

# Biomedical Signal Processing and Control

## Capturing Waveforms in Polysomnography

--Manuscript Draft--

<b>Manuscript Number:</b>	
<b>Article Type:</b>	VSI: Biomed Signals as Images
<b>Keywords:</b>	PSG; EEG; SIFT; Waveforms; SW
<b>Corresponding Author:</b>	Rodrigo Ramele, Ph.D Instituto Tecnológico de Buenos Aires: Instituto Tecnológico de Buenos Aires Ciudad de Buenos Aires, Buenos Aires ARGENTINA
<b>First Author:</b>	Giulia Carbonari, B.S.
<b>Order of Authors:</b>	Giulia Carbonari, B.S.
	Juliana Gambini, PhD
	Juan Miguel Santos, PhD
	Cecilia Forcato, PhD
	Rodrigo Ramele, Ph.D
<b>Abstract:</b>	<p>In this work, we propose a method to analyze and classify the detection of Slow Waves (SW) that are based on the extraction of descriptors of visually relevant characteristics from images of the signal plots. This procedure has the advantage that the characteristics used for classification are visually relevant and meaningful to a human observer. The characterization of these waves is essential because they play a major role in memory formation as well as in cleansing the brain of aberrant proteins, such as the precursors of Alzheimer's disease. A registered sleep technologist (RST) can identify an SW by visually observing its shape. However, this procedure is time-consuming and the agreement rate between different experts is often very low. Considering that sleep studies are typically several hours long, manual SW detection can be very slow, and human error can occur due to examiner fatigue. For this reason, it is of great interest the automatic detection of SW, minimizing the error and the classification time. Results on an annotated public dataset of Polysomnography (PSG) shows that the visual characterization of SW can be effectively mapped into a feature descriptor and that several classifiers used in EEG research are able to classify the signals with an accuracy similar to a visual observer.</p>
<b>Suggested Reviewers:</b>	Aykut Erdamar aerdamar@baskent.edu.tr
	Christoph Nissen christoph.nissen@upd.ch
	Fengyu Cong cong@dlut.edu.cn



ITBA – Instituto tecnológico de Buenos Aires  
Departamento de Ingeniería Informática  
Lavarden 315  
Buenos Aires, Argentina

April 4th, 2021

Dear Editors,

We are submitting a Full Paper, entitled Capturing Waveforms in Polysomnography which we believe is within the scope of Special Issue on Analysis of 1D Biomedical Signals through AI-based Approaches for Image Processing for the Biomedical Signal Processing and Control Journal. This article encompasses work performed in collaboration between the CiC Laboratory and the Laboratorio del Sueño of the ITBA University in Buenos Aires, Argentina in the context of research on Electroencephalographic Processing. This is an ongoing project, which we summarize in the manuscript and apply specifically to Sleep Research.

The work expands a method that we propose to analyze EEG signals based on the structure of the waveforms that are obtained by creating plotting images of the signals. The method is an extension of the Scale Invariant Feature Transform (SIFT) method used in Computer Vision. We used the method to analyze a public dataset of Polysomnography (PSG) where we use it to recognize Slow Waves (SW) which are very important in Sleep Research.

We believe this is an interesting work and we think it is very appropriate for this special issue, which emphasizes methods that are grounded in 2D image processing and used to analyze 1D signals.

Sincerely Yours  
Rodrigo Ramele

In this work, we propose a method to analyze and classify the detection of Slow Waves (SW) that are based on the extraction of descriptors of visually relevant characteristics from images of the signal plots. This procedure has the advantage that the characteristics used for classification are visually relevant and meaningful to a human observer. The characterization of these waves is essential because they play a major role in memory formation as well as in cleansing the brain of aberrant proteins, such as the precursors of Alzheimer's disease. A registered sleep technologist (RST) can identify an SW by visually observing its shape. However, this procedure is time-consuming and the agreement rate between different experts is often very low. Considering that sleep studies are typically several hours long, manual SW detection can be very slow, and human error can occur due to examiner fatigue. For this reason, it is of great interest the automatic detection of SW, minimizing the error and the classification time. Results on an annotated public dataset of Polysomnography (PSG) shows that the visual characterization of SW can be effectively mapped into a feature descriptor and that several classifiers used in EEG research are able to classify the signals with an accuracy similar to a visual observer.

# Capturing Waveforms in Polysomnography

Giulia Carbonari, Juliana Gambini, Juan Miguel Santos, Cecilia Forcato,  
Rodrigo Ramele\*

*Instituto Tecnológico de Buenos Aires*

---

## Abstract

In this work, we propose a method to analyze and classify the detection of Slow Waves (SW) that are based on the extraction of descriptors of visually relevant characteristics from images of the signal plots. This procedure has the advantage that the characteristics used for classification are visually relevant and meaningful to a human observer. The characterization of these waves is essential because they play a major role in memory formation as well as in cleansing the brain of aberrant proteins, such as the precursors of Alzheimer's disease. A registered sleep technologist (RST) can identify an SW by visually observing its shape. However, this procedure is time-consuming and the agreement rate between different experts is often very low. Considering that sleep studies are typically several hours long, manual SW detection can be very slow, and human error can occur due to examiner fatigue. For this reason, it is of great interest the automatic detection of SW, minimizing the error and the classification time. Results on an annotated public dataset of Polysomnography (PSG) shows that the visual characterization of SW can be effectively mapped into a feature descriptor and that several classifiers used in EEG research are able to classify the signals with an accuracy similar to a visual observer.

*Keywords:* SIFT, EEG, waveform, NREM, Sleep, PSG

---

---

\*Corresponding author

Email address: [rramele@itba.edu.ar](mailto:rramele@itba.edu.ar) (Rodrigo Ramele)

## 1. Introduction

A regular practice in image processing is to analyze images as bidimensional signals. As a one-dimensional signal can be understood as a dependent quantity that varies in time, a black-and-white image can be interpreted as having  
5 quantity values that vary for two independent variables, height and width. The method proposed in this work entails a different approach, where the opposite is proposed and one-dimensional signals are studied by how they are represented on images, how they look on an image plot. This is specifically tailored for processing Electroencephalographic (EEG) signals, and can be used to analyze  
10 them based on the shape of their waveforms, graphoelements that conceive cognitive meaning or are of clinical relevance. This provides a quantitative mapping of EEG components which at the same time convey meaning to the practitioners clinician, physician or registered sleep technologist (RST) who are studying these signals, and who traditionally analyze them visually by studying  
15 their waveforms [1, 2, 3].

This work expands the method previously published in [4, 5, 6]. It establishes its modelling and extends its usage to the study of polysomnographic signals, particularly slow-wave sleep signals which are particularly suitable to be analyzed in this way.

