Received April 2, 2020, accepted April 16, 2020, date of publication April 20, 2020, date of current version May 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2988722

Automated Method for Detecting Acute Insomnia Using Multi-Night Actigraphy Data

MAIA ANGELOVA[1], CHANDAN KARMAKAR[1], (Member, IEEE), YE ZHU[1], (Member, IEEE), SEAN P. A. DRUMMOND[2], AND JASON ELLIS[3]

[1]School of Information Technology, Deakin University, Geelong, VIC 3125, Australia
[2]Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Melbourne, VIC 3800, Australia
[3]Department of Psychology, Northumbria University, Newcastle upon Tyne NE1 8ST, U.K.

Corresponding author: Maia Angelova (maia.a@deakin.edu.au)

This work was supported by the Newton Advanced Fellowship through The Academy of Medical Sciences U.K. The work of Ye Zhu was supported by the Deakin University.

ABSTRACT In this paper we propose a new machine learning model for classification of nocturnal awakenings in acute insomnia and normal sleep. The model does not require sleep diaries or any other subjective information from the individuals who took part of the study. It is based on nocturnal actigraphy collected from pre-medicated individuals with acute insomnia and normal sleep controls. We have derived dynamical and statistical features from the actigraphy time series data. These features are combined using two machine learning techniques namely Random Forest (RF) and Support Vector Machine (SVM). RF shows better performance (accuracy - 84%) than SVM (73%) in classifying individuals with insomnia from healthy sleepers. The developed model provides a signature of the condition of acute insomnia obtained from actigraphy only and is very promising as a tool to detect the condition in a non-invasive way and without sleep diaries or any other subjective information.

INDEX TERMS Acute insomnia, actigraphy, machine learning, insomnia detection, dynamical features.

I. INTRODUCTION

Insomnia is characterised by the inability to fall asleep or stay asleep and/or waking too early and being unable to fall back asleep. The middle of the night awakenings could be one long awakening or several short awakenings (generally at least a few minutes in length). The break-in sleep is not a micro-arousal and as a result, the person remains awake for an extended period of time. Although the prevalence of insomnia is high, many individuals with insomnia do not seek medical attention for their sleep difficulties.

Thus, insomnia remains under-diagnosed like other sleep disorders such as sleep apnoea. Insomnia comes in many different forms including acute and chronic (based on duration), but in this paper, we focus our attention only on acute insomnia [10], [11], [13]. Acute insomnia, in this study, was defined on the basis of meeting all the criteria for a case of DSM-5 for insomnia disorder, which is characterised as an insomnia disorder lasting between two weeks and three months [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Wen-Sheng Zhao.

Actigraphy is the monitoring of physical activity patterns over time. It is non-invasive and cost-effective method of monitoring activity over several days up to several weeks. It has been also found to be an economic and ecologically valid method to study sleep-wake patterns in humans. Although the American Academy of Sleep Medicine (AASM) suggests actigraphy is not required for a diagnosis of insomnia, they suggest it as an optional and complementary method to examine other suspected sleep disorders [21], [26]. Recently, researchers have recorded actigraphy in addition to sleep diaries for monitoring sleep outside the laboratory.

Van Hees *et al.* [28] have studied sleep quality from wrist-worn raw-data-accelerometers in the absence of sleep diaries for healthy subjects. Their heuristic algorithm uses the variance in estimated z-axis angle and makes basic assumptions about sleep interruptions. Natale *et al.* [19] have studied primary insomnia using actiwatches. However, studies using actigraphy to understand insomnia have been largely inconsistent in terms of the choice of metrics and detailed analysis of the signal [16], [18]–[20], [30]. Some of the potential reasons for the lack of consistency are that the methodology,

and in particular, the statistical methods of analysis, may not be sensitive and have never been done in acute insomnia except for our studies [10], [11], [13]. Problems in identifying and eliminating artifacts in the data is another reason that actigraphy is not solely indicated at present for the diagnosis of many sleep disorders [25]. Acute insomnia is not well-studied, our papers [10], [11], [13] are the only works that investigated the phenomenon. In addition, individuals with insomnia usually demonstrate night-to-night variability in their sleep and as such one or two nights of actigraphy or even polysomnography (PSG) are unlikely to be representative [6], [22], [27]. Thus, although actigraphy contains rich information that can be used to extract insights about sleep, actigraphy data are still under-utilised to develop a robust model for assessing insomnia.

In this study, we aim to develop an automated method to detect acute insomnia using nocturnal actigraphy data over 7 nights. We will base our conclusions entirely on results from actigraphy, without the use of sleep diaries, Insomnia Severity Index or other subjective metrics for the evaluation of sleep problems. In actigraphy-based sleep research, a major challenge is to derive sleep parameters without additional information from sleep diaries [2], [8], [9], [28], [29]. Thus, the use of actigraphy as a cheap and non-invasive secondary measure of insomnia without sleep diaries requires the development of novel and consistent methods to derive indicators of sleep patterns.

In our previous work [13], fractal analysis techniques were implemented for the first time to study acute insomnia from actigraphy. Power spectrum and detrended fluctuation analysis were applied to search for $1/f$ scaling, meaning that the power of the signal is dominated by an inverse power law with the frequency f. This appears as a linear relationship in the log-log plot of the power as a function of the frequency of the time series. $1/f$ scaling, known also as pink noise, is associated with long-range correlation of the time series and high complexity, measured by complexity parameter $\sim$ 1. Our study concluded that there are variations in $1/f$ - type of scaling in the nocturnal signal of individuals with acute insomnia compared to healthy controls. The subjects with acute insomnia displayed complexity parameter in the range [0.75,1.25] corresponding to long-range correlations in the time series, owing to increased night-time arousals. The healthy controls displayed a complexity parameter in the range of (0.5-0.75), associated with positive but weaker correlations in the time series. Our second study [11] investigated the effect of circadian rhythms on the population of individuals with acute insomnia compared to healthy controls using complete day-night actigraphy. The paper proposed a novel adaptive technique for decomposition of the experimental time series, including day and night actigraphy, in a model-free way, into a trend, quasi-periodic components and noise fluctuations. However, none of the existing studies developed a robust model for detecting acute insomnia and healthy sleep from actigraphy data.

In this paper, we will develop a machine learning model using multi-night actigraphy data for classifying acute insomnia. The contributions and novelty of this study are as follows:

- Seven nights of nocturnal actigraphy data have been used to develop a model for classifying sleep quality into good and bad sleep nights. The subjective classification of the subjects in the study (individuals with Acute Insomnia or Healthy Sleepers) is then performed using the same model.
- Nocturnal actigraphy for each subject is labelled based on information derived from actigraphy data rather than using a sleep diary. Thus, the proposed method is only dependent on the actigraphy data.
- To the best of our knowledge, this is the first machine learning model developed using only the actigraphy data to assess acute insomnia.

