IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

*Digital Object Identifier 10.1109/ACCESS.2024.DOI*

# USSD: Unsupervised Sleep Spindle Detector

**E. Ramírez[1], P.A. Estevez[1,2,3], (Fellow, IEEE), M. Adams[1], (Senior Member, IEEE), C.A. Perez[1,2,3], (Senior Member, IEEE), M. Garrido[4,5] and P. Peirano[4]**

[1]Department of Electrical Engineering, Universidad de Chile, Santiago, Chile
[2]Millennium Institute of Intelligent Healthcare Engineering, Santiago, Chile
[3]IMPACT, Center of Interventional Medicine for Precision and Advanced Cellular Therapy, Santiago, Chile
[4]Sleep Laboratory, Institute of Nutrition and Food Technology, Universidad de Chile, Santiago, Chile
[5]UDP Facultad de Educación, Universidad Diego Portales, Santiago, Chile

Corresponding author: Pablo Estévez (e-mail:pestevez@cec.uchile.cl).

**ABSTRACT** Sleep spindles (SSs) appear in electroencephalogram (EEG) recordings during sleep stage N2, and they are usually detected through visual inspection by an expert. Labeling SSs in large datasets is time-consuming and depends on the expert criteria. In this work, we propose an unsupervised SS detector based on dictionary learning called the Unsupervised Sleep Spindle Detector (USSD). The proposed detector learns prototype SSs of different lengths (called atoms). An unsupervised adaptive threshold method based on the distribution of the automatically detected SS lengths is developed, which allows the adaptation of the USSD algorithm to different datasets in an unsupervised way. For each detection, the USSD estimates the probability of being an SS. The USSD performances on the labeled MASS-SS2 and INTA-UCH datasets yield F1-scores of $0.72 \pm 0.02$ and $0.72 \pm 0.04$, respectively. The USSD outperforms the A7 and LUNA detectors, which are traditional unsupervised models. Next, we fine-tune the resulting USSD model with $20\%$ of the labeled MASS-SS2 and INTA-UCH datasets, achieving F1 scores of $0.78 \pm 0.06$ and $0.75 \pm 0.05$, respectively. In addition, the SSs detected by USSD on the unlabeled CAP dataset are used to pre-train a supervised deep learning method, which after fine-tuning with $20\%$ of the MODA dataset, reaches an F1-score of $0.81 \pm 0.02$.

**INDEX TERMS** EEG, Sleep spindle, Unsupervised learning, Dictionary learning.

## I. INTRODUCTION

SLEEP patterns have been increasingly recognized as relevant factors in brain health [1]. Human sleep is divided into two main states: nonrapid eye movement (NREM) sleep and rapid eye movement (REM) sleep [2]. NREM sleep is further subdivided into three stages: N1, N2, and N3. Each stage serves distinct functions and is characterized by unique brain wave patterns observable in electroencephalogram (EEG) recordings. The importance of one factor or another varies depending on the age [3]. In short, N1 is the lightest stage of sleep, serving as the transition from wakefulness to sleep. N2 is marked by the presence of sleep spindles and K-complexes, which play a crucial role in sleep maintenance, information processing, and memory reinforcement capabilities from early infancy to aging. N3 is essential for physical restoration and recovery, strengthening immune function [4], and providing optimal time for clearing the brain of cellular debris and neurotoxins [5]. Diurnal sleep is also a key player, even more relevant considering that is mainly made up of NREM sleep [6]. Last but not least, REM sleep, which is characterized by rapid eye movements and dreaming, is critical for emotional regulation, memory processing, and cognitive function [7]. The time course of REM sleep development (and decline) in humans corresponds well with critical periods of brain maturation [8]. Indeed, REM sleep provides endogenous stimulation to promote brain development at ages when wake-related stimulation is low [9]. Figure 1 shows examples of EEG signals in the REM and NREM states (N1,N2, and N3).

In this work, we focus on detecting sleep spindles (SSs), which are transient events that occur mainly during the N2 stage of the NREM sleep. Figure 1 shows three examples of SSs in stage N2, with one of them shown in greater detail. SSs are burst-like sequences of sinusoidal cycles with a frequency of 11-16 Hz and a characteristic spindle shape that explains their name [10]. These events evolve throughout

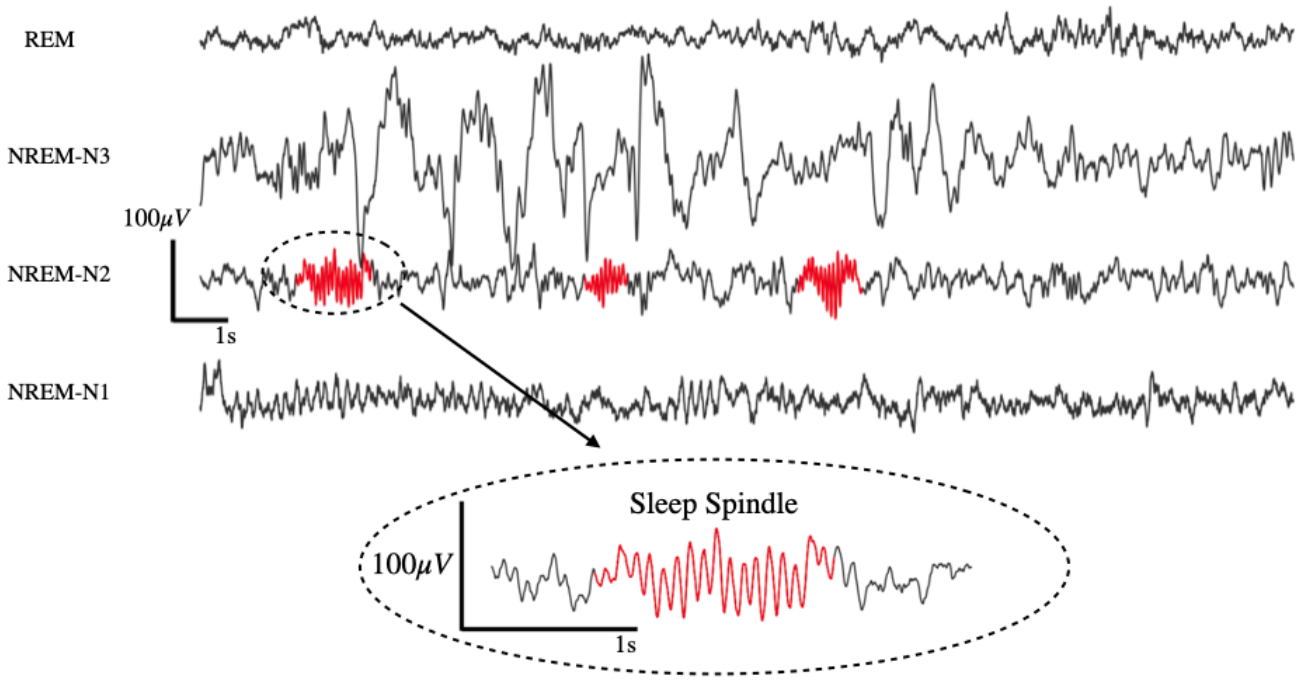



**Figure 1. Example of EEG signals during different sleep stages.**
Raw EEG signals in REM and NREM states (N1, N2, and N3). In NREM sleep-N2, three sleep spindles can be observed, marked in red. At the bottom, a close-up of a sleep spindle is shown. In NREM-N3, the characteristic slow waves sleep (SwS) can be observed.

life in parallel with brain maturity, and their density has been correlated with different assessments of intelligence, memory consolidation, and hippocampal-dependent learning [11].

Even though most studies have been performed in adults, a growing body of evidence supports the hypothesis that information processing and memory reinforcement capabilities during sleep occur throughout the human lifespan, i.e., from early infancy to aging [12]. Furthermore, adverse conditions in infancy could alter the normal developmental progression [13]. In addition, several studies suggest that sleep spindles (or a coupling between spindles and slow oscillations) are key mediators for sleep-associated memory consolidation [14], [15], [16], [17].

SSs also facilitate the synaptic plasticity necessary for transferring information from the hippocampus to the neocortex during sleep [11], [18]. SSs are involved in strengthening neural connections, which underlie the retention and recall of information [10], [11]. Furthermore, abnormalities in spindle activity have been associated with adverse effects on several health domains. In adults, this includes sleep disorders and neuropsychiatric conditions, such as insomnia [19], [20], schizophrenia [21], [22], and epilepsy [23], [24], highlighting their significance in both normal cognitive function and performance, and pathological states.

The state of the art of automatic SS detection is based on supervised deep learning [25], [26], [27]. However, supervised learning requires human experts to label the SSs in EEG recordings. This approach has two main difficulties: 1) Experts usually differ in their criteria for detecting SSs, making it difficult to automate the detection and characterization of SSs [28], and 2) this is a very time-consuming task. In addition, only a few public datasets with labeled SSs, having tens or hundreds of EEG recordings, are available, such as MASS [29] and MODA [30].

On the other hand, there are several SS detection algorithms that do not require labeled datasets, such as A7 [31], LUNA [32], and MPP [33], which are based on handcrafted features, Morlet wavelets, and dictionary learning, respectively. The advantage of these algorithms is that they can be applied to large unlabeled datasets such as NSRR [34]. However, there is usually a large performance gap between supervised and unsupervised approaches. An interesting point is that unsupervised approaches can be used to leverage supervised algorithms through the use of unlabeled datasets. For example, in [25], a deep learning-based detector called SEED was pre-trained using labels obtained automatically by applying A7 to the unlabeled CAP dataset. Then, this pre-trained network was fine-tuned with 10% of the MODA dataset, achieving state-of-the-art performance on the complete MODA dataset (an F1-score of 78.8).

In this work, we propose an unsupervised SS detector based on dictionary learning called the Unsupervised Sleep Spindle Detector (USSD). The main contributions of this work are the following:

- An unsupervised method for detecting SSs of variable lengths based on dictionary learning is proposed, which assigns a probability of being an SS to each detection.
- A novel unsupervised method is developed for setting

the single threshold of the USSD algorithm based on the distribution of the lengths of the detected SSs. This allows the USSD method to be automatically adapted to subjects of different ages and from different EEG datasets.
- The proposed USSD method outperforms the A7 and LUNA algorithms on three SS-labeled datasets.
- A fine-tuned version of the USSD model is provided, which improves the USSD algorithm performance when trained with 10-20% of the labeled SSs from a dataset.
- The SSs detected by the USSD model on the unlabeled CAP dataset are used to pre-train SEED, a supervised deep learning method, which reaches state-of-the-art performance after fine-tuning with 20% of the MODA dataset.

The remainder of this article is organized as follows. Section 2 reviews the related work, highlighting the main unsupervised algorithms for SS detection. Section 3 provides a description of the proposed USSD model, based on dictionary learning. Next, the labeled and unlabeled datasets used in this work are described, followed by the EEG signal preprocesing and the metrics employed to evaluate the models. Section 4 provides a description of the proposed fine-tuned USSD model. Section 5 presents the results and compares the performance of the proposed USSD model with the A7 and LUNA detectors, across different datasets. The results of the fine-tuned USSD model are also presented, as well as the results of transfer learning. Section 6 presents a discussion of the limitations and cost of the USSD algorithm. Finally, section 7 presents the conclusions, highlighting the contributions of the proposed USSD model in SS detection across three sleep EEG datasets.

## II. RELATED WORK

This section describes the related work on SS detection using unsupervised and transfer learning methods.

### A. UNSUPERVISED ALGORITHMS

Unsupervised algorithms for sleep spindle detection can be categorized as follows: signal processing-based, dictionary learning, and deep learning. The algorithms developed by Mölle et al. [35], Ferrarelli et al. [36], and Andrillon et al. [37] employ signal processing tools, such as the FFT and the Hilbert transform.

