Unobtrusive Inference of Affective States in Virtual Rehabilitation from Upper Limb Motions: A Feasibility Study

Jesús Joel Rivas, Felipe Orihuela-Espina, Lorena Palafox, Nadia Bianchi-Berthouze, María del Carmen Lara, Jorge Hernández-Franco, and Luis Enrique Sucar, *Senior Member, IEEE*

Abstract—Virtual rehabilitation environments may afford greater patient personalization if they could harness the patient's affective state. Four states: anxiety, pain, engagement and tiredness (either physical or psychological), were hypothesized to be inferable from observable metrics of hand location and gripping strength -relevant for rehabilitation-. Contributions are; (a) multiresolution classifier built from Semi-Naïve Bayesian classifiers, and (b) establishing predictive relations for the considered states from the motor proxies capitalizing on the proposed classifier with recognition levels sufficient for exploitation. 3D hand locations and gripping strength streams were recorded from 5 post-stroke patients whilst undergoing motor rehabilitation therapy administered through virtual rehabilitation along 10 sessions over 4 weeks. Features from the streams characterized the motor dynamics, while spontaneous manifestations of the states were labelled from concomitant videos by experts for supervised classification. The new classifier was compared against baseline support vector machine (SVM) and random forest (RF) with all three exhibiting comparable performances. Inference of the aforementioned states departing from chosen motor surrogates appears feasible, expediting increased personalization of virtual motor neurorehabilitation therapies.

Index Terms—Affective issues in user interaction, posture, hand movements, fingers pressure, rehabilitation, stroke, semi-Naïve Bayesian classifier.

◆

1 INTRODUCTION

DETECTION of user's affective states permit the implementation of algorithmic empathic interaction strategies. A growing area of application of Affective Computing (AC) is medical therapy, where affective-aware medical informatics technologies can help in monitoring and personalizing therapy to the patients' needs. In this paper we investigate the possibility to automatically detect the affective states of stroke patients during occupational rehabilitation.

Stroke is a leading cause of motor impairment [1], [2]. A common sequelae to stroke survivors is motor disability of the upper limb. Motor rehabilitation therapies, by fostering functionally targeted repetition, help patients to almost fully or partially recover their functional abilities by means of improving their post-stroke movements [3], [4]. Virtual rehabilitation (VR) [5], [6] is an alternative to promote motor reha-

bilitation exercises following stroke whereby motor training is gamified. For example, VR platforms favour engagement by simulating tasks that resemble everyday activity to which people with stroke aim to reengage in [2]. Patient motivation is crucial to physical therapy adherence [7] and affective states play an important role in it [8]. A systematic review by Luker et al. [9] on physical rehabilitation in stroke patients shows how motivation to engage in the physical rehabilitation tasks is affected by fatigue, pain, and anxiety among many other affective factors. Patients reported to be overwhelmed by pain and fatigue as well as being anxious about further injury. Pain in itself is an "unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage" [10]. Chronic pain (or maladaptive pain) and post-stroke fatigue are also typical consequences of stroke triggering related anxiety and fear towards movements [11]. Customizable and adaptable training virtual environments automatically change to the different patients' needs to optimize therapy outcome but thus far, these adaptive decisions are mostly based on observable performance metrics [12]. Automatic decisions should ideally be guided by both the observable performance and the hidden cognitive-affective state of the user. It follows that being able to access to the user's affective state can be exploited to design highly engaging and motivating VR sessions.

Incorporating AC into the VR systems enriches the system by measuring hidden variables i.e. affective states, often requiring the definition of observable surrogates. When not consciously inhibited, our behavioural gestures may convey information about our affective state. For instance,

- *Jesús Joel Rivas is with the Department of Computer Science, Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla, 72840 Mexico and the Computer Science Department, Science and Technology Faculty (FACYT), Universidad de Carabobo (UC), Valencia, Venezuela. E-mail: jrivas@ccc.inaoep.mx, jrivas@uc.edu.ve*
- *Felipe Orihuela-Espina is with the Department of Computer Science, Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla, Mexico.*
- *Lorena Palafox is with the Instituto Nacional de Neurología y Neurocirugía (INNN), Mexico City.*
- *Nadia Bianchi-Berthouze is with the UCL Interaction Centre, University College of London (UCL), London.*
- *María del Carmen Lara is with the Benémerita Universidad Autónoma de Puebla (BUAP), Puebla, Mexico.*
- *Jorge Hernández-Franco is with the Instituto Nacional de Neurología y Neurocirugía (INNN), Mexico City.*
- *Luis Enrique Sucar is with the Department of Computer Science, Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla, Mexico.*

it may be conjectured that gripping force exerted during the rehabilitation task may be affected by anxiety and by psychological and physical tiredness [13]. Our hypothesis is that anxiety, pain, engagement (as a being engrossed in the current activity) and tiredness are recoverable from basic observations of hand locations and pressure exerted upon gripping while the patient practices its rehabilitation in the virtual environments. If successfully recovered, the new knowledge may, in the near future, be capitalized upon for adapting the virtual therapy to the patient's affective state. In this paper, we use the term tiredness to indicate both physical or mental tiredness as in this condition they may be particularly interlinked.

A feasibility pilot is presented here whereby 5 post-stroke patients participated longitudinally in a virtual rehabilitation platform and information of their hand movements and fingers pressure (experimental units) were given as entries to some classifiers to establish whether recovery of the aforesaid states was viable from the dynamic behaviour of hand motion and fingers pressure. While an explanatory relation between the hidden affective state and the observable surrogates might be desirable, for the purposes of informing the decision maker in the virtual rehabilitation platform, a predictive relation suffices. Yet, even finding such predictive relation is far from trivial as we shall exemplify here by means of reporting results from the application of baseline classifiers. Hence, in our quest to uncover the predictive association we here further propose a novel classification strategy, the so called Multiresolution Semi-Naïve Bayesian classifier (MSNB). Since adaptation is often patient-based, we chose to train the classifiers independently for each patient. From the point of view of the classifier, the observations correspond to the local dynamics of hand movements and fingers pressure of the patient during the affective episodes. In other words, the sample size is not 5 (patients), but the number of affective episodes as labelled by experts. The registered affective states were manifested spontaneously by the patients whilst they participated in the virtual rehabilitation program, i.e. they were not acted. At this point, it suffices for our purposes to achieve a predictive power well above random choice. We do not intend to control every aspect of the game flux such as online requests of changing game behaviour according to the affective state but to provide the adaptation algorithms, when the technology is matured, with an additional information channel.

The main contributions of this paper are two: a) A novel classifier, Multiresolution Semi-Naïve Bayesian classifier (MSNB), which exploits structural improvement and temporal multiresolution to overcome the limitations of the naïve Bayesian classifier when dealing with the varying dynamics of the process, and b) Leveraging speed and acceleration of hand displacements and fingers pressure as a means to recognize affective states. While the evaluation of the approach is based only on five stroke patients, their data were recorded longitudinally over a period of 4 weeks throughout which their physical capabilities improved bringing hence more complexity to the recognition of the affective states.

This paper is organized as follows, section 2 summarizes related work. Sections 3, 4 and 5 describe the methodolog-ical steps to this study, explaining data collection which includes the affective states selection and labelling step; the design of the feature vector and the design of the proposed classification model MSNB. Section 6 highlights the results obtained with MSNB, including comparisons with SVM, SNB and RF. Section 7 contains the discussion and, finally section 8 summarizes main findings and describes future approaches to follow.

2 RELATED WORK

2.1 Naturalistic Everyday Affective States Recognition

In recognizing naturalistic everyday affective states and understanding human behaviour, there are several open research challenges [14]. There is a diversity of affective states and it is difficult to fully separate or discriminate them [15]. Also, people differ in the way they express those states as their expressions are affected by idiosyncrasy [16]. Further, the duration of emotions is highly variable [17]. Scherer K.R. and Ekman P. reported durations 0.5 to 4 seconds for some emotions in their works [18], [19], but there is an ongoing debate about how long emotions last and this is hindered by the lack of consensus about the definition of emotion [18], [20], [21]. Despite all, gaining a partial solution is still valuable and helpful, and it may contribute to intelligent interactions that benefit people's health and productivity [15].

2.2 Machine Learning Approaches in Affective Computing

A number of machine learning approaches have been used to continuously track information over time and recognize affective states [22], [23], [24], [25], [26], [27]. The temporal variation is an important issue in recognizing naturalistic everyday affective states [28]. SVM is a popular choice in many affective recognition systems, so it is used here as baseline [29]. Other classifiers that have been employed include neural networks [22], recurrent neural networks [24], [30], dynamic Bayesian networks [27], hidden Markov models [25], [26], [31], [32], latent-dynamic conditional random fields [33], [34]. The Naïve Bayes classifiers has been studied and compared with others classifiers and has often been more effective than sophisticated rules [35], [36], [37], [38]. We here proposed to use a classifier derived from Naïve Bayes (Semi-Naïve Bayes) for its efficiency, simplicity and because it deals with dependent features [39]. Anticipating the variability in the emotion durations, our Semi-Naïve Bayes multiresolution variant embeds the dynamic behaviour of hand movements and fingers pressure in windows of consecutive points of the hand trace and the fingers pressure trace over time. We consider recognizing the moment when the affective state begins, when it is manifesting and when it finishes.

