

Digital Object Identifier

Sleep Posture Monitoring using a Single Neck-Situated Accelerometer: A Proof-of-Concept

RAWAN S. ABDULSADIG (Member, IEEE), and ESTHER RODRIGUEZ-VILLEGAS

Department of Electrical and Electronic Engineering, Imperial College London, London, SW7 2BT UK

Corresponding author: Rawan S. Abdulsadig (e-mail: r.abdulsadig@imperial.ac.uk).

This work was supported by the European Research Council (ERC) for the NOSUDEP Project, grant agreement no. 724334.

ABSTRACT Sleep position identification and monitoring is important in the context of certain healthcare conditions, such as obstructive sleep apnoea and epilepsy. Many studies have thoroughly investigated automatic sleep detection using various sensing channels located in optimum body locations. However, this has not been the case for detection using physiological data acquired from a single sensing channel on the neck. In certain healthcare contexts the neck can, however, be an attractive location despite being suboptimal for position monitoring; the reason being that it enables better extraction of more critical biomarkers from other sensing modalities, making possible multimodal monitoring using just one wearable. This work focuses on investigating methods of automatic sleep position detection using one wearable channel of accelerometry data sensed on the neck. Three different models are explored. These are based on: decision trees (DT), extra-trees classifier (ET) and long-short term memory neural networks (LSTM-NN). The paper also investigates for the first time what would be optimum design choices when considering that wearables are power and memory-constrained, but performance in the type of healthcare applications where a single location multimodal sensing is important must not be compromised. This includes looking into how changing the sampling rate and window sizes would affect the performance of the different models. It is demonstrated that a sampling rate as low as 5 Hz, and a window size as short as 1 second, still lead to high classification performance (around 0.945, 0.975 and 0.965 mean f1-score when using the DT, ET and LSTM-NN models, respectively, and at least 98% average accuracy in all three models); and that the DT model occupies the least memory space (1.765 KB) and takes the least mean prediction time across all window sizes (around 0.8 ms).

INDEX TERMS Accelerometry, machine learning, neck, sampling rate, sleep posture, window size

I. INTRODUCTION

SLEEP posture is linked to general sleep quality and is also often associated with health and clinical outcomes. For example, in nursing homes and hospitals, monitoring posture is important for the elderly and unconscious patients to prevent the development of pressure ulcers by changing their lying posture every 2 hours [1]. Furthermore, sleeping in the supine position is known to cause certain health complications, for instance, it is known to influence the severity of obstructive sleep apnoea (OSA). It was shown that in a sample of 574 adults suffering from OSA, 55.9% had at least twice as many apneas/hypopneas in the supine position compared to the lateral position [2]. On the other hand, the prone position is known to be a risk factor for sudden unexpected death in epilepsy (SUDEP) and sudden infant

death syndrome (SIDS), with nearly 75% of the documented sudden death tragedies occurring in this position, in both conditions [3], [4]. This gives rise to the importance of accurate monitoring of sleep postures for high-risk patients, and extra care should be given to the prone posture when the focus is sudden death prevention for epilepsy patients (SUDEP).

When it comes to wearable monitoring technologies with a form factor constrained by the usability in a specific patient population, power consumption and detection latency are two important factors for the efficacy of the system. The rate at which a sensor is acquiring data is directly proportional to the energy consumed [5], while the size of the processing window contributes greatly to the latency of detection. The processing and prediction time required by the detection

model also contribute towards the detection latency, however, their impact is generally less. To this end, this work investigates the performance in terms of the detection of the four possible sleeping positions (supine, lateral right, lateral left and prone) of three different models using accelerometry data collected by a single wearable device placed on the neck as input; whilst also exploring what would be the minimal computational requirements that would make them suitable for a low-power wearable monitoring device. The effect of changing the sampling rate and window size on the detection performance of the models is investigated for the first time, with special care towards the sensitivity of detecting the prone posture, since this is the one with the highest associated level of risk. The first model type is a threshold-based detection algorithm, where the thresholds were derived using a decision tree classifier (representing an ultra-lightweight approach). This model type was chosen due to its lightweight quality, in addition to its data-driven threshold-search nature which is considered a superior alternative to arbitrarily set thresholds. The second is an extra-trees classifier (representing an ensemble-based machine learning approach) which is often found to outperform other classical machine learning algorithms in similar applications. The third is a long-short term memory neural network model (representing a deep learning approach) since deep learning models are widely used in all sorts of applications, however, in this work, the architecture was empirically simplified to minimise the number of learnable weights needed to be stored in memory. The simplification of all three models was made to account for the possibility of embedding any of these models into the wearable system, where memory and processing constraints exist. In addition to the classification metrics, pre-processing time, prediction time and model size were also measured and reported for completeness.

In a nutshell, the novelty of this work is twofold: the first is that it deals with accelerometry signals obtained from a body location which not only has not been investigated but is an unusual choice for sleep posture detection while being a rich location for multimodal wearable sensing, which makes this investigation worthwhile; the second is that it investigates for the first time data-driven modelling approaches while taking into account the hardware limitations that primarily exist in low-power monitoring wearable devices.

II. RELATED WORK

Many studies have explored sleep posture monitoring using various sensing modalities. The span of sensory mediums includes ECG electrodes arranged in a conductive sheet placed on the bed [1], fabricated conductive textile sheets [6], RFID-based system with under-sheet tags [7], infrared cameras [8], [9], a combination of under-mattress force-sensing resistors (FSR) and infrared array sensor [10] and under mattress pressure sensors [11]–[14].

Accelerometers are also utilised in the literature. Although many studies used accelerometers placed on multiple body locations [15]–[21], fewer studies explored using a single

accelerometer for this purpose. A recent study presented a diaper cover system called NAPPA wearable, which was designed to monitor infants' respiration and position during naps, embedding an accelerometer and a gyroscope used to estimate sleep posture. However, detailed sleeping posture detection methods and results were not provided [22]. A wearable device placed over the nose called MORFEA was also proposed in previous work, where the domestic pre-screening of sleep-related breathing disorders (SRBD) was the main goal. The device integrates a tri-axial accelerometer sensor to analyze movement. Sleep posture was estimated using head rotation and inclination compared against an arbitrary look-up table where certain conditions are checked. In this study, the prone posture was not covered due to the wearable device's placement [23]. Another wearable reported in the literature was a smart mandibular advancement device (MAD) where an integrated accelerometer sensor was used to estimate sleep positions. However, the detection performance was not provided, only an arbitrary approach was presented [24].

In a study investigating the use of a patch-type accelerometer, a high level of overall accuracy (99.16%) was reported when the sensor was attached to the left side of the chest, and pre-defined conditions along the X, Y and Z dimensions were used [25]. However, there were no occurrences of prone postures during the study. Another later study investigating accelerometry data obtained from the chest area used a smartphone placed over the sternum using a fixation system, and evaluated the detection performance of a set of arbitrary angular conditions. The reported results showed perfect sensitivity for the supine posture, and moderate sensitivity to the lateral positions but no detection of the prone posture [26]. The authors then later presented a sleep monitoring mHealth app 'SleepPos' [27], where they reported an overall accuracy of 98.2%, but only 38.9% sensitivity for the prone position, making it inapplicable for sleep position monitoring where the target population is especially endangered by the prone posture. Furthermore, the bulky hardware design of their system would make it uncomfortable and unreliable for use on a daily basis. A neck-worn wireless wearable sensor was proposed recently for audio and motion sensing for sleep apnea, where the pitch and roll of the motion were compared against an arbitrary set of thresholds to estimate the posture [28]. Although high accuracy was reported in distinguishing lying/non-lying positions and supine/non-supine postures (99.9% and 97.3% respectively), the accuracy of detecting the other main sleep postures was not reported.

It is noticeable that the majority of the studies in the literature resort to establishing intuitive and/or empirically developed conditions rather than taking data-driven approaches to building detective models. However, the authors in [29] investigated the use of 5 types of machine learning models for detecting the 4 main sleep postures, while investigating the best single sensor position using a combination of publicly available datasets (Class-Act dataset and Daily and Sports Activities dataset). The investigated locations were the left

and right thigh, left and right wrist and the chest. The combined dataset provided accelerometry data sampled at 25 Hz, and the processing window was chosen to be 3.8 s (due to it being the minimum length of labelled segments in the dataset). The study concluded the thighs and the chest were the best body positions (given the data), while the extra-trees classifier and LSTM-based neural network model as the best-performing model types (achieving mean F1-scores between 0.906 and 0.973, using leave-one-subject-out cross-validation). Although the study was comprehensive, it lacked the investigation of the neck, as a potentially optimal location in certain healthcare contexts. Furthermore, the effect of choosing different window sizes and/or sampling rates was not explored. In fact, to the best of our knowledge, the exploration of these parameters (i.e. window sizes and sampling rate) for sleep posture monitoring was never presented in the literature.

