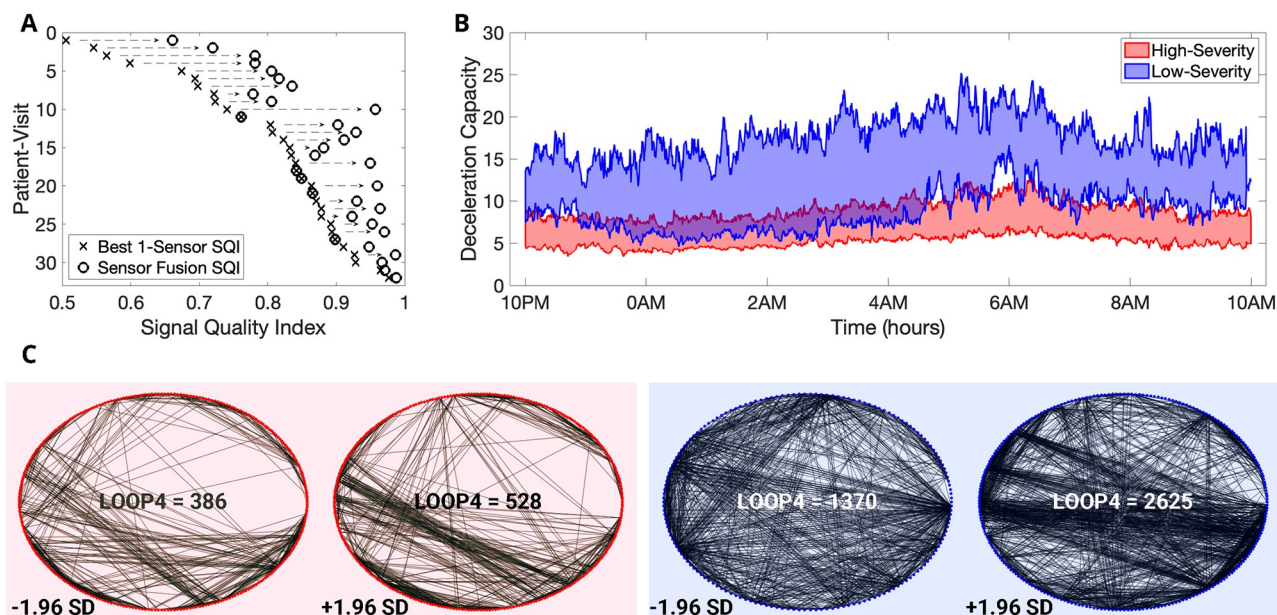
Fig 5. Data imputation techniques combined with first and second-order statistics improve classification performance. (A) An illustration of the effects of novel signal quality index-based ECG data imputation. We show the improvements in the average signal quality index of the electrocardiogram records for all 32 patient-visits. (B) The deceleration capacity (DC) values between 10 PM and 10 AM for the low-severity and high-severity groups. We show the DC values between the 75th and 25th percentile for low-severity and high-severity patient-visits. (C) We depict all edges corresponding to 4-loops in the networks constructed from trivariate time series (HR, DC, and Activity count signal) for a high-severity (left-red-panel) and a low-severity (right-blue-panel) patient at the 3rd coarse-graining time scale. For each patient, we show two surrogate LOOP4 networks at the following values for LOOP4: (1) $\mu_{LOOP4^{(3)}} - (1.96 \times \sigma_{LOOP4^{(3)}})$, (2) $\mu_{LOOP4^{(3)}} + (1.96 \times \sigma_{LOOP4^{(3)}})$. From the image, it is clear that both the number of 4-loops and the variance of the number of 4-loops are smaller for the low-severity patient compared to the high-severity patient.