2.3 Automatic Affect Recognition in Physical Rehabilitation

A particular area that has seen an increased attempt to study naturalist expressions in ecological or semi-ecological contexts is rehabilitation. Aung et al. [40] for example, investigated recognition of pain-related affective states in

chronic pain physical rehabilitation within the Emo&Pain project. The purpose was to develop automated coaching systems capable of recognizing pain related affective states. A multimodal database consisting of facial expressions, vocal expression, body movement and muscle activity was created and labelled according to pain levels, and protective behaviour (e.g., fear of pain, pain, anxiety towards movement). Automatic recognition of expressions was attempted with SVM and they report a Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) of 0.658 ± 0.170 ($mean \pm std$) detection rate [40]. The performance of the automatic detection of protective behaviour varied according to the exercise being performed. The analysis also showed that protective behaviour is mostly expressed during the exercise whereas facial expressions are less frequent and generally expressed at the end of the exercise, possibly with a communication purpose. This suggest that body behaviour exhibited during physical rehabilitation is critical to assess the psychological state of the patient especially when dealing with chronic conditions rather than acute pain [41].

In stroke rehabilitation, Kan et al. [42] built a partially observable Markov decision process (POMDP) to modify exercise parameters for rehabilitation of the upper limb. The automatic estimation of patient fatigue was included in the decision model so that the robotic system adapt to the patient's specific needs. A therapist and a patient were recruited for the study. The POMDP decisions were in agreement with the therapist 65% of the time. Further, the patient reported satisfaction and interest in using the system. Bonarini et al [43] focused on studying stress during robotic rehabilitation of upper limbs in post-stroke patients. Biological signals, such as blood pressure, skin conductance, electrocardiogram (EKG), respiratory rate, electromyogram (EMG) and temperature were sensed and fed to a k-Nearest Neighbour ($k = 11$) classifier. Five levels of stress were discriminated. Data from 6 healthy people was clustered and 88.09% accuracy was reported.

Also within the rehabilitation domain, our group previously developed an adaptation module for a VR platform based on a Markov decision process (MDP) and reinforcement learning (RL) to adjust the therapy to the patient's progress [44]. Capabilities measurements based on patient speed to reach the game targets and the steadiness control of the upper limb while approaching game targets were monitored. The adaptation consisted in optimizing game challenge (suit game difficulty) according to patient's performance [2]. Congruence of model decisions against the therapist decisions was deemed high when the algorithm training period length was limited to realistic constraints of available event information. Although the decision model exhibited good performance, we believe that the system can be further improved if we incorporate the automatic detection of the patient's affective states. This work explores this research line, and represents our initial efforts to make the adaptation engine of the VR platform aware of the affective state of the patient. In particular, this study differs from the ones presented above on stroke, as we only used hand displacement and gripping pressure from the affected upper limb to infer the affective states, to avoid additional sensors not already available in the rehabilitation platform

as detailed below.

3 COLLECTING DATA OF SPONTANEOUS AFFECTIVE STATES OF FIVE POST-STROKE PATIENTS

3.1 Virtual Rehabilitation Platform: Gesture Therapy

Gesture Therapy (GT) [2], developed by our group, is a virtual rehabilitation (VR) alternative for helping post-stroke patients in their upper limbs rehabilitation exercises. Similar to other VR platforms, exercises are presented as serious games. Distinctively, some advantages of GT are portability (it only needs a computer with a web camera and an ad hoc controller referred to as gripper), it is inexpensive, and it contains artificial intelligence supporting adaptability (it has an adaptation module that considers user's performance to adjust the games with therapist supervision). During a regular rehabilitation session, GT gripper is hold with the paretic side hand to reach some game targets (see Fig. 1), and its 3D location is tracked using a fixed webcam. Depth is estimated exploiting previous knowledge of the size of the gripper's topping ball. The gripper also incorporates a force sensor on the front to sense gripping strength. If our hypothesis is plausible, we should be able to recover some of the aforementioned states capitalizing only on the 3D trace of hand displacements and on the pressure trace as detected by the gripper. Inference of these states may later be used to design empathic interfaces which adapt to the user's affective conditions.
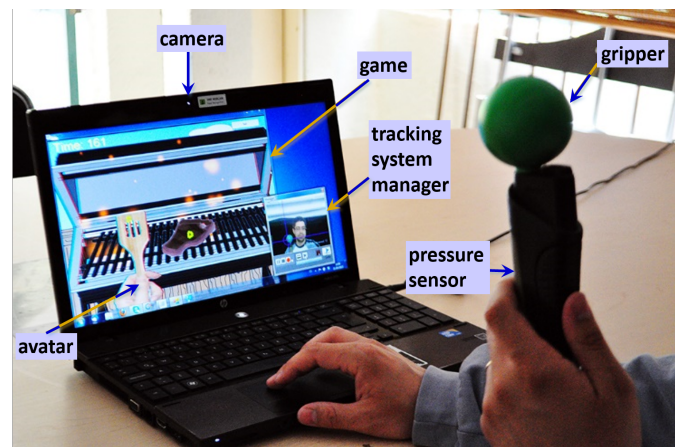

Fig. 1. The Gesture Therapy platform. The gripper, held with the left hand here, serves to control an avatar on the virtual environment (in this case, the hand with the kitchen palette on the screen). As the user interacts with the rehabilitation oriented games, the 3D location of the hand, and the gripping force are sensed and sent to the computer for processing the user's performance. Recognition of some affective states of the patient using only hand movements and fingers pressure should help to adjust the games challenges.

3.2 Patient Recruitment and Data Collection

We recorded data along longitudinal rehabilitation sessions from 5 post-stroke patients whilst interacting with GT. This data includes instantaneous hand location proxied by the gripper's ball, gripping strength and a frontal video. The stream of 3D coordinates of the hand motions and fingers pressure, sensed through the gripper, were taken as independent variables; and the state as labelled by psychiatrists

from the frontal video, as dependent variable. After the tagging process was developed, the inter-annotator agreement was computed. The set of states of interest were chosen by a group of experts in psychology, motor neurorehabilitation and affective computing. In this initial effort we are intentionally skipping emotion recognition from facial expression (available as it is from the video stream), which we aim to consider in future models.

Stroke survivor volunteers were recruited from the Instituto Nacional de Neurología y Neurocirugía (INNN) in Mexico. The demographic information is summarized in Table 1. Following brief explanations of the research, they give their consent to participate, and agreed that their rehabilitation sessions were video recorded and their motions and pressure data were used for scientific purposes.

TABLE 1
COHORT DEMOGRAPHICS

	P1	*P2*	*P3*	*P4*	*P5*
Age [years]	55	57	67	44	41
Gender	M	F	M	M	M
Stroke date	Apr, 2014	May, 2014	Jan, 2013	Dec, 2016	Nov, 2015
Therapy onset	May 8th, 2014	Sep 24th, 2014	Jul 5th, 2016	Jul 10th, 2017	Jul 7th, 2017
Paretic side	Left	Left	Right	Right	Right
# of sessions	6	10	10	8	5

Patient's data was recorded from a total of 10 rehabilitation sessions in a period of 4 weeks. Each session took place in a different day (max 3 per week) and each one was 45 mins long on average. Playing time is necessarily shorter; the rest of the time consisted of patient's welcome, therapist's instructions, stretching exercises and game switching. All the sessions were supervised by a qualified occupational therapist that had previous experience with the GT platform.

The platform automatically saves the following data through its tracking system while patients are interacting with the games:

- 3D coordinates of the coloured gripper ball, proxy for hand location at 15 Hz. Coordinates are referenced to camera position. The recovery of depth from monoscopic vision is achieved by exploiting previous knowledge of the gripper's ball size as described in [45].
- Gripping pressure exerted on the gripper frontal force sensor at 15 values per second (synchronized with the 3D coordinates of hand location).
- Frontal digital video at 15 frames per second. The video displays face expressions, hand movements and posture of the upper torso (see Fig. 2). Here, this information was only used for labelling the patient's affective states; but not used for classification purposes.

In addition, the participants were also asked to answer the Intrinsic Motivation Inventory (IMI) questionnaire [46], [47], [48], [49] at the end of every rehabilitation session.

Fig. 2. A patient during a rehabilitation session with GT platform during this feasibility study. Reproduced with permission [50].

3.3 Selection of states to monitor

A variety of sensorial, affective and cognitive phenomena can hinder physical rehabilitation in many conditions including stroke. A systematic review [9] provides a summary of the phenomena identified by the literature. Whilst addressing all of them in this study would not be possible, it was decided to address the ones that were directly related to physical activity sessions rather than general ones such as frustration due to lack of acceptance of the medical condition or general post-stroke traits (e.g., anxiety trait) and mood. We focus on affective and physical responses to physical activity as they occur, with the aim to build a system able to react to the in-the-moment person's needs. The set of states were chosen through discussion with a group formed by a therapist, psychiatrists and an affective computing expert involved in the project. Pain (*dolor* in Spanish), fatigue (*cansancio*) and anxiety (*ansiedad*) were considered primary factors that should be taken into account to personalize at run-time the level of the game or provide psychological support. Working in a state of pain and fatigue or anxiety does not only lead to an increase in the levels of those same states (e.g., due to increase muscle tension and hence to the movement control difficulty) but also to a reduction in therapy efficacy (e.g., due to use of compensatory movements) and in the long term desire to engage in the therapy. Anxiety could be due to the fear of not being capable of performing the task or of increase pain. Continue exposure to anxiety and pain may also contribute to the possibility to develop chronic pain [51]. The automatic detection of fatigue or pain could trigger adjustments in the game difficulty level or even suggest a break, whilst the detection of anxiety could trigger breathing exercise during the game or breaks. The group also recommended to track engagement as an affective-motivational state (*motivación* in Spanish) as it could provide guidelines of when to increase the task challenge or help identify patient's preferences.