This work is an extension of our previous preliminary work where the classification performance of the four main sleep postures using the decision tree (DT) and extra-trees classifier (ET) model types was reported, while keeping the sampling rate and window size fixed at 100 Hz and 5 seconds, respectively. In that study, the concept of sleep posture detection through data-driven modelling approaches using accelerometry data collected from the neck via a wearable device was first proposed, and its feasibility was highlighted [30]. This paper extends the previous work in two main ways: The first is the exploration of the impact of changing the sampling rate and the window size on the models' classification performance, detection and prediction time as well as memory consumption. The second is the addition of the LSTM-NN model type.

III. METHODS

A. DATA ACQUISITION

1) Acquisition system

Accelerometry data were obtained using a custom PCB integrating a triaxial accelerometer (LIS2DH12, ST Electronics) with an NRF5232 microcontroller (Nordic Semiconductor) and a rechargeable 3.8V 80mAh lithium polymer battery. The accelerometry data were sampled and transmitted wirelessly via Bluetooth low energy (BLE) at 100 Hz to an accompanying custom data acquisition iOS app. The PCB was housed in a 3207.23mm^3 additive manufactured enclosure, which was then placed approximately 1 inch above the suprasternal notch on the neck using a double-sided adhesive tape. The use of adhesive and the placement on the neck was shown to be highly durable and comfortable during sleep in previous studies [31], [32]. Fig. 1 illustrates the placement of the device on the neck.

Signals were collected from 18 participants (2:1 male-to-female ratio) from a study approved by the Local Ethics Committee of Imperial College London (ICREC reference number: 18IC4358). Table 1 lists the main characteristics of those participants.

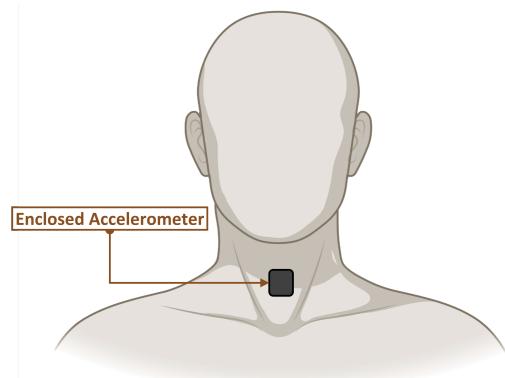


FIGURE 1. Visual illustration of the placement of the accelerometer sensor on the neck.

TABLE 1. Demographic details of the subjects Participating in the experiment

	Mean Value \pm std
Age (years)	27 ± 3.2
Height (cm)	177.33 ± 11.17
Weight (kg)	72.18 ± 12.49
BMI (kg/m^2)	22.78 ± 1.87

2) Study Protocol

All participants were directed, via verbal cues, to perform 4 main postures: supine, prone, right and left. Participants spent at least 30 seconds in each position before moving to the next one. Fig. 2 shows the progression of the experimental protocol throughout the duration of each subject's run. The protocol was designed to have three main phases: a discrete phase where each posture was performed for 30 seconds before going back to rest at the supine position (coloured blue); a 360° rotational phase where subjects changed through the 4 postures clockwise then counter-clockwise, staying at each posture for 15 seconds and each rotation was done three times (coloured orange); and a 180° rotational phase where subjects shifted between lateral right and lateral left, then between supine and prone postures, holding each posture for 20 seconds, with each rotation being done four times (coloured green).

B. BUILDING THE DETECTION MODELS

In this work, three modelling approaches were used: an ultra light-weight threshold-based method (Decision Tree DT), an ensemble-based machine-learning model (Extra-Trees classifier ET), and a neural network model (Long-short Term Memory Neural Network LSTM-NN).

The subjects were split into two groups. One group consisting of 10 randomly-selected subjects was used for building the models (training set), and the remaining 8 subjects were reserved for estimating the performance of the models on out-of-sample data (testing set). Fig. 3 represents an overall methodology map of the processing pipelines implemented to train the three model types.

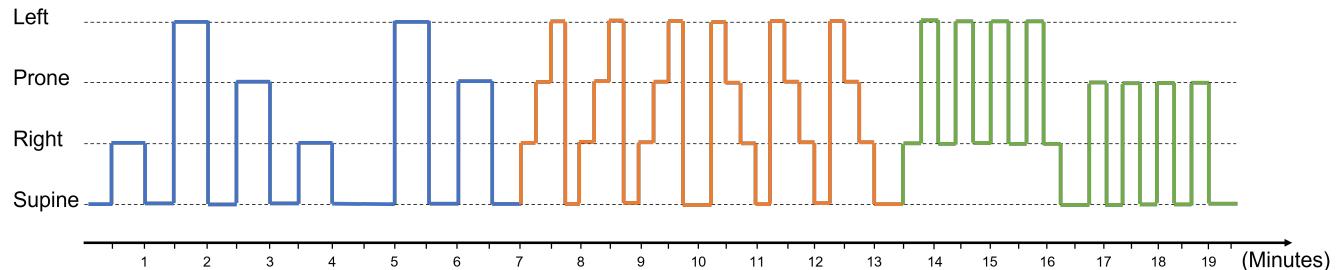


FIGURE 2. Experimental protocol showing the three main experiment phases: a discrete phase shown in blue, a 360° rotational phase shown in orange and a 180° rotational phase shown in green.

1) Data Pre-Processing

The pre-processing steps carried out on each subject's data for all model types included:

- Downsampling the signals as needed.
- Establishing the reference point and subtracting it from all sensor values of the whole run.

Data were downsampled from 100 Hz to the sampling rate being investigated (50, 25, 10 or 5 Hz) when training the ET classifier and LSTM-NN models. This was also done when evaluating all three model types on the testing subjects.

A reference point was established for each subject independently at the first supine position occurrence. This was considered as the relative starting point, based on which the angle of acceleration was calculated. This reference point was obtained by taking the mean of the x , y and z values of a 2-second window in the middle of that first supine posture. The reference point's x , y and z values were then subtracted from the values of the whole run of each subject independently, in order to eliminate the subtle inter-subjects variations in the signals that could influence the construction and performance of the detection models.

When building models using extracted features, models were provided with features calculated from the angles in the XZ, XY and YZ planes with respect to the reference point. The angles in those planes at any point were calculated using the following equation:

$$\theta_{\alpha\beta} = \arccos \frac{\text{Ref}_\alpha \times \text{Ref}_\beta}{|\text{Ref}_\alpha| \times |\text{Ref}_\beta|} - \arccos \frac{\mathbf{P}_\alpha \times \mathbf{P}_\beta}{|\mathbf{P}_\alpha| \times |\mathbf{P}_\beta|}, \quad (1)$$

where, $\alpha \in \{X, Y\}$, $\beta \in \{Y, Z\}$, $\alpha \neq \beta$, Ref refers to the reference point and P is the point at which the angle is calculated.

In the case of using the decision tree thresholds and the ET classifier models, pre-processing time included the mean time spent in digital to analogue conversion (DAC), angles calculation and feature extraction. When using the ET classifier models, pre-processing time, in addition to the aforementioned, included the mean standardisation time of the features in the window. On the other hand, when using the LSTM-NN models, pre-processing time only included the mean time spent in converting the digital values to voltage (DAC).

2) Feature Extraction

Time-domain features are often used for applications involving human movement sensed using accelerometry data [17], [33]. It was shown in a previous study that the median, mean, maximum and minimum values are the most informative time-domain features for sleep posture detection, after examining 18 different feature types [29]. Therefore, the mean and median features were used in this work. The intuitive choice of the mean measure is to obtain an aggregated estimation of the steady-state values contained in each window, while the median gives an aggregated estimation of those steady-state values without taking into account the effect of sudden movements or noise which can cause outliers in the signals. The maximum and minimum features were not used in this work due to the signals being unfiltered, which could result in a locally minimum or maximum value that is erroneous. 12 features in total were extracted per window, namely: Mean of X, Median of X, Mean of Y, Median of Y, Mean of Z, Median of Z, Mean of θ_{XY} , Median of θ_{XY} , Mean of θ_{XZ} , Median of θ_{XZ} , Mean of θ_{YZ} and Median of θ_{YZ} . These features were then used for fitting the decision tree model and extra-trees classifier.