II. METHODS

Figure 1 shows the workflow of the proposed method for detecting acute insomnia from nocturnal actigraphy data. The following subsections will discuss the stages of the workflow in detail.

A. DATA

Our analysis uses actigraphy time series data from a publicly available data source of the publication [13] (supplementary material). The data collection for the original study was approved by the University of Glasgow Ethics Committee. Actigraphy offers an objective measure of the level of physical activity using movement counts per sample time interval. In modern devices, the movement counts are usually taken per 30-second or 1-minute basis, called epochs. The data were recorded with an Actiwatch device, worn at all times throughout day and night, for a period of 2 weeks. However, not all subjects completed the entire 2-week period. The paper [13] compared night-time physical activity patterns, based on complexity parameters and 1/f-type of scaling, from acute insomnia subjects with those of asymptomatic controls and included individuals of all age ranges. The following study [11] was based on a subset of the same data, focusing on young individuals, and one-week continuous and complete day-and-night actigraphy time series data.

This study comprised 45 young adults (age: 18-40 years), including 24 asymptomatic controls (age: 28 $\pm$ 6 years) and 21 acute insomnia subjects (age: 25 $\pm$ 6 years). The control cohort was composed of self-declared healthy populations with no known problem with sleeping and the insomnia cohort (clinically assessed) did not have any known co-morbidity. All subjects were requested to remain in bed between 10pm and 8am next day. However, this did not prohibit the subject from going to bed before 10pm or leave the bed after 8am. As a result, traditional sleep parameter such as sleep latency, the difference between time to bed and fall asleep, can not be calculated since the Actiwatch used to

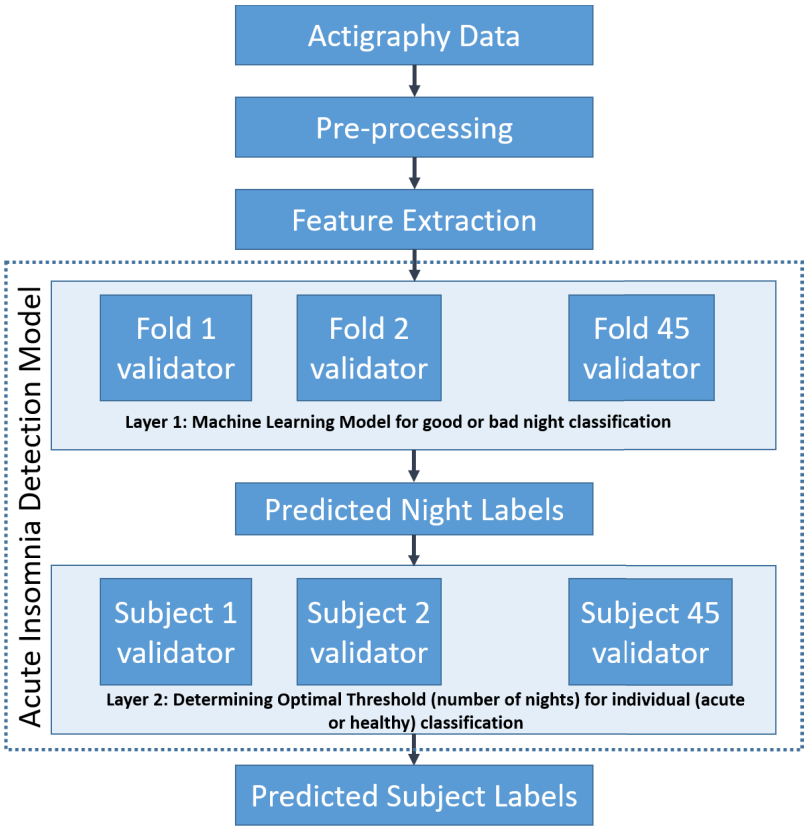


FIGURE 1. Workflow of the proposed machine learning model for classifying individuals with acute insomnia from healthy sleepers.

collect the data was unable to record the point of sleep onset automatically.

B. PRE-PROCESSING

The data were pre-processed and prepared for feature extraction. Only nocturnal actigraphy was used to extract features. The nocturnal period analysed in this study is 10pm to 8am (inclusive), since subjects were requested to be in bed during this period in this study. Actigraphy data are a measurement of the number of times (summed over 1 minute epochs) the acceleration of an activity exceeds a reference value. The clinician or the analyst could set the wake sensitivity threshold as high (20 counts per epoch), medium (40 counts per epoch), low (80 counts per epoch). However, there are no specific recommendations for how to use these thresholds [19]. In addition, the use of threshold act as a low-pass filter, which reduces the fluctuations from the nocturnal actigraphy signal. Therefore, features extracted after applying different thresholds will not be the same and may contain diverse information. As a result, the filtering process will enrich the feature set extracted from the nocturnal actigraphy signal that will be used by the machine learning model to classify subject with acute insomnia from healthy sleepers. In this study, we have used four different wake thresholds zero, low, medium, and high with activity counts of 0, 20, 40 and 80 respectively. These thresholds can be explained as very high, high, medium and low wake sensitive representation of nocturnal actigraphy signal and similar threshold-based filtering was also reported by Domingues and Sanches [1]. After extracting nocturnal actigraphy signal, the counts are set to zero if their values are less than or equal to a specific threshold value and pass to next stage for feature extraction. Since there are four different threshold values, this process results in four different sets of feature values from one nocturnal actigraphy signal.

C. FEATURE EXTRACTION

Linear and nonlinear features of nocturnal actigraphy signal were extracted using traditional statistical operators (mean, standard deviation (SD)), Poincaré plot (*SD*1, *SD*2, *Ratio*, *Area*, *CCM*), defined by equations (1a)-(1d), and entropy measure (sample entropy *SampEn*), given by equation (2). In addition, total sleep time (TST), wake after sleep onset (WASO) and sleep-wake ratio (SWR) were extracted automatically from the actigraphy signal. These three features were chosen as they can be derived without sleep diaries. Detailed definitions of these features are given in Section II-C.4. Therefore, a total of 10 different features were extracted from the actigraphy signals (Table 1). Since we used four filters for defining sleep or wake, it generated four different sleep-wake patterns from the nocturnal actigraphy data. Therefore, we extracted the same set of features after each

TABLE 1. List of the name of features extracted from Actigraphy. Definitions and measurement methods of feature subsets {1, 2}, {3, 4, 5, 6}, {7}, and {8, 9, 10} are detailed in subsections II-C.1, II-C.2, II-C.3 and II-C.4 respectively.

Index	Feature Name
1	Mean (Arithmetic)
2	SD (Standard Deviation)
3	SD1 (Poincaré Map)
4	SD2 (Poincaré Map)
5	Ratio (SD1/SD2 Poincaré Map)
6	CCM (Complex Correlation Measure)
7	SampEn (Sample Entropy)
8	TST (Total Sleep Time)
9	WASO (Wake time After Sleep Onset)
10	SWR (Sleep Wake Ratio)

filtering, which resulted in $10 \times 4 = 40$ features for each night of actigraphy data.