3.4 Labelling

Three psychiatrists visually inspected the rehabilitation session videos. They were blind to each other and labelled consecutive video frames as intervals where they considered the patient had manifested tiredness, anxiety, pain or engagement. In addition, these raters had the patient's IMI answers for the corresponding session at hand.

It should be noted that we were not interested in anxiety traits as measured by standard questionnaire (e.g., STAI) or general daily levels of pain and fatigue but on tracking the fluctuation of the patient's affective and physical states

to enable personalized support when needed. Whilst the ability to recognize expressions of pain or anxiety in others is not an easy task, our long term aim is to build a system that has at least the ability of an expert observer to detect relevant states in a patient. We could have used continuous self-report from patients but this approach would have been disruptive and possibly led to increase anxiety. In addition, patients often report not to be aware of their anxiety until it is very high given the cognitive and physical effort required by the task. Finally, it was decided not to use a dimensional approach (e.g., valence/arousal or PANAS) as it would not be sufficiently specific to inform interventions as each state (e.g., pain, fatigue, anxiety, frustration, low mood) may need a different type of intervention.

For the labelling process, raters made use of software ELAN - Linguistic Annotator 4.7.0 [52] tagging the four states in different tiers as illustrated in Fig. 3.
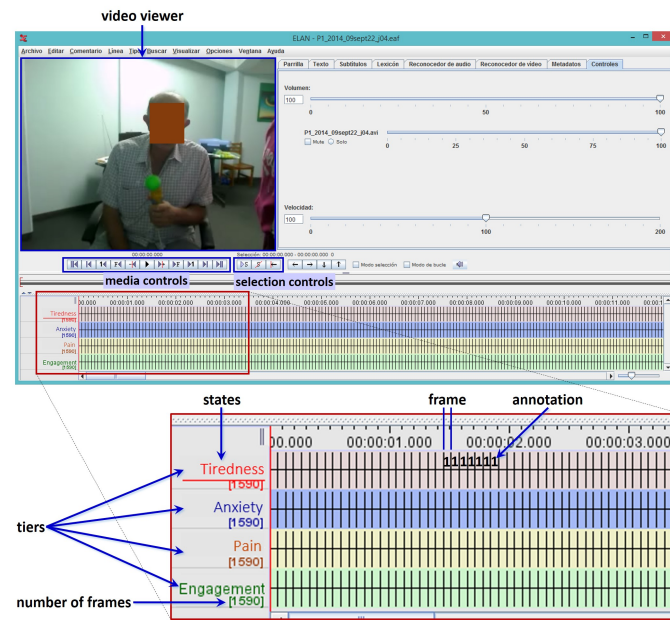

Fig. 3. Raters used ELAN to tag the interval of frames where (s)he considered the patient showed a state: tiredness, anxiety, pain and engagement. The videos were displayed in the upper left side (in the video viewer) and the rater had the media control buttons to play, stop, go backward or forward one frame, etc. The label lines or tiers are located in the whole lower side. A coloured tier identifies each state and these tiers were synchronized with video frames.

Every rater annotated with '1' in the tier of the respective state in the interval of frames where (s)he estimated the presence of the state. The system further filled with '-1' the rest of frames of that tier, considering that the rater was expressing that situation, the absence of the state. For every video frame, GT registered the associated respective 3D coordinates of the representative point of hand location and registered the pressure value at that instant. The final annotation for each frame was assigned as the majority agreement among the raters. Finally, the inter-rater agreement was calculated. When raters tagged a frame with the presence or absence of a state, the system retrieves the corresponding tag for 3D coordinates and pressure value.

4 DESIGNING THE FEATURE VECTOR

4.1 Feature Extraction

From the raw data collected, eight features were extracted to characterize the dynamic behaviour of the motions and pressures upon the gripper offline. GT recorded the stream video at 15Hz. The frame time in GT video $frT = \frac{1}{15}$ s. We denote 2 consecutive hand location points (3D coordinates in real world that were normalized to the interval [0,1]) as $p_i = (x_i, y_i, z_i)$ and $p_{i+1} = (x_{i+1}, y_{i+1}, z_{i+1})$, where $p_i, p_{i+1} \in [0,1]^3 \subset \mathbb{R}^3$. Also we denote P_i, the pressure recorded at frame i, $P_i \in [0,1] \subset \mathbb{R}$ (p_i, p_{i+1}, P_i are normalized). Table 2 presents the eight extracted features.

TABLE 2
EXTRACTED FEATURES CHARACTERIZING MOTOR RESPONSES

Feature name	Operationalization	
Hs: hand speed [*meters/second*]	$Hs_{i+1} = \frac{\sqrt{(x_{i+1}-x_i)^2+(y_{i+1}-y_i)^2+(z_{i+1}-z_i)^2}}{frT}$	(1)
Ha: hand acceleration [*meters/second*^{2}]	$Ha_{i+1} = \frac{|Hs_{i+1}-Hs_i|}{frT}$	(2)
Dx: differential location (distance travelled) of the hand along the x axis [*meters*]	$Dx_{i+1} = |x_{i+1} - x_i|$	(3)
Dy: differential location (distance travelled) of the hand along the y axis [*meters*]	$Dy_{i+1} = |y_{i+1} - y_i|$	(4)
Dz: differential location (distance travelled) of the hand along the z axis [*meters*]	$Dz_{i+1} = |z_{i+1} - z_i|$	(5)
AvP: average pressure [*kiloPascals*]	$AvP_{i+1} = \frac{(P_{i+1}+P_i)}{2}$	(6)
Ps: pressure speed [*kiloPascals/second*]	$Ps_{i+1} = \frac{|P_{i+1}-P_i|}{frT}$	(7)
Pa: pressure acceleration [*kiloPascals/second*^{2}]	$Pa_{i+1} = \frac{|Ps_{i+1}-Ps_i|}{frT}$	(8)

4.2 Feature Vector

The time series from hand varying location TrM (Trace of hand Movements) and fingers pressure TrP (Trace of Pressure) are collected synchronously. The consecutive labels assigned by the raters to the video frames can also be seen as a concomitant time series: TrL (Trace of Labels). It is possible to shift a window of length W on the series TrM and TrP and calculate the local values of the features described. It is convenient that the window has odd length so that it may be centred in a certain time sample p_i along these concurrent series. W represents the neighbourhood of p_i as depicted in Fig. 4. Shifting of W on TrM permit extracting the five features of hand motions (average speed, average acceleration and average of differential location along each axis: x, y and z), while doing so on TrP leave

us the features characterizing the exerted pressure (average pressure, average pressure speed and average pressure acceleration).
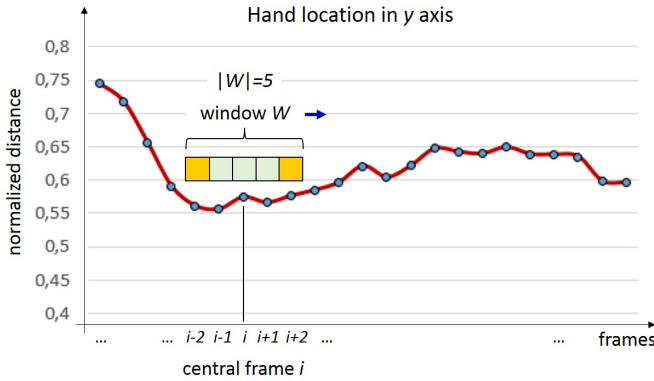

Fig. 4. Exemplification of the shifting of a window W of size 5 time samples centred around sample p_i. The example corresponds to changes in TrM along the y axis. For simplicity, we represented here the process with the hand movements only in y axis, but the actual TrM series is multivariate with x, y and z displacements.

Given a window sized $|W| = 2k + 1$, $k \in \mathbb{N}$ then we have $2k + 1$ consecutive points p_{i-k}, p_{i-k+1}, $...$, p_{i-1}, p_i, p_{i+1}, $...$, p_{i+k-1}, p_{i+k}. The windowed features are calculated as indicated in Table 3. Since at least 3 points are needed to calculate acceleration, the minimum window size is $|W| = 3$. All features represent averages in window W. Then the feature vector $\vec{F}_i$ for point p_i in W (using equations (10) to (17)) is:

$$\vec{F}_i = (AHs_i, AHa_i, ADx_i, ADy_i, ADz_i, AvP_i, APs_i, APa_i) \tag{9}$$

4.3 Assigning the Classes

For every feature vector, $\vec{F}_i$, there is an assignment of 4 binary values, each one for the respective class of every affective state. Independent classifiers are built for each affective state. Classes are binary and correspond to the presence or absence of the affective state being classified. Classes were therefore coded separately using the time series of the labels, TrL. Let $e \in E$, where $E = \{$tiredness, anxiety, pain, engagement$\}$, be any of the patient states; then shifting a window W centred at p_i over TrL give vectors of 4 binary labels, 1 or -1, corresponding to the presence or absence of each affective state as appreciated by the majority expert raters at each sample of the series. The final class labels for each affective state at p_i is the majority class among the labels (1 or -1) assigned by the raters to the frames covered by window W around the current point p_i.