3) Model Fitting

a: Decision Tree (DT)

Finding optimal thresholds across different feature dimensions that best segregate different classes in the data is what decision trees are made for, they search for the optimal splits that result in leaves with the lowest impurity index possible [34]. Any optimal splitting condition is found by searching all possible threshold values t along each of the extracted features F_i (where $i \in 1, 2, 3 \dots N_{\text{features}}$), evaluating the level of split class impurity then choosing the splitting threshold that results in the minimum impurity value [35].

The *Gini* index was used as the impurity criterion in this work. *Gini* criterion measures the divergence of the probability distributions of the classes in a node [35], the lower the *Gini* index the lower the impurity.

The splitting process typically takes place recursively until a regularisation condition is met such as the maximum tree depth or the minimum samples per leaf node, or it can continue until each leaf node is pure resulting in a fully grown tree. In this work, it was deemed appropriate to limit

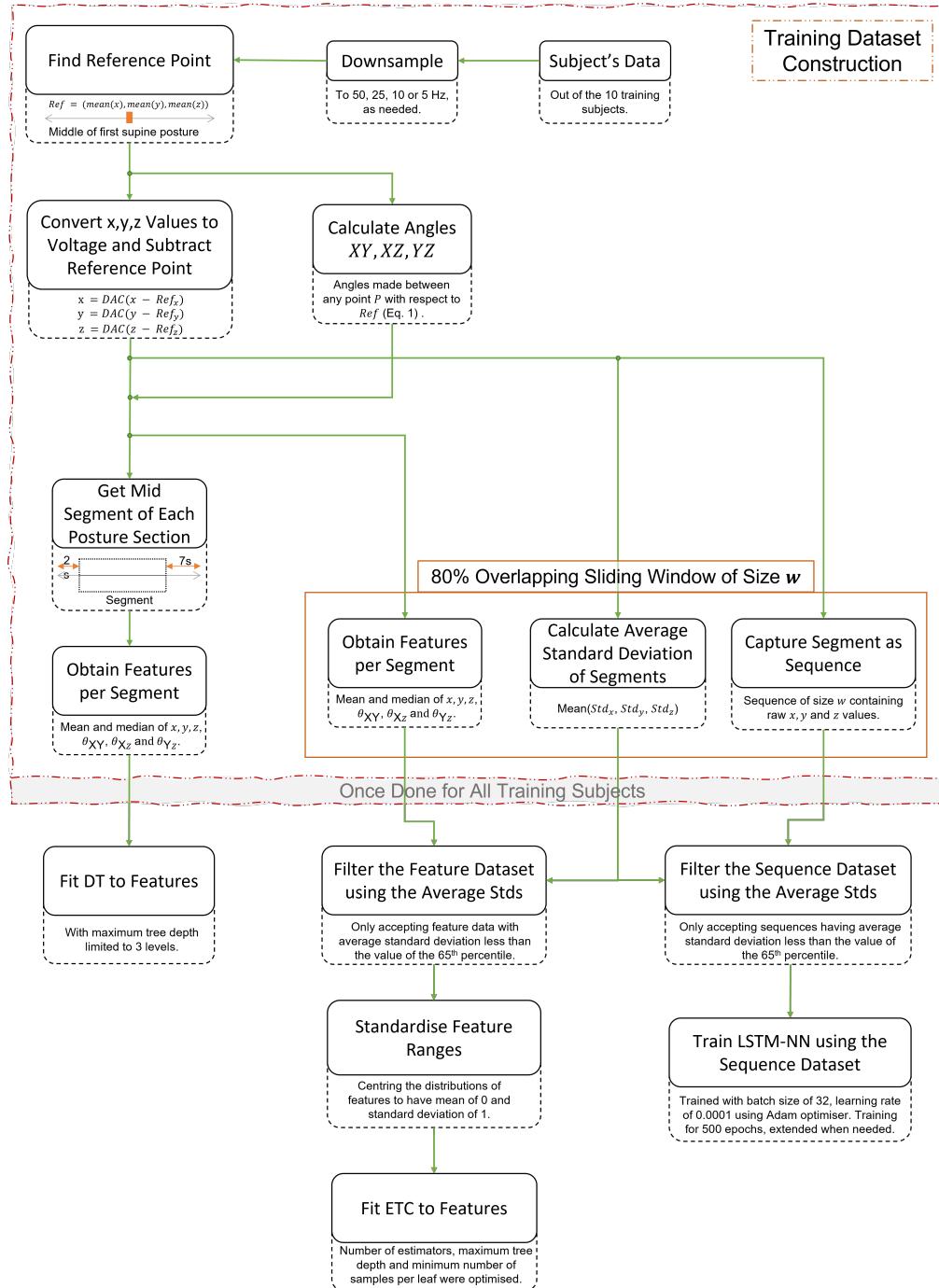


FIGURE 3. Overall methodology implemented to construct the training dataset and build the three models (DT, ET classifier and LSTM-NN). Each step is provided with a brief explanation or illustration.

the maximum tree depth to 3 levels since the task is to find minimal and optimal splitting thresholds that separate the four different postures.

Features obtained per posture section were provided for fitting the decision tree to the training data in order to find the optimal segregation thresholds without exposing the model to temporal fluctuations. Therefore, the middle section of each performed posture by the training subjects was obtained by

discarding the last 7 seconds which contained the transition movement, and the first 2 seconds which may contain some residual movements from the previous transition.

b: Extra-Trees Classifier (ET)

This model type is an extension of the decision tree model where an ensemble of decision trees work together to infer a prediction. The aggregation of the results obtained by all

decision trees in the ensemble is often done by averaging the individual estimations, resulting in a unified prediction. Fig. 4 illustrates the general structure of this model type.

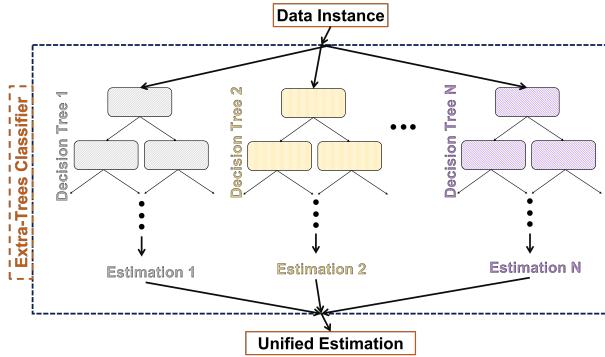


FIGURE 4. An illustration of the general structure of the Extra-Trees Classifier model type, which consists of an ensemble of decision trees providing their individual estimations, from which a unified estimation can be deducted.

At training time, extensive randomisation in terms of the choice of splitting conditions as well as the subset of features used for fitting each individual tree is key in this model type. This is why it is known as extremely-randomised trees classifier [36]. This randomisation is done to help prevent over-fitting by maintaining low variance, which ensures high generalisation performance.

Features were extracted using a sliding window with 80% overlap, this amount of overlap was chosen empirically as it was found to help provide the model with as many unique training data points as possible out of each run. The minimum number of samples in a leaf node, the number of estimators in the ensemble and the maximum tree depth are the three main hyper-parameters that were tuned using leave-one-subject-out grid search cross-validation. The specified range of values searched was [10, 50] for the number of estimators and the minimum number of samples in a leaf and [5, 50] for the maximum tree depth. Standardised features were provided to construct the model. Each iteration in the leave-one-subject-out cross-validation process consisted of a set of validation data points belonging to 1 distinct subject out of the 10 training set subjects, while the rest of the training set was used to fit the model using the selected combination of hyper-parameters for that particular iteration. The performance of the model at each iteration was then evaluated using the f1-score. At the end of the cross-validation process, the model was fitted to the entire training set using the hyper-parameter combination that resulted in the maximum f1-score.

c: Long-short Term Memory Neural Network (LSTM-NN)

Recurrent neural networks with long-short term memory (LSTM) units have been used extensively for time-series data where long-term temporal dependencies occur [37]. A typical LSTM unit consists of 4 computational gates: input gate i_t , output gate o_t , cell/update gate \tilde{C}_t and forget gate f_t , as illustrated in Fig. 5. Each gate is accompanied by a weight matrix W of size (input \times hidden dimension, hidden

dimension \times hidden dimension) and a bias vector b of size (1, hidden dimension), where hidden dimension is a fixed parameter chosen according to the problem at hand. Both W and b are trainable parameters optimised via backpropagation through time [37].

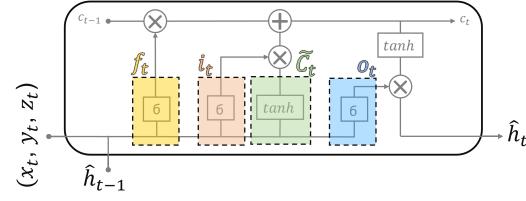


FIGURE 5. The typical structure of an LSTM unit, mainly comprising of a forget gate f_t , an input gate i_t , a cell/update gate \tilde{C}_t and an output gate o_t .