The Delay Differential Analysis (DDA) algorithm [38], based on delay embedding and delay differential equations, is applied to detect SSs in human intracranial sleep data. The Deep Dynamic Neural Bayesian Network (DNDBN) model [39] uses the Hidden Semi-Markov Model (HSM), which is a variation of a Markov model, and has supervised and unsupervised versions (S-HSM and NHSM). Both [38] and [39] were evaluated on the DREAMS Sleep Spindles database, which is a small dataset consisting of 8 excerpts, each 30 min long [40].

The algorithm in [41] uses Matching Pursuit, a dictionary learning-based algorithm, and Least Square Boosting, to obtain an F1-score of 70.8% on the small DREAMS dataset. In [42], a procedure for the atomic decomposition of multichannel EEGs is proposed based on matching pursuit and dipolar inverse solution. The basic algorithm achieved an F1-score of 47% in the MODA dataset [30], which improved to 63% after rejecting detections with a local signal-noise ratio lower than a given threshold.

In what follows, we describe the three main unsupervised algorithms related to this work:

- Marked Point Process (MPP), Loza & Príncipe [33], [43] is an unsupervised learning algorithm based on dictionary learning [44]. This model is based on the idea that a single EEG recording is the result of transient, reoccurring patterns over time added to a noisy background [43]. The objective is to find the precise duration (starting and ending points), amplitude, and shape of SSs. An alternating approximation of finding a sparse approximation and learning a dictionary of atoms (SS prototypes) is performed. This algorithm takes as input a filtered EEG signal in the sigma band and then decomposes it into background noise and a set of SS candidates. It uses a single threshold for SS detection. Figure 2 shows a diagram with the main steps of the MPP algorithm. In this work, we propose an SS detector that is inspired by the MPP algorithm. In order to achieve state-of-the-art performance, significant contributions are made to detect variable-length SSs and to set the threshold parameter in an unsupervised way.
- A7, Lacourse et al. [31], is an algorithm that uses four features to detect sleep spindles in EEG recordings. In this algorithm, $EEG_{bf}$ is defined as the broadband filtered EEG signals (0.3–30) Hz, and $EEG_s$ as the EEG signal filtered in the sigma band. The four characteristics are the following: Absolute sigma power, which represents the power of $EEG_s$; Relative sigma power, which is the ratio between the power of $EEG_s$ and $EEG_{bf}$ power; sigma covariance, which is the covariance between ($EEG_s$, $EEG_{bf}$); and Sigma correlation, which is the correlation between ($EEG_s$, $EEG_{bf}$). This algorithm has four thresholds, each associated with one of the four features. When all four features are above their respective thresholds, an event is considered a sleep spindle. This algorithm is unsupervised, although fine-tuning of its thresholds is possible.
- The LUNA algorithm, detailed in Purcell et al. [32], is an open-source software package for detecting sleep spindles, using Morlet wavelets [45], [46]. Each sleep spindle event is identified as such if the candidate SS exceeds the algorithm's threshold (defined as 4.5 times the mean of the artifact-free signal). The event duration is calculated as the width of the half-height of the energy represented in the wavelet spectral decomposition.

In this work, A7 and LUNA are used as baseline algorithms for comparison purposes.

Many of these algorithms use one or more thresholds to

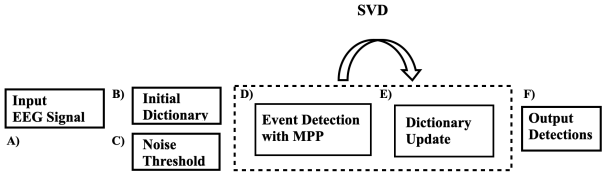



**Figure 2. MPP Block Diagram.**
The MPP algorithm is divided into 4 blocks: A) Input, EEG signal filtered in the sigma band (11-16 Hz). B) Initial generation of the dictionary from the input. C) Determination of the MPP threshold, where any signal below the threshold is considered noise. D) Generation of a set of detections from the dictionary. E) Singular value decomposition (SVD) is applied to the detections to update the dictionary and then step D is repeated to obtain a new set of detections. The number of repetitions of stages D and E is a parameter of the MPP algorithm. F) The output of the algorithm corresponds to the detections found in block D, when the dictionary updates stop.

adjust their parameters. Examples with one threshold are [32], [33], [38], two thresholds [36], [37], three thresholds [35] and four thresholds [31]. This could lead to a problem known as overfitting, where the algorithm's parameters are over adjusted to a given training set. However, when the algorithm is applied to a new dataset, it performs less effectively. Some algorithms [32], [33] adjust their thresholds subject-by-subject to avoid parameter overfitting. However, adjusting these thresholds is not sufficient to compensate for the EEG variability between subjects of different age ranges, such as adults and children. This variability is caused by the thalamic origin of sleep spindles changing as the thalamo-cortical system develops [47]. Another variability is the different criteria used by experts to label the EEG signal [28].

### B. DICTIONARY LEARNING

The proposed USSD method is based on dictionary learning, the principles of which are explained in this subsection. Dictionary learning (also known as sparse coding) is a representation learning method that aims at finding a sparse representation [44]. Sparse representation aims to express a signal $Y$ of dimension $\eta$ as a linear combination of signals taken from a dictionary. The dictionary elements are typically unit norm functions called atoms [44]. In accordance with this, a signal Y can be described by the following equation:

$$Y \approx \mathbf{\Phi}\alpha = \sum_{k=1}^{N} \alpha_k \phi_k, \tag{1}$$

where $\phi_k$ are the atoms in the dictionary, with $K = 1, ..., N$, and $N$ is the size of the dictionary. $\alpha_k$ are the scalars that are associated with each $\phi_k$ in the linear combination of the signal $Y$. $\mathbf{\Phi}$ is the vector composed of $\phi_k$ and $\alpha$ is the vector composed by $\alpha_k$. The dictionary is overcomplete ($N > \eta$) when it spans the signal space, and its atoms are linearly dependent. In that case, every signal can be represented as a linear combination of atoms in the dictionary. Because the dictionary is overcomplete, $\alpha$ in Eq. 1 is not unique. The requirement of finding an exact representation is generally relaxed to achieve efficient and sparse representations.

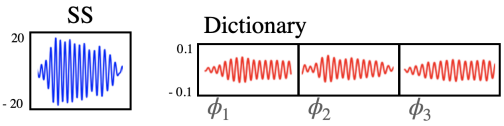



**Figure 3. Example of dictionary learning.**
Signals of the same duration (0.7s) are shown; the one on the left is the filtered SS in the sigma band with a peak-to-peak amplitude of $40\mu V$, and the three signals on the right are the normalized atoms of a dictionary.

Figure 3 shows an example that considers the signal $Y$ as a filtered SS in the sigma band and a dictionary formed by three atoms. The dictionary is overcomplete since it satisfies $(N > \eta) = (3 > 1)$. In the MPP method, only a single atom is selected to approximate an SS event, which is carried out via convolutions. In this example, $\phi_2$ is the closest atom when scaled up by $\alpha_2$, thus the SS event is approximated as follows:

$$SS \approx \phi_{\mathbf{2}}\alpha_{\mathbf{2}}. \tag{2}$$

### C. TRANSFER LEARNING

As mentioned in the introduction, unsupervised models can be used to pre-train a deep learning-based model. Transfer learning is a technique that uses knowledge from a pre-trained model to fine-tune another on a related task. In [48], a neural network model trained on sleep spindle data of healthy subjects is applied to the data of subjects with insomnia and then fine-tuned to achieve a classification accuracy of 92.80% on a private dataset. In [49], an unsupervised transfer learning model for sleep stage classification is proposed, where the EEG recordings in the target domain are unlabeled. In this work, the Sleep EEG Event Detector (SEED) [25] will be pre-trained with labels automatically obtained by applying the USSD model on the unlabeled CAP dataset. Then the pre-trained model will be fine-tuned with a different dataset containing expert labels.

SEED is a supervised algorithm that takes an EEG signal and estimates the probability of each sample being part of an event of interest (SS or k-complex). The local feature extraction consists of a stack of convolutional layers with an initial number of filters Nf, a kernel of size 3, zero-padding, ReLU activation, and batch-normalization after each convolution and before ReLU. Downsampling is performed after two convolutions using a pooling layer of size 2, and the number of filters is doubled after pooling. Batch-normalization is also applied before the first convolutional layer. Two recurrent LSTM layers are concatenated to form a Bidirectional LSTM (BiLSTM). By-sample classification is achieved by two 1D convolutional layers with kernels of size 1. The classification stage maps each time step to the probability of belonging to the background (class 0) or to an event of interest (class 1). SEED achieved state-of-the-art performance on the MASS-SS2 dataset with an F1-score of 0.805 and on the MODA dataset with an F1-score of 0.818.

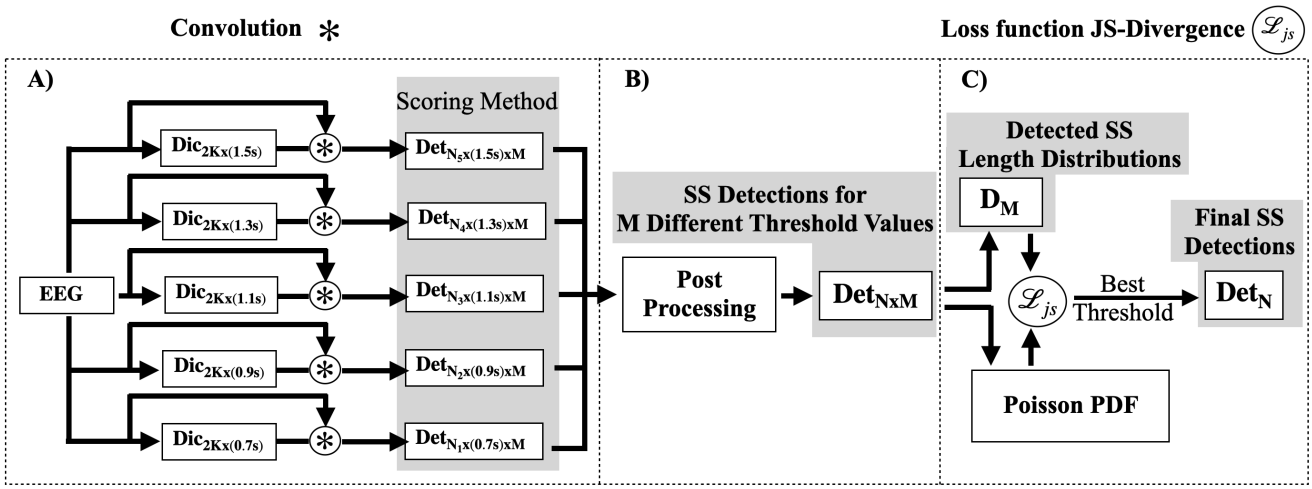



**Figure 4. Diagram of the USSD model.**
The diagram shows the USSD algorithm, which is made up of three stages as follows: **In Stage A,** the filtered EEG signal in the sigma band is convolved with the atoms of five dictionaries (Dic) of different lengths, where the convolution operator is denoted by the symbol $*$. By thresholding the resulting convolutions five sets of detections (Det) are generated with dimensions $(n, L \times freq, M)$, where n is the number of detections, L is the length of the dictionary atoms, freq is the EEG sampling frequency, and M are the values that the single threshold can take. **In Stage B,** post-processing is applied to the detection sets, which involves adjusting the length or eliminating detections that do not meet the frequency and amplitude criteria of the official manual of the AASM [2] that characterizes SSs. **In stage C,** the best threshold is set by minimizing the Jensen-Shannon divergence $\mathscr{L}_{js}$ between the Poisson distribution and the detected SS length distributions obtained for the M possible values of the single threshold. The algorithm's output is the set of SS detections associated with the best threshold.