20 Sleep Research is devoted to understanding the inner workings of the brain during sleep and its strong connection with memory formation. Sleep is defined as a natural and reversible state of reduced responsiveness to external stimuli and relative inactivity, accompanied by a loss of consciousness. Sleep occurs in regular intervals and is homeostatically regulated and in mammals consists of  
25 two core sleep stages: non-rapid-eye-movement (NREM) and REM sleep, which alternate in a cyclic manner [7, 8]. During human NREM, the EEG shows predominant slow-wave activity, which is defined by the 0.5 to 4.0 Hz frequency bands and includes the  $< 1$  Hz slow oscillations with a peak frequency of 0.8 Hz. Slow oscillations comprise alterations between periods of neuronal membrane  
30 depolarization accompanied by sustained firing (up-states) and periods of membrane

hyperpolarization associated with neuronal silence (down-state). In the scalp EEG, the negative peak of the slow oscillation coincides with the beginning of the down-to-up state transition, whereas the depolarizing phase of sustained firing correlates with the positive EEG deflection [7, 8]. The characterization  
35 of these waves is essential because they play a major role in memory formation as well as in the brain cleansing of aberrant proteins, such as the precursors of Alzheimer’s disease [7, 8].

The traditional approach to glimpse what is happening inside the brain while sleeping has been the Polysomnography (PSG), which is mainly based  
40 on Electroencephalography, Electromyography (EMG) and Electrooculography (EOG) [9]. Sleep research studies rely heavily on EEG graphoelements [10] like K-Complexes and Slow Oscillations. Particularly, Slow Waves (SW) are signal components known to be involved in memory formation processes as well as are understood to be relevant in cleaning beta-amyloid protein which is a precursor  
45 of Alzheimer’s disease [11]. These component’s identification is performed by visually marking waveforms in long-running PSG, looking for patterns based on standardized guidelines [2, 12, 8]. This visual characterization entails an expert’s subjective decision.

This makes PSG signal analysis particular relevant for this methodology  
50 because it can be used to derive a metric that can describe quantitatively the similarities between signal shape components. Sleep research requires long hours of visual inspection, hence, it will be greatly aided by automation tools that at the same time identify the signal structure in the same way as those experts who traditionally analyze these signals.

55 Here, we first provide the general layout of the model, detailed in section 2 which presents the details of the feature extraction procedure which is implemented in an accompanying software. For the experiment section, we analyzed a database of Sleep Research to identify Slow Waves based on their waveform structure. In the section 3, results are expounded and conclusions and future research  
60 directions are outlined in the last section 4.

## 2. Materials and Methods

### 2.1. Converting 1D signals to Images: Plotting

An EEG signal can be considered as a multichannel time point sequence. This sequence is obtained by a digitalization process, at a certain sampling  
65 frequency  $F_s$  which depends on the electrophysiological device [13]. The original EEG stream can be divided into segments or epochs  $\tilde{x}(n, c)$  of fixed length and zero media. These segments are extracted to be further analyzed to determine the presence or not of signal markers that are particularly relevant to EEG signal processing or sleep research.

70 Since a century ago, the study of these signals involved the creation of a representation in the form of a signal plot [14]. Plotting entails a digitalization process as well and produces a binary image per EEG channel with the trace representing the time-varying signal.

This binary image  $\mathcal{I}^{(c)}$  can be constructed based on Equation 1,

$$\mathcal{I}^{(c)}(z_1, z_2) = \begin{cases} 255 & \text{if } z_1 = \gamma_t n \quad \text{and} \quad z_2 = \lfloor \gamma \tilde{x}(n, c) \rfloor + z(c) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

75 where  $1 \leq c \leq C$  and  $1 \leq n \leq N$ , with  $C$  as the number of available channels and  $N$  as the length of each segment. The coordinate  $z_1$  is on the horizontal axis of the image and  $z_2$  is the vertical coordinate, increasing from top to bottom, so the  $(0, 0)$  position is on the upper-left corner of the image. The amplitude scale factor  $\gamma$  and time scale factor  $\gamma_t$  are used to determine the image size  
80 and at the same time the image resolution. In order to complete the trace of the signal plot, the isolated points produced by Equation 1 are connected using the Bresenham [15] algorithm, which performs a linear discrete interpolation between the pixels. This scheme produces a black-and-white plot of the signal with 255 being white and 0 black. There is one image per channel per segment.

85 Finally, the value  $z(c)$  corresponds to the position on the image where the signal is zero, the zero-level, and it is calculated as

$$z(c) = \left| \left[ \min_n \tilde{x}(n, c) \right] \right|.$$

This reveals that the plot of the signal is upside down on the image, where the amplitude is positive increasing to the bottom of the image, which by the way, is a regular procedure to plot signals in neuroscience [3].

## 90 2.2. The Scale Invariant Feature Transform Method

Recognizing structures from visual observation pertain to two different problems at the same time. The first is to identify which areas of the visual scene are conveying the higher amount of information. In a sense, what are the salient features of an image that are important to emphasize and that help to  
95 characterize what is being seen. The second problem is precisely to give some meaning to these particular regions, to understand when they are similar to each other, or when they are different.

The first problem, the detection, found inspiration in the studies performed by [16] on how human stereo vision works. The work [17, 18] emphasized  
100 the importance of *scale*, and that signal features that are resilient to scaling and smoothing transformations are key elements that convey important information. These results pave the way to the development of the classical scale space methodology [19, 20, 21, 22, 23, 24].

On the other hand, the seminal work of [25], expanded with the work of [26]  
105 on how the visual cortex of cats sense shapes based on the orientation of objects that were being displayed to them, pointed out towards a way to describe visual features based on how they are oriented on the visual scene. This was further elaborated in the Theory of Receptive Fields [27, 28].

The Scale Invariant Feature Transform (SIFT) method was proposed by  
110 [29] and it is the confluence of these two ideas. The SIFT Detector is the procedure to identify relevant areas of an image, based on the detection of certain regions that are obtained after applying several changes in scale and successive Gaussian filters to smooth the images [30, 31]. Given an image, the SIFT



Detector provides a set of candidates keypoints, locations on the image with  
115 relevant features in different scales. On the other hand, the SIFT Descriptor  
is the procedure to describe and characterize the regions around each keypoint,  
called patches, by creating a histogram of gradient orientations. This histogram  
is summarized in a feature vector which effectively represents the structure of  
the information contained within the patch.