Formally, for a window W such that $|W| = 2k + 1$, $k \in \mathbb{N}$, $k \geq 1$, centred around point p_i of the series TrL of an affective state $e \in E$, the class for affective state e in point p_i is given by:

$$class_{e,i} = \underset{t \in \{-1,1\}}{\arg\max} |\{j | W[i+j] = t \land j \in \mathbb{Z} \land -k \leq j \leq k\}| \tag{18}$$

TABLE 3
WINDOWED FEATURES OF W, $|W| = 2k + 1$, $k \in \mathbb{N}$

Feature name	Formula	
Average of hand speed in W (using Hs given by (1))	$AHs_i = \frac{\sum_{j=-k+1}^{k} Hs_{i+j}}{2k}$	(10)
Average of hand acceleration in W (using Ha given by (2))	$AHa_i = \frac{\sum_{j=-k+2}^{k} Ha_{i+j}}{2k-1}$	(11)
Average of differential location along the x axis for W (using Dx given by (3))	$ADx_i = \frac{\sum_{j=-k+1}^{k} Dx_{i+j}}{2k}$	(12)
Average of differential location along the y axis for W (using Dy given by (4))	$ADy_i = \frac{\sum_{j=-k+1}^{k} Dy_{i+j}}{2k}$	(13)
Average of differential location along the z axis for W (using Dz given by (5))	$ADz_i = \frac{\sum_{j=-k+1}^{k} Dz_{i+j}}{2k}$	(14)
Average pressure in W (as (6))	$AvP_i = \frac{\sum_{j=-k}^{k} P_{i+j}}{2k+1}$	(15)
Average of pressure speed in W (using Ps given by (7))	$APs_i = \frac{\sum_{j=-k+1}^{k} Ps_{i+j}}{2k}$	(16)
Average of pressure acceleration in W (using Pa given by (8))	$APa_i = \frac{\sum_{j=-k+2}^{k} Pa_{i+j}}{2k-1}$	(17)

5 MULTIRESOLUTION SEMI-NAÏVE BAYESIAN (MSNB) CLASSIFIER

5.1 Semi-Naïve Bayesian (SNB) Classifier

SNB classifier is based on Naïve Bayes classifier [53], a probabilistic classifier based on Bayes' theorem. Given a sample $s_a = (a_1, a_2, \cdots, a_n)$, the decision rule for a two class problem (class variable C takes values in $\{-1, 1\}$), A_i represents the ith attribute, is expressed as:

$$class(s_a) = \underset{c \in \{-1,1\}}{\arg\max}(prob(C = c) \prod_{i=1}^{n} prob(A_i = a_i | C = c)) \tag{19}$$

The product in (19) is based on the strong (naïve) assumption that all attributes A_i are independent given the class C [54]. To alleviate this assumption, the SNB classifier executes a structural improvement [54], [55], [56] to eliminate and/or to join attributes. The structural improvement, as illustrated in Fig. 5, is based on mutual information and conditional mutual information calculations [57]. After each structural modification operation, whether elimination or join, the new structure is tested to estimate whether classification accuracy is improved. The process is repeated until all attributes have been checked. Upon successful structural improvement, an enhanced feature representation is obtained.
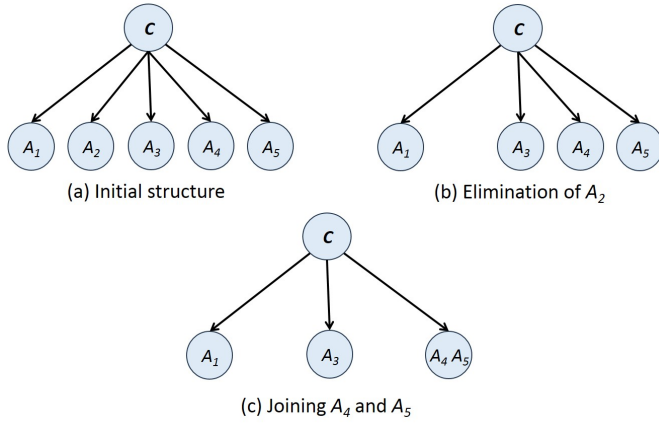
Fig. 5. Example of structural improvement method: (a) example of an original Naïve Bayes structure with 5 attributes (all the attributes are assumed independent), (b) Attribute A_2 is removed because the mutual information value between A_2 and the class C is insignificant, (c) Attributes A_4 and A_5 are joined into one, as they are considered dependent base upon the mutual information value between A_4 and A_5 given the class C.
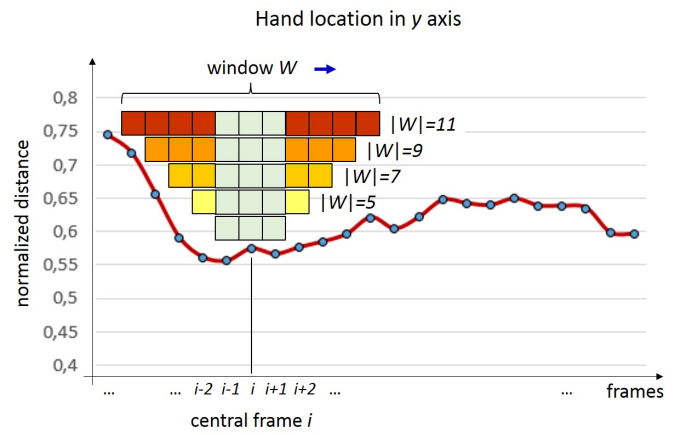

Fig. 6. Schematic depiction of the multiresolution process with five windows W sized 3, 5, 7, 9 and 11 respectively centred around p_i. Exemplification corresponds to TrM along y axis. A SNB classifier is trained for each window size to infer the presence (1) or absence (-1) of the affective state into consideration. Afterwards, each of the SNB models (5 SNB for this example) return a class label in each sample p_i of the series, and the MSNB assigns the final class label (1 or -1) to p_i by majority voting.

For handling continuous data inputs in the SNB classifier, we discretized feature values using the Proportional k-Interval Discretization (PKID) [37]. PKID has been suggested to be a suitable discretization alternative for Bayesian classifiers [37].

In a previous effort, SNB classifier was used with promising results [58], and hence we continue to employ it in this research for comparison purposes. Here, SNB classifiers were independently trained to predict absence or presence of each affective state, and independently for each patient, so we had a total number of models of (number of affective states) × (number of patients) × (number of window sizes) which is 100. Each one was a binary classifier (with classes -1 and 1), which received as inputs the discretized samples of the feature vector (expressed in (9)), calculated according to window size $|W|$ and the corresponding classes (defined in (18)) for these samples.

5.2 Multiresolution SNB Classifier

As aforementioned, emotions have a complex dynamics with non stationary episode length. Consequently, new to this effort, we propose MSNB classifier trying to detect the presence of some patient states (tiredness, anxiety, pain and engagement) during virtual rehabilitation, through a strategy of shifting a set of parallel windows W of different odd sizes along the TrM, TrP and TrL, but all concurrently centred around the same time sample p_i of these parallel time series. The rationale is to recognize the presence of an affective state with simultaneous estimators at different resolutions in p_i as schematically exemplified in Fig. 6.

MSNB has been designed using SNB as the basis classifier for each window of odd size. The purpose of each SNB is to infer the presence (1) or absence (-1) of the affective state (into consideration) at every sample p_i of these parallel series. MSNB calculates the SNB result for each window size and assigns the class to sample p_i, by majority voting of the SNBs (see Fig. 6). Although, MSNB was designed using SNB as the basis classifier, in principle other classifiers could

be used as the basis classifier instead of SNB under the multiresolution strategy. MSNB is indeed an ensemble of classifiers that tries to synthesize different detectors in one meta-recognizer.

Our data was recorded for each patient whilst (s)he was interacting with the virtual rehabilitation platform during the sessions and labelled by experts as previously described. The dataset contains the parallel time series from where are extracted the feature vectors and classes that are passed to train the MSNB classifier. Since the concurrent multiresolution windows W are centred around the same sample p_i, a group of underrepresented samples at the beginning and final of the whole series introducing a boundary bias. This situation occurs because the concentrical windows do not use some first or last points of the series. For example, in Fig. 6, window $|W| = 3$ do not use the first and last 4 points of the series. But this effect losses relevance as the time series length increases.

MSNB was independently trained for each affective state and for each patient, so we had as many MSNB models as (number of affective states) × (number of patients) which is 20. In Fig. 7, the operation of MSNB is charted. For every window size, three processes are executed: (i) feature extraction and class assignment, (ii) discretization and (iii) construction of a SNB model to infer the presence (1) or absence (-1) of an affective state. Afterwards, these SNB models yield their estimations for p_i, and the MSNB assigns the final class label (1 or -1) to p_i by majority voting.