## III. METHODS

### A. UNSUPERVISED SLEEP SPINDLE DETECTOR

Figure 4 shows the diagram of the proposed USSD model. This is divided into three stages: A) Involves generating a set of initial detections. The model input is the filtered EEG signal in the sigma band (11-16 Hz), which is used to generate five dictionaries with lengths varying between 0.7 and 1.5s. Next, the convolution (*) is calculated between the dictionary atoms and the EEG signal filtered in the sigma band. An event is detected when the convolution surpasses a given threshold. B) Post-processing of the initial detections involves merging detections made by multiple dictionaries and adjusting the lengths of the SS detections using the American Academy of Sleep Medicine (AASM) criteria [2] such as amplitude and instantaneous frequency. C) An adaptive threshold algorithm is applied to the variable-length detections resulting from block B. This adaptive threshold utilizes the distribution of detection durations as its main feature, a feature that is assumed to exhibit low variability across different datasets.
The three blocks that comprise the USSD algorithm are described in detail below.

#### Stage A

This stage takes the EEG signal, filtered in the sigma band, as input. This signal is used to generate five dictionaries with fixed atom durations, defined as L = [0.7, 0.9, 1.1, 1.3, 1.5] s. The number of atoms in each dictionary is 2K, where K is an input parameter for our algorithm, typically K=10. The first K atoms have an average frequency $\leq$ 13 Hz, and the last K atoms have an average frequency > 13 Hz. This division of atoms is carried out to distinguish between slow and fast sleep spindles.

A set of detections is generated by convolving the atoms of a dictionary with the EEG signal filtered in the sigma band; see Figure 5. A detection is made when a peak in the convolution exceeds the threshold of our algorithm, as shown in Fig. 5 C). For each subject, the threshold can take one out of M-values between the minimum and maximum peaks of the convolution, with $M$ being a user-defined parameter. Typically $M = 14$. In Stage C, we describe a method to select the best threshold among the M possible values.

Let $X$ be the filtered EEG signal in the sigma band and $X[n]$ a segment of $X$, which is decomposed into the background activity $\eta_m$ and $y_m$, a set of SSs, where $m$ indicates the value of the single threshold. Each SS is defined by parameters $\alpha, \tau$, and $\omega$, where $\alpha$ is the amplitude of the SS, $\tau$ is the position at the center of the SS in the EEG signal, $\omega$ indicates the index of the atom in the dictionary $D$. $N$ is the number of atoms in each $D$, the subscript $l$ denotes the dictionary index, and $L_l$ stands for the length of the SS. The corresponding equations are:

$$X_m[n] = \eta_m[n] + y_m[n], \tag{3}$$

$$y_m[n] = \sum_{j=1}^{N_l} \sum_{p=-L_l/2}^{L_l/2} \alpha_j \delta[n - \tau_j - p] D_{l,\omega_j}[p]. \tag{4}$$

We use five dictionary atoms, each one having a different fixed duration. After adjusting the detected SS durations (Stage B), their lengths will vary with respect to the initial

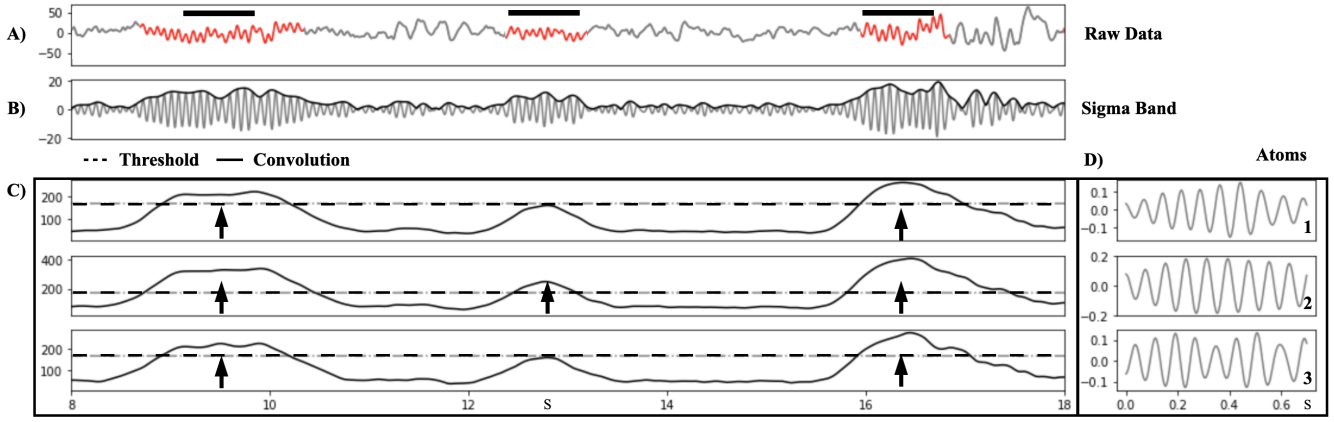



**Figure 5. Convolution between atoms and EEG filtered signal.**
In A), the EEG signal (raw data ) is shown with the sleep spindles highlighted in red. The black bar indicates the preliminary detection associated with the dictionary and the threshold. In B), the EEG signal filtered in the sigma band, which corresponds to the input, is shown. In C), the convolution between the input B) and the atoms of the dictionary D) can be observed. D) shows a dictionary formed by three atoms, each 0.7 s long, In C), the dashed line represents the algorithm threshold (Thr), and the vertical black arrows indicate the candidates for sleep spindles. All three atoms shown in D) detect the first and third sleep spindles, only $atom_2$ detects the second sleep spindle.

durations. We restrict the atom length to the interval [0.7 - 1.5] seconds. We use dictionaries whose atom lengths are 0.2 seconds apart to reduce the events detected by multiple dictionaries. The duration of the SS detections is defined in the range [0.4,2] s.

### 1) Learning atoms

The USSD algorithm learns atoms for each dictionary. To achieve this, we consider a dictionary and its corresponding set of initial detections as follows:

- Initial detections are assigned to the closest atom in the dictionary.
- Singular Value Decomposition (SVD) is applied to the set of initial detections associated with an atom, and the principal value is identified.
- The principal component ($PC$) of the SVD and its corresponding $atom$ are compared using the Euclidean distance metric ($L_2$).
- A parameter $\epsilon = 10^{-4}$ is defined to determine whether the $atom$ is replaced. If $L_2 \leq \epsilon$, the $atom$ remains unchanged; if $L_2 > \epsilon$, the $atom$ is replaced by the $PC$ of the SVD as follows,

$$L_2(PC, atom) = \begin{cases} L_2 \leq \epsilon, & \text{No change} \\ L_2 > \epsilon, & atom = PC. \end{cases} \quad (5)$$

A new set of detections is generated if the atoms have been changed. The number of repetitions is a parameter of the algorithm.

### 2) Event scoring

For each event detected, we compute the likelihood of being a sleep spindle. The scoring method is based on determining how many dictionaries detect a given event, and its duration. Each event is assigned a high, medium, or low score, as follows:

- Events of longer duration are assigned a high probability. Specifically, SSs lasting more than $\lambda + 0.5$ seconds and that are simultaneously detected by the five dictionaries, where $\lambda$ stands for the mean of the length distribution of SSs. They follow a Poisson distribution with a mean close to 0.7s.
- Events of short duration of less than 0.5 seconds, detected only by the 0.7s dictionary, are assigned a low probability.
- Events with medium probability include all detections not classified as having high or low probability.

This procedure of assigning high and low scores is biased towards events of longer duration. This is done because the shorter the event, the higher the probability of confusion with non-SS events. For a high score, the detection must have a duration longer than the mean of the distribution of detections, which is approximately 0.7s. For a low score, the detection must be less than 0.5 seconds. It is important to note that the event duration is different from the fixed duration associated with a dictionary, since the SS lengths detected in Stage A can be modified in Stage B, as explained below.

## Stage B

This stage applies post-processing to the set of detections obtained by each dictionary in Stage A. The post-processing checks if the SS detections comply with the standard definition of SS made by the AASM, where each SS must have a peak-to-peak amplitude >20 mV and a range of frequencies of 11 -16 Hz. This involves the following steps:

1) Merge repeated detections made by multiple dictionaries.

2) Eliminate detections that do not meet the standard criteria.
3) Adjust the duration of events based on the standard AASM criteria of frequency and amplitude.

These steps transform the set of fixed-length detections into a set of variable-length detections for SSs. Let $X_m$ be the EEG filtered in the sigma band, which decomposes into background noise $\eta_m[n]$ and a set of SS detections, $\hat{y}_m[n]$, for different threshold values $(m)$. Let $\hat{y}_m$ be the set of SS detections of variable length, with $m = 1, ..., M$ being the index of the $m$th threshold value. $\hat{y}_m$ is composed of detections of SSs from all dictionaries. After applying this procedure, Eqs. (3) - (4) are changed to:

$$X_m[n] = \eta_m[n] + \hat{y}_m[n] \tag{6}$$

$$\hat{y}_m[n] = \sum_{j=1}^{N} \sum_{p=-L_l/2}^{L_l/2} \hat{\alpha}_j \delta[n - \hat{\tau}_j - p] \hat{D}_{l,w_j}[p], \tag{7}$$

where $\hat{\alpha}, \hat{\tau}$ and $\hat{D}$ are the amplitude, position at the center of an SS of length $L_l$, and dictionary, respectively, after post-processing each event detection, for the different threshold values $(m)$. $\hat{D}$ is composed of atoms of variable length obtained from the five initial dictionaries, with $N$ being the total number of atoms in $\hat{D}$.

**Stage C**

In this stage, we determine the best threshold among the $M$ possible values. This is based on the assumption that the distribution of SS lengths follows a Poisson distribution, whose formula is as follows:

$$P(x, \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}, \tag{8}$$

where $\lambda$ represents the average number of events. To select the threshold, we minimize a cost function to find the distribution of durations among the $M$ thresholds that best resemble a Poisson distribution. This is an unsupervised method for setting the threshold. First, for each threshold value, we generate a normalized distribution of the duration of detections, as shown in Figure 6 A). Second, we determine the best-fitting Poisson distribution to the average of the duration distributions for the M threshold values, see Figure 6 B). The parameter $\lambda$ is constrained within the range of [0.5-1.1] s.

The Jensen-Shannon divergence $\mathscr{L}_{js}$ is minimized, to find the Poisson distribution that best fits the average distribution obtained from the sets of detections, see Figure 6 B). The Jensen-Shannon divergence is defined as follows:

$$\mathscr{L}_{js}(P||Q) = \frac{D_{kl}(P||\mathcal{M}) + D_{kl}(Q||\mathcal{M})}{2} \tag{9}$$

$$\mathcal{M} = \frac{P + Q}{2} \tag{10}$$

$$D_{kl}(P||Q) = \sum_x P(x) log\left(\frac{P(x)}{Q(x)}\right) \tag{11}$$

$$\hat{P} = \min_{\lambda}(\mathscr{L}_{js}), \tag{12}$$

where $\hat{P}$ is the best Poisson distribution, $D_{kl}$ is the Kullback-Leibler divergence, $P$ is the Poisson distribution for different $\lambda$, and $Q$ is the average length distribution of the $M$ sets of SS detections.

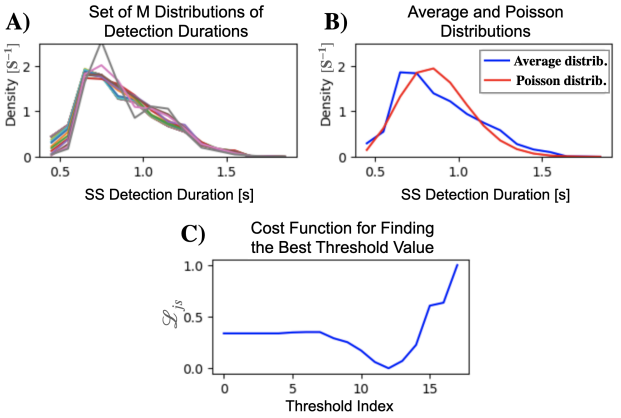



**Figure 6. Adaptive threshold procedure.**
In A), the distributions of SS durations obtained with M different threshold values are shown. This corresponds to the output of Stage B of the USSD algorithm in Figure 4. In B), the average of the M distributions (blue line), and the Poisson distribution (red line) that best fits the average distributions are shown. In C) the cost function $\mathscr{L}_{js}$ against the threshold indexes (from 1 to 20) is shown, where the minimum value of $\mathscr{L}_{js}$ is obtained with a threshold index of 12, in this example.