### 120 2.3. Waveform Extraction Procedure

The method proposed here is grounded on an extension and modification  
of the SIFT Descriptor. The procedure to capture the waveform of the signal  
using this method is thus summarized in three steps:

1. Keypoints **kp** are located on an image of a signal plot.
- 125 2. A region of an image, a patch, is established using keypoints as centers.  
Each patch has a horizontal  $St$  and vertical scale  $Sv$ , which determines the  
size in pixels  $\mathbf{S}_x$  and  $\mathbf{S}_y$ , along the horizontal and vertical axis respectively.
3. From each patch, a descriptor **d** is derived which is used as a representation  
of the graphical information contained within the patch.

130 The keypoint **kp** represent the center location of the patch and is placed  
on a pixel  $(x_{kp}, y_{kp})$  over the image plot. A local image patch of size  $\mathbf{S}_x \times \mathbf{S}_y$   
pixels is constructed around the keypoint and it is divided in 16 blocks  $B_{i,j}$ . It  
is arranged in a  $4 \times 4$  grid and the pixel **kp** is the patch center. Figure 1(a)  
shows a plot of a signal, a keypoint at the center and the surrounding patch in  
135 green.

Pixel intensity gradients can be obtained from an image by applying the  
Sobel filter [31] and using finite differences to obtain pixel differences on both  
the  $x$  and  $y$  direction. Composing them as vectors, oriented gradients on each  
pixel can be calculated. Figure 2 shows a vector field of oriented gradients over  
140 a region of the SW shown on the Figure 1.

With this, a local representation of the EEG signal shape within the patch  
can be described by obtaining the gradient orientations on each of the 16 blocks



Figure 1: Two Slow Waves (SW) are shown, with the superimposed 4x4 patch in green and the patch center as a black dot. The Slow Waves shown here have their positive peaks upwards.

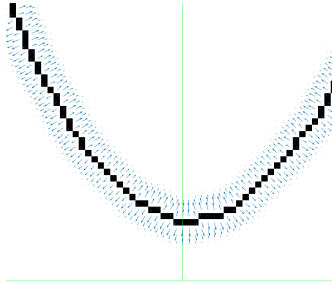


Figure 2: A vector field of oriented gradients around the minimum peak of the Slow Wave shown on the left on Figure 1. Each vector is shown as a blue arrow and points towards the direction of maximum change in pixel intensities. These arrows are grouped in 8 direction bins on each block  $B_{i,j}$ .

$B_{i,j}$  and creating a histogram of gradients. In order to calculate the histogram, the interval  $[0 - 360]$  of possible angles is divided in 8 bins, each one at 45 degrees. Figure 3 shows a sample histogram obtained for eight orientations.

Hence, for each spatial bin  $i, j = \{0, 1, 2, 3\}$ , corresponding to the indexes of each block  $B_{i,j}$ , and for each one of the eight angle orientations, the gradients

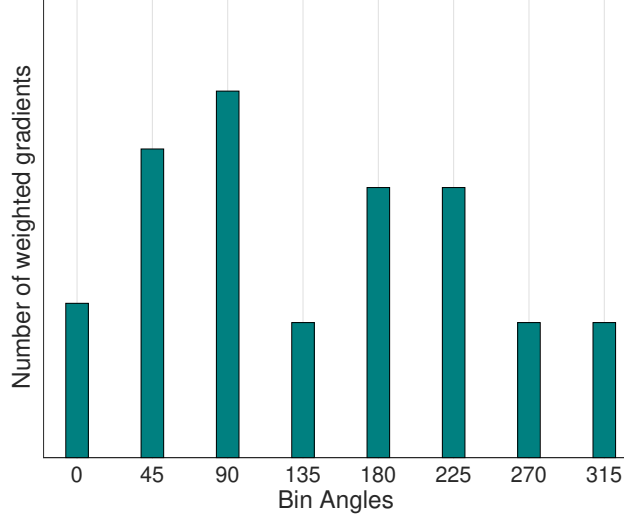


Figure 3: A sample histogram of the 8 angle orientations bins that is used to group all the different gradient vectors calculated on each block  $B_{i,j}$ .

are accumulated in a 3-dimensional histogram  $h$  through the following equation:

$$h(\theta, i, j) = \sum_{\mathbf{p}} \omega_{\text{ang}}(\angle J(\mathbf{p}) - \theta) \omega_{ij}(\mathbf{p} - \mathbf{k}\mathbf{p}) \|J(\mathbf{p})\| \quad (2)$$

where  $\theta$  is the angle bin with  $\theta \in \{0, 45, 90, 135, 180, 225, 270, 315\}$ ,  $\mathbf{p}$  is a pixel from within the patch,  $\|J(\mathbf{p})\|$  is the norm of the gradient vector in the pixel  $\mathbf{p}$ , computed using finite differences, and  $\angle J(\mathbf{p})$  is the angle of the gradient vector. The scalar  $\omega_{\text{ang}}(\cdot)$  and vector  $\omega_{ij}(\cdot)$  functions are linear interpolations proposed by [29] and [32] to provide a weighting contribution to eight adjacent bins. They are calculated as

$$\omega_{ij}(\mathbf{v}) = \omega\left(\frac{5}{\Delta s} \frac{v_x}{St} - x_i\right) \omega\left(\frac{5}{\Delta s} \frac{v_y}{Sv} - y_i\right) \quad (3)$$

$$\omega_{\text{ang}}(\alpha) = \sum_{r=-1}^1 \omega\left(\frac{8\alpha}{2\pi} + 8r\right) \quad (4)$$

where  $x_i$  and  $y_i$  are the spatial bin centers located in  $x_i, y_i \in \{-\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}\}$ .

The function parameter  $\mathbf{v} = (v_x, v_y)$  is a vector variable and  $\alpha$  a scalar variable. The value of  $\Delta s$  is the length of the patch in pixels for the unit scale, which is described in the section 2.5. The values of  $\frac{5}{\Delta s \, S_t}$  and  $\frac{5}{\Delta s \, S_v}$  allow a conversion from pixels to units of separation from the patch center. On the other hand,  $r$  is an integer that can vary freely in the set  $\{-1, 0, 1\}$  and allows the argument  $\alpha$  to be unconstrained in terms of its values in radians. The interpolating function  $\omega(\cdot)$  is defined as:

$$\omega(z) = \max(0, 1 - |z|). \quad (5)$$

These binning functions conform to a trilinear interpolation that has a combined effect of sharing the contribution of each oriented gradient between their eight adjacent bins in a tridimensional cube in the histogram space and zero everywhere else. This procedure is important to avoid quantization issues that may appear with the histogram (i.e. smoothing transitions between different values).

As the patch has 16 blocks and 8 bin angles are considered, a feature  $\mathbf{d}$  called *descriptor* of 128 dimension is obtained. This technique is a subtle adaptation of Lowe’s SIFT Descriptor method.