5.3 Baseline Classifiers: Support Vector Machine (SVM) and Random Forests (RF)

Support vector machine (SVM) models served as baseline for concurrent validity against SNB and MSNB. The SVM with RBF kernel models ($K(x,y) = exp^{\gamma * <x-y, x-y>}$) were trained in Weka 3.8.1 [59] with the same number of models as SNB. The parameters were optimized through
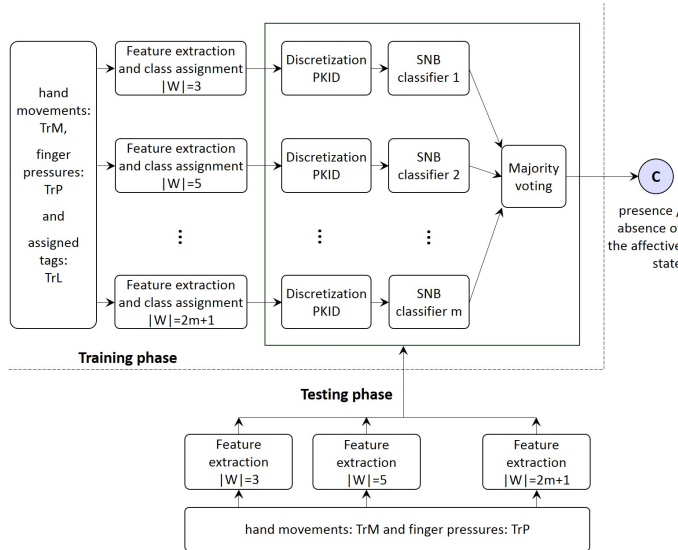
Fig. 7. Flowchart of MSNB classifier operations. Input data (hand location: TrM, gripping pressures: TrP and, in training phase, the corresponding affective states tags: TrL) were processed with windows of different old sizes. At each window size, features vectors are extracted and, for training phase, class labels are assigned. Afterwards, a discretization process transforms data for the SNB classifiers. The presence or absence of the affective state is estimated by SNB for each window size. Finally, MSNB assigns the final class (1 or -1) to each point p_i by majority voting.

grid search for each patient, affective state and window size. Linear SVM models ($K(x, y) =< x, y >$) were trained too with the same number of models as SNB, and the parameters (c and $epsilon$) were optimized using grid search for each patient, affective state and window size, but the results were lower than SVM with RBF kernel.

Random forests (RF) [60] models were also used as baseline for comparing against SNB and MSNB. In the same way, the RF models were trained in Weka 3.8.1 with the same number of models as SNB. The number of trees was changed with 10, 100 and 300 instances, trying to optimize the parameters for the model of each patient, affective state and window size.

5.4 Metrics

Internal validity of the SVM, SNB, RF and MSNB classifier models was established using the stratified 10 cross folding replication mechanism across all the rehabilitation sessions. Three classical metrics associated with the confusion matrix (accuracy, F-measure and ROC AUC[1]) were used to evaluate the results of each classification model and to compare them.

6 EXPERIMENTAL RESULTS

We intended to record ten sessions from each patient ($P1$, $P2$, $P3$, $P4$ and $P5$) over a four weeks period. Patients $P2$ and $P3$ attended all the 10 sessions while patient $P1$ only

---

1. TP: true positive, TN: true negative, FP: false positive and FN: false negative; accuracy = (TP+TN)/(TP+FP+TN+FN); sensitivity = TP/(TP+FN); specificity = TN/(TN+FP); precision = TP/(TP+FP); F-measure = 2(precision*sensitivity)/(precision+sensitivity); ROC AUC = (sensitivity + specificity)/2.

attended 6 sessions, patients $P4$ attended 8 sessions, and $P5$ attended 5 sessions respectively. The platform version of GT used for this study includes a set of 5 serious games: steak, fly_killers, clean_window, wash_dishes and spider [2]. Pressing was only needed for fly_killers. A video clip per game was retrieved for every patient during his/her playing time. Each video clip lasted about 1 minute and 10 seconds. During regular virtual rehabilitation sessions, each game is played for up to 3 minutes approximately as dictated by the therapist. The games are switched frequently to avoid boredom. In each rehabilitation session, the patients played all the 5 games, so 5 video clips were obtained per session, with an exception of one session of $P1$ in which he only played 4 games. For this reason, $P1$ had 29 (=(6x5)-1) video clips, $P2$ and $P3$ had 50 (=10x5) video clips, $P4$ had 40 (=8x5) video clips, and $P5$ had 25 (=5x5) video clips. The video clips of interest for this study were the fly_killers videos because they had critical information about the pressure exerted by the fingers. Note that any presence of pressure in the other games, or even in this game but out of timing, can be regarded as hints of affective interactions but we decided to skip this at this time. From the video clips for fly_killers game along all sessions, the total number of samples i.e. feature vectors, were 5826 for $P1$, 8935 for $P2$, 7334 for $P3$, 6068 for $P4$ and 3814 samples for $P5$, respectively.

Fleiss' κ [61] was run to determine agreement between the three raters (psychiatrists) regarding the assessment of the affective states independently in the video clips. Fleiss' κ were 0.6107 for tiredness, 0.2271 for anxiety, 0.2420 for pain, and 0.4607 for engagement. This suggests (according with [62]) substantial agreement for tiredness, moderate agreement for engagement and fair agreement for anxiety and pain. Indeed, despite their experience, raters particularly expressed their concerns about tagging anxiety through a short video due to its complexity.

During the labelling process, pain was not appreciated by the raters for patient $P2$ and therefore there are no results for this state associated for $P2$.

6.1 Total Number of Models

Five levels were considered for multiresolution. Feature vectors timecourses were processed at window sizes $|W| = 3, 5, 7, 9, 11$ obtaining the corresponding patterns covering a range from 0.2 seconds to 0.73 seconds temporal windows. A different SNB model for each state (4: tiredness, anxiety, pain and engagement) except for $P2$ that we had 3 states (3: tiredness, anxiety and engagement), and each window value (3, 5, 7, 9, 11) was developed for a total of 20 SNB models for $P1$, $P3$, $P4$ and $P5$, and 15 SNB models for $P2$. In the same way, 20 SVM models and 20 RF models for $P1$, $P3$, $P4$ and $P5$, and 15 SVM and 15 RF for $P2$ were obtained. There were MSNB models for each patient and for each affective state, in total 4 MSNB models for $P1$, $P3$, $P4$ and $P5$, and 3 MSNB models for $P2$.

6.2 Results

Table 4 summarizes the classification results of SVM with RBF kernel, SNB, RF and MSNB for all the patients: $P1$, $P2$, $P3$, $P4$ and $P5$, respectively, for the 4 states (3 states for $P2$).

TABLE 4

CLASSIFICATION RESULTS ($\mu \pm \sigma$) FOR SVM WITH RBF KERNEL, SNB, RF AND MSNB ACROSS THE 10 FOLDS (AND ACROSS THE FIVE WINDOW SIZES $|W|$=3,5,7,9,11 FOR SVM WITH RBF KERNEL, SNB AND RF). AVERAGE RESULTS OF ALL THE CONSIDERED STATES OF A PATIENT ARE SHOWN IN THE LAST ROW OF THAT PATIENT. THE BEST RESULTS FOR CLASSIFIER AND ROC AUC AVERAGE ARE SHOWN IN BOLDFACE TYPE.