The last step consists in determining which of the $M$ distributions of detections most closely resembles the Poisson distribution, as follows:

$$\min_{m}(\mathscr{L}_{js}(\hat{P}||\hat{Q})), \tag{13}$$

where $\hat{Q}$ is the distribution of the lengths of the SSs and $m$ is the index that corresponds to the $m$th threshold value, see Figure 6 C.

### B. DATASETS

Table 1 shows five polysomnographic (PSG) sleep datasets: The Institute of Nutrition and Food Technology of the University of Chile (INTA-UCH) dataset [50]; the Montreal Sleep Studies (MASS) dataset [29], which is divided into the cohort 2 ($SS_2$) and cohorts $SS_{1,3,4,5}$; the MODA dataset [30], and the Cyclic Change Allocation Patterns (CAP) dataset [51], [52]. The INTA-UCH, MASS-$SS_2$, and MODA datasets have labeled sleep spindles. The MASS-$SS_{1,3,4,5}$, and CAP datasets are unlabeled, but their sleep stages are annotated. Table 1 includes the details of the five datasets regarding age, gender, number of subjects, number of sleep spindles, and the EEG channel used. All datasets use the international 10-20 EEG recording system [53]. According to the AASM standard criteria, slow SSs of maximal amplitude are obtained in the central derivations of the EEG, i.e., channels C3/C4, and fast SSs in the frontal brain region corresponding to the

F3/F4 derivations. As shown in Table 1, the MASS, MODA, and CAP datasets use the C3 channel, and INTA-UCH uses the F4 channel. The INTA-UCH dataset was recorded with a Cadwell Easy II PSG/EEG system, and the MASS dataset was recorded with a Grass Model 12 or 15 PSG/EEG system, depending on the cohort. The massive online data annotation (MODA) dataset [30] corresponds to an artifact-free segment extraction of the PSGs from the 5 MASS cohorts. The MASS-SS2 and INTA-UCH datasets contain 9990 and 12237 annotated sleep spindles, respectively, while the MODA dataset includes 5272 annotated sleep spindles, totaling 27,499 SSs.

The subjects from the INTA-UCH, MASS cohorts 2-5, and 16 subjects from the CAP datasets are free from neurological disorders and medications that affect the central nervous system. The MASS-SS1 cohort includes 15 subjects, who were diagnosed with Mild Cognitive Impairment (MCI). The CAP dataset includes subjects with the following pathologies: 2 with bruxism, 9 with insomnia, 5 with narcolepsy, 40 with nocturnal frontal lobe epilepsy, 10 with periodic leg movements, 22 with REM sleep behavior disorder, and 4 with sleep-disordered breathing.

**Table 1. Sleep EEG Datasets.**
Five datasets are used in this work: the first three datasets are labeled for SSs, covering a wide age range from 10 to 76 years old. MODA used segments of 115s from the MASS dataset to label SSs with the consensus of 31-42 experts [30]. The datasets listed in the last two columns are unlabeled, where the subscripts of MASS indicate the corresponding cohorts. WN stands for Whole-night. The channel within brackets is the alternative one in case the specified channel is not available, where the reference is linked ear reference (LER) or computerized linked ear (CLE). Not available (NA). Sleep states were classified according to the Rechtschaffen and Kales (RK) [54] or American Academy of Sleep Medicine (AASM) [2] rules.

| | Labeled datasets | | | Unlabeled datasets | |
|---|---|---|---|---|---|
| | **INTA-UCH** | **MASS-SS$_2$** | **MODA** | **MASS-SS$_{1,3,4,5}$** | **CAP** |
| Subjects | 11 | 19 | 180 | 181 | 104 |
| Male | 6 | 8 | 92 | 89 | 66 |
| Female | 5 | 11 | 88 | 92 | 42 |
| Age [Yr] | 10 | 18-33 | 18-76 | 18-76 | 14-82 |
| Expert | 1 | 2 | 31-42[1] | NA | NA |
| Sampling rate [Hz] | 200 | 256 | 256 | 256 | 100-512 |
| Stage sleep | N2 | N2 | N2 | N2 | N2 |
| Scoring sleep | RK | RK | RK-AASM | RK-AASM | RK |
| Duration | WN | WN | 115s[2] | WN | WN |
| Channel | F4-C4 (F3-C3) | C3-CLE (C3-LER) | C3-CLE (C3-LER) | C3-CLE (C3-LER) | C3 (C4) |
| SSs | 12237 | 9990 | 5272 | NA | NA |

[1] The labels were carried out by a consensus of 31-42 experts.
[2] The duration corresponds to multiple segments of 115s.

### C. EEG SIGNAL PRE-PROCESSING

All the datasets used in this work, MASS, MODA, CAP, and INTA-UCH, contain the expert's annotation of the sleep stages. In the sleep-staging process, the muscular artifacts have been rejected [29], [30], [50], [51], [52]. We used sleep stage N2, where there are no rapid nor slow eye movements [2], so there was no need for removing eye artifacts. To confirm what was mentioned above, we applied the ICA functions for automatic eye artifact detection and muscle artifact detection from the MNE-Python Package [55] to the N2 stages of the INTA-UCH and MASS-SS2 datasets. No eye artifacts were detected in both datasets. A few muscle artifacts were detected in MASS-SS2. The latter were removed as we applied a FIR/firwin2 bandpass filter of 0.3-30 Hz to all EEG recordings ($\text{EEG}_{0.3-30}$) in the PSG. The filter with a 0.3 Hz lower cutoff frequency eliminates very slow waves and aids in sleep spindle detection. The records were segmented into epochs (windows) of 20 or 30 s depending on the predetermined interval in the PSG dataset, to which an extra 1 s was added at the beginning and end of each epoch. These generate a superposition at the boundaries of the analyzed epoch with the previous and subsequent epochs, respectively, to avoid edge effects such as cutting detections at the boundaries of the analyzed epoch. Extreme peak-to-peak amplitudes in the $EEG_{0.3-30}$ signal were limited to the range (-200, 200) $\mu V$, which was determined using the training set from the MASS-SS2 dataset. It is worth noting that the input to the USSD algorithm is the EEG filtered in the sigma band (11-16 Hz), therefore, any remaining general artifact outside this band is automatically removed.

### D. SOFTWARE IMPLEMENTATION

For the development of our USSD algorithm, we utilized Python 3.8.18, with key libraries including NumPy for array manipulation, SciPy for scientific analysis, pyEDFlib for reading and writing EDF (EEG data format), and MNE [55] for neurophysiological data analysis. The A7 code is available at https://github.com/swarby/A7_LacourseSpindleDetector, the LUNA code is available at https://zzz.bwh.harvard.edu/luna/ref/spindles-so/, and the MPP code is available at https://github.com/carlosloza/MPP-EEG.

### E. PERFORMANCE METRICS FOR SS DETECTION

To evaluate the proposed algorithm's performance in detecting SSs, we consider two kinds of metrics: metrics for labeled datasets using the golden standard, i.e., events labeled by human experts, and metrics for unlabeled datasets.

#### Metrics for labeled datasets

Typically, event-based metrics based on the parameters of a confusion matrix are used: true positives ($TP$), true negatives ($TN$), false positives ($FP$), and false negatives ($FN$). However, we do not consider $TN$ because we focus on SS detections, which are rare events in an EEG recording. To assess the length of the detections, we use the **Intersection over Union** ($IoU$) metric [28], [56] as follows:

$$IoU = \frac{A \cap B}{A \cup B}, \tag{14}$$

where A corresponds to the set of detections and B to the set of expert annotations. Following [28] an event is considered a $TP$ when the $IoU \geq 0.2$. This metric computes the percentage of overlap between an SS event detected by

the algorithm and the expert's label. Under this condition, the Recall, Precision, and F1-score metrics are defined as follows:

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{15}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{16}$$

$$\text{F1-score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}. \tag{17}$$

The **mean Intersection over Union** $(mIoU)$ metric will also be used, which corresponds to the average of the non-zero $IoU$ value for a set of detections ($\Gamma$) of cardinality $|\Gamma|$, defined as follows:

$$mIoU = \frac{1}{|\Gamma|} \sum_{\Gamma} IoU. \tag{18}$$

**Metrics for unlabeled datasets**

To evaluate unsupervised algorithms on unlabeled databases, the Jensen-Shannon divergence [57] (see Eq. 9) is proposed as a metric. The Jensen-Shannon divergence is applied between the length distribution of the SS detections and the Poisson distribution [28] that minimizes this metric for each dataset. This is because the length distribution of SSs has a known pattern that follows a Poisson distribution [32].

## IV. FINE-TUNING THE USSD MODEL

Machine learning models trained with unlabeled datasets, such as the proposed USSD algorithm, can be fine-tuned with a small, labeled dataset to improve the classification results. In this section, a fine-tuning stage is added to the USSD algorithm. In Figure 7 blocks A, B, and C comprise the USSD algorithm, and block D is the new fine-tuning stage, the output of which is used to update blocks B and C. The steps that describe the fine-tuning USSD model are the following:

- The input of block D is the set of SS detections made by the USSD algorithm.
- First, $n\%$ of the epochs with SS detections are selected, where $n$ is a percentage between 10 and 20 $\%$ (see Figure 7 D). The selected epochs must have at least one detected SS with a low or medium score of being an SS (see section III.A, Event Scoring), and a duration of less than 1 s.
- Second, the SS detections on the selected epochs are corrected, adjusting their duration according to the expert's labels, and those that are non-SSs according to the expert's labels are eliminated.
- In the update process, for each set of detections obtained with different threshold values, the corrected detections are updated with the expert's labels.
- Finally, the adaptive threshold procedure is applied, and block C in Figure 7 is updated to generate the final set of detections with the fine-tuned model.

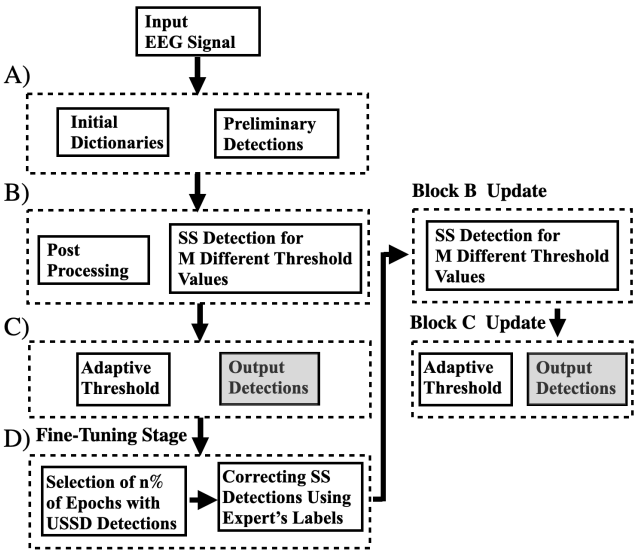



**Figure 7. USSD with a fine-tuning stage.**
The USSD algorithm consists of blocks A to C, with the Output Detections as the final result. Block D is added in the fine-tuning stage, where n% of epochs with SS detections made by USSD are selected, and expert labels are used to adjust all detections in those epochs. Then, Block B updates the detections based on the M-threshold values, and Block C recalculates the adaptive threshold to generate a new set of detections.

## V. RESULTS

In this section, we present the results of the proposed USSD model on both the labeled and unlabeled datasets shown in Table 1.

Using dictionary learning tools allows us to associate sets of detections with different atoms, as shown in Figure 8 A). The atoms shown in red are learned from the filtered EEG signal in the sigma band and are grouped into five dictionaries with durations [0.7-0.9-1.1-1.3-1.5] seconds. Each dictionary is composed of 20 atoms, although in the figure, only a single atom per dictionary is shown. In blue, detections of variable lengths associated with each atom are shown in the EEG filtered with a band pass filter between [0.3-30 Hz]. Figure 8 B) shows the distance of event detections to their closest atom, using the Longest Common Subsequence Similarity Measure (LCSS) metric [58] between time series. The width of each curve corresponds to the approximate frequency of data points for each distance.