In Figure 4 the possible orientations on each patch are illustrated. The first eight orientations of the first block  $B_{1,1}$ , are labeled from 1 to 8 clockwise. The orientations of the second block  $B_{1,2}$  are labeled from 9 to 16. This labelling continues left-to-right, up-down until the eight orientations for all the sixteen blocks are assigned. They form the corresponding descriptor  $\mathbf{d}$  of 128 dimensions.

#### 2.4. Keypoint Location

The keypoint  $\mathbf{kp}$  location must be accurately specified in order to establish in which part of the image, the waveform should be captured.

For the horizontal position, the time axis, the localization of the keypoint is based on a priori information, based on the characteristics of the event under study. In general, as EEG signal segments have a very specific length which

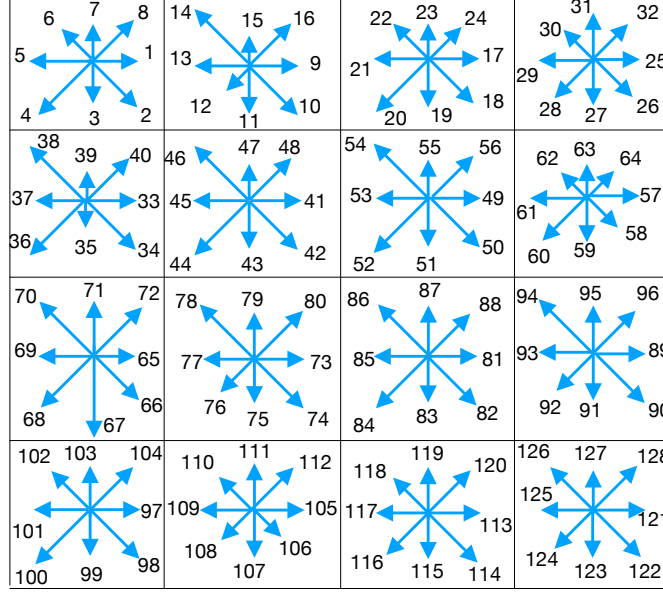


Figure 4: A scheme of the orientation's histogram computation. The first eight orientations of the first block  $B_{1,1}$ , are labeled from 1 to 8 clockwise. The orientation of the second block  $B_{1,2}$  is labeled from 9 to 16. This labelling continues left-to-right, up-down until the eight orientations for all the sixteen blocks are assigned. They form the corresponding descriptor of 128 coordinates. The length of each arrow represents the value of the histogram in each direction for each block.

translates to a fixed width, the center of the image is the more appropriate  
choice. Certain EEG components have a specific timing that can be explored  
to elucidate in which position the expected signal pattern is ought to be found,  
and that can be used to localize the keypoint.

Once the horizontal location is established, the vertical location can be  
obtained by localizing it along with the signal plot, exactly on a corresponding  
sample point calculated by Equation 1.

Additionally, there can be more than just one keypoint and patch located  
over the signal plot. This is particular important for oscillatory processes where  
many waveforms are contained within the same signal segment.

## 2.5. Patch Geometry

195 The standard implementation of the SIFT Descriptor uses a squared-size patch, and there is only one scale parameter  $S$ . However, this is not appropriate to capture waveforms that may expand horizontally on the time scale. This has been modified on this proposal, allowing a rectangular patch geometry that can be used to cover an entire waveform, regardless of their span, which we call  $\lambda$ .  
 200 The original SIFT scale is modified in this implementation to allow two scale parameters, one per axis.

The Horizontal Patch Scale  $St$  determines the size of the patch on the image horizontally, and it is related to the span  $\lambda$  of the waveform to analyze according to

$$St = \frac{\lambda F_s \gamma_t}{\Delta_s} \quad (6)$$

205 where  $F_s$  is the sampling frequency,  $\gamma_t$  is the time scale factor and  $\Delta_s$  is the unit length of the patch which determines the pixel conversion factor. This value depends on the actual implementation, and for this work and for the accompanying software, its value is  $\Delta_s = \sqrt{2} \cdot 3 \cdot 5$ , where 3 is the fixed magnification factor, and 5 correspond to the number of blocks in which the patch is divided, plus half the size of the block on each direction. Section 2.6  
 210 provides more details on the SIFT method implementation.

On the other hand, on the vertical axis, the vertical patch scale depends on the peak-to-peak amplitude  $\Delta\mu V$ , and the amplitude scale factor  $\gamma$ , as

$$Sv = \frac{\Delta\mu V \gamma}{\Delta_s}. \quad (7)$$

The vertical scale can be dynamically adjusted according to the peak-to-peak  
 215 amplitude of each segment, or it can be set fixed. This is more appropriate if the underlying signal is bounded which is the case if the signal is normalized or standardized.

Figure 5 shows the different parameters of the patch and how they are related to the underlying signal. Once these parameters are set, the size in pixels of the

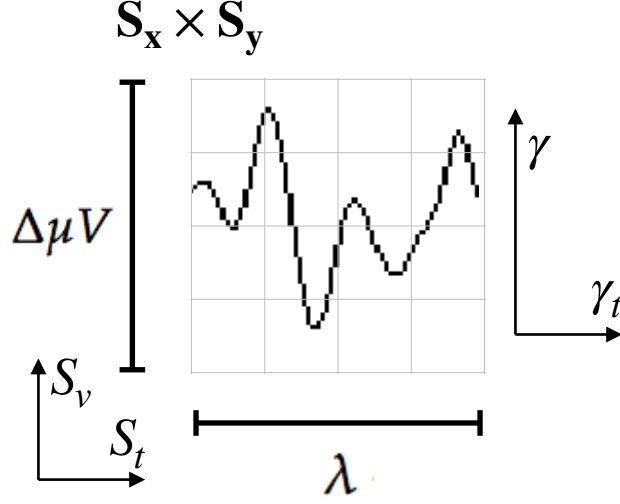


Figure 5: The scale of the local patch is selected in order to capture the whole waveform, which can be scaled in the time  $\gamma_t$  and amplitude  $\gamma$  direction. This determines appropriate horizontal  $S_t$  and vertical  $S_v$  patch scales. The size of the patch is  $\mathbf{S}_x \times \mathbf{S}_y$  pixels. The vertical size consists of 4 blocks of size  $\mathbf{S}_y$  pixels which should be high enough as to contain the signal  $\Delta\mu V$ , the peak-to-peak amplitude of the signal component. The horizontal size includes 4 blocks, up to  $\mathbf{S}_x$  pixels, and should cover the entire duration in seconds of the signal waveform,  $\lambda$ .