need for objective measures that can be used in clinical trials increases. Neurological disorders, including Rett, suffer from difficult-to-measure symptoms. Most efficacy assessments are based on parent, clinician (and in some cases, patient) interpretation of symptom severity, which by nature introduces bias and often, exemplified by high placebo rates, results in a high noise to signal ratio. Hence, the need to establish objective measurements in patients directly, especially for symptoms that are difficult or impossible to observe, would open the door to evaluate therapeutics in novel ways and has the potential to expedite therapeutic development in multiple ways. Namely, (1) It would reduce bias; (2) It would help reduce clinical trial sample size by reducing the noise to signal ratio; and (3) It would facilitate shorter trial duration by capturing continuous data at home. The measurement of autonomic function could be an early biomarker of therapeutic efficacy and may be particularly relevant for curative therapeutics such as gene therapies that theoretically should improve or restore baseline function. If trials focus on efficacy measures that require learning and implementation (like mobility, communication, and hand use), this may take significantly longer to detect than an autonomic function correction that should not require learning. Thus, in the current study, we attempted to address these unmet needs by capturing physiological (ECG) and body activity (three-axis accelerometer data) from a 20-patient cohort. We chose to regress features extracted from the ECG signal and body activity measurements against the binarized CGI-S to correlate objective measures attained from a wearable biosensor to an overall symptom severity scale. We have shown that the inclusion of multiscale features (MSTE and MSNR) along with HRV-metrics provided a performance improvement of 21% in terms of the AUC (AUC = 0.92) when compared to using HRV-metrics alone (AUC = 0.76). This improvement is attributable to the

information embedded in the higher-order interactions between the HR, DC, and activity count signals. While the transfer entropy-based features enabled us to study the 2-dimensional variations between all combinations of HR, DC, and activity count signals in both directions, the network representations-based features enabled us to study the 3-dimensional interactions between these signals. Further, the computations of these features at multiple coarse-graining scales provided a means to analyze the signal modulations occurring due to different physiological phenomena (at different timescales). The coarse-graining scales we used in this study ranged from 1 to 10, which corresponded to a variation of sampling rate from 1/30 Hz to 1/300 Hz (i.e., from a sample every 30 seconds to a sample every 5 minutes). It allowed us to study the effect of different physiological processes, including respiration, vagal activity, and circadian rhythm, on the interactions between HR, DC, and activity count signals.

Our analysis suggested that the mean value of the DC of the HR signal captured on the BioStamp® nPoint between 10 PM and 10 AM was the most popular feature to classify low-severity patients from high-severity patients. It was the most popular feature in all the eight classifier models in which it was used, with a consistent and highest feature popularity score of 1. The computation of DC involved synchronizing the phases of all periodic components of the signal irrespective of their timescales [47]. The DC captured the overall deceleration of the sinus rhythm due to physiological processes that occurred at different timescales, including the vagal (parasympathetic) activity, the baroreflex, and the circadian rhythm. In Fig 5B, we depicted the variation of the 5-minute averages of DC between the times of 10 PM and 10 AM for both low-severity and high-severity patient visits. Apart from the mean value of the DC, we observed the emergence of the variance of transfer entropy at 8th coarse-graining scale from (1) Activity count signal to the HR signal and (2) Activity count signal to the DC signal among the most popular features. Further, the variance of the number of loops of size 4 in the network representations at coarse-graining scales 3 and 5 were among the top−5 most popular features in the best classifier model (AUC = 0.92). Thus, in Fig 5C, we illustrated an example of the network representations for a low-severe and a high-severe Rett patient at the 3rd, 5th, and 8th coarse-graining scales. In addition to demonstrating the viability of a wearable biosensor to estimate disease severity based on objective measurements directly in patients, another novelty in our work was handling missing data in signals captured using wearables for patient state analysis. We proposed three different techniques for this purpose. The first method of combining data from multiple channels boosted the amount of available data by 13.4% when compared to using a single channel (medial chest) and improved the average SQI by 10%. In Fig 5A, we illustrated this improvement in SQI for each of the 32 patient-visits due to the usage of the novel SQI based ECG imputation. The second method helped impute the activity count data and reduce missingness. The third and final technique of generating stochastic surrogates enabled us to compute the MSTE and MSNR features. It was specifically developed for this study, without which no multiscale analysis could have been performed. We computed the sample mean and sample variance of all MSTE and MSNR metrics across all surrogate representations and used them as features for our classification models. The sample mean is a measure of the samples' central tendency, and the sample variance is a measure of the spread of the samples. Since our method used a normal distribution to sample the random variables, and a normal distribution can be completely defined by a mean value and variance, it was rational to use the sample mean and variance to characterize the underlying MSTE and MSNR values.