Patient state	Patient P1 accuracy	F-measure	ROC AUC	Patient P2 accuracy	F-measure	ROC AUC	Patient P3 accuracy	F-measure	ROC AUC	Patient P4 accuracy	F-measure	ROC AUC	Patient P5 accuracy	F-measure	ROC AUC
SVM with RBF kernel															
tiredness	0.910±0.096	0.891±0.120	0.906±0.099	**0.888±0.100**	**0.875±0.118**	**0.886±0.100**	0.704±0.061	0.690±0.073	0.704±0.061	0.779±0.059	0.785±0.060	0.778±0.059	0.708±0.082	0.683±0.115	0.709±0.081
anxiety	0.864±0.089	0.836±0.111	0.860±0.091	0.754±0.124	0.739±0.140	0.754±0.124	0.717±0.053	0.754±0.048	0.708±0.053	0.753±0.063	0.747±0.071	0.753±0.063	0.723±0.084	0.737±0.067	0.723±0.085
pain	0.897±0.106	0.847±0.162	0.883±0.121	–	–	–	0.838±0.094	0.868±0.075	0.831±0.114	0.812±0.096	0.848±0.077	0.803±0.108	0.885±0.170	0.900±0.149	0.885±0.184
engagement	**0.929±0.031**	**0.918±0.034**	**0.934±0.028**	0.710±0.074	0.718±0.069	0.710±0.074	0.688±0.071	0.691±0.071	0.689±0.071	0.744±0.073	0.751±0.064	0.744±0.073	0.728±0.069	0.680±0.102	0.722±0.072
average	0.900±0.088	0.873±0.120	**0.896±0.095**	0.784±0.126	0.777±0.132	**0.783±0.126**	0.737±0.092	0.751±0.099	**0.733±0.097**	0.772±0.078	0.783±0.079	**0.770±0.081**	0.761±0.130	0.750±0.143	**0.760±0.135**
SNB															
tiredness	0.907±0.100	0.925±0.075	0.896±0.116	**0.864±0.120**	**0.877±0.107**	**0.862±0.119**	0.628±0.109	0.672±0.076	0.629±0.109	**0.719±0.116**	**0.708±0.127**	**0.719±0.115**	0.643±0.064	0.676±0.048	0.645±0.062
anxiety	0.846±0.139	0.874±0.104	0.842±0.144	0.639±0.063	0.710±0.053	0.625±0.065	0.682±0.072	0.630±0.103	0.677±0.071	0.639±0.087	0.664±0.097	0.639±0.086	0.701±0.126	0.710±0.129	0.702±0.125
pain	**0.924±0.149**	**0.928±0.161**	**0.919±0.156**	–	–	–	**0.890±0.118**	**0.881±0.130**	**0.902±0.108**	0.706±0.094	0.621±0.117	0.696±0.087	**0.927±0.151**	**0.935±0.136**	**0.938±0.128**
engagement	0.901±0.164	**0.928±0.114**	0.889±0.186	0.579±0.066	0.619±0.066	0.579±0.065	0.572±0.056	0.626±0.062	0.571±0.056	0.600±0.102	0.635±0.102	0.601±0.101	0.653±0.139	0.706±0.119	0.644±0.145
average	0.895±0.142	0.914±0.119	**0.887±0.154**	0.694±0.151	0.735±0.133	**0.689±0.152**	0.693±0.152	0.702±0.142	**0.695±0.154**	0.666±0.111	0.657±0.116	**0.664±0.108**	0.731±0.169	0.756±0.153	**0.732±0.170**
RF															
tiredness	0.897±0.099	0.870±0.135	0.891±0.106	**0.825±0.105**	**0.801±0.143**	**0.823±0.107**	0.729±0.075	0.724±0.076	0.729±0.075	0.777±0.061	0.788±0.059	**0.777±0.061**	0.724±0.070	0.720±0.071	0.724±0.071
anxiety	0.840±0.079	0.810±0.098	0.835±0.081	0.720±0.079	0.696±0.084	0.718±0.079	0.740±0.069	0.772±0.058	0.734±0.071	0.749±0.057	0.744±0.056	0.749±0.058	0.730±0.086	0.737±0.082	0.730±0.086
pain	0.828±0.148	0.741±0.249	0.816±0.163	–	–	–	**0.828±0.096**	**0.855±0.086**	**0.825±0.104**	**0.782±0.102**	**0.826±0.084**	0.765±0.111	**0.752±0.241**	**0.778±0.216**	**0.753±0.260**
engagement	**0.923±0.029**	**0.910±0.033**	**0.927±0.029**	0.737±0.064	0.739±0.061	0.737±0.064	0.718±0.085	0.720±0.083	0.718±0.085	0.753±0.075	0.758±0.073	0.752±0.075	0.749±0.075	0.719±0.087	0.745±0.076
average	0.872±0.105	0.833±0.163	**0.867±0.114**	0.761±0.096	0.745±0.110	**0.759±0.096**	0.754±0.092	0.768±0.094	**0.752±0.095**	0.765±0.077	0.779±0.075	**0.761±0.079**	0.739±0.137	0.739±0.130	**0.738±0.146**
MSNB															
tiredness	**0.925±0.114**	**0.945±0.079**	0.913±0.137	**0.929±0.082**	**0.935±0.074**	**0.929±0.082**	0.664±0.116	0.715±0.072	0.666±0.115	0.752±0.118	**0.743±0.123**	0.752±0.117	0.698±0.074	0.738±0.046	0.702±0.069
anxiety	0.905±0.133	0.926±0.095	0.900±0.139	0.674±0.075	0.760±0.043	0.655±0.083	0.741±0.060	0.692±0.079	0.733±0.062	0.685±0.082	0.719±0.076	0.684±0.082	0.771±0.115	0.786±0.102	0.772±0.115
pain	0.922±0.166	0.930±0.164	0.921±0.173	–	–	–	**0.950±0.097**	**0.951±0.091**	**0.958±0.081**	**0.788±0.097**	0.727±0.112	**0.780±0.091**	**0.967±0.105**	**0.967±0.105**	**0.975±0.079**
engagement	0.916±0.171	**0.944±0.112**	0.902±0.200	0.613±0.089	0.670±0.076	0.614±0.088	0.601±0.074	0.673±0.064	0.600±0.074	0.636±0.130	0.676±0.127	0.637±0.129	0.676±0.157	0.742±0.115	0.664±0.167
average	0.917±0.142	0.936±0.113	**0.909±0.158**	0.738±0.160	0.788±0.129	**0.733±0.164**	0.739±0.158	0.757±0.136	**0.739±0.159**	0.715±0.120	0.716±0.110	**0.713±0.117**	0.778±0.161	0.808±0.132	**0.778±0.164**

Results are summarized as mean and standard deviation (STD) across the 10 folds (and across the five window sizes $|W|$=3,5,7,9,11 for SVM with RBF kernel, SNB and RF).

For each classifier, recognition rates varied across patients, but in general, the predictive relations got higher discriminative capacity for tiredness and pain. The results for tiredness are consistent with the highest agreement of psychiatrists in Fleiss' κ for this state. In contrast, anxiety was not the best classified for any patient and any classifier, and this is consistent with the raters' opinion.

Results for $P1$ were higher than the other patients. Psychiatrists that participated in this study considered that $P1$ is an extrovert subject, and this situation could have helped in the labelling process and could have an influence on these results.

Average ROC AUC results were higher for MSNB for $P1$ and $P5$, but SVM with RBF kernel got the best average ROC AUC results for patients $P2$ and $P4$; and RF had the better average for $P3$.

The results in Table 4 reveal that all the computational models learned predictive relations with ROC AUC ≥ 0.60, except for $P2$ and $P3$, in engagement using SNB whose results were ≥ 0.57.

Group comparisons of ROC AUC results of the 10-fold cross validation for each classifier (SVM with RBF kernel, SNB, RF and MSNB) with respect to each patient and each affective state were performed using Friedman test at 5% significance level. Post-hoc paired comparisons (SVM vs SNB, SVM vs RF, SVM vs MSNB, SNB vs RF, SNB vs MSNB and RF vs MSNB) were performed using Wilcoxon signed-rank tests with Bonferroni correction threshold (significance level: $p < 0.008$). It is important to highlight that we obtained 5826 samples for $P1$, 8935 samples for $P2$, 7334 for $P3$, 6068 for $P4$ and 3814 samples for $P5$ of hand movements and fingers pressure associated with affective states to test the classifiers.

Statistically significant differences in ROC AUC were found between the four classifiers, Friedman: $\chi^2(3) = 84.608$; $p < 0.05$. The post hoc tests report is presented in Table 5 and the differences were significant when SNB classifier was compared with any other of the three remaining classifiers. These results suggest that either MSNB, SVM with RBF kernel or RF may be adequate choices without significant differences in their performances as determined by the ROC AUC.

These results are promising as they reveal a mechanism

TABLE 5
STATISTICAL SIGNIFICANCE ANALYSIS USING FRIEDMAN TEST AND WILCOXON SIGN RANK POST-HOC. SIGNIFICANT P VALUES ARE HIGHLIGHTED IN BOLD.

Friedman: $\chi^2(3) = 84.608$, p < 0.05

Post-Hoc Wilcoxon Sign Rank:

	SVM vs SNB	SVM vs RF	SVM vs MSNB	SNB vs RF	SNB vs MSNB	RF vs MSNB
W	-5.789	-1.113	-0.941	-4.396	-9.998	-0.162
p	< 0.008	0.266	0.347	< 0.008	< 0.008	0.871

(patterns of hand motions) that could help estimating non-basic affective states for some stroke patients in the scope of virtual rehabilitation.

7 DISCUSSION

The proposed computational model, MSNB, was competitive with SVM and RF. The results give evidence that the information at several scales of time (multiresolution strategy) can help to estimate the aforesaid states in time course. The multiresolution idea can be implemented using another classifier as base classifier, for example, we can use SVM as base classifier to obtain MSVM (Multiresolution using SVM), but the advantage of using SNB classifier as base classifier is that SNB is simpler than SVM with RBF kernel.

Comparison of SNB and RF indicates that RF got significant higher ROC AUC values than SNB classifier, in this work. However, when SNB operates as base classifier at the multiresolution strategy, the new classifier MSNB is competitive with RF.

The research of Aung et al. (2014) [40] that addressed the problem of automatic recognizing of affective states in rehabilitation of patients with chronic pain has important elements related with this work. They included the labelling process of pain observed in videos of patients, but the tags were continuous values in the interval [0,1], although afterwards they converted them to binary and used SVM to build the classification model. So, in a certain way, the work of Aung et al. (2014) served as reference for our results. As it was presented, the MSNB can have better results than SVM in the estimation of pain.

In Bonarini et al. (2010) [43], KNN classifier was used and they obtained 0.88 of accuracy detecting stress in post-stroke patients, but they employed obstructive sensors for registering various biological signals, as blood volume pressure, electrocardiogram and respiration, among others. An important issue of our investigation is to use non-obstructive measurements for facilitating the system to be used in daily rehabilitation at home. Another necessity for our purpose is to avoid expensive devices thus the system can be accessible to low and middle income people. Robotic systems are employed in [42], [43] but these systems can be expensive.

The aforementioned works focused in one affective state, and this investigation addressed the problem of four states that were recommended for rehabilitation of post-stroke patients by therapists, psychiatrists and an affective computing expert.

Unsurprisingly, anxiety was consistently the most difficult state to identify. This is in line with the experts remarks about the challenge to label this affective state and their apparent disagreement during the labelling.

According to experts who participated in this study, the labelling process using only the frontal video of the patients and IMI is complex, particularly for anxiety and pain. Patient $P1$ has characteristics that allowed the experts to make better discrimination in the labelling process and his movements allowed all classifiers achieve favourable results for his states. The argument that arose was that $P1$ was more spontaneous and variable in his expressions than the other patients. Personal differences can affect the results, for that reason it is necessary to get a good labelling and to build models tailored to each patient.

In rehabilitation, pain is a particularly critical state; painful exercises may be harmful to patients' recovery. The patients involved in this study reported low levels of pain using the intrinsic motivation inventory (IMI) questionnaire at the end of each session.

Gesture Therapy is an accessible virtual rehabilitation platform that is designed low-cost for intended use in low and middle income countries. The gripper and accompanying software can be installed in any regular low specification computer. Incorporating the recognition of affective states to the therapy plan through its adaptation engine, it could further enhance the tailoring of rehabilitation tasks to the patient needs. This would represent an important technological advance for the system and its users.