1) STATISTICAL FEATURES

Arithmetic Mean and SD of nocturnal actigraphy signal were calculated to measure average amplitude and amount of variation in nocturnal actigraphy signal. It is expected that both, the average amplitude and variation should be greater for insomnia subjects than for healthy controls.

2) POINCARÉ PLOT

The Poincaré plot is a well-known method for quantifying lagged correlation of any time-series signal. It is most frequently used in quantifying heart rate variability. It is taken from nonlinear dynamics, to visualise the heart rate variability [31] i.e., the nature of fluctuation in heart rate time series. It has also been used to analyse other time series signal including an accelerometer to extract feature for seizure detection [17]. Standard Poincaré plot is a 2D plot, in which successive sample of a time series represents a point in the plot (Figure 2). Therefore, i^{th} point in the plot represents the sample (T_i, T_{i+1}), where $\mathbf{T}$ is a vector that represents the time-series signal. The standard descriptors or parameters used to measure the variability of the plots are $SD1$ (short-term variability), $SD2$ (long-term variability) and their *Ratio* ($SD1/SD2$), defined by equations (1a)-(1c) [14]. In addition, the Poincaré plot has also been used to measure the dynamical properties of time-series signal such as the complex correlation measure (CCM), introduced in [15] by equation (1d).

For a given time series $\mathbf{T} = (T_1, T_2, \ldots, T_N)$ of length N, the above mentioned parameters are calculated using equations (1a)- (1d).

$$SD1 = \sqrt{\frac{1}{\sqrt{2}(N-1)} \sum_{i=1}^{N-1} (T_i - T_{i+1})^2} \tag{1a}$$

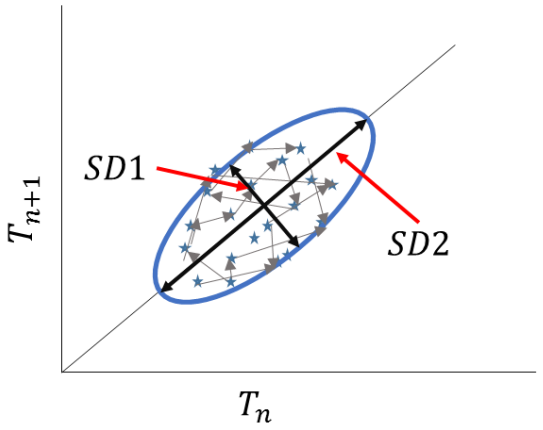


FIGURE 2. Example Poincaré plot with standard descriptors *SD1* and *SD2*. *SD1* is called short-term variability measure and quantifies the variability along the minor axis of the fitted ellipse. On the other hand, *SD2* is called long-term variability measure and quantifies the variability along the major axis of the fitted ellipse.

$$SD2 = \sqrt{\frac{1}{\sqrt{2}(N-1)} \sum_{i=1}^{N-1} (T_i + T_{i+1} - \overline{T})^2} \tag{1b}$$

$$Ration = \frac{SD1}{SD2} \tag{1c}$$

$$CCM = \frac{1}{N-2} \sum_{i=1}^{N-2} |A_i|. \tag{1d}$$

Here, $\overline{T}$ is the arithmetic mean of T_i, $i = 1, \ldots, N-1$. $|A_i|$ is the determinant of the matrix used to compute the area of triangle for the i^{th} window of the time series. For more details see [15].

3) SAMPLE ENTROPY (*SampEn*)

Sample Entropy is a measure of irregularity of a time series signal. It has been widely applied in extracting biomarkers from physiological signals. *SampEn* is an approximation of the conditional probability of two segments matching at a length of $m+1$ if they match at m, where the match is decided by the tolerance parameter r [24]. In *SampEn*, self-matches are excluded when calculating the percentage of the vectors $T_m(j)$ that are within r of $T_m(i)$, i.e., $A_i^{(m)}(r) = \frac{N_i^{(m)}(r)}{N-m-1}$, where $N_i^{(m)}(r)$ indicates the number of j's that meet $d_{i,j} \leq r$, and $1 \leq j \leq N-m, j \neq i$. The average of the percentage $A_i^{(m)}(r)$ over $1 \leq i \leq N-m$ is then defined by $\Psi^{(m)}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^{(m)}(r)$. In similar means, $\Psi^{(m+1)}(r)$ is defined after increasing the dimension to $m+1$. The SampEn value of the actigraphy time-series can be calculated by equation (2):

$$\text{SampEn}(N, m, r) = -\ln \frac{\Psi^{(m+1)}(r)}{\Psi^{(m)}(r)}. \tag{2}$$

where N is the length of the time series.

4) SLEEP FEATURES

In this study, we empirically selected the Time In Bed (TIB) as 10 hours (10pm to 8am) for all subjects, which determines

the length of nocturnal actigraphy data. (Note, the design of the experiment required participants to stay in bed for this period of time). The following sleep features such as Total Sleep Time (TST), Wake After Sleep Onset (WASO) and Sleep-Wake Ratio (SWR) were determined from the nocturnal actigraphy signal.

TST is the duration in minutes of all sleep epochs (when activity count is zero) between Sleep Onset (SO) time and sleep end (8am in our case). In this study, we used a 10-minute period of inactivity after the start of bedtime (10pm) to determine the SO time and used it for calculating TST.

WASO is the duration in minutes calculated by summing of all wake epochs between SO time and sleep end (8am in our case).

SWR is the ratio of TST and WASO given by equation (3),

$$SWR = \frac{TST}{WASO}. \tag{3}$$

D. BUILDING ACUTE INSOMNIA DETECTION MODEL

In this study, we have proposed a two-layered model for acute insomnia detection, as shown in Figure 1. At the first layer, a supervised machine learning model is developed with k-fold cross-validation technique for predicting night labels of all subjects. Since we did not use sleep diaries, we labelled each night of healthy sleepers as a "good night sleep" and night of insomnia subjects as "bad night sleep". These labels were used to build supervised machine learning models for night classification. In this study, we have used $k = 45$, which means that the original dataset is randomly partitioned into 45 equal size subsets/folds. During each iteration, the classifier sequentially picked one of the folds as the test dataset and gets the classification results based on the trained the model using the remaining 44 folds. The reason for splitting 45 folds is to predict 7 nights per model. Therefore, after 45 iterations the aggregated predicted night labels were equal to the entire data set. Figure 3 illustrates how 45 fold cross-validation works. Two supervised machine learning models namely Support Vector Machine (SVM) and Random Forests (RF) were used to build the nigh level prediction model.