The small sample size (n = 20) presented us with the challenge of choosing between binary and multi-class (seven-class) classification. The number of patients were not evenly distributed across the seven levels of the scale, and therefore it made regressing on these items very hard (and likely to skew to the more well represented classes). Moreover, with so few individuals, a binary task provided us with more datapoints per class (n≥10 datapoints per class). Thus,

separating patients by general severity into two categories, with the more mild patients in one group and the more severe patients in another group, was the best way to organize patients in a clinically meaningful way that would provide sufficient numbers. Identifying a binary signal that can distinguish more severe from less severe patients is meaningful for clinical development. We intend the binary outcome to result in a binary intervention (e.g. increase medication or not), and that a continuous (integer) scale makes it hard to know whether to treat someone or not. Ideally if there were enough patients it would be great to tease apart differences between each gradation on the severity scale but this was not feasible in such a small study and unlikely to be accomplished in a rare disorder. Further, all the models validated in our experiments involved training via a leave-one-patient out cross-validation. The small number of participants in the study and consequently the validated models potentially may not extend with a similar high performance on an external dataset. This is a limitation of our work and the further validation of trained models and features learned in our study on an external and larger dataset is part of our future work. However, Rett syndrome being a rare disorder (affects 1 in 10000 females) introduces the practical constraint of difficulty in recruiting large cohorts (n>100) for a single study. A potential solution to address this issue could be to create simultaneous ECG and three-axis acceleration datasets from Rett syndrome patients and to create a corpus of these datasets on open source repositories such as Physionet [48]. We did not compare our L1-regularized logistic regression method with different classifiers including random forests, k-nearest neighbors and XGBoost for binary classification or feature importance measurements. The comparison of various classifiers for a given classification task can be useful in determining the best model as well as improve our understanding of the underlying features. This is a limitation of our current study and is something we will explore in the future. Additionally, due to the large number of potential features, the feature importance measurement procedure can be made statistically robust by performing repeated test correction [49]. However, as the purpose of the study was exploratory in nature [50], and the cohort size (N = 20) was rather small, we did not perform the repeated test correction.

Conclusion

We developed a machine learning model that utilized features characterizing HRV, body movement, and the interaction between the two to estimate Rett syndrome severity in a group of 20 female Rett patients. For this, we developed a novel stochastic method for biosignal data imputation. We obtained the highest pooled-AUC equal to 0.92 utilizing the feature combination of HRV-metrics, MSTE-features, and MSNR-features. Further, the proposed approach provided us with physio-motor biomarkers that could be used in clinical trials as objective metrics to quantify the improvement in overall patient state. Specifically, the mean DC of the HR signal captured between 10 PM and 10 AM using the BioStamp® nPoint biosensor was the most popular feature with a feature popularity score equal to 1. In conclusion, our study (1) Implemented a novel data imputation technique for physiological signals, (2) Developed a machine learning model to estimate Rett disease severity, and (3) Developed objective measures that characterize the autonomic function in Rett syndrome.

Supporting information

S1 Fig. The novel feature popularity scores-based visualizations assist in determining the top-5 features for each of the 15 Rett severity classifiers. We show the normalized feature values for all 32 patient-visits for top-5 most popular features in each subplot. The data points with the red marker correspond to high-severity Rett patient-visits, and the data points with the blue marker correspond to low-severity Rett patient-visits. Further, we provide the feature

popularity (ρ) scores for these five features. The individual subplots correspond to different feature combinations of the following feature sets: (1) HRV—Heart Rate Variability; (2) Actigraphy; (3) MSTE—Multiscale Transfer Entropy; (4) MSNR—Multiscale Network Representation.
(TIF)

S1 Dataset.
(XLSX)

Author Contributions

Conceptualization: Pradyumna Byappanahalli Suresha, Daniel C. Tarquinio, Jana Von Hehn, Gari D. Clifford.

Data curation: Pradyumna Byappanahalli Suresha, Heather O'Leary, Daniel C. Tarquinio.

Formal analysis: Pradyumna Byappanahalli Suresha.

Funding acquisition: Daniel C. Tarquinio, Jana Von Hehn, Gari D. Clifford.

Investigation: Pradyumna Byappanahalli Suresha.