8 CONCLUSIONS

This study confirmed the feasibility to estimate the presence of four states in patients under upper extremity rehabilitation by means of classification models that only use as inputs the information of their hand motions and fingers pressure (non-obstructive measurements that can be sensed in daily activities) to a degree and which could be useful for later leverage in adaptive systems. This possibility was enhanced with the proposed computational model that tries to detect more precisely the manifestation of the states by integrating different temporal resolutions simultaneously. More specifically, we have obtained predictive models for decoding specific states from gripping measurements (hand motions and fingers pressure) of 5 post-stroke patients while they were interacting with a virtual rehabilitation system. These results suggest that at least tiredness and pain are susceptible of exploitable classification from observable data streams. The results are promising because they reveal a mechanism of affective state recognition in virtual rehabilitation scopes and this is useful for the adaptation of the system to the patient. Anxiety recovery was also above the random decision levels. Nevertheless, anxiety and engagement may require more aggressive models before satisfactory recognition levels can be claimed. A larger trial should confirm whether this apparent tendency can be further generalized to the population.

The dynamic problem of recognizing affective states through hand movements and fingers pressure was treated

with static machine learning classifiers using a feature vector which includes averages of speed and acceleration of the movements and pressures in the time course (the dynamic behaviour). Considering hand movements and fingers pressure with this feature vector is an important contribution of this work. The other contribution is the proposed computational model, MSNB, that for this problem is competitive with SVM and RF.

At the present, patient's observable performance is the only factor to adjust game difficulty in GT. The adjustment of the game difficulty levels could be done with the combination of performance indicators and detected affective states.

As part of future work, we consider to improve MSNB performance implementing other strategies in the base classifier SNB as changing the calculation of mutual information. We also consider exploit transfer learning strategies to migrate population-based models to specific patient-based models. Another aspect is to study obstrusive and non-obstrusive measurements of observable surrogates to estimate affective states in serious games environments; and analyze other affective states that were suggested by specialists for future studies, such as stress and depression. We are also beginning to study multi-classification models.

ACKNOWLEDGMENTS

REFERENCES

[1] V. L. Feigin, C. M. Lawes, D. A. Bennett, S. L. Barker-Collo, and V. Parag, "Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review," *The Lancet Neurology*, vol. 8, no. 4, pp. 355–369, April 2009.

[2] L. E. Sucar, F. Orihuela-Espina, R. L. Velazquez, D. J. Reinkensmeyer, R. Leder, and J. Hernández-Franco, "Gesture therapy: An upper limb virtual reality-based motor rehabilitation platform," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 22, no. 3, pp. 634–643, 2014.

[3] C. E. Lang, J. R. MacDonald, and C. Gnip, "Counting repetitions: an observational study of outpatient therapy for people with hemiparesis post-stroke," *Journal of Neurologic Physical Therapy*, vol. 31, no. 1, pp. 3–10, 2007.

[4] J. W. Krakauer, S. T. Carmichael, D. Corbett, and G. F. Wittenberg, "Getting neurorehabilitation right what can be learned from animal models?" *Neurorehabilitation and neural repair*, vol. 26, no. 8, pp. 923–931, 2012.

[5] M. K. Holden, "Virtual environments for motor rehabilitation: review," *Cyberpsychology & behavior*, vol. 8, no. 3, pp. 187–211, 2005.

[6] G. Saposnik, M. Levin, S. O. R. C. S. W. Group *et al.*, "Virtual reality in stroke rehabilitation a meta-analysis and implications for clinicians," *Stroke*, vol. 42, no. 5, pp. 1380–1386, 2011.

[7] R. Colombo, F. Pisano, A. Mazzone, C. Delconte, S. Micera, M. C. Carrozza, P. Dario, and G. Minuco, "Design strategies to improve patient motivation during robot-aided rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 4, no. 1, pp. 3–15, 2007.

[8] M. M. Bradley, M. Codispoti, B. N. Cuthbert, and P. J. Lang, "Emotion and motivation I: defensive and appetitive reactions in picture processing." *Emotion*, vol. 1, no. 3, pp. 276–298, 2001.

[9] J. Luker, E. Lynch, S. Bernhardsson, L. Bennett, and J. Bernhardt, "Stroke survivors' experiences of physical rehabilitation: a systematic review of qualitative studies," *Archives of physical medicine and rehabilitation*, vol. 96, no. 9, pp. 1698–1708, 2015.

[10] "Pain terms: a list with definitions and notes on usage. recommended by the IASP subcommittee on taxonomy." *Pain*, vol. 6, no. 3, pp. 249–249, 1979.

[11] G. E. DAniello, F. Scarpina, A. Mauro, I. Mori, G. Castelnuovo, M. Bigoni, S. Baudo, and E. Molinari, "Characteristics of anxiety and psychological well-being in chronic post-stroke patients," *Journal of the neurological sciences*, vol. 338, no. 1, pp. 191–196, 2014.

[12] F. Orihuela-Espina and L. E. Sucar, "Adaptation and customization in virtual rehabilitation." IGI Global, 2016, pp. 141–163.

[13] T. U. Nguyen-Oghalai, K. J. Ottenbacher, C. V. Granger, and J. S. Goodwin, "Impact of osteoarthritis on the rehabilitation of patients following a stroke," *Arthritis Care & Research*, vol. 53, no. 3, pp. 383–387, 2005.

[14] D. Cernea and A. Kerren, "A survey of technologies on the rise for emotion-enhanced interaction," *Journal of Visual Languages & Computing*, vol. 31, pp. 70–86, 2015.

[15] R. W. Picard, "Affective computing: challenges," *International Journal of Human-Computer Studies*, vol. 59, no. 1, pp. 55–64, 2003.

[16] B. Romera-Paredes, M. S. Aung, M. Pontil, N. Bianchi-Berthouze, A. C. d. C. Williams, and P. Watson, "Transfer learning to account for idiosyncrasy in face and body expressions," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–6.

[17] P. Verduyn, P. Delaveau, J.-Y. Rotgé, P. Fossati, and I. Van Mechelen, "Determinants of emotion duration and underlying psychological and neural mechanisms," *Emotion Review*, vol. 7, no. 4, pp. 330–335, 2015.

[18] K. R. Scherer, "What are emotions? and how can they be measured?" *Social science information*, vol. 44, no. 4, pp. 695–729, 2005.

[19] K. R. Scherer and P. Ekman, *Approaches to emotion*. Psychology Press, 2014.

[20] M. Cabanac, "What is emotion?" *Behavioural processes*, vol. 60, no. 2, pp. 69–83, 2002.

[21] P. R. Kleinginna Jr and A. M. Kleinginna, "A categorized list of emotion definitions, with suggestions for a consensual definition," *Motivation and emotion*, vol. 5, no. 4, pp. 345–379, 1981.

[22] H. Meng, N. Bianchi-Berthouze, Y. Deng, J. Cheng, and J. P. Cosmas, "Time-delay neural network for continuous emotional dimension prediction from facial expression sequences," *IEEE transactions on cybernetics*, vol. 46, no. 4, pp. 916–929, 2016.

[23] A. Savran, H. Cao, A. Nenkova, and R. Verma, "Temporal Bayesian fusion for affect sensing: Combining video, audio, and lexical modalities," *IEEE transactions on cybernetics*, vol. 45, no. 9, pp. 1927–1941, 2015.

[24] M. Wöllmer, F. Eyben, S. Reiter, B. W. Schuller, C. Cox, E. Douglas-Cowie, R. Cowie *et al.*, "Abandoning emotion classes-towards continuous emotion recognition with modelling of long-range dependencies." in *Interspeech*, vol. 2008, 2008, pp. 597–600.

[25] M. Yeasin, B. Bullot, and R. Sharma, "From facial expression to level of interest: a spatio-temporal approach," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2. IEEE, 2004, pp. II–II.

[26] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences: temporal and static modeling," *Computer Vision and image understanding*, vol. 91, no. 1, pp. 160–187, 2003.

[27] X. Li and Q. Ji, "Active affective state detection and user assistance with dynamic Bayesian networks," *IEEE transactions on systems, man, and cybernetics-part a: systems and humans*, vol. 35, no. 1, pp. 93–105, 2005.

[28] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1113–1133, 2015.

[29] A. T. Sohaib, S. Qureshi, J. Hagelbäck, O. Hilborn, and P. Jerčić, "Evaluating classifiers for emotion recognition using EEG," in *International Conference on Augmented Cognition*. Springer, 2013, pp. 492–501.

[30] F. Eyben, S. Petridis, B. Schuller, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Audiovisual classification of vocal outbursts in human conversation using long-short-term memory networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 5844–5847.

[31] T. L. Nwe, S. W. Foo, and L. C. De Silva, "Speech emotion recognition using hidden Markov models," *Speech communication*, vol. 41, no. 4, pp. 603–623, 2003.

[32] Z. Zeng, J. Tu, B. Pianfetti, M. Liu, T. Zhang, Z. Zhang, T. S. Huang, and S. Levinson, "Audio-visual affect recognition through multi-stream fused HMM for HCI," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 967–972.

[33] G. A. Ramirez, T. Baltrušaitis, and L.-P. Morency, "Modeling latent discriminative dynamic of multi-dimensional affective signals," in *Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 396–406.

[34] T. Baltrušaitis, N. Banda, and P. Robinson, "Dimensional affect recognition using continuous conditional random fields," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–8.

[35] D. J. Hand and K. Yu, "Idiot's Bayesnot so stupid after all?" *International statistical review*, vol. 69, no. 3, pp. 385–398, 2001.

[36] J. Cheng and R. Greiner, "Comparing Bayesian network classifiers," in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1999, pp. 101–108.