220 patch can be obtained in both dimensions. Hence, the horizontal patch size in pixels is

$$\mathbf{S}_x = \lfloor \Delta_s S_t \rfloor + 1 \quad (8)$$

and the vertical patch size in pixels can be calculated from

$$\mathbf{S}_y = \lfloor \Delta_s S_v \rfloor + 1 \quad (9)$$

where  $\Delta_s$  being the unit length of the patch. The parameters  $S_t$  and  $S_v$  are the horizontal and vertical patch scale. This region is arranged in a  $4 \times 4$  grid  
 225 and the pixel  $\mathbf{k}\mathbf{p}$  is the patch centre. For instance, for a given set of values of

$Sv = 1$  and  $St = 1$ , the patch is a squared region on the image of size 22 pixels.

The patch size cannot be bigger than the image itself, whose width is  $Wx$  and its height is  $Hy$ . This is reflected by the following two inequalities that restrict the size of the patch according to

$$\frac{Wx - 1}{\Delta_s} \geq St, \quad (10)$$

230 on the horizontal axis, and on the vertical axis,

$$\frac{Hy - 1}{\Delta_s} \geq Sv. \quad (11)$$

## 2.6. Variations on the SIFT Method

The details provided here are implemented in the open source software package described in the repository [33].

### 2.6.1. SIFT Detector and Custom Patch

235 The SIFT Detector is not being used in this implementation. Hence, the keypoint location and patch parameters are directly calculated based on the method provided in 2.4. A *frame* is a data structure composed of keypoint center location  $(x_{kp}, y_{kp})$ , patch scale  $S$  and patch orientation  $\theta$ :  $(x_{kp}, y_{kp}, S, \theta)$ . These parameters are the output of the SIFT Detector. The code provided in  
240 Repository [33] calculates these frames based on the provided parameters and on the structure of the EEG signal.

### 2.6.2. Patch Scale

Whereas in the standard SIFT implementation the patch is a squared region and there is only one SIFT scale parameter, in this implementation the scale  
245 is divided in two:  $St$  and  $Sv$ . This is a very important modification because otherwise signal plots, which may extend only in the horizontal direction, would not have been able to be mapped. Using a rectangular frame there isn't any constraint on its size and it can be adjusted at will to map any waveform.



This modification forces the frame data structure to be also altered to  
 250 incorporate one extra parameter, the extra scale. The frame is thus composed  
 of:  $(x_{kp}, y_{kp}, St, Sv, \theta)$ .

### 2.6.3. Patch Orientation

The value  $\theta$  is the patch orientation which does not provide any extra utility  
 so far for the extraction of characteristics from plots. The orientation of the  
 255 patch is fixed to zero (vertical, pointing upwards or towards the horizontal axis  
 of the image coordinate system).

### 2.6.4. Patch Size in Pixels

The nominal size of a patch in the SIFT implementations proposed by [29]  
 is  $4 \times m \times S$ , with  $m$  being the magnification factor and  $S$  being the SIFT scale.  
 260 However, due to the trilinear interpolation, pixels that are located outside the  
 nominal patch size are also considered to calculate the histogram, hence the  
 effective size of the patch must be considered. This size is indeed based on the  
 unit-scale constant and it is  $\Delta s = \sqrt{2} m 5$ . The magnification factor  $m$  is set  
 as 3 and 5 is the result of using 4 blocks on each direction plus half a block on  
 265 each side. For a unit scale on vertical and horizontal direction, this gives an  
 effective patch size of 22 pixels in both directions.

### 2.6.5. Octave Selection

A gradient image is used to calculate the oriented gradients and reckon the  
 histogram of gradient orientations. SIFT calculates different octaves downsampling  
 270 the original image and applying a Gaussian smoothing operation increasing the  
 sigma parameter of the Gaussian window step by step. SIFT calls *octave* to  
 each downsampling level [29, 30]. In this implementation, only the zero octaves  
 is used which means that the gradient image has the same size as the original  
 patch, without any image downsampling.

275 *2.6.6. First Octave Smoothing*

Additionally, the standard SIFT implementation performs a Gaussian blurring on the gradient image regardless of the octave. This is disabled in this implementation.

*2.6.7. Rotations*

SIFT was designed to allow affine invariance, i.e. to be robust to rotations  
280 and scale modifications. This feature has not been disabled in this implementation, but it is avoided by using a patch orientation equals to 0. This the reason why the  $\sqrt{2}$  constant is kept in the definition of the unit-scale constant  $\Delta s$ .

*2.6.8. Gaussian Smoothing*

Gaussian smoothing is performed on the SIFT patch to increase the importance  
285 of the gradients from pixels closer to the centre of the patch. In this case, this is found to be to the detriment of the waveform characterization and is disabled in this implementation.

*2.6.9. SIFT Descriptor Values*

The SIFT descriptor  $\mathbf{d}$  is a 128-dimension feature vector, as described in  
290 Section 2.2. Histogram values are double-precision floating-point numbers, all positive, and they are accumulated on each coordinate. Once the gradients are calculated, the following operations are performed:

- The descriptor is  $\ell_2$  normalized (i.e all the values are divided by the euclidean norm of the descriptor).
- 295 • Each value is clamped to 0.2. This means that any value above 0.2 is set to 0.2.
- The descriptor is  $\ell_2$  re-normalized again [30].

This generate a 128-vector of double precision floating point numbers, between  $[0 \cdots 1]$ . This implementation was modified to allow the following representations [34]:

- Discrete: The vector is rescaled to  $[0 \cdots 511]$  and clamped at 255. Output values are cast to integer representations in 8-bit precision. This yields an effective 128-vector of integer values between  $[0 \cdots 255]$ .
- Euclidean: The vector is rescaled to  $[0 \cdots 511]$ . Output values are cast to single-precision floating point numbers (i.e. floats). This yields an effective 128-vector of floats between  $[0 \cdots 255]$ .
- Cosine: The vector is rescaled to  $[-1 \cdots 1]$ . Output values are cast to single-precision floating point numbers (i.e. floats). This yields an effective 128-vector of floats between  $[-1 \cdots 1]$ .
- Hellinger: The vector is rescaled to  $[0 \cdots 1]$ . A  $\ell_1$  normalization is applied (i.e. each vector values are divided by the absolute value of the summation of all the values). The square-root on each coordinate is applied. This yields an effective 128-vector of floats between  $[0 \cdots 1]$ .