Methodology: Pradyumna Byappanahalli Suresha, Gari D. Clifford.

Project administration: Heather O'Leary, Jana Von Hehn, Gari D. Clifford.

Resources: Daniel C. Tarquinio, Jana Von Hehn, Gari D. Clifford.

Software: Pradyumna Byappanahalli Suresha.

Supervision: Daniel C. Tarquinio, Jana Von Hehn, Gari D. Clifford.

Validation: Pradyumna Byappanahalli Suresha.

Visualization: Pradyumna Byappanahalli Suresha.

Writing – original draft: Pradyumna Byappanahalli Suresha, Jana Von Hehn, Gari D. Clifford.

Writing – review & editing: Pradyumna Byappanahalli Suresha, Heather O'Leary, Daniel C. Tarquinio, Jana Von Hehn, Gari D. Clifford.

References

1. Rett A. [On a unusual brain atrophy syndrome in hyperammonemia in childhood]. Wiener medizinische Wochenschrift (1946). 1966; 116:723–6.

2. Chahrour M, Zoghbi HY. The story of Rett syndrome: from clinic to neurobiology. Neuron. 2007; 56:422–37. https://doi.org/10.1016/j.neuron.2007.10.001 PMID: 17988628

3. Tarquinio DC, Percy AK. Rett Syndrome: Clinical Aspects; 2016.

4. Chapleau CA, Lane J, Pozzo-Miller L, Percy AK. Evaluation of current pharmacological treatment options in the. management of Rett syndrome: from the present to future therapeutic alternatives. Current clinical pharmacology. 2013; 8:358–369.

5. Neul JL, Glaze DG, Percy AK, Feyma T, Beisang A, Dinh T, et al. Improving treatment trial outcomes for Rett syndrome: the development of Rett-specific anchors for the clinical global impression scale. Journal of child neurology. 2015; 30:1743–1748. https://doi.org/10.1177/0883073815579707 PMID: 25895911

6. Busner J, Targum SD. The clinical global impressions scale: Applying a research tool in clinical practice. Psychiatry (Edgmont (Pa: Township)). 2007; 4:28–37. PMID: 20526405

7. Julu POO, Kerr AM, Apartopoulos F, Al-Rawas S, Engerström IW, Engerström L,. et al. Characterisation of breathing and associated central autonomic dysfunction in the Rett disorder. Archives of disease in childhood. 2001; 85:29–37. https://doi.org/10.1136/adc.85.1.29 PMID: 11420195

8. Guideri F, Acampa M, DiPerri T, Zappella M, Hayek Y. Progressive cardiac dysautonomia observed in patients affected by classic Rett syndrome and not in the preserved speech variant. Journal of Child Neurology. 2001; 16:370–373. https://doi.org/10.1177/088307380101600512 PMID: 11392524

9. Guideri F, Acampa M, Blardi P, Lalla AD, Zappella M, Hayek Y. Cardiac dysautonomia and serotonin plasma levels in Rett syndrome. Neuropediatrics. 2004; 35:36–38. https://doi.org/10.1055/s-2004-815789 PMID: 15002050

10. Kumar A, Jaryal A, Gulati S, Chakrabarty B, Singh A, Deepak KK, et al. Cardiovascular autonomic dysfunction in children and adolescents with Rett syndrome. Pediatric neurology. 2017; 70:61–66. https://doi.org/10.1016/j.pediatrneurol.2017.01.010 PMID: 28351539

11. Singh J, Santosh P. Key issues in Rett syndrome: emotional, behavioural and autonomic. dysregulation (EBAD)-a target for clinical trials. Orphanet journal of rare diseases. 2018; 13:1–17. https://doi.org/10.1186/s13023-018-0873-8

12. Tarquinio DC, Hou W, Neul JL, Berkmen GK, Drummond J, Aronoff E, et al. The course of awake breathing disturbances across the lifespan in Rett syndrome. Brain and Development. 2018; 40:515–529. https://doi.org/10.1016/j.braindev.2018.03.010 PMID: 29657083

13. Guideri F, Acampa M. Sudden death and cardiac arrhythmias in Rett syndrome. Pediatric cardiology. 2005; 26:111. https://doi.org/10.1007/s00246-005-0784-z PMID: 15136902