### A. TRAINING AND PARAMETER ADJUSTMENT

For the training and evaluation of the USSD model, we utilized the labeled datasets MASS-SS2 and INTA-UCH. We employed a training, validation, and test set of 5,6, and 13 subjects, respectively. The training set comprises five

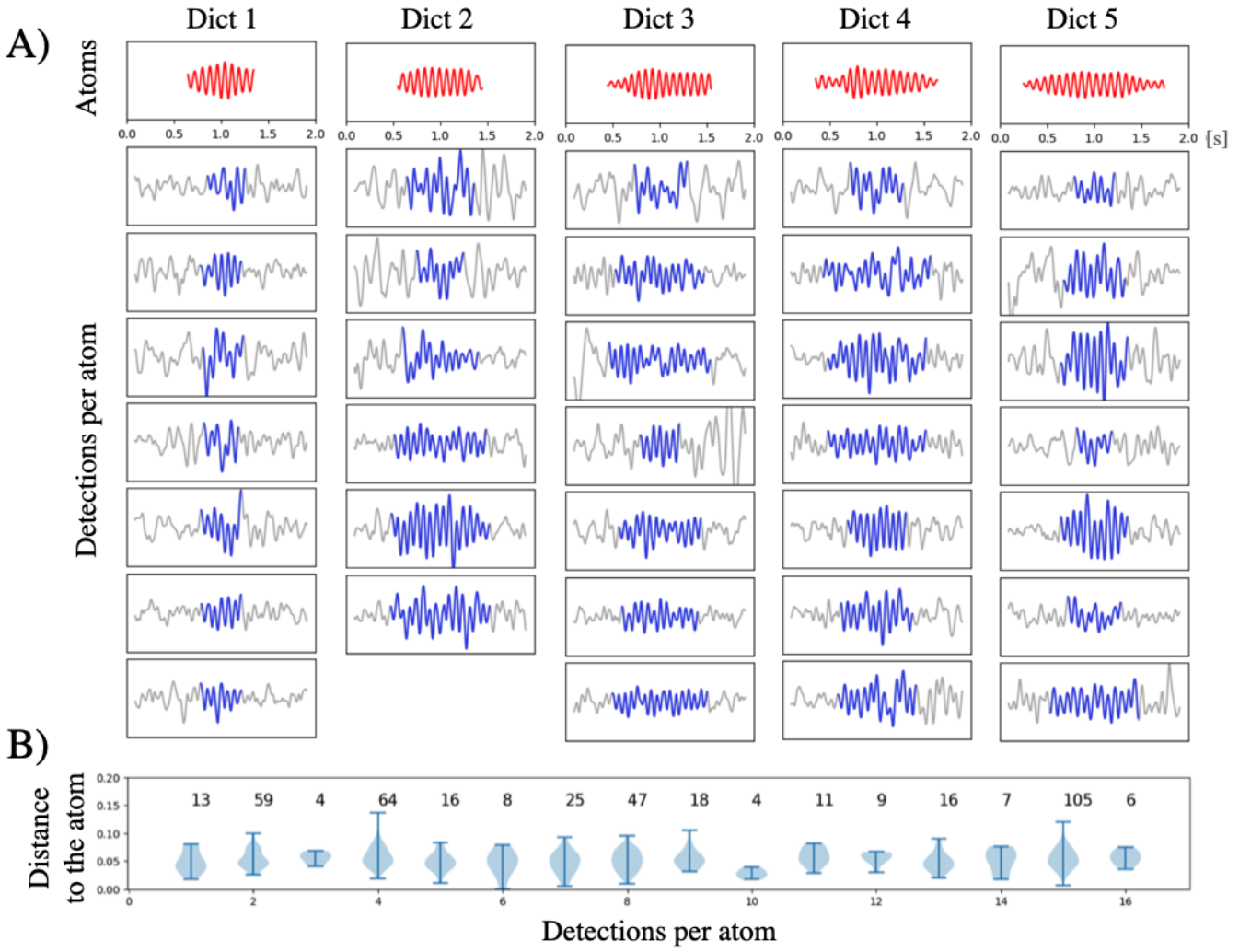



**Figure 8. Example showing a single atom per dictionary and a set of SS detections.**
In A, we show five columns, each associated with a dictionary. The first element of each column corresponds to a dictionary atom (red), whose duration is [0.7, 0.9, 1.1, 1.3, 1.5]s for dictionaries 1 to 5, respectively. The remaining elements of each column correspond to a segment of the EEG signal of 2s duration filtered between 0.3-30 Hz, where each segment contains a detection of SS (blue) obtained by USSD. The detections shown in each column are the variable-length SSs obtained from the last stage of the USSD model (see Figure 4 C). B) Violin plots showing the distribution of the distances of the detections with their respective atom. The 16 atoms associated with SS detections in dictionary #5 are shown. Note that 4 atoms in this dictionary did not obtain any detection. The numbers above each violin plot indicate the number of detections per atom.

subjects from the MASS-SS2 dataset. The validation set includes six subjects divided into {3,3}={MASS-SS2, INTA-UCH}, and the test set with 13 subjects is composed of {7,6}={MASS-SS2, INTA-UCH}.

For the state-of-the-art models, 4 parameters were tuned for A7, while only 1 parameter was adjusted for the LUNA and MPP models. A grid system with 92 combinations of the 4 parameters was employed for A7, and 10 values were tested for the LUNA and MPP models. For the MPP model, the duration of the dictionary atoms was adjusted, using 4 values between 0.7 and 1.5s. This process was applied to the training set for the MASS-SS2 dataset and to the validation set for the INTA-UCH dataset.

### B. RESULTS ON THE SS LABELED DATASETS

Table 2 shows the performance of USSD, MPP, A7, and LUNA under the recall, precision, and F1-score metrics. For MASS-SS2, the proposed USSD model achieved the best performance with an F1-score of $0.72 \pm 0.02$, with a p-value of 0.003 in the permutation-test [59] compared to A7 and a p-value less than 0.0002 compared to LUNA and MPP. In the INTA-UCH dataset, the USSD obtained the best performance, with an F1-score of $0.72 \pm 0.04$, and p-values less than 0.005 when compared to MPP, A7, and LUNA.

The top row in Figure 9 illustrates the trade-off between recall and precision, where the proposed USSD model obtains the best precision-recall relationship. The bottom row shows

**Table 2. Performance of unsupervised algorithms on the test set.**
The USSD, A7, and LUNA models were evaluated on the respective test subsets of the MASS-SS2 and INTA-UCH datasets, for which an IoU value $\geq 0.2$ was considered. The p-value was computed using a permutation test to compare the performance of USSD with the rest of the models.

| Dataset | Method | Recall | Precision | F1-score | p-value (F1-score) |
|---|---|---|---|---|---|
| MASS-SS2 | **USSD** | $0.72 \pm 0.03$ | $0.72 \pm 0.03$ | $\mathbf{0.72 \pm 0.02}$ | |
| | MPP | $0.37 \pm 0.11$ | $0.88 \pm 0.05$ | $0.52 \pm 0.11$ | $< 0.0002$ |
| | A7 | $0.66 \pm 0.09$ | $0.72 \pm 0.09$ | $0.69 \pm 0.05$ | 0.003 |
| | LUNA | $0.57 \pm 0.08$ | $0.64 \pm 0.05$ | $0.60 \pm 0.02$ | $< 0.0002$ |
| INTA-UCH | **USSD** | $0.72 \pm 0.03$ | $0.73 \pm 0.07$ | $\mathbf{0.72 \pm 0.04}$ | |
| | MPP | $0.47 \pm 0.05$ | $0.71 \pm 0.06$ | $0.57 \pm 0.05$ | 0.0005 |
| | A7 | $0.83 \pm 0.04$ | $0.60 \pm 0.07$ | $0.70 \pm 0.04$ | 0.005 |
| | LUNA | $0.50 \pm 0.04$ | $0.73 \pm 0.09$ | $0.59 \pm 0.02$ | 0.0005 |

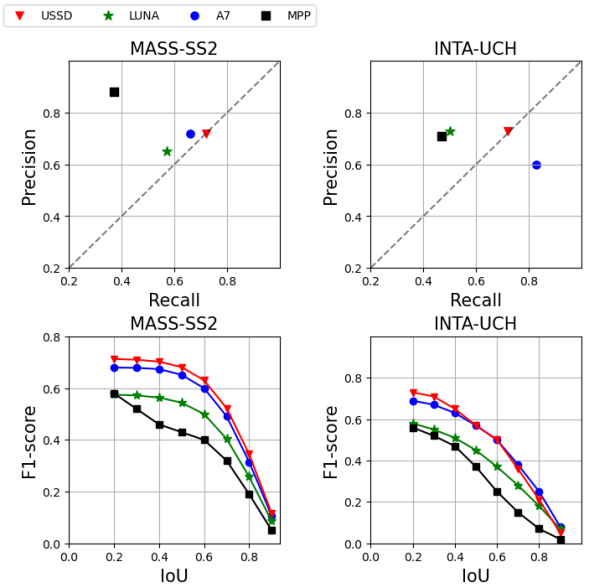



**Figure 9. Precision-Recall and F1-score vs. IoU curves.**
Top row) Precision vs Recall plots, illustrating their trade-off for both datasets. Bottom row) F1-score vs $IoU$ plots, with $IoU \geq 0.2$ showcasing the performance differences between the two datasets. All three evaluated models exhibit better performance in the adult MASS-SS2 dataset.

the F1-score vs. IoU plot for an IoU $\geq 0.2$, where USSD achieves the best performance in the MASS-SS2 and INTA-UCH datasets. The INTA-UCH presents a slightly worse performance due to the extra variability posed by children's EEG recordings, which were not included in the training set.

Table 3 shows the performance of the fine-tuned USSD model on the MASS-SS2 and INTA-UCH datasets using different percentages [10, 15, 20] % of labeled detections. The best performance is obtained when using 20% of labeled detections, reaching an F1-score of 0.78 and 0.75 for the MASS-SS2 and INTA-UCH datasets, respectively. This means an increase of 6% and 3% for the MASS-SS2 and INTA-UCH datasets, respectively.

**Table 3. Performance of the fine-tuned USSD on the test set.**
Performance of the fine-tuned USSD model when using different percentages of expert-labeled data from the MASS-SS2 and INTA-UCH datasets.

| Dataset | Labeled Data % | Recall | Precision | F1-score |
|---|---|---|---|---|
| MASS-SS2 | 10 | $0.66 \pm 0.09$ | $0.87 \pm 0.05$ | $0.75 \pm 0.06$ |
| | 15 | $0.68 \pm 0.11$ | $0.86 \pm 0.05$ | $0.76 \pm 0.07$ |
| | 20 | $0.69 \pm 0.09$ | $0.90 \pm 0.02$ | $0.78 \pm 0.06$ |
| INTA-UCH | 10 | $0.79 \pm 0.03$ | $0.69 \pm 0.08$ | $0.74 \pm 0.05$ |
| | 15 | $0.80 \pm 0.02$ | $0.69 \pm 0.08$ | $0.74 \pm 0.05$ |
| | 20 | $0.81 \pm 0.02$ | $0.69 \pm 0.08$ | $0.75 \pm 0.05$ |

### C. EVALUATING THE ADAPTIVE THRESHOLD METHOD

In section III, stage C, we proposed an unsupervised method to automatically and adaptively determine the single threshold of the USSD algorithm, which is adjusted by utilizing the distribution of detection durations. In this subsection, we compare the results of our proposed unsupervised method with those obtained using SS-labeled data. For simplicity, we refer to the optimal threshold as that which maximizes the F1-score when using labeled data. The latter is used as a gold standard, for evaluation purposes only. Figure 10 shows the performance of the adaptive threshold on the labeled datasets, where we compare the F1-score metric obtained with the proposed adaptive threshold and the optimal threshold. With this

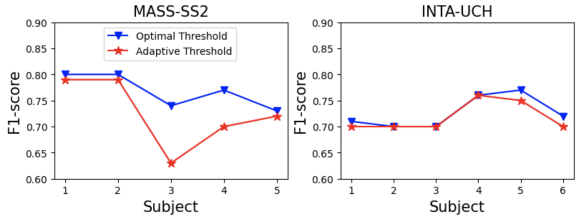



**Figure 10. Adaptive threshold performance.**
A) Comparison of the F1-score performance obtained with the optimal threshold (blue) versus the one obtained with the proposed adaptive threshold on 5 subjects from the test set of the MASS-SS2 dataset. The optimal threshold is computed using labeled data, while the adaptive threshold is computed using unlabeled data. B) F1-score obtained with the optimal and the adaptive thresholds for 6 subjects from the test set of the INTA-UCH dataset.

aim, the final set of detections obtained with the USSD model (the output of stage C in Figure 4) and their corresponding expert labels are used to compute a confusion matrix and the F1-score for the M-threshold values. The threshold value that maximizes the F1-score is chosen.