In the second layer of the model (Figure 1), the frequency of bad nights (predicted by the ML model at layer 1) is used as a feature for individual classification. We used leave-one-subject-out approach to optimise the threshold value (number of "bad nights" for a subject to be classified as "Individuals with insomnia") for best subjective classification and then use that threshold for classifying the test subject as either "Healthy" or "Individuals with insomnia". In this approach, at each iteration 44 subjects' predicted night labels were used for selecting the optimal threshold and then that was used for classifying the test subject. Therefore, after 45 iterations the predicted labels of all subjects were aggregated and used for performance analysis.

1) SUPPORT VECTOR MACHINES (SVM)

SVM [7] are powerful supervised learning models which have been widely used for two-group classification problems

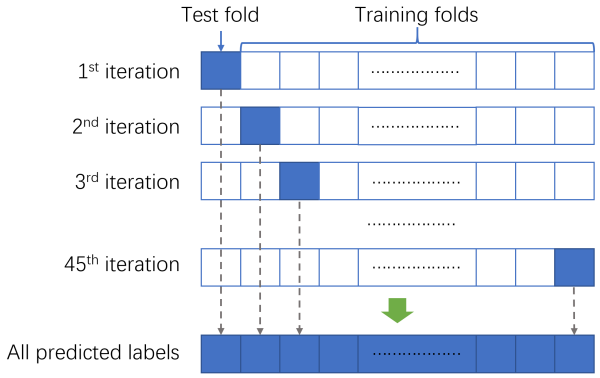


FIGURE 3. Demonstration of N-fold cross-validation process, where N=45 is the total number of subjects used in this study. As shown above, during each iteration one subject is left out for testing and the model is build using features from N-1 (i.e., 45) subjects. After completing all iterations, we get predicted labels for all subjects that are used for performance evaluation.

in a variety of domains. An SVM model maps each instance from the training data in a high/infinite-dimensional space such that the instances of different categories are able to be divided by a hyperplane or decision boundary. Then a class of test/new instance is predicted by mapping the instance into the same space and find which category it locates. The mathematical formulation is outlined as follows [12]:

Given the training data x_j along with their binary categories y_j, the equation of a hyperplane is

$$f(x_j) = \beta x_j^T + b = 0 \tag{4}$$

where $\beta \in R_d$, b is a real number and x_j^T is the transpose of the vector x_j.

The best separating hyperplane is to find β and b which minimise $||\beta||$ such that for all data points (x_j, y_j),

$$y_j f(x_j) \geq 1 \tag{5}$$

where the point x_j on the boundary are the support vectors, i.e., $y_j f(x_j) = 1$.

Searching for the optimal hyperplane is a quadratic programming problem. When the data may not be linearly separated using a hyperplane, SVM can use different kernel functions for nonlinear transformation. In this paper, for simplicity, we use the hyperplane (linear kernel) and standardise the predictors before training the classifier in SVM for two-class learning.[1]

2) RANDOM FORESTS (RF)

RF [4] are one of the most popular ensemble learning methods for classification. The method constructs a large set of decision trees independently using the training dataset such that the nodes of each decision tree are a randomly selected subset of features from the dataset.

A decision tree [5] is a flowchart-like structure where each internal node represents an evaluation or condition on one

[1]We use the function "fitcsvm" in Matlab with parameters settings: 3 (default) for "PolynomialOrder", "auto" for "KernelScale" and "1" for "BoxConstraint".

feature of the data, each branch represents the result of the evaluation, and each leaf node represents a response or label. The paths from the root node to a leaf node represent a classification rule. When predicting a response, the decisions follow the path along the tree from the root node down to a leaf node.

RF constructs a multitude of decision trees as a forest. Given testing data, RF uses the mode of the predicted classes of these individual trees as the final classification result. By averaging multiple deep decision trees, it greatly boosts the performance in the final model. In this paper, we use 100 random trees in RF and set the maximum depth of the trees to the size of training data.

E. EVALUATION METRICS

After getting all classification results (from both layers of the acute insomnia detection model), three standard measures [23] are applied to evaluate the performance of each classifier, including Sensitivity, Specificity and Accuracy. First, we compare the predicted classes with the ground truth (true labels) and calculate the true positives (TP), false negatives (FN), true negatives (TN) and false positives (FP). The three standard measures can be defined as:

(i) $Sensitivity = \frac{TP}{TP+FN} \times 100$, i.e., the true positive rate in percent.
(ii) $Specificity = \frac{TN}{TN+FP} \times 100$, i.e., the true negative rate in percent.
(iii) $Accuracy = \frac{TN+TP}{TP+TN+FP+FN} \times 100$.

In case of subject-level classification results, Sensitivity refers to the rate of correctly classified acute insomnia and Specificity refers to the rate of correctly classified healthy sleepers. Accuracy refers to the rate of correctly classified both acute insomnia and healthy sleepers.

F. AUC AND OTHER STATISTICAL ANALYSIS

Area Under the receiver-operating Curve (AUC) was calculated for each feature as well as during developing a machine learning model for feature ranking. AUC value of a feature can vary between 0.5 and 1, where the value 0.5 indicates that the feature has no capacity to differentiate two groups and the value 1.0 indicates complete separability.

Differences between Healthy and Insomnia groups for each feature were determined by non-parametric Mann-Whitney U Test. We use non-parametric test since the number of samples was small. $p < 0.05$ was accepted as statistically significant difference.

All statistical analyses were performed with MATLAB Statistics and Machine Learning Toolbox (Ver R2018a, The MathWorks Inc., Natick, MA, USA).

III. RESULTS AND DISCUSSION

This study uses the 7-night of actigraphy data for automatic detection of acute insomnia subject using machine learning models. A two-level model is developed to detect acute insomnia, where at the first level quality of sleep (good or bad) is predicted and at the second level, a threshold on seven nights is used to classify the subjects into acute insomnia

TABLE 2. *Mean±SD* **(standard deviation) values of features for Healthy and Insomnia groups are summarised in this table for (a) Intensity threshold = 0; (b) Intensity threshold = 20; (c) Intensity threshold = 40; and (d) Intensity threshold = 80. *p*-value is calculated using non-parametric Mann-Whitney U Test and $p < 0.05$ indicates that the feature values are different between Healthy and Insomnia groups. AUC (Area under the receiver-operating curve) value indicates the capacity of the feature in differentiating Insomnia subjects from Healthy and AUC=1.0 indicates the groups are fully separable.**

(a) Intensity threshold=0

Feature	Healthy	Insomnia	p	AUC
Mean	245.61±102.1	282.85±174.64	0.500	0.52
SD	296.67±115.7	336.71±207.28	0.304	0.53
SD1	107.02±40.86	140.85±88.46	0.001	0.61
SD2	266.96±123.22	335.54±245.97	0.169	0.55
Ratio	0.43±0.12	0.46±0.1	0.014	0.58
CCM	0.17±0.04	0.18±0.04	0.012	0.58
SampEn	0.16±0.07	0.16±0.07	0.740	0.51
TST	375.36±90.34	322.76±122.9	0.000	0.63
WASO	102.08±55.34	109.95±119.97	0.005	0.59
SWR	0.29±0.32	0.55±1.16	0.035	0.57