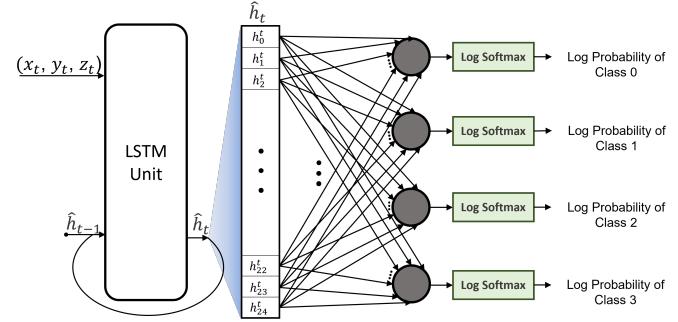


FIGURE 6. LSTM_NN architecture implemented consisting of an LSTM unit with the hidden dimension equal to 25 followed by a (25, 4) fully connected layer accompanied by a Log-Softmax non-linearity function.

The hidden dimension was chosen empirically to be 25, as it was found in initial experiments that increasing the hidden dimension beyond this value did not have any significant impact on the performance of the model, but only significantly increased the size of the model and the computation time. Following the LSTM unit, a fully-connected layer with an input size of 100 and an output size of 4 was used. The output of the fully-connected layer was then passed through a Log Softmax non-linearity function producing log-likelihood measures per class, from which the negative log-likelihood loss was calculated and used to optimise the model using Adam optimiser [38]. Fig. 6 illustrates the LSTM-NN model architecture used in this work, while Table 2 lists the number of trainable parameters in the model, representing the weight and bias parameters of the LSTM unit at the input-hidden and the hidden-hidden interfaces, in addition to the weight and bias parameters of the fully-connected layer.

The model was fed with x , y and z data sequences contained in an 80% overlapping sliding window, which was empirically found to help provide the model with as many unique training data instances as possible out of each run. Data obtained from 2 training subjects were used as validation set while the remaining 8 subjects provided data to train the model. Training was done for 500 epochs and extended whenever needed for the model to reach plateau.

TABLE 2. LSTM-NN trainable parameters distribution

Model Parameter	Count
LSTM unit: input-hidden weight	300
LSTM unit: hidden-hidden weight	2500
LSTM unit: input-hidden bias	100
LSTM unit: hidden-hidden bias	100
Fully-Connected Layer: weight	100
Fully-Connected Layer: bias	4
Total Trainable Parameters	3104

The batch size was chosen to be 32 and the learning rate was set to 0.0001.

4) Handling Transitions

Given the experimental protocol and the frequent shifts between postures, subjects were spending around 35% of the duration of their run moving from one position or adjusting to the new one. The average standard deviation across the X, Y and Z axes was used to detect whether a transitional movement had taken place within the processing window. Detecting transitions was done by comparing the mean of standard deviations of the signal contained in the window to a pre-specified filtering cutoff value, which was set to be equal to the 65th percentile value of the mean of standard deviations acquired from the training subjects. Any segment of data with an average standard deviation greater than, or equal to that cutoff value was regarded as transitional movement and discarded when training the ET classifier and LSTM-NN models and when testing all three model types. Those filtering thresholds were established per window size and sampling frequency combination.

C. SEMI-REALTIME EVALUATION OF THE DETECTION MODELS

In order to imitate a real-time data acquisition system when evaluating the three types of models, data from each test subject was taken one sample at a time and entered into a queue-type processing window, which was then individually processed and passed to the fitted detection model being evaluated. Furthermore, the beginning of each test subject's run was set to be at the centre of the first supine posture in order to resemble the system being deployed, in which the user would be instructed to start the monitoring session at that position. Therefore, the reference point was obtained from the first 2 seconds of the run, once the deduction was applied.

D. DETECTION PERFORMANCE EVALUATION METRICS

Accuracy, sensitivity (recall), specificity, precision and f1-score are the main performance metrics that are often used when evaluating models that are conducting a classification task. Those evaluation metrics are derived from 4 basic measures: the number of true positive instances, the number of true negative instances, the number of false positive instances and the number of false negative instances. These quantities are defined as:

- True positive (TP): correctly estimating a truly positive instance as being positive
- True negative (TN): correctly estimating a truly negative instance as being negative
- False positive (FP): incorrectly estimating a truly negative instance as being positive
- False negative (FN): incorrectly estimating a truly positive instance as being negative

In a multi-class task such as the one presented in this work, positive generally refers to the instance belonging to the class of concern, while negative refers to it not belonging to the class of concern, in a one-vs-all fashion.

Using these quantities, the accuracy of classifying instances belonging to a class of concern is calculated as:

$$\text{Accuracy} = \frac{\text{TP} + \text{FP}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (2)$$

Recall or sensitivity (also called true positive rate (TPR) or hit rate), which is a measure of the extent to which the model can correctly capture the truly positive instances of the class of concern, is defined as follows:

$$\text{Recall (sensitivity)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

Specificity, which is a measure of the ability of the model to correctly reject the truly negative instances of the class of concern (also known as true negative rate (TNR)) is measured as:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (4)$$

Precision (also known as positive predictive value (PPV)) is a measure of how precise the model is in its estimation of the positive instances being actually true (belonging to the class of concern), it is calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

F1-score is a measure that equally combines sensitivity and precision, and it is defined as the harmonic mean of both measures as follows:

$$\text{F1-Score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

IV. RESULTS

A. DETECTION PERFORMANCE

Fig. 7, Fig. 8 and Fig. 9 show the results of evaluating the thresholds developed by the decision tree, the ET classifier models and LSTM-NN models on the test subjects' data, respectively. Mean precision, sensitivity (recall), specificity, accuracy and F1-score across the 4 postures, averaged across the 8 test subjects along with their 95% confidence interval (a), in addition to the sensitivity score of detecting the prone position for each test subject (b) are shown per window size and for each examined sampling rate (shown in different rows of figures).