The percentage similarity between the two F1-score values is 97 % for MASS-SS2 and 99% for INTA-UCH, as shown in Table 4. This indicates that the proposed adaptive threshold method is near optimal, and adapts well to changes in the dataset and variability in EEGs despite differences in age ranges and variability between experts and datasets.

**Table 4. Adaptive threshold performances.**
Average F1-score performance obtained with the proposed adaptive threshold (unsupervised method that optimizes the USSD threshold) and those obtained with the optimal threshold (using expert labels) for the MASS-SS2 and INTA-UCH test sets. The ratio between the performances of the adaptive and optimal thresholds is shown.

| Dataset | Optimal threshold | Adaptive threshold | Ratio |
|---|---|---|---|
| MASS-SS2 | $0.74 \pm 0.04$ | $0.72 \pm 0.04$ | 0.97 |
| INTA-UCH | $0.73 \pm 0.09$ | $0.72 \pm 0.03$ | 0.99 |

### D. SLEEP SPINDLE DETECTION SCORING

In Section III, Stage A, we introduced a method to compute the probability that a detected event is a sleep spindle, where each event is assigned a high, medium, or low score. Figure 11 shows the percentage of unsupervised SS detections in each score category [High - Medium - Low] for 5 subjects of the MASS-SS2 dataset. Table 5 shows the average percentage of classifications for the MASS-SS2 dataset, which are 24, 38, and 38 % for the high, medium, and low scores, respectively, and for the INTA-UCH dataset are 41, 47, and 12 % for the high, medium, and low scores, respectively.

Table 6 shows the accuracy metric for each score category [High - Medium - Low] in the MASS-SS2 and INTA-UCH datasets. In the case of the high score, an average accuracy of $0.96 \pm 0.02$ is obtained with an SS detection percentage of 24 % on the test set of the MASS-SS2 dataset. An average accuracy of $0.93 \pm 0.01$ is achieved with a 41 % detection percentage of SSs in the test set of the INTA-UCH dataset. As expected, the lowest performance is obtained with the set of SS detections belonging to the low score category, reaching an accuracy of only 56 % and 62 % for the MASS-SS2 and INTA-UCH datasets, respectively.

**Table 5. Average percentage of SS detections per score category.**
Average percentage of SS detections per each score category for the test sets of the MASS-SS2 and INTA-UCH datasets.

| Dataset | High | Medium | Low |
|---|---|---|---|
| MASS-SS2 | $24 \pm 8$ | $38 \pm 2$ | $38 \pm 8$ |
| INTA-UCH | $41 \pm 5$ | $47 \pm 5$ | $12 \pm 5$ |

### E. EVALUATION OF SS LENGTH DISTRIBUTIONS

As mentioned in section III.A, Stage C, our main assumption is that the SS length distribution must follow a Poisson distribution. In section III.E, we introduced the Jensen-Shannon divergence (JSD) as a metric for unlabeled datasets. This metric indicates which algorithm has an SS length distribution closest to a Poisson distribution. In Table 7, the JSD metric was applied to the unlabeled CAP and MASS-SS$_{1,3,4,5}$ datasets and the labeled MASS-SS2 and INTA-UCH datasets. The labeled datasets are used for comparison purposes. It can be observed that USSD achieves the lowest JSD in all datasets. Figure 12 shows the distributions of the length of SSs in the CAP dataset for the USSD, A7, and LUNA models.

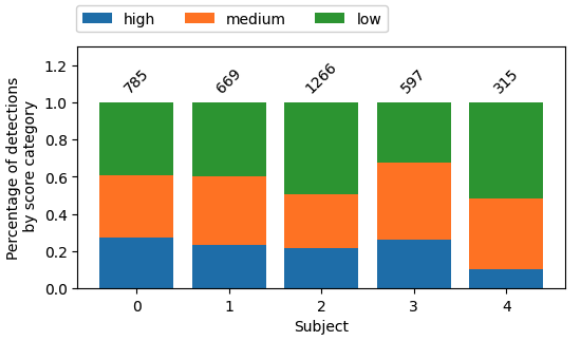



**Figure 11. Percentage of detections per score category per subject.**
Scores generated in an unsupervised manner by the proposed USSD model, on the test set of the MASS-SS2 dataset. The SS detection set is divided into three score categories [high-medium-low]. The number on top of each bar is the total number of detections per subject.

**Table 6. Accuracy metric per each score category.**
Average accuracy for each score category in the MASS-SS2 and INTA-UCH test sets.

| | Score Category | | |
|---|---|---|---|
| Dataset | High | Medium | Low |
| MASS-SS2 | $0.96 \pm 0.02$ | $0.80 \pm 0.13$ | $0.56 \pm 0.21$ |
| INTA-UCH | $0.93 \pm 0.01$ | $0.72 \pm 0.07$ | $0.62 \pm 0.11$ |

**Table 7. Jensen-Shannon Divergence.**
Jensen-Shannon divergence between the distributions of the SS detection durations made by USSD, LUNA, and A7 models, with respect to the best Poisson distribution.

| Dataset | USSD | LUNA | A7 |
|---|---|---|---|
| CAP | **0.11** | 0.27 | 0.21 |
| MASS-SS $_{1345}$ | **0.14** | 0.24 | 0.29 |
| MASS-SS2 | **0.14** | 0.17 | 0.33 |
| INTA-UCH | **0.12** | 0.29 | 0.32 |

### F. TRANSFER LEARNING

In this section, we evaluate the performance of pre-training the SEED supervised deep learning model [25] on the CAP dataset, using the SSs detected by the unsupervised algorithms USSD, A7, and LUNA as labels. After pre-training SEED with the automatically labeled CAP dataset, it was fine-tuned using the labeled MODA dataset. SEED was fine-tuned using 10 and 20 % of the training set from the MODA

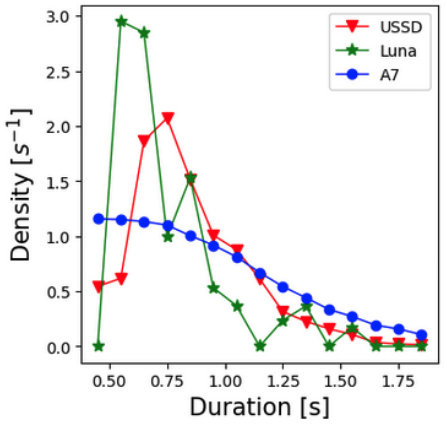



**Figure 12. SS detection length distributions.**
The length distributions of SS detections in the CAP dataset obtained by USSD, A7, and LUNA are used to compute the Jensen-Shannon divergence with the best Poisson distribution.

dataset. The performance of SEED without pre-training was also evaluated in the CAP dataset.

Table 8 shows that fine-tuning a pre-trained SEED model with 10% of MODA reached an F1-score of approximately 0.77. This result is independent of the unsupervised model that automatically labels the CAP dataset. Using 20% of the MODA dataset, a state-of-the-art performance is achieved with an F1-score of 0.81, with no significant difference between the three unsupervised models, where the permutation test yields a p-value $\geq 0.09$.

The results indicate that there is no difference between the three models, but there is a clear advantage in using automatically labeled datasets with unsupervised algorithms to pre-train SEED.

**Table 8. Transfer learning results.**
SEED was pre-trained with the CAP dataset automatically labeled by the unsupervised algorithms, and then fine-tuned with different percentages of the MODA training set. The performance was evaluated with the MODA test set. The term $None$ indicates that SEED was not pre-trained with CAP and was only trained with the MODA training set. The p-values show that there are no statistical differences between the mean results of the A7 and LUNA models and those of the USSD model.

| MODA % | Pre-train Method | F1-score | p-value | mIoU | p-value |
|---|---|---|---|---|---|
| 100% | $None$ | $0.82 \pm 0.01$ | | $0.83 \pm 0.01$ | |
| 10% | $None$ | $0.74 \pm 0.04$ | | $0.75 \pm 0.04$ | |
| | USSD | $0.77 \pm 0.02$ | | $0.78 \pm 0.01$ | |
| | A7 | $0.78 \pm 0.02$ | 0.09 | $0.76 \pm 0.02$ | 0.15 |
| | LUNA | $0.76 \pm 0.02$ | $\geq 0.09$ | $0.77 \pm 0.01$ | 1.0 |
| 20% | $None$ | $0.78 \pm 0.02$ | | $0.79 \pm 0.01$ | |
| | USSD | $0.81 \pm 0.02$ | | $0.80 \pm 0.01$ | |
| | A7 | $0.81 \pm 0.01$ | 0.09 | $0.79 \pm 0.01$ | 0.37 |
| | LUNA | $0.81 \pm 0.02$ | $\geq 0.09$ | $0.80 \pm 0.01$ | 0.59 |

## VI. DISCUSSION

In this section, we discuss the limitations, drawbacks, and costs of the proposed USSD algorithm. The main limitation is the size of the available SS-labeled datasets, which implies that a small number of subjects were used for training, validation, and testing. Despite that, the proposed USSD algorithm outperforms the A7 and LUNA algorithms on the tested datasets. Our algorithm performs very well in detecting events of medium and long duration, but the performance drops to approximately 60% for short-duration events of less than 0.5s. It is worth mentioning that according to the AASM standard criteria, the shortest duration of a spindle is 0.5 s. However, it is a common practice for experts to annotate shorter events, too. Improving the detection of short-duration events could be addressed in future work by considering additional information, e.g., the spectral power of events in the original signal. To compute the proposed adaptive threshold, the current version of the algorithm requires a continuous EEG recording of at least one hour. The drawback is that the current version of the USSD algorithm cannot process datasets composed of small window samples. The proposed fine-tuning algorithm (see Section IV) corrects the output of the USSD algorithm using a small percentage of labeled SSs, but it does not adjust the whole algorithm. This is left for future work. All unsupervised algorithms were run on a 2.2GHz 6-core Intel Core i7 MacBook Pro, with 16 GB of memory. The computational time of running the unsupervised algorithms one iteration on a 240-minute-long EEG recording are as follows: USSD 259.8 s, MPP 254.1 s, A7 237.7 s, and LUNA 9.5 s. As can be observed, LUNA is approximately 25 times faster than USSD, MPP, and A7. The USSD algorithm has not been optimized, and there is room for improvement by using GPUs and parallelism.

## VII. CONCLUSIONS

We have proposed an unsupervised algorithm for detecting sleep spindles in sleep EEG recordings based on dictionary learning. Our algorithm incorporates a novel unsupervised adaptive threshold method based on the distribution of detection durations. The proposed USSD method can automatically adapt to EEG recordings from subjects of all ages and coming from different datasets, addressing a typical transfer-learning challenge, especially for unsupervised algorithms.