(b) Intensity threshold=20

Feature	Healthy	Insomnia	p	AUC
Mean	295.87±110.5	320.84±178.11	0.954	0.50
SD	305.73±120.54	345.04±212.29	0.285	0.54
SD1	107.23±40.86	141.08±88.4	0.001	0.61
SD2	267.18±123.28	335.75±246.01	0.170	0.55
Ratio	0.43±0.12	0.46±0.11	0.014	0.58
CCM	0.17±0.04	0.18±0.04	0.011	0.58
SampEn	0.16±0.07	0.16±0.07	0.825	0.51
TST	411.37±86.78	350.7±118.85	0.000	0.66
WASO	88.18±47.06	101.03±112.1	0.020	0.58
SWR	0.22±0.18	0.43±0.84	0.094	0.56

(c) Intensity threshold=40

Feature	Healthy	Insomnia	p	AUC
Mean	329.22±112.23	351.01±181.07	0.872	0.51
SD	309.3±122.15	351.01±216.84	0.227	0.54
SD1	107.68±40.8	141.58±88.32	0.001	0.61
SD2	267.43±123.42	336.13±246.15	0.165	0.55
Ratio	0.44±0.13	0.46±0.11	0.014	0.58
CCM	0.17±0.04	0.18±0.04	0.010	0.58
SampEn	0.15±0.07	0.15±0.07	0.656	0.51
TST	432.38±78.75	371.98±116.56	0.000	0.66
WASO	83.27±50.19	96.8±106.15	0.061	0.56
SWR	0.21±0.21	0.36±0.66	0.333	0.53

(d) Intensity threshold=80

Feature	Healthy	Insomnia	p	AUC
Mean	389.48±116.52	409.61±182.43	0.920	0.50
SD	313.02±125.88	360.96±224.78	0.098	0.55
SD1	108.67±41.11	142.97±88.3	0.001	0.61
SD2	267.86±123.97	336.85±246.71	0.160	0.55
Ratio	0.44±0.13	0.47±0.11	0.011	0.58
CCM	0.17±0.05	0.18±0.04	0.012	0.58
SampEn	0.12±0.06	0.14±0.07	0.022	0.58
TST	462.67±73.76	410.38±111.99	0.000	0.64
WASO	74.17±43.92	92.16±96.97	0.362	0.53
SWR	0.17±0.14	0.3±0.52	0.944	0.50

and healthy groups. Since there is no specific sensitivity threshold of actigraphy data for wake detection [19], we have used four different threshold values {0, 20, 40 and 80} for feature extraction. Our study appears to be first in detecting acute insomnia using only the actigraphy data. The overall accuracy of 84% with 76% sensitivity and 92% specificity in this regard is an excellent beginning for automated analysis of acute insomnia using data collected in the home environment.

Our study is also different from existing studies with respect to validation against sleep diaries. Since the sleep diary is a low-cost subjective report of sleep and vastly reported as highly effective in differentiating good sleepers from insomnia [18], [19], previous studies compared actigraphy based measures or performances with measures derived from sleep diaries. However, in this study, we have compared the performance of our model with clinical diagnosis. Therefore, this study investigates nocturnal awakenings in acute insomnia and normal sleep that are independent of sleep diary.

The *Mean*±*SD* (standard deviation) values of features for healthy and insomnia subjects are summarised in Table 2. The feature values are higher in Insomnia than Healthy subjects except for TST (total sleep time), which shows the opposite trend. This indicates that other features are measure of activity which is positively correlated with the wake time. This trend of values remains the same or similar for all different thresholds of wake used in this study. Among all features, TST shows the maximum differential capacity among two groups that are measured as AUC value. The AUC value varied with different thresholds of wake and maximum AUC (0.66) is obtained for $threshold = 40$. However, the maximum AUC is very low compared to the standard AUC value reported in other studies. This low value of AUC can be attributed to the blind labelling of night actigraphy data (every night of a Healthy sleeper is good sleep and every night of an Insomnia subject is bad), which is unusual. In reality, every Healthy subject can have one or two bad sleep nights and every Insomnia subject can have one or two good sleep nights. This explains the reason behind the higher variance of feature values in each group and lower AUC value.

This low value of AUC means that we cannot separate well the normal and insomnia subject based only on the individually extracted features. Thus, we use advanced machine learning models which incorporate the relationship between multiple independent variables to obtain higher classification performance.

Table 3a shows the results for evaluation on individual nights, i.e., correctly predicting the quality of the sleep either good or bad. Although the accuracy (69%) of RF classifier for the individual night was higher than the SVM (63%), both were relatively low for potential clinical applications. This weak performance of individual night model can be contributed to the gross labelling of healthy subjects' night as good night (label 0) and insomnia subjects' night as bad night (label 1).

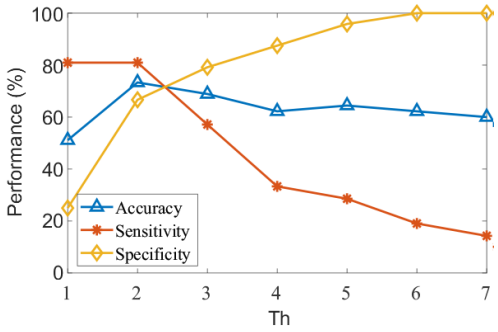
TABLE 3. Classification performances (accuracy, sensitivity and specificity) for the individual night (a) and subject (b) are summarised.

(a) Evaluation results on individual nights

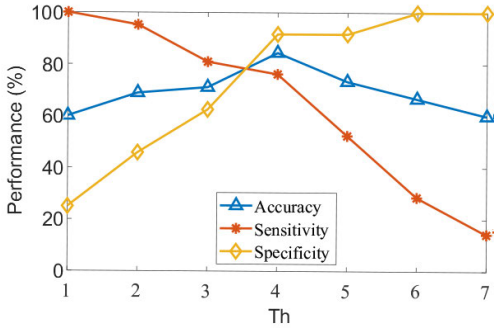
Algorithm	Accuracy	Sensitivity	Specificity
SVM	63%	45%	79%
RF	69%	64%	74%

(b) Evaluation results on subjects

Algorithm	Accuracy	Sensitivity	Specificity
SVM and Th	73%	81%	67%
RF and Th	84%	76%	92%



(a) Performance obtained using support vector machine (SVM) classifier at layer 1 (Figure 1).



(b) Performance obtained using random forest (RF) classifier at layer 1 (Figure 1).