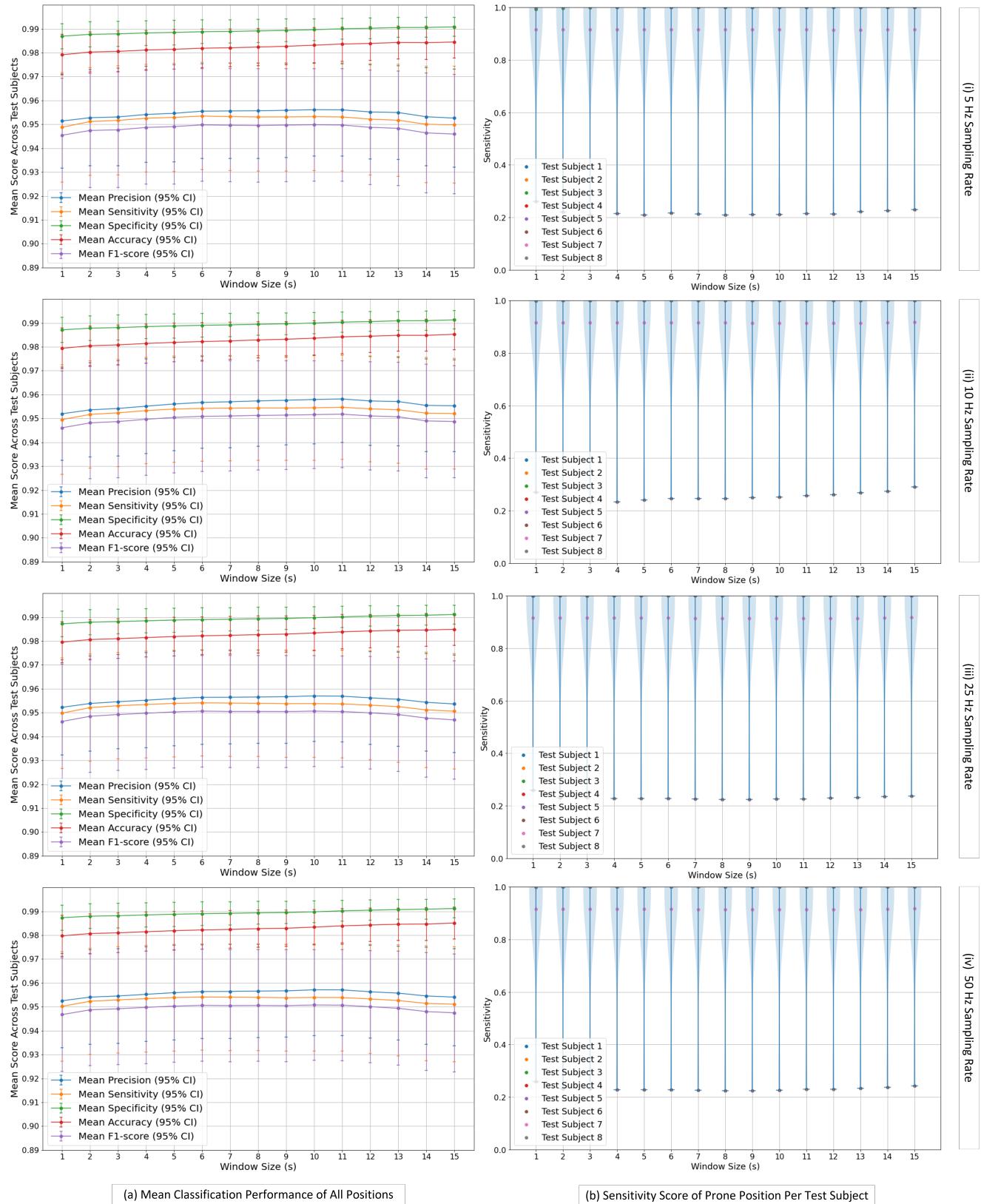


FIGURE 7. Evaluation of the detective thresholds produced by the DT model on the test set. (a) shows the mean classification performance across all classes and averaged across all test subjects against the different window sizes (1s to 15s), along with the 95% confidence interval bars. (b) shows the sensitivity score of detecting the prone posture for each test subject (shown in coloured dots), in addition to the distribution of the scores shown in transparent blue. (i), (ii), (iii) and (iv) represent the 4 different sampling rates: 5 Hz, 10 Hz, 25 Hz and 50 Hz, respectively.

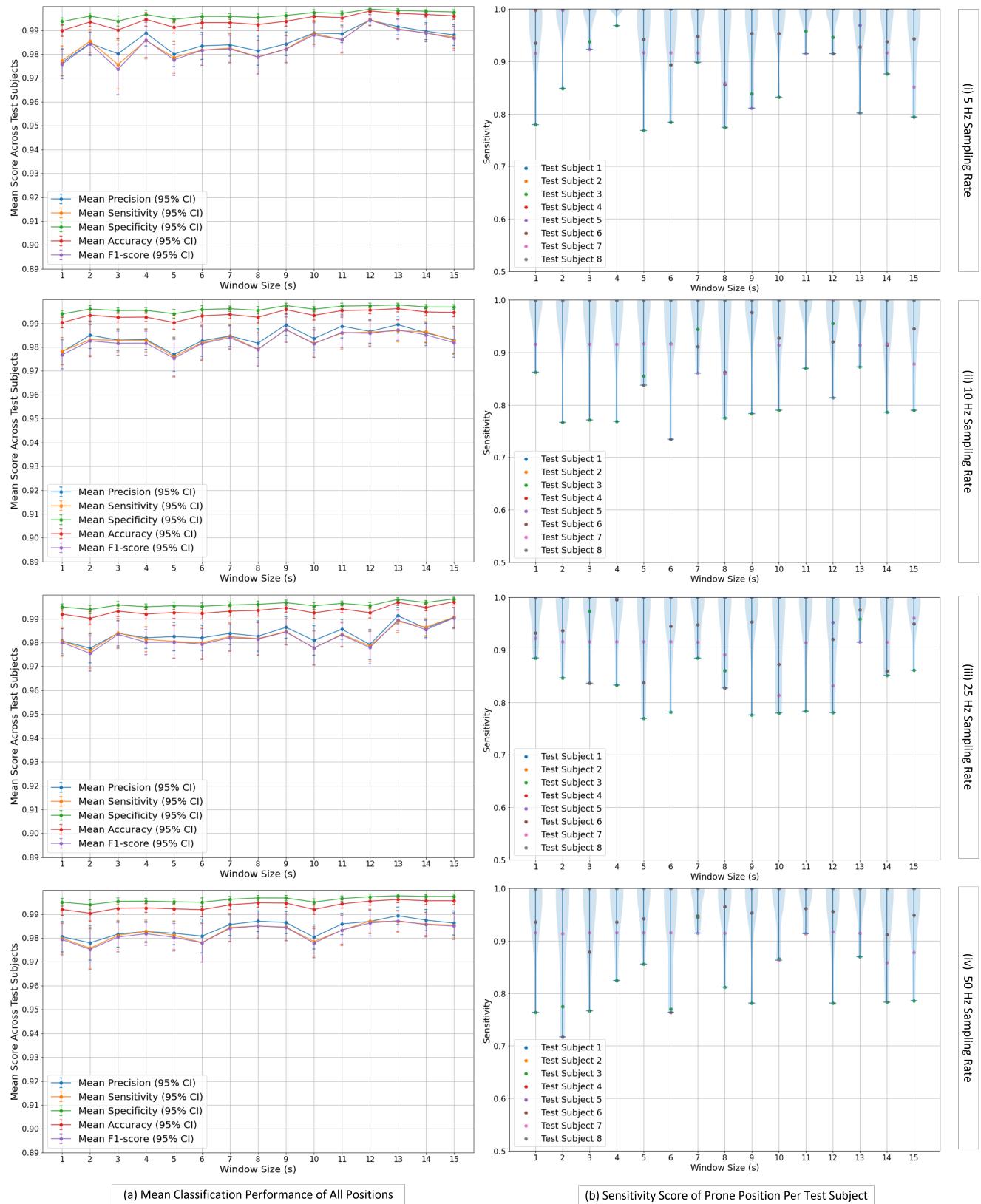


FIGURE 8. Evaluation of the performance of the detective ET classifier models on the test set. (a) shows the mean classification performance across all classes and averaged across all test subjects against the different window sizes (1s to 15s), along with the 95% confidence interval bars. (b) shows the sensitivity score of detecting the prone posture for each test subject (shown in coloured dots), in addition to the distribution of the scores shown in transparent blue. (i), (ii), (iii) and (iv) represent the 4 different sampling rates: 5 Hz, 10 Hz, 25 Hz and 50 Hz, respectively.