[37] Y. Yang and G. I. Webb, "Proportional k-interval discretization for naive-Bayes classifiers," in *12th European Conference on Machine Learning (ECML)*. Springer, 2001, pp. 564–575.

[38] D. Novak, M. Mihelj, and M. Munih, "A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing," *Interacting with computers*, vol. 24, no. 3, pp. 154–172, 2012.

[39] L. E. Sucar, *Probabilistic Graphical Models*. Springer, 2015.

[40] M. Aung, S. Kaltwang, B. Romera-Paredes, B. Martinez, A. Singh, M. Cella, M. Valstar, H. Meng, A. Kemp, M. Shafizadeh, A. Elkins, N. Kanakam, A. Rothschild, N. Tyler, P. Watson, A. Williams, M. Pantic, and N. Bianchi-Berthouze, "The automatic detection of chronic pain-related expression: requirements, challenges and a multimodal dataset," 2015.

[41] P. OSullivan, "Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanism," *Manual therapy*, vol. 10, no. 4, pp. 242–255, 2005.

[42] P. Kan, R. Huq, J. Hoey, R. Goetschalckx, and A. Mihailidis, "The development of an adaptive upper-limb stroke rehabilitation robotic system," *Journal of Neuroengineering and Rehabilitation*, vol. 8, no. 1, pp. 1–18, 2011.

[43] A. Bonarini, M. Garbarino, M. Matteucci, and S. Tognetti, "Affective evaluation of robotic rehabilitation of upper limbs in post-stroke subjects," in *Proceedings 1st Workshop on the Life Sciences at Politecnico di Milano*, 2010, pp. 290–293.

[44] S. Ávila-Sansores, F. Orihuela-Espina, and L. Enrique-Sucar, "Patient tailored virtual rehabilitation," in *Converging Clinical and Engineering Research on Neurorehabilitation*. Springer, 2013, pp. 879–883.

[45] L. E. Sucar, A. Molina, R. Leder, J. Hernández, and I. Sánchez, "Gesture therapy: a clinical evaluation," in *2009 3rd International Conference on Pervasive Computing Technologies for Healthcare*. IEEE, 2009, pp. 1–5.

[46] R. M. Ryan and E. L. Deci, "Self-Determination theory organization, intrinsic motivation inventory (IMI)," url http://www.selfdeterminationtheory.org/intrinsic-motivation-inventory/, 2015, [Web; accessed on 09/08/2015].

[47] R. M. Ryan and E. Deci, "Intrinsic and extrinsic motivations: Classic definitions and new directions," *Contemporary Educational Psychology*, vol. 25, no. 1, pp. 54–67, 2000.

[48] J. Choi, T. Mogami, and A. Medalia, "Intrinsic motivation inventory: an adapted measure for schizophrenia research," *Schizophrenia Bulletin*, pp. 1–11, 2009.

[49] C. t. D. M. Fisher, "What is intrinsic motivation inventory?" url http://yourbusiness.azcentral.com/intrinsic-motivation-inventory-25561.html, [Web; accessed on 09/08/2015].

[50] J. J. Rivas, F. Orihuela-Espina, L. E. Sucar, L. Palafox, J. Hernández-Franco, and N. Bianchi-Berthouze, "Detecting affective states in virtual rehabilitation," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2015 9th International Conference on*. IEEE, 2015, pp. 287–292.

[51] C. R. Chapman and C. J. Vierck, "The transition of acute postoperative pain to chronic pain: an integrative overview of research on mechanisms," *The Journal of Pain*, vol. 18, no. 4, pp. 359–e1, 2017.

[52] B. Hellwig, D. Van Uytvanck, and M. Hulsbosch, "ELAN - Linguistic Annotator, version 4.7.0," *Manual. Online publication http://www.mpi.nl/corpus/manuals/manual-elan.pdf*, 2014. [Online]. Available: https://tla.mpi.nl/tools/tla-tools/elan/

[53] R. O. Duda, P. E. Hart *et al.*, *Pattern classification and scene analysis*. Wiley New York, 1973, vol. 3.

[54] M. J. Pazzani, "Searching for dependencies in Bayesian classifiers," in *Learning from Data*. Springer, 1996, pp. 239–248.

[55] M. Martínez-Arroyo and L. E. Sucar, "Learning an optimal naive Bayes classifier," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3. IEEE, 2006, pp. 1236–1239.

[56] M. Martínez Arrollo, "Aprendizaje de clasificadores bayesianos estáticos y dinámicos," Ph.D. dissertation, Instituto Tecnológico y de Estudios Superiores de Monterrey, 2007.

[57] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *Information Theory, IEEE Transactions on*, vol. 14, no. 3, pp. 462–467, 1968.

[58] J. J. Rivas, P. Heyer, F. Orihuela-Espina, and L. E. Sucar, "Towards incorporating affective computing to virtual rehabilitation; surrogating attributed attention from posture for boosting therapy adaptation," in *Tenth International Symposium on Medical Information Processing and Analysis*. International Society for Optics and Photonics, 2015, pp. 92 870Y–1–92 870Y–8.

[59] E. Frank, M. Hall, and I. Witten, "The WEKA workbench. online appendix for "data mining: Practical machine learning tools and techniques"," *Morgan Kaufmann*, vol. Fourth Edition, 2016.

[60] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[61] J. L. Fleiss, "Measuring nominal scale agreement among many raters." *Psychological bulletin*, vol. 76, no. 5, pp. 378–382, 1971.

[62] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: the kappa statistic," *Family Medicine*, vol. 37, no. 5, pp. 360–363, 2005.

Jesús Joel Rivas received the M.Sc. degree in computer science from the National Institute of Astrophysics, Optics and Electronics (INAOE), Puebla, Mexico, in 2015. He is currently studying a Ph.D. in computer science at the INAOE.

He has been working in affective computing applied to stroke rehabilitation. His research interest includes affective computing, probabilistic graphical models and machine learning. He has been a Lecturer at the Universidad de Carabobo, Valencia, Venezuela.

Felipe Orihuela-Espina received his Ph.D. degree from the University of Birmingham, Birmingham, U.K.

He has been a Lecturer in the Autonomous University of the State of Mexico and a Post-doctoral Research Associate at Imperial College London from 2007 to 2010 and at the National Institute of Astrophysics, Optics and Electronics (INAOE), Puebla, Mexico, from 2010 to 2012. He is currently a Senior Lecturer at INAOE and member of the National Research System. His current research interests are in neuroimage understanding and interpretation.

Lorena Palafox is graduated as occupational therapist from the National Rehabilitation Institute, Mexico City, in 2009; and was certified in neurological rehabilitation by the Universidad Autónoma de Barcelona and the Guttmann Institute, Spain, in 2012.

Since 2009, she has collaborated in investigations of different neurological damages, she is the only occupational therapist at the National Neurology and Neurosurgery Institute MVS, Mexico City. She has been certified in different rehabilitation techniques.

Nadia Bianchi-Berthouze is a Professor in Affective Computing and Interaction at the University College London, UK. She received the Laurea degree with honors in computer science in 1991 and the Ph.D. degree in science of biomedical images in 1996 from the University of Milano, Milano, Italy. Her research interests include the study of body movement, muscle activity and touch behaviour as ways to automatically recognize and steer the quality of experience of humans interacting and engaging with/through whole-body technology. She has been pioneering the analysis of affective body expressions in the context of physical rehabilitation. She was the Principal Investigator on an EPSRC funded project on Pain rehabilitation: E/Motion-based automated coaching (Emo-pain.ac.uk). She is now investigating wellbeing, movement and affect in a variety of real-life situations such as factory work, education and textile design.

María del Carmen Lara is a psychiatrist who received a Ph.D. in Medical Sciences from the National University of Mexico. She is a Professor-Investigator at the Benemerita Universidad Autonoma de Puebla. Her main field of interest is the measurement of clinical phenomena.

Jorge Hernández-Franco graduated as medical doctor from the Universidad Autónoma de México, Mexico City, Mexico, in 1985, got his speciality in rehabilitation medicine, in 1989, and was certified in neurological rehabilitation by the Newcastle University, Newcastle upon Tyne, U.K., in 1999.

Since 1991, he is the Head of the rehabilitation ward at the National Neurology and Neurosurgery Institute MVS, Mexico City, Mexico, where he lectures on neurological rehabilitation. He has further lectured on physical therapy in neurological rehabilitation at the American British Cowdray Hospital since 1996. He is member of the editorial board of the journal *Developmental Neurorehabilitation* since 2005.

Dr Hernández-Franco is the Vice-President for Mexico, Central America and the Caribbean of the World Federation for Neurologic Rehabilitation.

Luis Enrique Sucar (Senior Member IEEE) has a Ph.D. in Computing from Imperial College, London, 1992; a M.Sc. in Electrical Engineering from Stanford University, USA, 1982; and a B.Sc. in Electronics and Communications Engineering from ITESM, Mexico, 1980. He is currently Senior Research Scientist at the National Institute for Astrophysics, Optics and Electronics, Puebla, Mexico. He has been an invited professor at the University of British Columbia, Canada; Imperial College, London; INRIA, France; and CREATE-NET, Italy. He has more than 300 publications and has directed 21 Ph.D. thesis. Dr. Sucar is Member of the National Research System and the Mexican Science Academy. He is associate editor of the Pattern Recognition journal, and has served as president of the Mexican AI Society and as member of the Advisory Board of IJCAI. In 2016 he received the National Science Prize from the Mexican government. His main research interest are in graphical models and probabilistic reasoning, and their applications in computer vision, robotics, energy, and biomedicine.