14. Acampa M, Guideri F. Cardiac disease and Rett syndrome. Archives of Disease in Childhood. 2006; 91:440–443. https://doi.org/10.1136/adc.2005.090290 PMID: 16632674

15. Carroll MS, Ramirez JM, Weese-Mayer DE. Diurnal variation in autonomic regulation among patients with genotyped Rett syndrome. Journal of Medical Genetics. 2020; 57:786–793. https://doi.org/10.1136/jmedgenet-2019-106601 PMID: 32156713

16. Weese-Mayer DE, Lieske SP, Boothby CM, Kenny AS, Bennett HL, Silvestri JM,. et al. Autonomic nervous system dysregulation: breathing and heart rate perturbation during wakefulness in young girls with Rett syndrome. Pediatric research. 2006; 60:443–449. https://doi.org/10.1203/01.pdr.0000238302.84552.d0 PMID: 16940240

17. Weese-Mayer DE, Lieske SP, Boothby CM, Kenny AS, Bennett HL, Ramirez JM. Autonomic dysregulation in young girls with Rett Syndrome during nighttime in-home recordings. Pediatric pulmonology. 2008; 43:1045–1060. https://doi.org/10.1002/ppul.20866 PMID: 18831533

18. FitzGerald PM, Jankovic J, Percy AK. Rett syndrome and associated movement disorders. Movement disorders: official journal of the Movement Disorder. Society. 1990; 5:195–202.

19. Roze E, Cochen V, Sangla S, Bienvenu T, Roubergue A, Leu-Semenescu S, et al. Rett syndrome: an overlooked diagnosis in women with stereotypic hand movements, psychomotor retardation, Parkinsonism, and dystonia?. Movement disorders: official journal of the Movement Disorder Society. 2007; 22:387–389. https://doi.org/10.1002/mds.21276 PMID: 17216643

20. Temudo T, Maciel P, Sequeiros J. Abnormal movements in Rett syndrome are present before the regression period: A case study. Movement disorders: official journal of the Movement Disorder. Society. 2007; 22:2284–2287.

21. Temudo T, Ramos E, Dias K, Barbot C, Vieira JP, Moreira A, et al. Movement disorders in Rett syndrome: an analysis of 60 patients with detected MECP2 mutation and correlation with mutation type. Movement disorders: official journal of the Movement Disorder Society. 2008; 23:1384–1390. https://doi.org/10.1002/mds.22115 PMID: 18512755

22. Oakley NR. Validation with polysomnography of the Sleepwatch sleep/wake scoring algorithm used by the Actiwatch activity monitoring system. Bend: Mini Mitter, Cambridge Neurotechnology. 1997;.

23. Cakmak AS, Reinertsen E, Taylor HA, Shah AJ, Clifford GD. Personalized heart failure severity estimates using passive. smartphone data. IEEE; 2018. p. 1569–1574.

24. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet. Circulation. 2000; 101. https://doi.org/10.1161/01.CIR.101.23.e215 PMID: 10851218

25. Lee J, Nemati S, Silva I, Edwards BA, Butler JP, Malhotra A. Transfer entropy estimation and directional coupling change detection in biomedical time series. Biomedical engineering online. 2012; 11:1–17. https://doi.org/10.1186/1475-925X-11-19 PMID: 22500692

26. Shashikumar SP, Li Q, Clifford GD, Nemati S. Multiscale network representation of physiological time series for early prediction of sepsis Physiological measurement. 2017; 38:2235. https://doi.org/10.1088/1361-6579/aa9772 PMID: 29091053

27. Tarquinio DC. Treatment of Mitochondrial Dysfunction in Rett Syndrome With Triheptanoin;. Available from:. https://clinicaltrials.gov/ct2/show/NCT02696044.

28. Outcome Measures and Biomarkers Development Initiative;.

29. Burns C. MC10 "BioStamp" connects to your body, shares data;. Available from: https://www.slashgear.com/mc10-biostamp-connects-to-your-body-shares-data-06421694/.