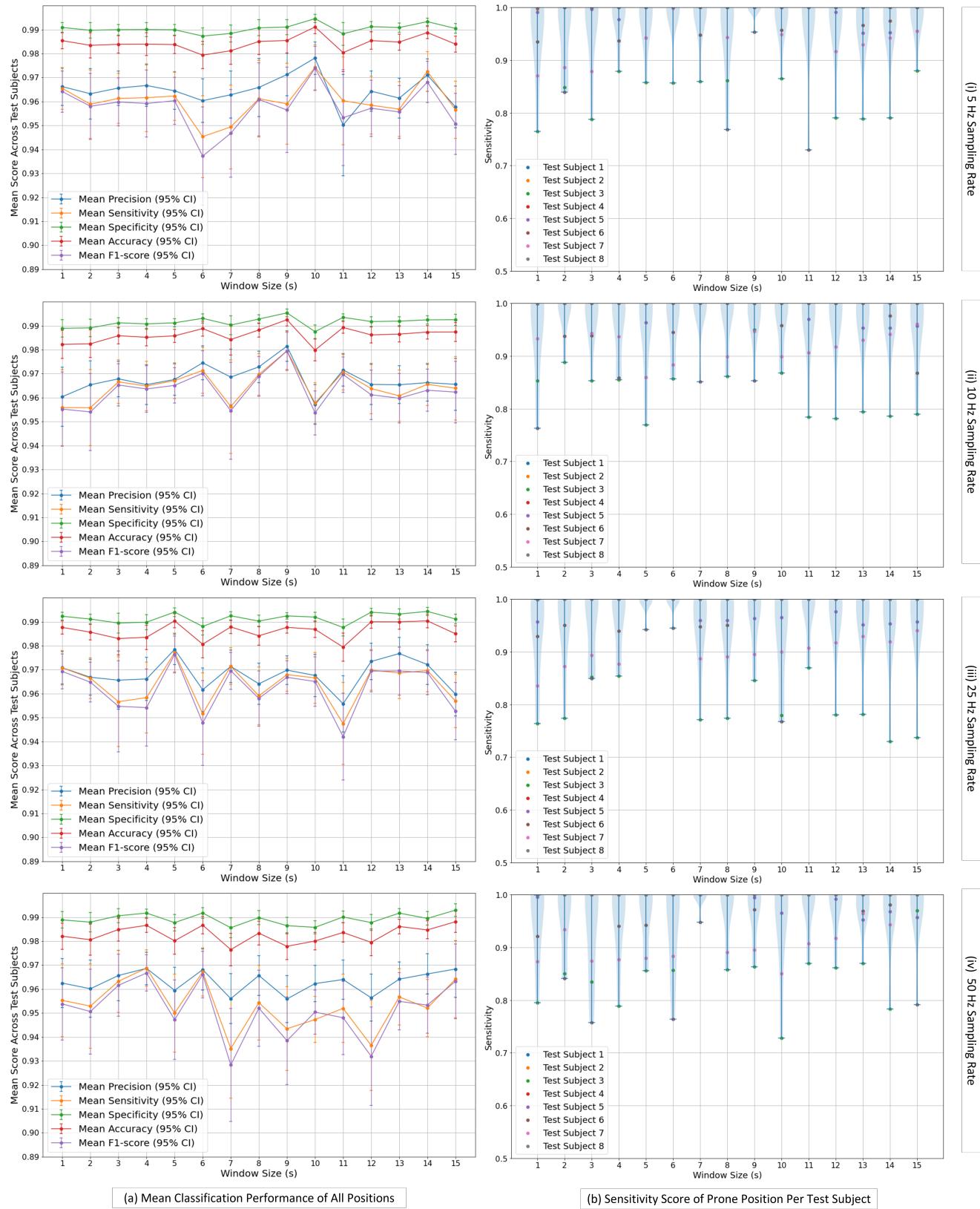


FIGURE 9. Evaluation of the performance of the detective LSTM-NN models on the test set. (a) shows the mean classification performance across all classes and averaged across all test subjects against the different window sizes (1s to 15s), along with the 95% confidence interval bars. (b) shows the sensitivity score of detecting the prone posture for each test subject (shown in coloured dots), in addition to the distribution of the scores shown in transparent blue. (i), (ii), (iii) and (iv) represent the 4 different sampling rates: 5 Hz, 10 Hz, 25 Hz and 50 Hz, respectively.

By observing the results illustrated in (a) of Fig. 7, 8 and 9, it was revealed that, in general, changing the window size did not have any significant effect on the mean classification performance of sleep postures, in any of the three model types. Although a small increase in the overall accuracy can be observed in the cases of the DT and ET classifier models as the window size increases, this increase is very little in value (maximum of 0.01). Furthermore, it was also shown that the reduction of the sampling rate did not have any negative effect on the detection performance.

Overall, when using the thresholds generated by the decision tree (Fig. 7.a), the range of the mean specificity score was found to be between 0.985 and 0.991 and the mean accuracy was between 0.978 and 0.985, both of which showed little variation across the different subjects illustrated by the tight 95% confidence interval. On the other hand, the mean precision, sensitivity and F1-score showed relatively large confidence bounds, and mean values ranging between 0.945 and 0.954. This could be explained by observing the detection performance of the prone position (Fig. 7.b), which reveals that the model consistently failed to detect this posture when performed by test Subject 6, with a sensitivity score of around 0.2 only, which is likely due to the irregular execution of the test instructions with a tendency to lean more towards the lateral positions. However, it was able to detect all prone postures of 6 test subjects perfectly, while test Subject 7 had a sensitivity score of around 0.9.

Looking at the mean classification metrics obtained when evaluating the ET classifier model type (Fig. 8.a), it can be immediately seen that the overall results are consistent and with tight confidence intervals relative to the decision tree model, with all metrics having mean scores above 0.975. Observing the sensitivity of the prone posture per test subject in Fig. 8.b exposes that the fitted ET classifier models generally struggled to detect the prone position of test subjects 3, 6 and 7, while perfectly detecting all prone postures of the remaining 5 test subjects. However, the sensitivity of all subjects at all window sizes and sampling rates was no less than 0.7. Granular variations in sensitivity errors, between sampling frequencies and across the different window sizes, were due to differences in the fitted models in each combination and the randomised nature of the ET classifier model fitting.

In the case of using the LSTM-NN model, the classification results shown in Fig. 9.a illustrate a significant amount of random variations in the mean performance measures of the different trained models, which could indicate the difficulty of training the models optimally and consistently with the amount of data available, and/or the architecture used. However, the general trend of the mean accuracy and specificity scores was well above 0.98 in most cases, while the mean sensitivity, precision and f1-scores were ranging between 0.93 and 0.98. Fig. 9.b shows similar general observation to the ET classifier model type results, with certain test subjects likely to have sensitivity scores below 0.9 (namely test subjects 3, 6 and 7). Also similar to the ET classifier model type results, the sensitivity of all subjects at all window

sizes and sampling rates was greater than 0.7.

B. PRE-PROCESSING TIME, PREDICTION TIME AND MODEL SIZES

Fig. 10 shows the mean time (in ms) taken to pre-process and to produce an estimation of the sleep posture when using the decision tree thresholds (a), ET classifier models (b) and LSTM-NN models (c) at the different window sizes and sampling frequencies.

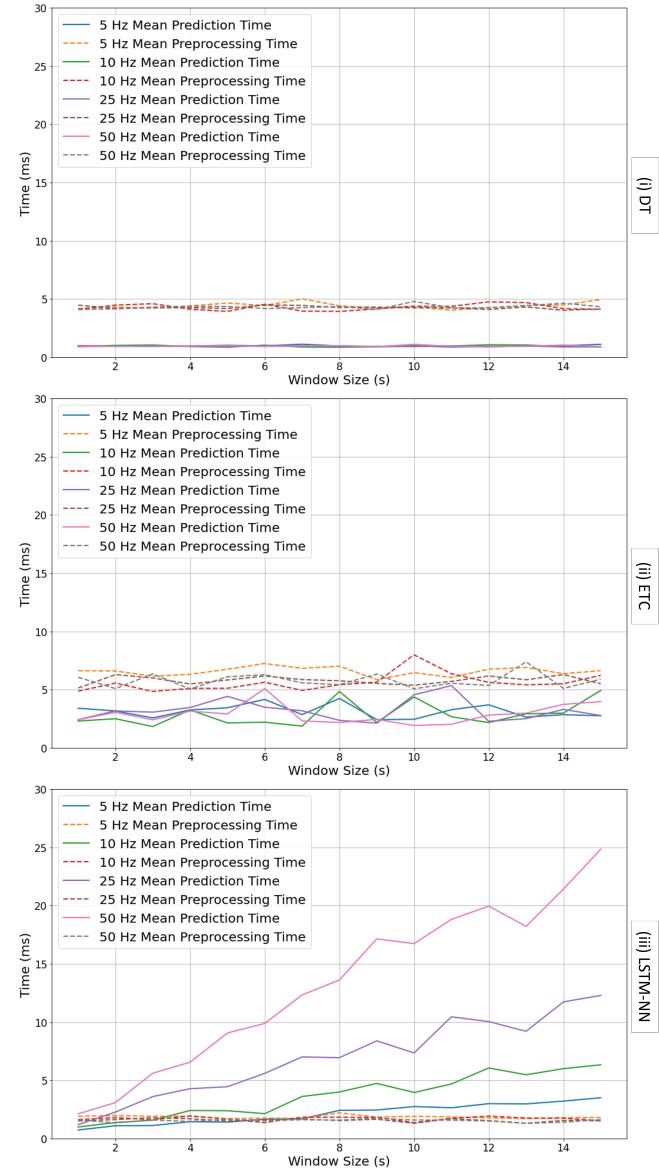


FIGURE 10. Line plots of mean time (in ms) spent in pre-processing (dashed lines) and in prediction (solid lines) when using the DT thresholds (a), ET classifier (b) and LSTM-NN (c) models. Time is shown per window size and for each sampling rate.

Fig. 10.a and Fig. 10.b show that the change in window size and sampling rate had no effect on the pre-processing time and the prediction time, when using the DT and ET classifier models, respectively. Pre-processing time was generally greater than the prediction time in both cases (less than

10 ms in both). However in the case of using the LSTM-NN models, Fig. 10.c shows that while the pre-processing time was low in value and unaffected by the window size and sampling rate changes (less than 3 ms), the prediction time was directly proportional to the window size and amplified by the sampling rate, reaching a maximum of 27 ms at window size equal to 15 s and sampling rate equal to 50 Hz. This was due to the LSTM-NN model type performing computations for each sample of the window. The number of samples in the window grows with the increase in window size, as well as with the increase in sampling frequency.

Regarding the size of the detective models, since the structure of the ET classifier model type was optimised for each window size, the ET classifier model size (in Kilobytes) was varying in each case. Fig. 11 gives an estimation of the model size as the window size increases. This was shown as the mean model size over all sampling rates. DT and LSTM-NN model sizes are shown as well.