FIGURE 4. Accuracy, Sensitivity and Specificity with varying threshold values (*Th*), which is defined as the number of bad night sleep. The sensitivity indicates the percentage rate of correctly detected insomnia, Specificity indicates the percentage rate of correctly detected healthy and Accuracy indicates the percentage rate of correctly detected healthy or insomnia subject. Performances at any threshold value *Th* is defined based on the relation that any correctly classified insomnia subject has ≥ *Th* number of bad night sleep. The intersecting point of three performance lines indicates the threshold value for which the balanced performance, rather than skewed sensitivity or specificity, can be obtained. Subplot (a) shows the performance obtained using support vector machine (SVM) classifier and (b) shows the performance obtained using random forest (RF) classifier at layer 1 of two-layer model (Figure 1).

Based on all predicted night labels from layer 1 (Figure 1), we can set a bad night threshold (Th) to distinguish acute insomnia subjects from healthy sleepers, i.e., a subject having at least Th bad nights out of 7-nights can be classified as an

acute insomnia subject. To clarify this statement, let consider the following example. Consider the case with 100 healthy subjects participating in the study. Now, let the specificity is 33% with $Th = 1$, this indicates that 33 of them are classified correctly as healthy. The remaining 67 healthy subjects have at least one bad night and are mis-classified as individuals with insomnia.

At layer 2 (Figure 1), we used leave-one-out cross validation strategy to find the best threshold distribution on subjects, i.e., leave one subject out (test subject), we train our model based on the predicted night labels of other 44 subjects, and then identify the best *Th* for getting the highest accuracy score. After that, we apply this threshold on the test subject to classify whether he/she is an acute insomnia subject or not.

Table 3b shows the evaluation results on subjects when learning the best *Th*. It is interesting to mention that the best thresholds learned by SVM and RF are 2 and 4 respectively, for all iterations of training. Although the overall accuracy of the model for the individual night is low (max accuracy 69% using RF classifier), it increases to 84% using RF classifier and optimised *Th* for subject classification (Table 3b). In addition, RF classifier has shown consistently better performance than SVM for both night and subject level classifications.

To further explore the effect of *Th* on the subject level performance of the proposed model, we have varied *Th* from 1 to 7 and calculated all performances from the predicted night labels at layer 1 (Figure 1). Figure 4 shows the results on subjects with different *Th* values on night labels provided by SVM and RF (from the layer one of the proposed model), respectively. The balanced performance (sensitivity, specificity and accuracy are equal) was obtained at $2 \leq Th \leq 3$ and $3 \leq Th \leq 4$ for the SVM and RF model outcome at layer 1, respectively. This provides a better insight to build a model with an expected level of sensitivity or specificity.

One limitation of this study was the sleep analysis period chosen (i.e. 10pm to 8am). This was chosen to account for the sleep period of all included subjects and as an exclusion criterion for those with erratic sleep.

IV. CONCLUSION

Acute insomnia is a condition that can go undetected and untreated that lead to general fatigue, decrease of productivity. The early and correct detection of acute insomnia is very important to identify and manage the condition. Laboratory investigation of acute insomnia with the current methods of PSG gives the opportunity to detect the condition only in the nights when PSG is performed. However, there are cases when the condition does not manifest itself in the sleep laboratory when the PSG investigation is done, and the condition may be undetected in this cases. At present, clinical assessment based on sleep diaries is the approved method of detection. Actigraphy is a cheap and non-invasive complementary method, which is becoming popular and widely used in sleep research. The ability to establish markers for acute insomnia from actigraphy signals only has a huge potential as an alternative method of early and objective diagnosis of the condition.

In this paper, we have made the following contributions to the methodology and insomnia research. We have established markers for nocturnal awakenings in acute insomnia derived

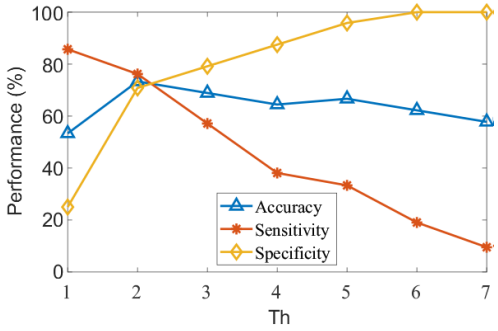
TABLE 4. Accuracy(%), sensitivity(%) and specificity(%) of support vector machine (SVM) and random forest (RF) models are summarised with different cross evaluation technique.

(a) 10-fold cross evaluation

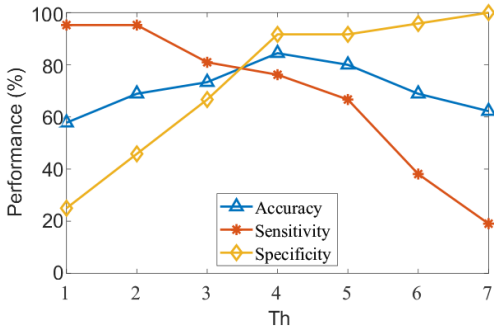
Algorithm	Accuracy	Sensitivity	Specificity
SVM	76%	81%	71%
RF	82%	71%	92%

(b) 20-fold cross evaluation

Algorithm	Accuracy	Sensitivity	Specificity
SVM	73%	76%	71%
RF	84%	76%	92%



(a) Performance obtained using support vector machine (SVM) classifier at layer 1 (Figure 1).



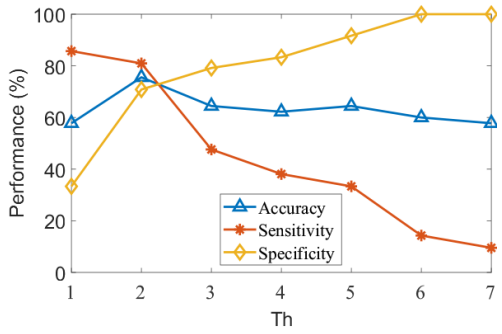
(b) Performance obtained using random forest (RF) classifier at layer 1 (Figure 1).

FIGURE 5. Evaluation results on subjects with 10-fold cross evaluation. Accuracy, Sensitivity and Specificity with varying threshold values (*Th*), which is defined as the number of bad night sleep. The sensitivity indicates the percentage rate of correctly detected insomnia, Specificity indicates the percentage rate of correctly detected healthy and Accuracy indicates the percentage rate of correctly detected healthy or insomnia subject. Performances at any threshold value *Th* is defined based on the relation that any correctly classified insomnia subject has ≥ *Th* number of bad night sleep. The intersecting point of three performance lines indicates the threshold value for which the balanced performance, rather than skewed sensitivity or specificity, can be obtained. Subplot (a) shows the performance obtained using support vector machine (SVM) classifier and (b) shows the performance obtained using random forest (RF) classifier at layer 1 of two-layer model (Figure 1).