Automated methods provide robust detection of sleep spindles, eliminating the variability among human experts inherent to manual evaluation. Unsupervised methods enable the analysis of unlabeled datasets. Consequently, USSD emerges as an algorithm that could benefit research on sleep spindles by analyzing large unlabeled datasets such as NSRR [34]. USSD assigns a probability of being a sleep spindle to each detection, which could be useful for diagnosing and monitoring both sleep disorders and sleep-related disorders from early to late periods of the vital cycle. The proposed USSD algorithm performed better than A7 and LUNA, with an average F1-score of 0.72. On the other hand, the proposed adaptive threshold method demonstrated excellent performance compared to the optimal threshold when maximizing the F1-score on the MASS-SS2 and INTA-UCH datasets.

The proposed USSD includes standard AASM criteria such as frequency and amplitude for adjusting detections and, at the same time, provides a score indicating how likely a detection is a sleep spindle, which is not available in other unsupervised algorithms. The USSD algorithm has good interpretability, since the atoms and the SS detections associated with each atom can be visualized.

A fine-tuned USSD model was also implemented, which increases the performance of the USSD from an F1-score of 0.72 to 0.78 in the MASS-SS2 dataset and from 0.72 to 0.75 in the INTA-UCH dataset when using 20 % of the labeled SS detections. While the performance of USSD is lower than that of supervised models such as SEED based on deep learning, our unsupervised model can be employed to pre-train deep learning-based models. Using our unsupervised model, we can automatically generate a labeled dataset for pre-training neural network-based models, which can be further fine-tuned. Pre-training with unsupervised models allows using $10 - 20\%$ of the labeled set for fine-tuning to achieve state-of-the-art performance. This is especially valuable because obtaining labeled SS data to train neural networks is expensive and challenging.

## ACKNOWLEDGMENTS

E. Ramírez acknowledges financial support from the National Agency for Research and Development (ANID) / Scholarship Program / DOCTORADO NACIONAL / 2018—21181277.

P. A. Estévez and C.A. Perez acknowledge financial support from ANID-Chile through Millennium Science Initiative Program ICN2021-004; Basal Funding for Scientific and Technological Center of Excellence, IMPACT #FB210024; and FONDECYT 1220829.

M. Adams acknowledges financial support from Agencia Nacional de Investigación y Desarrollo (ANID - National Research Agency, Chile)/Programa de Investigación Asociativa (PIA) Project AFB230001 and ANID Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) project 1231658.

• • •

## References

[1] S. Clocchiatti-Tuozzo, C. A. Rivier, D. Renedo, et al., "Suboptimal sleep duration is associated with poorer neuroimaging brain health profiles in middle-aged individuals without stroke or dementia," *Journal of the American Heart Association*, vol. 13, no. 1, Jan. 2024. DOI: 10.1161/JAHA.123.031514.

[2] R. B. Berry, R. Brooks, E. C. Gamaldo, et al., "The AASM manual for the scoring of sleep and associated events : rules, terminology, and technical specifications," *American Academy of Sleep*, vol. 28, no. 3, pp. 391–397, 2016.

[3] P. Peirano, C. Algarín, and R. Uauy, "Sleep-wake states and their regulatory mechanisms throughout early human development," in *Journal of Pediatrics*, vol. 143, Mosby Inc., 2003, pp. 70–79. DOI: 10.1067/s0022-3476(03)00404-9.

[4] E. G. Ibarra-Coronado, A. M. Pantaleón-Martínez, J. Velazquéz-Moctezuma, et al., "The Bidirectional relationship between sleep and immunity against infections," *Journal of Immunology Research*, vol. 2015, 2015. DOI: 10.1155/2015/678164.

[5] L. Xie, H. Kang, Q. Xu, et al., "Sleep drives metabolite clearance from the adult brain," *Science*, vol. 342, no. 6156, pp. 373–377, 2013. DOI: 10.1126/science.1241224.

[6] O. Lahl, C. Wispel, B. Willigens, and R. Pietrowsky, "An ultra short episode of sleep is sufficient to promote declarative memory performance," *Journal of Sleep Research*, vol. 17, no. 1, pp. 3–10, 2008. DOI: 10.1111/j.1365-2869.2008.00622.x.

[7] R. E. Dahl, "The Impact of Inadequate Sleep on Children's Daytime Cognitive Function," *Seminars in pediatric neurology*, vol. 3, no. 1, WB Saunders, Ed., pp. 44–50, 1996. DOI: 10.1016/s1071-9091(96)80028-3.

[8] M. Mirmiran, Y. G. Maas, and R. L. Ariagno, "Development of fetal and neonatal sleep and circadian rhythms," *Sleep Medicine Reviews*, vol. 7, no. 4, pp. 321–334, 2003. DOI: 10.1053/smrv.2002.0243.

[9] H. P. Roffwarg, J. N. Muzio, and W. C. Dement, "Ontogenetic development of the human sleep-dream cycle the prime role of "dreaming sleep" in early life may be in the development of the central nervous system," *Science*, vol. 152, no. 3722, pp. 604–619, 1966.

[10] S. Chokroverty and R. J. Thomas, *Atlas of sleep medicine: expert consult-Online and print.* Elsevier Health Sciences, 2013.

[11] L. M. Fernandez and A. Lüthi, "Sleep spindles: mechanisms and functions," *Physiological Reviews*, vol. 100, no. 2, pp. 805–868, 2020. DOI: 10.1152/physrev.00042.2018.

[12] L. M. O'Brien and D. Gozal, "Neurocognitive dysfunction and sleep in children: From human to rodent," *Pediatric Clinics of North America*, vol. 51, no. 1, pp. 187–202, 2004. DOI: 10.1016/S0031-3955(03)00184-6.

[13] C. Algarin, K. D. Karunakaran, S. Reyes, et al., "Differences on brain connectivity in adulthood are present in subjects with iron deficiency anemia in infancy," *Frontiers in aging neuroscience*, vol. 9, 2017. DOI: 10.3389/fnagi.2017.00054.

[14] M. A. Hahn, D. Heib, M. Schabus, K. Hoedlmoser, and R. F. Helfrich, "Slow oscillation-spindle coupling predicts enhanced memory formation from childhood to adolescence," *eLife*, vol. 9, pp. 1–21, 2020. DOI: 10.7554/eLife.53730.

[15] R. F. Helfrich, B. A. Mander, W. J. Jagust, R. T. Knight, and M. P. Walker, "Old brains come uncoupled in sleep: slow wave-spindle synchrony, brain atrophy, and forgetting," *Neuron*, vol. 97, no. 1, pp. 221–230, 2018. DOI: 10.1016/j.neuron.2017.11.020.

[16] A. K. Joechner, S. Wehmeier, and M. Werkle-Bergner, "Electrophysiological indicators of sleep-associated memory consolidation in 5- to 6-year-old children," *Psychophysiology*, vol. 58, no. 8, 2021. DOI: 10.1111/psyp.13829.

[17] M. Mölle, T. O. Bergmann, L. Marshall, and J. Born, "Fast and slow spindles during the sleep slow oscillation: disparate coalescence and engagement in memory processing," *Sleep*, vol. 34, no. 10, pp. 1411–1421, 2011. DOI: 10.5665/SLEEP.1290.

[18] N. Leresche and R. C. Lambert, "T-type calcium channels in synaptic plasticity," *Channels*, vol. 11, no. 2, pp. 121–131, 2017.

[19] T. T. Dang-Vu, A. Salimi, S. Boucetta, et al., "Sleep spindles predict stress-related increases in sleep disturbances," *Frontiers in Human Neuroscience*, vol. 9, 2015. DOI: 10.3389/fnhum.2015.00068.

[20] D. J. Buysse, A. Germain, M. L. Hall, et al., "EEG spectral analysis in primary insomnia: NREM period effects and sex differences," *Sleep*, vol. 31, no. 12, 2008. [Online]. Available: https://academic.oup.com/sleep/article-abstract/31/12/1673/2454125.

[21] A. Castelnovo, B. Graziano, F. Ferrarelli, and A. D'Agostino, "Sleep spindles and slow waves in schizophrenia and related disorders: main findings, challenges and future perspectives," *European Journal of Neuroscience*, vol. 48, no. 8, pp. 2738–2758, 2018. DOI: 10.1111/ejn.13815.

[22] M. S. Chan, K. F. Chung, K. P. Yung, and W. F. Yeung, "Sleep in schizophrenia: A systematic review and meta-analysis of polysomnographic findings in case-control studies," *Sleep Medicine Reviews*, vol. 32, pp. 69–84, 2017. DOI: 10.1016/j.smrv.2016.03.001.

[23] S. Ritter-Makinson, A. Clemente-Perez, B. Higashikubo, et al., "Augmented Reticular Thalamic Bursting and Seizures in Scn1a-Dravet Syndrome," *Cell Reports*, vol. 26, no. 4, p. 1071, 2019. DOI: 10.1016/j.celrep.2019.01.037.

[24] L. Feng, J. E. Motelow, C. Ma, et al., "Seizures and sleep in the thalamus: Focal limbic seizures show divergent activity patterns in different thalamic nuclei," *Journal of Neuroscience*, vol. 37, no. 47, pp. 11 441–11 454, 2017. DOI: 10.1523/JNEUROSCI.1011-17.2017.

[25] N. I. Tapia-Rivas, P. A. Estevez, and J. A. Cortes-Briones, "A robust deep learning detector for sleep spindles and K-complexes: towards population norms," *Scientific Reports*, vol. 14, no. 1, 2024. DOI: 10.1038/s41598-023-50736-7.

[26] S. Chambon, V. Thorey, P. J. Arnal, E. Mignot, and A. Gramfort, "DOSED: A deep learning approach to detect multiple sleep micro-events in EEG signal," *Journal of Neuroscience Methods*, vol. 321, pp. 64–78, 2019. DOI: 10.1016/j.jneumeth.2019.03.017.

[27] P. M. Kulkarni, Z. Xiao, E. J. Robinson, et al., "A deep learning approach for real-time detection of sleep spindles," *Journal of neural engineering,*, vol. 16, no. 3, 2019. DOI: 10.1088/1741-2552/ab0933.

[28] S. C. Warby, S. L. Wendt, P. Welinder, et al., "Sleep-spindle detection : crowdsourcing and evaluating performance of experts , non-experts and automated methods," *Nature methods*, vol. 11, no. 4, p. 385, 2014. DOI: 10.1038/nmeth.2855.

[29] C. O'Reilly, N. Gosselin, J. Carrier, and T. Nielsen, "Montreal archive of sleep studies: An open-access resource for instrument benchmarking and exploratory research," *Journal of Sleep Research*, vol. 23, no. 6, pp. 628–635, 2014. DOI: 10.1111/jsr.12169.

[30] K. Lacourse, B. Yetton, S. Mednick, and S. C. Warby, "Massive online data annotation, crowdsourcing to generate high quality sleep spindle annotations from EEG data," *Scientific Data*, vol. 7, no. 1, 2020. DOI: 10.1038/s41597-020-0533-4.

[31] K. Lacourse, J. Delfrate, J. Beaudry, P. Peppard, and S. C. Warby, "A sleep spindle detection algorithm that emulates human expert spindle scoring," *Journal of Neuroscience Methods*, vol. 316, pp. 3–11, 2019. DOI: 10.1016/j.jneumeth.2018.08.014.

[32] S. M. Purcell, D. S. Manoach, C. Demanuele, et al., "Characterizing sleep spindles in 11,630 individuals from the National Sleep Research Resource," *Nature Communications*, vol. 8, pp. 1–16, 2017. DOI: 10.1038/ncomms15930.

[33] C. A. Loza, M. S. Okun, and J. C. Principe, "A marked point process framework for extracellular electrical potentials," *Frontiers in systems neuroscience*, vol. 11, pp. 1–17, 2017. DOI: 10.3389/fnsys.2017.00095.

[34] G. Q. Zhang, L. Cui, R. Mueller, et al., "The national sleep research resource: towards a sleep data commons," *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1351–1358, 2018. DOI: 10.1093/jamia/ocy064.

[35] M. Mölle, L. Marshall, S. Gais, and J. Born, "Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep," *Journal of Neuroscience*, vol. 22, no. 24, pp. 10 941–10 947, 2002.