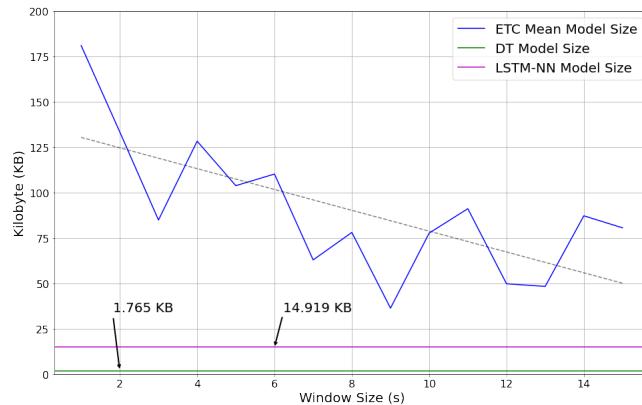


FIGURE 11. Mean model size in KB of the ET classifier model type against the different window sizes, shown in blue. DT and LSTM-NN model sizes are also shown in green and magenta, respectively.

Fig. 11 demonstrates the effect of the window size on the structure of the fitted ET classifier models. The high granularity of the data forces the ET classifier to grow deeper trees while reducing the granularity allows it to fit with shallower tree structures. Therefore, the mean model size was inversely proportional to the window size for that model type, with a maximum of ≈ 180 KB at 1-second window size. The LSTM-NN model architecture occupied 14.919 KB of memory space while the DT model only needed 1.756 KB which mostly contained model metadata.

V. DISCUSSION

A. MAIN FINDINGS

The work in this paper demonstrates that for sleep posture detection using data acquired from a single neck-located wearable accelerometer, high detection accuracy can be achieved with minimal data acquisition parameters. This work shows that the classification performance was unaffected by the change in the sampling rate when set to 5 Hz, 10 Hz, 25 Hz or 50 Hz. It also shows that there was no significant difference

in the values of the performance metrics when the window size ranges between 1 to 15 seconds. This opens the door to the possibility of power consumption optimisation at the system level, by choosing the minimum sampling rate (5 Hz), and also to minimising the detection latency by choosing a small processing window size (e.g. 1-5 seconds); without compromising the detection performance.

It also further proves that high classification performance is attainable without the need for a complex model. And for sleep position detection in the context of SUDEP prevention, the DT threshold-based approach could be the most appropriate one compared to the ET classifier and LSTM-NN model types due to its consistency and direct interpretability, as well as the ease of integration into the embedded systems of the wearable and the implications it could have in other system design factors, such as memory allocation. It was demonstrated that this model was able to perfectly capture all prone postures performed by 6 out of 8 test subjects, while having an overall mean accuracy no less than 98%.

B. LIMITATIONS AND BIASES

The main limitation of this work is the small number of participants, as well as the imbalanced gender ratio which might have imposed bias in the data. It is also worth noting that the recruiting of subjects could have been influenced by selection bias, which can impact the external validity of the classification results. However, this work is focused on the general trends and relative differences between results which mitigates these concerns and lessens their impact. Nonetheless, these limitations can be addressed in a future validation study where the subjects' gender ratio is better representing the target population, and where selection bias is accounted for. Another limitation is the data acquisition being carried out in a controlled manner where the subjects were instructed to perform which posture and at what time. This was done in order to have an equal distribution of all postures, especially the prone posture which is often under-represented in previous studies. However, a future overnight study should be done to represent free-living conditions, where the chosen model from this work can be validated. Finally, the exploratory purpose of this work required offline processing, modelling and measurement of the different evaluation metrics without the deployment of those models into the embedded system. This could be considered a limitation that can be addressed in future work.

VI. CONCLUSION

This paper provides evidence that sleep position monitoring of the four main postures can be efficiently done with a high level of accuracy and minimal data requirements using one source of data and from one body location: the neck. This finding is especially important when utilising wearable devices for SUDEP prevention, an area where sleep position carries significant weight in the likelihood of sudden death, which along with other more critical biomarkers such as respiratory rate and blood oxygen saturation, allow for an

effective multi-modal easy to use monitoring system for epilepsy patients. This work is a step further to having a low-power, easy-to-deploy and easy-to-use monitoring wearable device.

APPENDIX A THRESHOLD TREE-SEARCH - FITTED DECISION TREE

Fitting the decision tree algorithm to the training set resulted in the construction of the tree structure shown in Fig. 12.

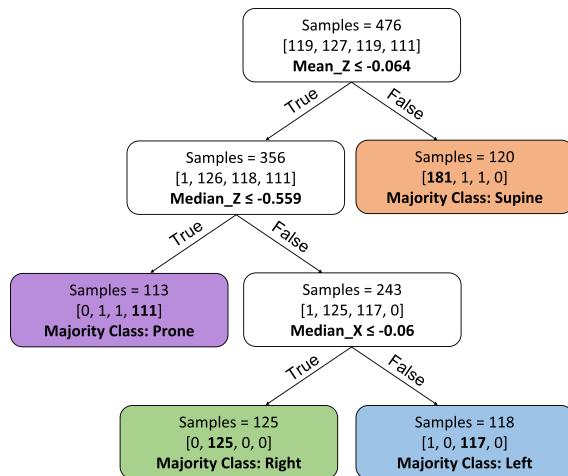


FIGURE 12. A visualisation of the fitted decision tree. External nodes (coloured boxes) represent the leaves which contain the resulting groups of samples while internal nodes (white branching boxes) represent the conditions at each level. The distribution of samples in each leaf node is shown between brackets in the class order: supine, right, left and prone.

ACKNOWLEDGMENT

The authors would like to thank Dr. Zaibaa Patel and Sukhpreet Singh for providing the experimental data used in this paper and for immensely supporting the work.