## 2.7. Dataset

The dataset used was recorded by standard polysomnography including EEG, EMG, and EOG recordings using a BrainAmp amplifier (Brain Products, Munich, Germany). EEG was recorded from six scalp electrodes (F3, F4, C3, C4, P3, and P4 according to the International 10-20 system) and two electrodes on the left and right mastoids serving as a combined reference. Data were recorded at a sampling rate of 200 Hz and bandpass-filtered between 0.16 and 35 Hz. This dataset contains information from 48 subjects (mean age:  $23.5 \pm 0.5$ ; 24 females). More information can be obtained [35]. To test the algorithm implemented in this work, the signal from channel C4 was annotated by experts in 4 subjects.

## 2.8. Processing Pipeline

### 2.8.1. Preprocessing

A spectral filter of the raw signal was implemented to narrow the frequency spectrum to Slow Waves which is between 0.5 Hz and 1 Hz. The filter is a band-

pass digital IIR filter with cut frequencies of 0.5 Hz and 4 Hz implemented in the MNE Software [36]; in this way, the frequency characteristics of the Slow Waves are preserved, discarding other irrelevant frequencies.

### 2.8.2. Segmentation

After the experts annotated the signals for 4 subjects, segments of 1 second duration are extracted (i.e. 200 time points). The Slow Wave is composed of three peaks (Figure 1), and the minimum value, the center negative peak can be used as a reference. Hence, this value was used as the center of the segment (the 100th time point), so that all the segments are properly aligned. In addition, the same amount of random non-SW segments of equal duration which were not marked by the experts are extracted from the EEG signal to achieve two balanced classes for the classification stage.

Once the signal segments have been obtained, they are fed to the feature extraction procedure described in Section 2.2, to produce descriptors  $\mathbf{d}$  that are used by the classifiers.

Table 1: Number of Slow Waves extracted from each Subject

Subject ID	Number of Extracted Segments
0	58
1	127
2	138
3	129

### 2.8.3. Classification Algorithms

We tested the performance of several classifiers in differentiating between the descriptors  $\mathbf{d}$  of the Slow Waves segments and the other EEG baseline segments extracted from the C4 channel of the four subjects' EEG recordings. A feed forward Neural Network (NN) with an input layer of 128 inputs (i.e. the dimension of descriptors  $d$ ), and a hidden layer of 256 units, with ReLU as

activation function, and a softmax final layer for the two binary classes [37].

350 Additionally, a classifier based on k-Near Neighbors (kNN) with  $k = 5$ , a Support Vector Machine (SVM) classifier with a linear kernel, a Decision Tree Classifier (DecTree), a Logistic Regression (LogReg) Classifier, a Linear Discriminant Analysis (LDA) Classifier and a Random Forrest (RF) Classifier [38].

### 3. Results and Discussion

355 In Figure 6 it can be observed the plot of True Positive Rate (sensitivity) versus False Positive Rate (1-specificity) across varying cut-offs generating a curve in the unit square called a ROC curve [39]. It can be seen that LDA produces better results in 2 of the subjects while RF works better for all the subjects where the area under the curve (AUC) is closer to one. Table 2 shows  
360 the obtained Area Under the Curve obtained for the RF classifier for the four different subjects.

Table 2: Obtained AUC and Accuracy values for the four differents subjects using a RF classifier.

Subject ID	AUC	Accuracy
0	0.760	0.77
1	0.748	0.71
2	0.824	0.74
3	0.814	0.80

On Figure 7 the confusion matrix using Random Forest, the best performing classifier, for the four subjects can be seen. The Accuracy metric, which is also shown on Table 2 can be obtained by summing the values for the diagonal  
365 (True Positives plus False Negatives) and dividing them on the total number of available samples to classify. The obtained performance was higher for subjects Number 0 and 3. In spite of that, it should be also noted that there is a low percentage of false negatives.

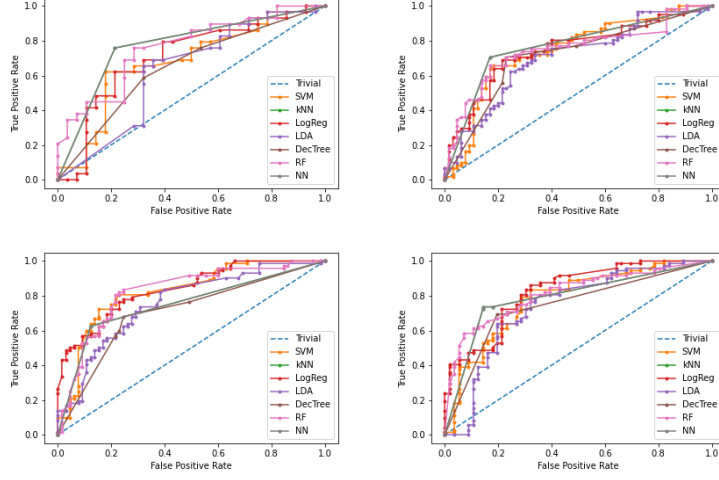


Figure 6: ROC curves obtained for the 7 classifiers for each one of the subjects. From top to bottom, left-to-right obtained values for Subject 0,1,2 and 3. The chance level boundary is shown as a *trivial*.

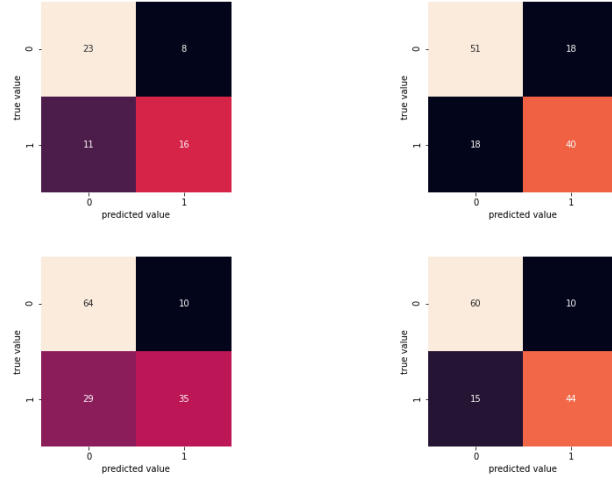


Figure 7: Confusion matrices obtained for the four different subjects. From top to bottom, left-to-right obtained values for Subject 0,1,2 and 3.

On the other hand, Figure 8 shows a transfer learning experiment where the  
 370 accuracy values obtained by training a RF classifier using a training subset of the

descriptors for one subject and testing on testing sets from the other subjects. Automatic EEG signal analysis is highly sensible to inter-subject variability, and models trained with the information obtained for one subject are unlikely to extrapolate easily to decode the signals for other subjects [40]. In this case, the feature extraction method described in 2.2 showed that, though the diagonal on Figure 8 shows the higher accuracy values, the obtained values for cross-subject classification showed similar accuracy values.