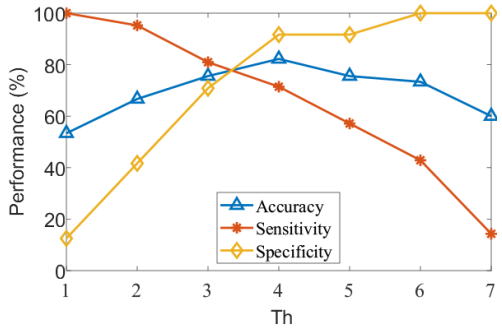
from wrist actigraphy data. We have combined these markers in the machine learning model and trained the models using data provided for 24 healthy sleepers and 21 subjects with acute insomnia aged between 20-45. Our results show that the machine learning model demonstrates significant accuracy, sensitivity and specificity. Thus, this model can be used to derive a robust signature for acute insomnia from wrist actigraphy data within one week. In the context of this paper, ''signature'' is a function objectively characterising acute insomnia, which combines and optimises the features.

We have not used sleep diaries deliberately in order to establish the accuracy of the models using secondary data only. We have used only night time actigraphy and selected complete data for one week from the available two-weeks data collection in order to avoid problems with missing data. determined from the actigraphy time series data. In addition, the algorithm based on RF is superior to those based on SVM in terms of accuracy and specificity and thus would be better suited for clinical purposes. Further study is under completion to compare these results with a study of individuals with chronic insomnia. The feature extraction combined with machine learning model and the following optimisation provides a hybrid machine learning tool for establishing a signature for acute insomnia.

Our markers and machine learning model for Acute Insomnia signature can provide a sound basis for building a tool for robust diagnostic and early treatment of this debilitating condition.



(a) Performance obtained using support vector machine (SVM) classifier at layer 1 (Figure 1).



(b) Performance obtained using random forest (RF) classifier at layer 1 (Figure 1).

FIGURE 6. Evaluation results on subjects with 20-fold cross evaluation. Accuracy, Sensitivity and Specificity with varying threshold values (*Th*), which is defined as the number of bad night sleep. The sensitivity indicates the percentage rate of correctly detected insomnia, Specificity indicates the percentage rate of correctly detected healthy and Accuracy indicates the percentage rate of correctly detected healthy or insomnia subject. Performances at any threshold value *Th* is defined based on the relation that any correctly classified insomnia subject has ≥ *Th* number of bad night sleep. The intersecting point of three performance lines indicates the threshold value for which the balanced performance, rather than skewed sensitivity or specificity, can be obtained. Subplot (a) shows the performance obtained using support vector machine (SVM) classifier and (b) shows the performance obtained using random forest (RF) classifier at layer 1 of two-layer model (Figure 1).

APPENDIX
CROSS VALIDATION RESULTS

Table 4a and Table 4b show the evaluation results on subjects using 10-fold and 20-fold cross evaluation for night modelling, respectively. Figure 5 and Figure 6 show the corresponding results on subjects with different *Th* values.

ACKNOWLEDGMENT

Ye Zhu would like to thank Deakin University for the funding of his RF position which preceded his current lectureship.

REFERENCES

[1] A. Domingues, T. Paiva, and J. M. Sanches, ''Sleep and wakefulness state detection in nocturnal actigraphy based on movement information,'' *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 426–434, Feb. 2014.
[2] K. N. Anderson, M. Catt, J. Collerton, K. Davies, T. von Zglinicki, T. B. L. Kirkwood, and C. Jagger, ''Assessment of sleep and circadian rhythm disorders in the very old: The newcastle 85+ cohort study,'' *Age Ageing*, vol. 43, no. 1, pp. 57–63, Jan. 2014.
[3] *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*, American Psychiatric Association, Philadelphia, PA, USA, 2013.
[4] L. Breiman, ''Random forests,'' *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[5] L. Breiman, *Classification Regression Trees*. Evanston, IL, USA: Routledge, 2017.
[6] D. J. Buysse, Y. Cheng, A. Germain, D. E. Moul, P. L. Franzen, M. Fletcher, and T. H. Monk, ''Night-to-night sleep variability in older adults with and without chronic insomnia,'' *Sleep Med.*, vol. 11, no. 1, pp. 56–64, Jan. 2010.
[7] C. Cortes and V. Vapnik, ''Support-vector networks,'' *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[8] I. C. da Silva, V. T. van Hees, V. V. Ramires, A. G. Knuth, R. M. Bielemann, U. Ekelund, S. Brage, and P. C. Hallal, ''Physical activity levels in three brazilian birth cohorts as assessed with raw triaxial wrist accelerometry,'' *Int. J. Epidemiol.*, vol. 43, no. 6, pp. 1959–1968, Dec. 2014.
[9] A. Doherty, D. Jackson, N. Hammerla, T. Plötz, P. Olivier, M. H. Granat, T. White, V. T. van Hees, M. I. Trenell, C. G. Owen, S. J. Preece, R. Gillions, S. Sheard, T. Peakman, S. Brage, and N. J. Wareham, ''Large scale population assessment of physical activity using wrist worn accelerometers: The UK biobank study,'' *PLoS ONE*, vol. 12, no. 2, 2017, Art. no. e0169649.
[10] J. G. Ellis, P. Gehrman, C. A. Espie, D. Riemann, and M. L. Perlis, ''Acute insomnia: Current conceptualizations and future directions,'' *Sleep Med. Rev.*, vol. 16, no. 1, pp. 5–14, Feb. 2012.
[11] R. Fossion, A. L. Rivera, J. C. Toledo-Roy, J. Ellis, and M. Angelova, ''Multiscale adaptive analysis of circadian rhythms and intradaily variability: Application to actigraphy time series in acute insomnia subjects,'' *PLoS ONE*, vol. 12, no. 7, 2017, Art. no. e0181762.
[12] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical learning*, vol. 1. New York, NY, USA: Springer, 2001.
[13] P. M. Holloway, M. Angelova, S. Lombardo, A. St Clair Gibson, D. Lee, and J. Ellis, ''Complexity analysis of sleep and alterations with insomnia based on non-invasive techniques,'' *J. Roy. Soc. Interface*, vol. 11, no. 93, Apr. 2014, Art. no. 20131112.
[14] P. W. Kamen and A. M. Tonkin, ''Application of the Poincaré plot to heart rate variability: A new measure of functional status in heart failure,'' *Austral. New Zealand J. Med.*, vol. 25, no. 1, pp. 18–26, Feb. 1995.