30. Azuaje F, Clifford GD, McSharry P. Advanced methods and tools for ECG data analysis; 2006.

31. Vest AN, Poian GD, Li Q, Liu C, Nemati S, Shah AJ, et al. An open source benchmarked toolbox for cardiovascular waveform and interval analysis. Physiological measurement. 2018; 39:105004. https://doi.org/10.1088/1361-6579/aae021 PMID: 30199376

32. Li Q, Mark RG, Clifford GD. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. Physiological measurement. 2007; 29(1):15. https://doi.org/10.1088/0967-3334/29/1/002 PMID: 18175857

33. Borazio M, Berlin E, Kücükyildiz N, Scholl P, Laerhoven KV. Towards benchmarked sleep detection with wrist-worn sensing units;. 2014. p. 125–134.

34. Cakmak AS, Clifford GD. Open Source Actigraphy Toolbox; 2020. Available from: https://github.com/cliffordlab/OS_Actigraphy_Toolbox.

35. Cakmak AS, Alday EAP, Suresha PB, Clifford GD. Open Source Actigraphy Toolbox; https://doi.org/10.5281/zenodo.4287769, 2020.

36. Clifford GD, Azuaje F, McSharry P, et al. Advanced methods and tools for ECG data analysis. vol. 10. Artech house Boston; 2006.

37. Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of biological signals. Physical review E. 2005; 71(2):021906. https://doi.org/10.1103/PhysRevE.71.021906 PMID: 15783351

38. Nemati S, Edwards BA, Lee J, Pittman-Polletta B, Butler JP, Malhotra A. Respiration and heart rate complexity: effects of age and gender assessed by band-limited transfer entropy. Respiratory physiology & neurobiology. 2013; 189:27–33. https://doi.org/10.1016/j.resp.2013.06.016 PMID: 23811194

39. Lachenbruch PA, Mickey MR. Estimation of error rates in discriminant analysis. Technometrics. 1968; 10(1):1–11. https://doi.org/10.1080/00401706.1968.10490530

40. Luntz A. On estimation of characters obtained in statistical procedure of recognition. Technicheskaya Kibernetica. 1969; 3.

41. Baumann K. Cross-validation as the objective function for variable-selection techniques. TrAC Trends in Analytical Chemistry. 2003; 22(6):395–406. https://doi.org/10.1016/S0165-9936(03)00607-1

42. Suresha PB, Cakmak AS, Da Poian G, Shah AJ, Vaccarino V, Bliwise D, et al. Obstructive Sleep Apnea Classification in a Mixed-Disorder Elderly. Male Population Using a Low-Cost Off-Body Movement Sensor. In: 2019 IEEE EMBS International Conference on Biomedical & Health. Informatics (BHI). IEEE; 2019. p. 1–4.

43. Feng D, Chen F, Xu W. Efficient leave-one-out strategy for supervised feature selection. Tsinghua Science and Technology. 2013; 18(6):629–635. https://doi.org/10.1109/TST.2013.6678908

44. Chandrashekar G, Sahin F. A survey on feature selection methods. Computers & Electrical Engineering. 2014; 40(1):16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024

45. Jović A, Brkić K, Bogunović N. A review of feature selection methods with applications. In: 2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO). Ieee;. 2015. p. 1200–1205.

46. Katz DM, Bird A, Coenraads M, Gray SJ, Menon DU, Philpot BD, et al. Rett syndrome: crossing the threshold to clinical translation. Trends in neurosciences. 2016; 39:100–113. https://doi.org/10.1016/j.tins.2015.12.008 PMID: 26830113

47. Bauer A, Kantelhardt JW, Barthel P, Schneider R, Mäkikallio T, Ulm K, et al. Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction: cohort study. The Lancet. 2006; 367:1674–1681. https://doi.org/10.1016/S0140-6736(06)68735-7 PMID: 16714188

48. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. circulation. 2000; 101(23):e215–e220. https://doi.org/10.1161/01.CIR.101.23.e215 PMID: 10851218

49. Noble WS. How does multiple testing correction work? Nature biotechnology. 2009; 27(12):1135–1137. https://doi.org/10.1038/nbt1209-1135 PMID: 20010596

50. Streiner DL, Norman GR. Correction for multiple testing: is there a resolution?. Chest. 2011; 140(1):16–18. https://doi.org/10.1378/chest.11-0523 PMID: 21729890