REFERENCES

- [1] H. J. Lee, S. H. Hwang, S. M. Lee, Y. G. Lim, and K. S. Park, "Estimation of body postures on bed using unconstrained ecg measurements," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, pp. 985–993, 2013.
- [2] A. Oksenberg, D. S. Silverberg, E. Arons, and H. Radwan, "Positional vs nonpositional obstructive sleep apnea patients: Anthropomorphic, nocturnal polysomnographic and multiple sleep latency test data," *Chest*, vol. 112, pp. 629–639, 9 1997.
- [3] J. A. Liebenthal, S. Wu, S. Rose, J. S. Ebersole, and J. X. Tao, "Association of prone position with sudden unexpected death in epilepsy," *Neurology*, vol. 84, pp. 703–709, 2 2015.
- [4] H. Ohta, Y. Oishi, T. Hirose, S. Nakaya, K. Tsuchiya, M. Nakagawa, H. Gima, I. Kusakawa, H. Yoda, T. Sato, T. Sasaki, H. Nishida, and T. Obonai, "Postural change for supine position does not disturb toddlers' nap," *Scientific reports*, vol. 10, 12 2020.
- [5] A. Tobola, F. J. Streit, C. Espig, O. Korpok, C. Sauter, N. Lang, B. Schmitz, C. Hofmann, M. Struck, C. Weigand, et al., "Sampling rate impact on energy consumption of biomedical signal processing systems," in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 1–6, IEEE, 2015.
- [6] Z. Zhou, S. Padgett, Z. Cai, G. Conta, Y. Wu, Q. He, S. Zhang, C. Sun, J. Liu, E. Fan, K. Meng, Z. Lin, C. Uy, J. Yang, and J. Chen, "Single-layered ultra-soft washable smart textiles for all-around ballistocardiograph, respiration, and posture monitoring during sleep," *Biosensors and Bioelectronics*, vol. 155, p. 112064, 5 2020.
- [7] J. Liu, X. Chen, S. Chen, X. Liu, Y. Wang, and L. Chen, "Tagsheet: Sleeping posture recognition with an unobtrusive passive tag matrix," *Proceedings - IEEE INFOCOM*, vol. 2019-April, pp. 874–882, 4 2019.
- [8] S. M. Mohammadi, S. Kouchaki, S. Khan, D. J. Diljk, A. Hilton, and K. Wells, "Two-step deep learning for estimating human sleep pose occluded by bed covers," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 3115–3118, 7 2019.
- [9] A. Y.-C. Tam, B. P.-H. So, T. T.-C. Chan, A. K.-Y. Cheung, D. W.-C. Wong, J. C.-W. Cheung, A. Psarrou, J. G. Rodriguez, S. Orts-Escalano, A. Garcia-Garcia, and C. Chen, "A blanket accommodative sleep posture classification system using an infrared depth camera: A deep learning approach with synthetic augmentation of blanket conditions," *Sensors* 2021, Vol. 21, Page 5553, vol. 21, p. 5553, 8 2021.
- [10] R. S. Hsiao, T. X. Chen, M. A. Bitew, C. H. Kao, and T. Y. Li, "Sleeping posture recognition using fuzzy c-means algorithm," *BioMedical Engineering Online*, vol. 17, pp. 1–19, 11 2018.
- [11] Y. Nam, Y. Kim, and J. Lee, "Sleep monitoring based on a tri-axial accelerometer and a pressure sensor," *Sensors* 2016, Vol. 16, Page 750, vol. 16, p. 750, 5 2016.
- [12] M. Enayati, M. Skubic, J. M. Keller, M. Popescu, and N. Z. Farahani, "Sleep posture classification using bed sensor data and neural networks," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2018-July, pp. 461–465, 10 2018.
- [13] Q. Hu, X. Tang, and W. Tang, "A real-time patient-specific sleeping posture recognition system using pressure sensitive conductive sheet and transfer learning," *IEEE Sensors Journal*, vol. 21, pp. 6869–6879, 3 2021.
- [14] M. Laurino, L. Arcarisi, N. Carbonaro, A. Gemignani, D. Menicucci, and A. Tognetti, "A smart bed for non-obtrusive sleep analysis in real world context," *IEEE Access*, vol. 8, pp. 45664–45673, 2020.
- [15] P. Jiang and R. Zhu, "Dual tri-axis accelerometers for monitoring physiological parameters of human body in sleep; dual tri-axis accelerometers for monitoring physiological parameters of human body in sleep," 2016.
- [16] S. Fallmann, R. V. Veen, L. Chen, D. Walker, F. Chen, and C. Pan, "Wearable accelerometer based extended sleep position recognition,"
- [17] R. M. Kwasnicki, G. W. Cross, L. Geoghegan, Z. Zhang, P. Reilly, A. Darzi, G. Z. Yang, and R. Emery, "A lightweight sensing platform for monitoring sleep quality and posture: a simulated validation study," *European Journal of Medical Research*, vol. 23, p. 28, 5 2018.
- [18] E. Sen-Gupta, D. E. Wright, J. W. Caccese, J. A. Wright, E. Jortberg, V. Bhatkar, M. Ceruolo, R. Ghaffari, D. L. Clason, J. P. Maynard, and A. H. Combs, "A pivotal study to validate the performance of a novel wearable sensor and system for biometric monitoring in clinical and remote environments," 2019.
- [19] S. Jeon, T. Park, A. Paul, Y. S. Lee, and S. H. Son, "A wearable sleep position tracking system based on dynamic state transition framework," *IEEE Access*, vol. 7, pp. 135742–135756, 2019.
- [20] E. P. Doheny, M. M. Lowery, A. Russell, and S. Ryan, "Estimation of respiration rate and sleeping position using a wearable accelerometer," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2020-July, pp. 4668–4671, 7 2020.
- [21] E. J. Smits, S. Salomoni, N. Costa, B. Rodríguez-Romero, and P. W. Hodges, "How reliable is measurement of posture during sleep: real-world measurement of body posture and movement during sleep using accelerometers," *Physiological Measurement*, vol. 43, p. 015001, 1 2022.
- [22] J. Ranta, E. Ilén, K. Palmu, J. Salama, O. Roienko, and S. Vanhatalo, "An openly available wearable, a diaper cover, monitors infant's respiration and position during rest and sleep," *Acta Paediatrica*, vol. 110, pp. 2766–2771, 10 2021.
- [23] A. Manoni, F. Loreti, V. Radicioni, D. Pellegrino, L. D. Torre, A. Gumiero, D. Halicki, P. Palange, and F. Irrera, "A new wearable system for home sleep apnea testing, screening, and classification," *Sensors* 2020, Vol. 20, Page 7014, vol. 20, p. 7014, 12 2020.
- [24] S. Nabavi and S. Bhadra, "Smart mandibular advancement device for intraoral monitoring of cardiorespiratory parameters and sleeping postures," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, pp. 248–258, 4 2021.
- [25] H. Yoon, S. Hwang, D. Jung, S. Choi, K. Joo, J. Choi, Y. Lee, D.-U. Jeong, and K. Park, "Estimation of sleep posture using a patch-type accelerometer based device," vol. 2015-November, pp. 4942–4945, IEEE, 8 2015.
- [26] I. Ferrer-Lluis, Y. Castillo-Escario, J. M. Montserrat, and R. Jane, "Analysis of smartphone triaxial accelerometry for monitoring sleep-disordered

- breathing and sleep position at home," IEEE Access, vol. 8, pp. 71231–71244, 2020.
- [27] I. Ferrer-Lluis, Y. Castillo-Escario, J. M. Montserrat, and R. Jané, "Sleep-pos app: An automated smartphone application for angle based high resolution sleep position monitoring and treatment," Sensors 2021, Vol. 21, Page 4531, vol. 21, p. 4531, 7 2021.
- [28] W. Kukwa, T. Lis, J. Łaba, R. B. Mitchell, and M. Młyńczak, "Sleep position detection with a wireless audio-motion sensor—a validation study," Diagnostics 2022, Vol. 12, Page 1195, vol. 12, p. 1195, 5 2022.
- [29] P. Alinia, A. Samadani, M. Milosevic, H. Ghasemzadeh, and S. Parvaneh, "Pervasive lying posture tracking," Sensors 2020, Vol. 20, Page 5953, vol. 20, p. 5953, 10 2020.
- [30] R. S. Abdulsadig, S. Singh, Z. Patel, and E. Rodriguez-Villegas, "Sleep posture detection using an accelerometer placed on the neck," 2022 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2430–2433, 7 2022.
- [31] P. Corbishley and E. Rodriguez-Villegas, "Breathing detection: towards a miniaturized, wearable, battery-operated monitoring system," IEEE Transactions on Biomedical Engineering, vol. 55, no. 1, pp. 196–204, 2007.
- [32] N. Devani, R. X. A. Pramono, S. A. Imtiaz, S. Bowyer, E. Rodriguez-Villegas, and S. Mandal, "Accuracy and usability of acupebble sa100 for automated diagnosis of obstructive sleep apnoea in the home environment setting: An evaluation study," BMJ open, vol. 11, no. 12, p. e046803, 2021.
- [33] A. Abdallah, R. S. Abdulsadig, and M. Amien, "A comparative study on human loco-motor activity recognition using wearable sensors," Proceedings of the International Conference on Computer, Control, Electrical, and Electronics Engineering 2019, ICCCEEE 2019, 9 2019.
- [34] J. Han, J. Pei, and M. Kamber, Data mining: concepts and techniques. Elsevier, 2011.
- [35] O. R. Lior and Maimon, "Decision trees," 2005.
- [36] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Machine learning, vol. 63, pp. 3–42, 2006.
- [37] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," IEEE Transactions on Neural Networks and Learning Systems, vol. 28, pp. 2222–2232, 10 2017.
- [38] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization."



ESTHER RODRIGUEZ-VILLEGAS is a Professor (Chair) of Low-Power Electronics at Imperial College London, originally known for her engineering techniques to drastically reduce power in integrated circuits. She subsequently focused her research on life-science applications, founding the Wearable Technologies Lab. This lab specialises on both: creating innovative wearable medical technologies to improve the management and diagnosis of chronic diseases; and neural interfaces to facilitate brain research whilst improving animals' welfare. Esther is also a founder, co-CEO/CSO, of two active life-sciences companies, Acurable and TainiTec. She has served in many prestigious international technical committees, including, but not limited, to the Administrative Committee of the IEEE Solid-State Circuit Society, IEEE ISSCC, IEEE ISCAS, IEEE ESSCIRC. Esther has received many international recognitions and awards. Including an IET Innovation Award (2009), a global XPRIZE-award (2014), AAALAC 3Rs award (2018), and a Silver Medal from the UK Royal Academy of Engineering (2020). She was also named the top scientist/engineer in Spain under the age of 36 in 2009 (Complutense award); and in 2020 she was elected as a Fellow of the UK Royal Academy of Engineering.

• • •



RAWAN S. ABDULSADIG (Member, IEEE) received her B.Sc. (Hons.) in Electrical and Electronic Engineering from the University of Khartoum, Khartoum, Sudan, (2018), where she specialised in Electronic Systems Software Engineering. She then received her M.Sc. in Data Science from Lancaster University, Lancaster, UK, (2020), where she was awarded the outstanding student prize offered by the Data Science Institute for consistently demonstrating high levels of achievement

throughout the academic year. She joined the early oncology R&D team in AstraZeneca for her placement project where she applied deep learning techniques into the digital pathology domain. She is currently working as a research assistant in the Wearable Technologies Lab, Department of Electrical and Electronic Engineering, Imperial College London, UK. Her current research interests are mainly about utilising data science and AI in improving healthcare.