[15] C. K. Karmakar, A. H. Khandoker, J. Gubbi, and M. Palaniswami, "Complex correlation measure: A novel descriptor for Poincaré plot," *Biomed. Eng. OnLine*, vol. 8, no. 1, p. 17, 2009.
[16] A. Khademi, Y. EL-Manzalawy, L. Master, O. M. Buxton, and V. G. Honavar, "Personalized sleep parameters estimation from actigraphy: A machine learning approach," *Nature Sci. Sleep*, vol. Volume 11, pp. 387–399, Dec. 2019.
[17] S. Kusmakar, C. K. Karmakar, B. Yan, T. J. O'Brien, R. Muthuganapathy, and M. Palaniswami, "Automated detection of convulsive seizures using a wearable accelerometer device," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 2, pp. 421–432, Feb. 2019.
[18] K. L. Lichstein, K. C. Stone, J. Donaldson, S. D. Nau, J. P. Soeffing, D. Murray, K. W. Lester, and R. N. Aguillard, "Actigraphy validation with insomnia," *Sleep*, vol. 29, no. 2, pp. 232–239, 2006.
[19] V. Natale, D. Léger, M. Martoni, V. Bayon, and A. Erbacci, "The role of actigraphy in the assessment of primary insomnia: A retrospective study," *Sleep Med.*, vol. 15, no. 1, pp. 111–115, Jan. 2014.
[20] V. Natale, G. Plazzi, and M. Martoni, "Actigraphy in the assessment of insomnia: A quantitative approach," *Sleep*, vol. 32, no. 6, pp. 767–771, Jun. 2009.
[21] T. Morgenthaler, M. Kramer, C. Alessi, L. Friedman, B. Boehlecke, T. Brown, J. Coleman, V. Kapur, T. Lee-Chiong, J. Owens, J. Pancer, and T. Swick, "Practice parameters for the psychological and behavioral treatment of insomnia: An Update. An American academy of sleep medicine report," *Sleep*, vol. 29, no. 11, pp. 1415–1419, Nov. 2006.
[22] M. L. Perlis, C. M. Swinkels, P. R. Gehrman, W. R. Pigeon, S. E. Matteson-Rusby, and C. R. Jungquist, "The incidence and temporal patterning of insomnia: A pilot study," *J. Sleep Res.*, vol. 19, no. 1, pp. 31–35, Mar. 2010.
[23] D. M. W. Powers, "Evaluation: From precision, recall and f-measure to roc., informedness, markedness correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.
[24] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiology-Heart Circulatory Physiol.*, vol. 278, no. 6, pp. H2039–H2049, Jun. 2000.
[25] A. Sadeh, "The role and validity of actigraphy in sleep medicine: An update," *Sleep Med. Rev.*, vol. 15, no. 4, pp. 259–267, Aug. 2011.
[26] A. Sadeh, P. J. Hauri, D. F. Kripke, and P. Lavie, "The role of actigraphy in the evaluation of sleep disorders," *Sleep*, vol. 18, no. 4, pp. 288–302, Jun. 1995.
[27] A. Vallieres, H. Ivers, C. H. Bastien, S. BEAULIEU-BONNEAU, and C. M. Morin, "Variability and predictability in sleep patterns of chronic insomniacs," *J. Sleep Res.*, vol. 14, no. 4, pp. 447–453, Dec. 2005.
[28] V. T. van Hees, S. Sabia, S. E. Jones, A. R. Wood, K. N. Anderson, M. Kivimäki, T. M. Frayling, A. I. Pack, M. Bucan, M. I. Trenell, D. R. Mazzotti, P. R. Gehrman, B. A. Singh-Manoux, and M. N. Weedon, "Estimating sleep parameters using an accelerometer without sleep diary," *Sci. Rep.*, vol. 8, no. 1, Dec. 2018, Art. no. 12975.
[29] M. Willetts, S. Hollowell, L. Aslett, C. Holmes, and A. Doherty, "Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK biobank participants," *Sci. Rep.*, vol. 8, no. 1, Dec. 2018, Art. no. 7961.
[30] J. M. Williams, D. J. Taylor, D. C. Slavish, C. E. Gardner, M. R. Zimmerman, K. Patel, D. A. Reichenberger, J. M. Francetich, J. R. Dietch, and R. Estevez, "Validity of actigraphy in young adults with insomnia," *Behav. Sleep Med.*, vol. 18, no. 1, pp. 91–106, Jan. 2020.
[31] M. A. Woo, W. G. Stevenson, D. K. Moser, R. B. Trelease, and R. M. Harper, "Patterns of beat-to-beat heart rate variability in advanced heart failure," *Amer. Heart J.*, vol. 123, no. 3, pp. 704–710, Mar. 1992.

MAIA ANGELOVA received the Ph.D. degree from the University of Sofia. She is currently a Professor of data analytics and machine learning at the School of Information Technology, Deakin University, VIC, Australia. Prior to that, she was a Lecturer in physics with the Somerville College, University of Oxford, and a Professor of mathematical physics with Northumbria University, U.K., until 2016. She is also the Director of the Data to Intelligence Research Centre and leads the Data Analytics Research Lab, Deakin University. Her research interests include mathematical modeling, data analytics, time series, machine learning, and dynamical systems with applications to health.

CHANDAN KARMAKAR (Member, IEEE) received the Ph.D. degree from The University of Melbourne, Australia, in 2012. He is currently working as a Lecturer at the School of Information and Technology, Deakin University, Australia. He has extensive research experience in biomedical devices and signal processing, the IoT in healthcare, health informatics, and AI in healthcare. He has published more than 120 refereed research articles (including one book). For further details of his publications, please explore his: Google Scholar and ResearchGate accounts.

YE ZHU (Member, IEEE) received the Ph.D. degree in artificial intelligence from Monash University, Australia, in 2017.

He has been a Lecturer at the School of Information Technology, Deakin University, Australia, since 2019. He joined Deakin University, in 2017, as a Research Fellow of complex system data analytics. His research work focuses on clustering analysis, anomaly detection, and their applications for pattern recognition and information retrieval. He received the Mollie Holman Medal for the Best Doctoral Thesis of the Year, for his Ph.D. degree.

SEAN P. A. DRUMMOND is currently a Professor of clinical neuroscience at the Turner Institute for Brain and Mental Health, Monash University. Prior to that, he held joint appointments at the University of California at San Diego and the VA San Diego Healthcare System. He is the Director of the Sleep and Circadian Rhythms Research Program within the Turner Institute and the Immediate Past President of the Sleep Research Society.

JASON ELLIS is currently a Professor of sleep science and the Director of the Northumbria Sleep Research Laboratory, U.K. He is a qualified Somnologist—Expert in Behavioural Sleep Medicine from European Sleep Research Society and a Practicing Health Psychologist under the HCPC. He splits his time between his basic research interests include the pathophysiology of sleep disorders (Insomnia, Restless Legs Syndrome, and Circadian Rhythm Disorders), the natural history of Insomnia, and his applied work on Cognitive Behavioral Therapy for Insomnia (CBT-I). Within the latter framework, he examines the impact of novel adjunct therapies, the influence of social factors on adherence, and the effective delivery of CBT-I in complex cases and environments.