[36] F. Ferrarelli, R. Huber, M. J. Peterson, et al., "Reduced sleep spindle activity in schizophrenia patients," *American Journal of Psychiatry*, vol. 164, no. 3, pp. 483–492, 2007. DOI: 10.1176/ajp.2007.164.3.483.

[37] T. Andrillon, Y. Nir, R. J. Staba, F. Ferrarelli, C. Cirelli, and G. Tononi, "Sleep spindles in humans : insights from intracranial EEG and unit recordings," *The Journal of Neuroscience*, vol. 31, no. 49, pp. 17 821–17 834, 2011. DOI: 10.1523/JNEUROSCI.2604-11.2011.

[38] A. L. Sampson, C. Lainscsek, C. E. Gonzalez, et al., "Delay differential analysis for dynamical sleep spindle detection," *Journal of Neuroscience Methods*, vol. 316, pp. 12–21, 2019. DOI: 10.1016/j.jneumeth.2019.01.009.

[39] C. A. Loza and L. L. Colgin, "Deep neural dynamic bayesian networks applied to EEG sleep spindles modeling," *In: de Bruijne, M., et al. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. MICCAI 2021. Lecture Notes in Computer Science(), vol 12905. Springer, Cham.*, vol. 12905, 2021. [Online]. Available: https://doi.org/10.1007/978-3-030-87240-3_53.

[40] S. Devuyst, T. Dutoit, J. F. Didier, et al., "Automatic sleep spindle detection in patients with sleep disorders," *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3883–3886, 2006.

[41] F. Wang, L. Li, Y. Wan, et al., "An efficient sleep spindle detection algorithm based on MP and LSBoost," *Computers, Materials and Continua*, vol. 76, no. 2, pp. 2301–2316, 2023. DOI: 10.32604/cmc.2023.037727.

[42] P. Durka, M. Dovgialo, A. Duszyk-Bogorodzka, and P. Biegański, "Two-stage atomic decomposition of multichannel EEG and the previously undetectable sleep spindles," *Sensors*, vol. 24, no. 3, Feb. 2024. DOI: 10.3390/s24030842.

[43] C. A. Loza and J. C. Principe, "Estimation and modeling of EEG amplitude-temporal characteristics using a marked point process approach," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 3720–3723, 2016. DOI: 10.1109/EMBC.2016.7591536.

[44] I. Tošic and P. Frossard, "Dictionary Learning," *Signal Processing Magazine, IEEE*, vol. 28.2, pp. 27–38, 2011. DOI: 10.1201/9781351261364-9.

[45] E. J. Wamsley, M. A. Tucker, A. K. Shinn, et al., "Reduced sleep spindles and spindle coherence in schizophrenia: Mechanisms of impaired memory consolidation?" *Biological Psychiatry*, vol. 71, no. 2, pp. 154–161, 2012. DOI: 10.1016/j.biopsych.2011.08.008.

[46] M. Akay, "Wavelet applications in medicine," *IEEE Spectrum*, vol. 34, no. 5, pp. 50–56, 1997. DOI: 10.1109/6.590747.

[47] R. Gruber and M. S. Wise, "Sleep spindle characteristics in children with neurodevelopmental disorders and their relation to cognition," *Neural Plasticity*, 2016. DOI: 10.1155/2016/4724792.

[48] J. Liang, A. N. Belkacem, Y. Song, et al., "Classification and transfer learning of sleep spindles based on convolutional neural networks," *Frontiers in Neuroscience*, vol. 18, 2024. DOI: 10.3389/fnins.2024.1396917.

[49] R. Li, H. Wei, and B. Wang, "An unsupervised transfer learning model based on joint information for sleep stage classification," in *In 2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, (CISP-BMEI)*, 2022, pp. 1–5. DOI: 10.1109/CISP-BMEI56279.2022.9980333.

[50] P. A. Estevez, C. M. Held, C. A. Holzmann, et al., "Polysomnographic pattern recognition for automated classification of sleep-waking states in infants," *Medical and biological engineering and computing*, vol. 40, pp. 105–113, 2002.

[51] M. G. Terzano, L. Parrino, A. Sherieri, et al., "Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep," *Sleep Medicine*, vol. 2, no. 6, pp. 537–553, 2021. [Online]. Available: www.elsevier.com/locate/sleep.

[52] A. L. Goldberger, L. A. Amaral, L. Glass, et al., "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.," *Circulation*, vol. 101, no. 23, 2000. DOI: 10.1161/01.cir.101.23.e215.

[53] R. W. Homan, J. Herman, and P. Purdy, "Cerebral location of international 10-20 system electrode placement," *Electroencephalography and clinical neurophysiology*, vol. 66, pp. 376–382, 1987.

[54] A. Rechtscahffen and A. Kales, *A manual of standardized terminology, techniques and scoring system for sleep stages in human subjects*. National Institutes of Health, Neurological Information Network, Publication 204, 1969.

[55] A. Gramfort, M. Luessi, E. Larson, et al., "MEG and EEG data analysis with MNE-Python," *Frontiers in Neuroscience*, 2013. DOI: 10.3389/fnins.2013.00267.

[56] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: a metric and a loss for bounding box regression," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 658–666, 2019. DOI: 10.1109/CVPR.2019.00075.

[57] M. L. Menendez, J. A. Pardo, L. Pardo, and M. C. Pardo, "The jensen-shannon divergence," *Journal of the Franklin Institute*, vol. 334, no. 2, pp. 307–318, 1997. DOI: https://doi.org/10.1016/S0016-0032(96)00063-4.

[58] M. Vlachos, G. Kollios, and D. Gunopulos, "Discovering similar multidimensional trajectories," in *In Proceedings of the 18th International Conference on Data Engineering (ICDE '02).*, 2002, p. 673.

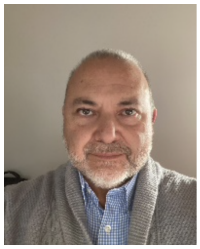
[59] M. D. Ernst, "Permutation methods: A basis for exact inference," *Statistical Science*, vol. 19, no. 4, pp. 676–685, 2004. DOI: 10.1214/088342304000000396.

EDGARDO RAMIREZ received a B.Sc. in Physics from the Pontifical Catholic University of Chile in 2006 and an M.Sc. in Medical Biophysics from the University of Chile in 2011. He is a PhD(c) in Electrical Engineering from the University of Chile. Research experience: From 2015 to 2017, he worked on three projects in soft robotics, two as a researcher and one as a principal researcher in the Intelligent Machine Synthesis Laboratory of the Mechanical Engineering Department of University of Chile. From 2010 to 2017, he was a research assistant in EEG analysis in general anesthesia at the Faculty of Sciences of the University of Chile. From 2011 to 2015 he was a researcher in magnetoreception in pigeons in the Laboratory of Biology of Cognition of the Faculty of Sciences of the University of Chile. Academic experience: From 2010 to 2011 he worked as the coordinator of the Physics Laboratory in the Faculty of Physics of Pontifical Catholic University of Chile. From 2002 to 2017, he worked as a laboratory assistant in the Pontifical Catholic University of Chile Faculty of Physics. His research interests are Machine Learning, Biological Signal Analysis, Neuroscience, and Robotics.

PABLO A. ESTEVEZ (Fellow, IEEE) received a B.Sc. in electrical engineering (EE) from the University of Chile in 1981, and an M.Sc. and Dr. Eng. degrees from the University of Tokyo, Japan, in 1992 and 1995, respectively. Prof. Estevez is a Full Professor at the Department of Electrical Engineering, University of Chile, Chile. He was president of the IEEE Computational Intelligence Society (2016-2017). He is currently an Associate Editor of the IEEE Transactions on Artificial Intelligence. Prof. Estévez is a Principal Investigator of the Millennium Institute of Astrophysics and the Millennium Institute for Intelligent Healthcare Engineering. He is also an adjunct investigator at the IMPACT Center of Interventional Medicine for Precision and Advanced Cellular Therapy. Prof. Estévez has been an Invited Researcher with the NTT Communication Science Laboratory, Kyoto, Japan; the Ecole Normale Supérieure, Lyon, France; and a Visiting Professor at the Pantheon-Sorbonne University, Paris, France; and the University of Tokyo, Tokyo, Japan. Prof. Estevez has coauthored more than 150 articles in journals and conferences. His research areas of interest include machine learning, deep learning, and information theory, with applications in astronomy, medicine, and energy.

MARTIN D. ADAMS (SM'08) is Professor of Electrical Engineering and director at the Dept. of Electrical Engineering, University of Chile. He is also a principal investigator in the industrially sponsored Advanced Mining Technology Centre (AMTC) and is currently Chair of the IEEE Robotics & Automation Society, Chilean Chapter. He obtained his first degree in Engineering Science at the University of Oxford, U.K., in 1988 and continued to study for a D.Phil. at the Robotics Research Group, University of Oxford, which he received in 1992. He continued his research in autonomous robot navigation as a project leader and part-time lecturer at the Institute of Robotics, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland. He was employed as a Guest Professor and taught control theory in St. Gallen (Switzerland) from 1994 to 1995. From 1996 to 2000, he served as a senior research scientist in robotics and control, in the field of semiconductor assembly automation, at the European Semiconductor Equipment Centre (ESEC), Switzerland. From 2000 to 2010, he was Associate Professor at the School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore. His research work focuses on autonomous robot navigation, space situational awareness, sensing, sensor data interpretation and control, and he has published many technical papers in these fields.

CLAUDIO A. PEREZ (Senior Member, IEEE) received the B.S. degree in electrical engineering, the P.E. degree in electrical engineering, and the M.S. degree in biomedical engineering from Universidad de Chile, in 1980 and 1985, respectively, and the Ph.D. degree from The Ohio State University, in 1991. He was a Fulbright Student with The Ohio State University, where he received a Presidential Fellowship, in 1990. He was a Visiting Scholar with UC Berkeley, in 2002, through the Alumni Initiatives Award Program from the Fulbright Foundation. He was the Department Chairperson, from 2003 to 2006, and the Director of the Office of Academic and Research Affairs, School of Engineering, Universidad de Chile, from 2014 to 2018. He is currently a Professor with the Department of Electrical Engineering, Universidad de Chile. His research interests include biometrics, image processing applications, convolutional neural networks, and pattern recognition. He is a Senior Member of the IEEE Systems, Man and Cybernetics Society and the IEEE Computational Intelligence Society.



MARCELO GARRIDO GONZÁLEZ is a professor of Psychology and Education and director of the School of Pedagogy in Basic Education at Diego Portales University. He is also a researcher associated with the Laboratory of Sleep and Functional Neurobiology at the Institute of Nutrition and Food Technology (INTA) of the University of Chile since its foundation, where he specialized in electrophysiological recordings and research protocols. He obtained the title of Professor of Biology and Sciences at the University of Playa Ancha in 1993, later he studied Neurobiology at the Faculty of Medicine of the University of Chile until 1996. In 1997, he began teaching neuropsychology and neurophysiology in the degree of Psychology at the Diego Portales University; starting in 2002 he joined the Basic Education program as a teacher. In 2014, he obtained a master's degree in Cognitive Development at Diego Portales University and since 2020 he is a candidate for the degree of Doctor of Education at the University of Christian Humanism. His research focuses on deepening the understanding of the relationship between the development of thinking and biological rhythms in university students.

PATRICIO PEIRANO after receiving his medical degree from the University of Chile in 1980, he completed training in Neurophysiology and Sleep medicine and obtained his Ph.D. in Paris (INSERM and Salpêtrière University Hospital, Paris VI, France). He then joined the Institute of Nutrition and Food Technology (INTA), University of Chile in 1990 as Founder and Head of the Sleep laboratory. His research interest focuses mainly on the interaction between sleep, nutrition, and health throughout the life span, to identify early neurofunctional markers that contributed to the understanding of chronic disease risk. The long-lasting relationship with the Biomedical Engineering laboratory of the same university has been key to progressively implementing automated methodologies for electrophysiological signals processing to extend thus our ongoing research perspectives on both sleep and neurocognitive domains.