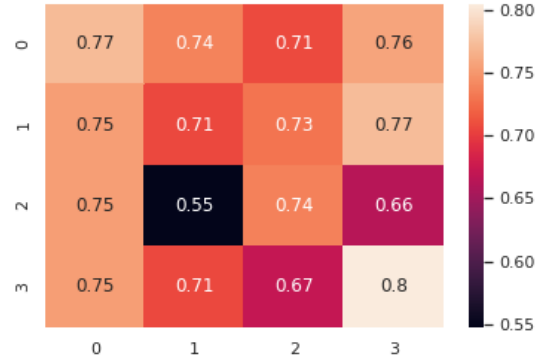


Figure 8: Transfer learning experiment, where a RF classifier is trained using a training set from the descriptors obtained for the Slow Waves extracted from the EEG records of subject, and tested on testing sets of the remaining three subjects. Each value represent the Accuracy obtained. Although the diagonal shows the higher values, accuracies values remain at higher levels.

#### 4. Conclusions and Future Work

In this work, a method to capture the waveform of Slow Waves as a feature extraction procedure is proposed. Their validity is shown by obtaining performance values similar to those obtained by experts performing visual observations. This method has the advantage that by analyzing waveforms it is feasible and direct to understand what is the classifier actually using to arrive to a decision [41]. This is a very important topic in health application of artificial intelligence and machine learning. At the same time, this process is highly parallelizable and

that can be used to take advantage of Graphic Process Units (GPU) which are a key component of the current trend in this area [37]. On the other hand, this system requires an excellent alignment of the signals, and this can be avoided by proposing a more accurate keypoint location method. In Sleep Research, the  
 390 localization of features within the signal stream is very relevant. The procedure, and the experiments performed in this work, assume that the location is known. At the same time, the scale-space theory started precisely on 1D signals [18], so further research could be performed to understand if Witkin's ideas can be reapplied again on PSG signals.

## 395 References

- [1] S. R. Cole, B. Voytek, Brain Oscillations and the Importance of Waveform Shape, *Trends in Cognitive Sciences* 21 (2) (2017) 137–149. doi:10.1016/j.tics.2016.12.008.  
 URL <http://dx.doi.org/10.1016/j.tics.2016.12.008>
- 400 [2] A. Rodenbeck, R. Binder, P. Geisler, H. Danker-Hopfe, R. Lund, F. Raschke, H. G. Weeß, H. Schulz, A review of sleep EEG patterns. Part I: A compilation of amended rules for their visual recognition according to Rechtschaffen and Kales, *Somnologie* 10 (4) (2006) 159–175. doi:10.1111/j.1439-054X.2006.00101.x.
- 405 [3] D. L. Schomer, F. L. D. Silva, *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, Vol. 1, Oxford University Press, 2010.
- [4] R. Ramele, A. Villar, J. Santos, BCI classification based on signal plots and SIFT descriptors, in: 2016 4th International Winter  
 410 Conference on Brain-Computer Interface (BCI), IEEE, 2016, pp. 1–4. doi:10.1109/IWW-BCI.2016.7457454.  
 URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7457454>



- [5] R. Ramele, A. J. Villar, J. M. Santos, EEG waveform analysis of p300  
ERP with applications to brain computer interfaces, *Brain Sciences* 8 (11)  
(2018) 199. doi:10.3390/brainsci8110199.  
URL <http://www.mdpi.com/2076-3425/8/11/199>
- [6] R. Ramele, A. J. Villar, J. M. Santos, Histogram of gradient orientations  
of signal plots applied to p300 detection, *Frontiers in Computational  
Neuroscience* 13. doi:10.3389/fncom.2019.00043.  
URL <https://www.frontiersin.org/article/10.3389/fncom.2019.00043/full>
- [7] B. Rasch, J. Born, About sleep's role in memory, *Physiological Reviews*  
93 (2) (2013) 681–766. doi:10.1152/physrev.00032.2012.  
URL [www.prv.org](http://www.prv.org)
- [8] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus, B. V.  
Vaughn, et al., The aasm manual for the scoring of sleep and associated  
events, Rules, Terminology and Technical Specifications, Darien, Illinois,  
American Academy of Sleep Medicine 176 (2012) 2012.
- [9] R. Yan, C. Zhang, K. Spruyt, L. Wei, Z. Wang, L. Tian, X. Li,  
T. Ristaniemi, J. Zhang, F. Cong, Multi-modality of polysomnography  
signals' fusion for automatic sleep scoring, *Biomedical Signal Processing  
and Control* 49 (2019) 14–23. doi:10.1016/j.bspc.2018.10.001.
- [10] R. Boostani, F. Karimzadeh, M. Nami, A comparative review on  
sleep stage classification methods in patients and healthy individuals,  
*Computer Methods and Programs in Biomedicine* 140 (2017) 77 – 91.  
doi:<https://doi.org/10.1016/j.cmpb.2016.12.004>.  
URL <http://www.sciencedirect.com/science/article/pii/S0169260716308276>
- [11] H. V. V. Ngo, J. Claassen, M. Dresler, Sleep: Slow Wave Activity Predicts  
Amyloid- $\beta$  Accumulation, *Current Biology* 30 (22) (2020) R1371–R1373.  
doi:10.1016/j.cub.2020.09.058.

- [12] A. K. Rechtschaffen, Manual of Standardized Terminology, Technique and Scoring System for Sleep Stages of Human Sleep, Brain Information Service, Brain Information Institute, UCLA: Los Angeles, CA, USA, .
- [13] A. F. Jackson, D. J. Bolger, The neurophysiological bases of EEG and EEG measurement: A review for the rest of us, *Psychophysiology* 51 (11) (2014) 1061–1071. doi:10.1111/psyp.12283.  
URL <http://doi.wiley.com/10.1111/psyp.12283>
- [14] J. Jestico, P. Fitch, R. W. Gilliatt, R. G. Willison, Automatic and rapid visual analysis of sleep stages and epileptic activity. A preliminary report, *Electroencephalography and Clinical Neurophysiology* 43 (3) (1977) 438–441. doi:10.1016/0013-4694(77)90268-1.  
URL <http://linkinghub.elsevier.com/retrieve/pii/0013469477902681>
- [15] J. E. Bresenham, Algorithm for computer control of a digital plotter, *IBM Systems Journal* 4 (1) (1965) 25–30.
- [16] D. Marr, T. Poggio, A computational theory of human stereo vision, *Proceedings of the Royal Society of London - Biological Sciences* 204 (1156) (1979) 301–328. doi:10.1098/rspb.1979.0029.  
URL <https://royalsocietypublishing.org/doi/abs/10.1098/rspb.1979.0029>
- [17] T. Iijima, Observation theory of two-dimensional visual patterns, Tech. rep., Electrotechnical Laboratory ETL, Tokyo Institute Of Technology (1962).
- [18] A. P. Witkin, Scale-space filtering, in: *Proceedings of the Eighth International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'83*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1983, p. 1019–1022.

- 470 [19] J. J. Koenderink, The structure of images, *Biological Cybernetics* 50 (5) (1984) 363–370. doi:10.1007/BF00336961.  
URL <https://link.springer.com/article/10.1007/BF00336961>
- [20] J. Weickert, S. Ishikawa, A. Imiya, Linear scale-space has first been proposed in Japan, *Tech. Rep. 3* (1999). doi:10.1023/A:1008344623873.
- 475 [21] T. Lindeberg, Scale-space for discrete signals, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (3) (1990) 234–254. doi:10.1109/34.49051.
- [22] T. Lindeberg, *Scale-Space Theory in Computer Vision*, 1st Edition, Springer US, Boston, MA, 1994. doi:10.1007/978-1-4757-6465-9\_1.  
480 URL [http://link.springer.com/10.1007/978-1-4757-6465-9\\_1](http://link.springer.com/10.1007/978-1-4757-6465-9_1)
- [23] B. ter Haar Romeny, B. M. ter Haar Romeny, L. Florack, J. Koenderink, M. Viergever, et al., *Scale-Space Theory in Computer Vision: First International Conference, Scale-Space’97, Utrecht, The Netherlands, July 2-4, 1997, Proceedings, Vol. 1252*, Springer Science & Business Media, 1997.
- 485 [24] Y. Jansson, T. Lindeberg, Exploring the ability of CNNs to generalise to previously unseen scales over wide scale ranges, *arXiv:2004.01536*.  
URL <http://arxiv.org/abs/2004.01536>
- [25] D. H. HUBEL, T. N. WIESEL, Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 and 19) of the Cat, *Journal of neurophysiology* 28 (1965) 229–289. doi:10.1152/jn.1965.28.2.229.  
490 URL <https://journals.physiology.org/doi/abs/10.1152/jn.1965.28.2.229>
- [26] S. Edelman, N. Intrator, T. Poggio, Complex cells and object recognition, *Unpublished manuscript* (1997) 1–12.  
495
- [27] O. Linde, T. Lindeberg, Composed complex-cue histograms: An investigation of the information content in receptive field based

- image descriptors for object recognition, *Computer Vision and Image Understanding* 116 (4) (2012) 538–560. doi:10.1016/j.cviu.2011.12.003.
- 500 URL <https://www.sciencedirect.com/science/article/pii/S1077314211002633>
- [28] T. Lindeberg, A computational theory of visual receptive fields, *Biological Cybernetics* 107 (6) (2013) 589–635. arXiv:1701.06333, doi:10.1007/s00422-013-0569-z.
- 505 URL <http://arxiv.org/abs/1701.06333>
- [29] G. Lowe, SIFT - The Scale Invariant Feature Transform, *International Journal* 2 (2004) 91–110.
- [30] I. Rey-Otero, M. Delbracio, Anatomy of the SIFT Method, *Image Processing On Line* (2014) 370–396doi:<http://dx.doi.org/10.5201/ipol.2014.82>.
- 510
- [31] R. Szeliski, *Computer Vision: Algorithms and Applications*, Vol. 24, Springer, 2010.
- URL <http://books.google.com.ar/books/about/Computer{ }Vision.html?id=8{ }2RNQEACAAJ{ }&pgis=1>
- 515
- [32] A. Vedaldi, B. Fulkerson, VLFeat - An open and portable library of computer vision algorithms, *Design* 3 (1) (2010) 1–4. doi:10.1145/1873951.1874249.
- URL <http://vision.ucla.edu/{ }brian/papers/vedaldi10vlfeat.pdf>
- 520
- [33] R. Ramele, Bci Sift Repository, <https://github.com/faturita/BciSift> (2018 (accessed March 31, 2021)).
- [34] R. Arandjelovic, A. Zisserman, Three things everyone should know to improve object retrieval, in: *2012 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2012, pp. 2911–2918. doi:10.1109/
- 525

CVPR.2012.6248018.

URL [doi.ieeecomputersociety.org/10.1109/CVPR.2012.6248018](https://doi.ieeecomputersociety.org/10.1109/CVPR.2012.6248018)

- [35] C. Forcato, J. G. Klinzing, J. Carbone, M. Radloff, F. D. Weber, J. Born, S. Diekelmann, Reactivation during sleep with incomplete reminder cues rather than complete ones stabilizes long-term memory in humans, *Communications Biology* 3 (1) (2020) 1–13. doi:10.1038/s42003-020-01457-4.

URL <https://www.nature.com/articles/s42003-020-01457-4>

- [36] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, M. Hämäläinen, MEG and EEG data analysis with MNE-Python, *Frontiers in Neuroscience* 7 (7) (2013) 267. arXiv:[1]A.Gramfort, MEGandEEGdataanalysiswithMNE-Python, *Front.Neurosci.*, vol.7, no.December, pp.113, 2013., doi:10.3389/fnins.2013.00267.

URL <http://journal.frontiersin.org/article/10.3389/fnins.2013.00267/abstract>

- [37] S. Gong, K. Xing, A. Cichocki, J. Li, Deep learning in eeg: Advance of the last ten-year critical period, arXiv preprint arXiv:2011.11128.

- [38] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, F. Yger, A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update, *Journal of Neural Engineering* 15 (3) (2018) 031005. doi:10.1088/1741-2552/aab2f2.

URL <http://stacks.iop.org/1741-2552/15/i=3/a=031005?key=crossref.9cd2b15ab65c8ad34b475584b43dc509>

- [39] T. Fawcett, An introduction to ROC analysis *Tom, Irbm* 35 (6) (2005) 299–309. arXiv:/dx.doi.org/10.1016/j.patrec.2005.10.010. doi:10.1016/j.patrec.2005.10.010.

URL <https://www.sciencedirect.com/science/article/abs/pii/S016786550500303X>

- 555 [40] Z. Wan, R. Yang, M. Huang, N. Zeng, X. Liu, A review on transfer learning  
in EEG signal analysis, *Neurocomputing* 421 (2021) 1–14. doi:10.1016/  
j.neucom.2020.09.017.
- [41] E. Khodabandehloo, D. Riboni, A. Alimohammadi, HealthXAI:  
Collaborative and explainable AI for supporting early diagnosis of cognitive  
560 decline, *Future Generation Computer Systems* 116 (2021) 168–189. doi:  
10.1016/j.future.2020.10.030.

**Declaration of interests**

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: