Extracting Fuzzy Rules From Polysomnographic Recordings for Infant Sleep Classification

Claudio M. Held*, *Member, IEEE*, Jaime E. Heiss, *Member, IEEE*, Pablo A. Estévez, *Senior Member, IEEE*, Claudio A. Perez, *Senior Member, IEEE*, Marcelo Garrido, Cecilia Algarín, and Patricio Peirano

Abstract—A neuro-fuzzy classifier (NFC) of sleep-wake states and stages has been developed for healthy infants of ages 6 mo and onward. The NFC takes five input patterns previously identified on 20-s epochs from polysomnographic recordings and assigns them to one out of five possible classes: Wakefulness, REM-Sleep, Non-REM Sleep Stage 1, Stage 2, and Stage 3-4. The definite criterion for a sleep state or stage to be established is duration of at least 1 min. The data set consisted of a total of 14 continuous recordings of naturally occurring naps (average duration: 143 ± 39 min), corresponding to a total of 6021 epochs. They were divided in a training, a validation and a test set with 7, 2, and 5 recordings, respectively. During supervised training, the system determined the fuzzy concepts associated to the inputs and the rules required for performing the classification, extracting knowledge from the training set, and pruning nonrelevant rules. Results on an independent test set achieved $83.9 \pm 0.4\%$ of expert agreement. The fuzzy rules obtained from the training examples without *a priori* information showed a high level of coincidence with the crisp rules stated by the experts, which are based on internationally accepted criteria. These results show that the NFC can be a valuable tool for implementing an automated sleep-wake classification system.

Index Terms—ANFIS, fuzzy rule extraction, knowledge discovery, neural nets and expert systems, rule pruning, sleep classification.

I. INTRODUCTION

A. Sleep Classification

SLEEP scoring is a fundamental tool for sleep research and sleep medicine [1]. In 1968, Rechtschaffen and Kales (R&K) [2] edited a manual for normal sleep classification which became the gold standard for classification and, despite its drawbacks [3], [4], is still in use in most sleep laboratories around the world [4], [5]. Since the R&K rules were originally

determined to score sleep in adults, they do not take into consideration the maturational development in infants. Thereafter, specific age-related scoring systems were introduced. In particular, Guilleminault and Souquet established a classification procedure for infants, also accepted as an international standard, considering subjects of age 3 mo and onward [6]. Thus, sleep-wake is classified in wakefulness (WA) or one of two sleep states: REM and Non-REM (NREM) sleep. The latter is divided in 4 stages numbered from 1 to 4. In this study, we grouped NREM stages 3 and 4 into a single stage NREM 3 and 4.

The sleep classification process can be divided into three steps: data acquisition, pattern identification and sleep-wake states classification. In the first step, several signals generated by bioelectrical and biomechanical activity of the infant's body are simultaneously recorded by a polygraph, generating a large number of pages with graphical data, i.e., the polysomnogram. The pattern identification process is performed for each page or epoch. In the manual approach, the expert determines the background predominant frequency range in the electroencephalogram (EEG) according to [2] and [6]; relevant for this study are the slow delta (SD) (0.5–2 Hz) and theta (TH) (3–7 Hz) frequency ranges. The EEG is also examined to detect sleep spindles (SS), which are in the sigma range (12–14 Hz). The electrooculogram (EOG) and the electromyogram (EMG) are used to determine the presence of rapid eye movements (REMov) and muscular tone (MT), respectively. They are analyzed in 20- or 30-s-long pages in a polysomnograph. In this paper, we will always refer to an epoch as a 20-s data frame.

After identifying the sleep patterns present within an epoch, the experts determine the sleep state or stage using a set of rules based on [2], [6] (Table I). It is widely accepted that a minimum duration of 1 min is required to establish a sleep class [7], [8], i.e., three consecutive epochs. If periods of at least 1 min do not match any of the rules, they are classified as indeterminate sleep (IS). The result of this scoring is a hypnogram, which represents the temporal profile of sleep-wake states evolution. The 1-min rule cannot be applied if fragmented sleep, characterized by rapid changes of sleep states, were to be analyzed, but this is beyond the current scope of our ongoing automated classification project.

Sleep classification has some degree of ambiguity. Studies have shown that even among expert coworkers there is less than 90% agreement in sleep classification [9]. Usually, experts from different research centers have slightly different approaches. In [5], a reliability study is mentioned where the test-retest agreement for the same experts was 79% for NREM 1, while 98%

TABLE I
EXPERT'S RULES FOR SLEEP-WAKE STATE CLASSIFICATION

Pattern	Sleep-Waking States & Stages				
	WA	NREM 1	NREM 2	NREM 3&4	REM
SD	A	A	A	P	A
TH	X	P	X	X	P
SS	A	A	P	X	A
REMov	P	A	A	A	P
MT	P	X	X	X	A

A: absent, P: present, X: irrelevant. Patterns: Rapid eye movements (REMov); theta (TH) and slow-delta (SD) background predominant frequency bands, and sleep spindles (SS) in the EEG; and muscle tone (MT). Sleep-waking states: Wakefulness (WA), Non-REM (NREM) and REM sleep.

for REM sleep. The same study reported the inter agreement rates among 3 different laboratories, under uniform criteria, obtaining 63% for NREM 1 and 95% for REM sleep. The other stages ranged from 82% to 93% of inter expert agreement. Similar results have been found in other studies [10].

The large amount of data, hence the time-intensive task, as well as the complexity of the classification procedure and the variability among human experts are reasons to develop an automated sleep classification system [11]–[17]. The computational tools developed for automatic sleep scoring can be grouped in three different categories. The first are the computational tools that facilitate the visual scoring by identifying the main sleep patterns and frequencies of the signals, like [18], [19]; most of the modern digital polysomnographs come with analysis software of this kind. The second category consists of classification systems [14]–[17], [20]–[22] that perform traditional (R&K-like) classification of the raw data, yielding a hypnogram. Finally, the third category includes sleep classifiers using other than traditional standards to characterize the sleep state of the patient [23]–[27] .

An evaluation of the computerized system ALICE 3 (Healthdyne Technologies, Marietta, GA) using 50 subjects [28], showed substantial differences between automated computer scoring and manually scored paper polysomnographies. A manual edition of the computer scoring enhanced the agreement with the paper polysomnography to 75.7%. In general, none of the existing automated scoring systems is reliable enough to be fully trusted by physicians or sleep experts. Nevertheless, they are being used in many sleep laboratories as a first general characterization that could detect possible abnormalities.

In [29], a pattern identification system for sleep-wake state classification which emulates the way the expert searches for each of the five relevant patterns was implemented. The quality of the pattern detection was evaluated by classifying the epochs in their corresponding sleep states and stages using an α-cut classifier which uses the information of Table I as its rule base.

It is desirable for a sleep classifier, or in general, for an automatic diagnosis system, to possess the following three characteristics in order to be useful and trustworthy: First, it must have at least a performance similar to the agreement achieved between two experts; second, it must be adaptable to the user's criteria and expertise and third, it must be able to give an acceptable explanation of the decisions it makes.

B. Neuro-Fuzzy Systems

The artificial neural networks (NN) [30], [31], and the fuzzy inference systems (FIS) [30], [32], are considered part of the "soft computing" family. Both approaches generate nonlinear systems that relate a set of inputs with their corresponding outputs. The main difference between them is that NN establish the input-output relations in a numerically quantitative way, while FIS do it in a symbolically qualitative way [33]. One of their main applications is to model the behavior of an expert, imitating its actions in order to solve complex problems. Neuro-fuzzy modeling allows the construction of FIS to model a problem domain using a linguistic model, whose parameters can be adjusted in the same way as the parameters of NN [34]. Since the early 90's, several neuro-fuzzy systems have been developed. Most research focused on the ability of neuro-fuzzy systems to approximate nonlinear functions or to perform fuzzy control without *a priori* information, partitioning the input space with clustering-like methods [36]–[42]. A few of them addressed the issue of classification [43]–[50]. Applications of neuro-fuzzy systems to biomedical problems can be found in [35], [51]–[53].

This paper presents the implementation of a new adaptive and robust automated neuro-fuzzy classifier (NFC), which is capable of discovering rules from examples, applied to the sleep classification problem. It illustrates how neuro-fuzzy systems can be very valuable in biomedical applications in general and in automatic diagnosis systems in particular. The NFC is based on ANFIS [30], [37], [43], one of the most well-known neuro-fuzzy systems. The initial rule base is usually composed of all possible rules to gain as many degrees of freedom as possible. In this case, two fuzzy concepts are associated to each of the 5 sleep patterns, i.e., it can be either present or absent. All possible combinations of inputs result in $2^5 = 32$ rules for each of the 5 classes, i.e., a total of 160 rules. Their parameters can be modified with supervised learning using gradient descent methods [31]. This number of rules does not convey a meaning for a human user. In order to achieve interpretability with the NFC, we developed a simple, yet effective, pruning mechanism based on the average contribution of the rules, to eliminate the irrelevant ones. Preliminary results of a first version of the proposed NFC were published in [54].

II. METHODS

A. Data Acquisition and Pattern Identification

Polygraphic recordings were obtained at the Sleep Laboratory, INTA, Universidad de Chile. The procedures were standardized to minimize the potential influences of the environment, circadian rhythms, and/or food intake on sleep-wake patterns and related physiological activities. Fourteen continuous recordings of naturally occurring daytime sleep (naps) were obtained from infants between 6 and 13 mo. of age, on a TECA 1A97 18-channel polygraph connected as follows: five EEG channels with electrode placement adapted for infants from the international 10–20 system [55] (FP1-C3, O1-C3, FP2-C4, O2-C4, and C3-C4), EOG for REMov detection, tonic chin and diaphragmatic EMGs, electrocardiogram (ECG), body

	Register Label	Gender	Age [month]	Nº of Epochs	WA [Epochs]	NREM 1 [Epochs]	NREM 2 [Epochs]	NREM 3&4 [Epochs]	REM [Epochs]	IS [Epochs]
Training Set	AC100394	M	7	558	78	136	83	195	66	0
	CR082995	F	12	459	217	25	65	135	17	0
	CV061493	M	12	668	116	82	239	121	102	8
	FH120594	M	13	395	82	30	76	180	27	0
	PG031693	F	12	390	64	91	182	0	53	0
	PG091592	F	6	194	50	39	80	0	21	4
	NL101994	F	12	298	77	21	14	144	10	32
Validation Set	DA091394	F	7	346	144	31	43	105	10	13
	FM041194	M	6	345	67	59	19	162	38	0
Test Set	AM102793	F	6	463	105	56	60	168	74	0
	BH072093	M	12	356	122	36	49	123	20	6
	KD032493	F	6	501	105	41	288	4	63	0
	SS051893	F	7	510	235	16	170	48	37	4
	SS101893	F	12	538	53	38	111	273	56	7

movement detection of upper and lower limbs using piezo-electric crystal transducers, abdominal respiratory movements using a mercury strain gauge; and nostrils airflow, by means of a thermistor. All data were simultaneously recorded on paper and on digital means at a 250-Hz sampling rate. The digital data were stored in laser media and divided in 20-s epochs for off-line analysis. Infant behavior was also observed directly and noted on the polygraph paper. The data were divided in a training set with 7 recordings, corresponding to 2962 epochs, a validation set with 2 recordings, corresponding to 691 epochs and test set with 5 recordings corresponding to 2368 epochs. Table II shows some characteristics of each recording of the database, corresponding to 12 different individuals. The first two letters in the register label identify the patient. The testing data came from different subjects than the training data or the validation data. The database separation into the different sets was decided using Kohonen maps, to have a better representation of the different combinations of input patterns in all sets.

The pattern detection system described in [29] was applied to obtain a level of presence for the SD, TH, REMov, SS, and MT sleep patterns. Its outputs are either percentage of presence, or quality indices of a given pattern per epoch in the [0,1] range. Fig. 1 shows examples of the patterns visually identified as present or absent in the polysomnogram. Both background activity patterns (SD and TH) were detected analyzing power spectra on EEG channel 4 (C3-O1) applying fast Fourier transform. For SD the spectra of all overlapping windows of 1024 samples (step size: 128) within an epoch were averaged. The method also considered the relation between SD and fast delta activity (2–3 Hz), as well as context information, such as limb movements (LM) and overall delta activity. For TH the index of presence was the percentage of windows that showed significant power in the theta range in a given epoch. The third EEG-based pattern, SS were detected in a four stage process analyzing the 4 EEG channels FP1-C3, O1-C3, FP2-C4, and O2-C4. The detection of three consecutive peaks was followed by the formation of trains of pulses, evaluated according to the amplitude and duration of the trains and finally, the level of presence of SS in the epoch was determined considering a 5-min context of not only the 4 EEG channels but also the indices of the other patterns as well as LM.
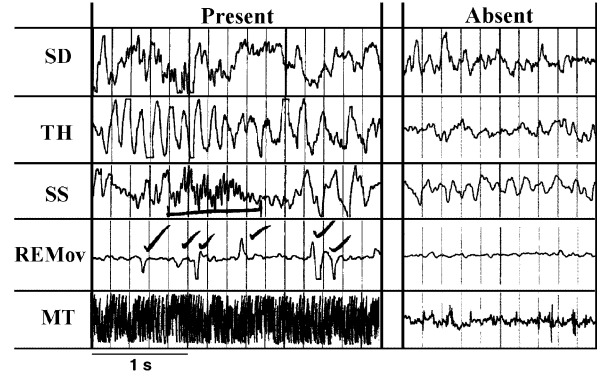

Fig. 1. Examples of signal plots taken from isolated segments of polysomnograms showing present and absent sleep patterns. The plots for SD and TH were taken from EEG channel O2-C4, the plot for SS from the EEG channel FP1-C3, the plot for REMov from the EOG channel and the plot for MT from the chin EMG channel. The experts mark the presence of sporadic patterns on the paper as they perform the visual scoring, underlining the episodes of SS and ticking the episodes of REMov.

The REMov pattern detection began by identifying a peak in the EOG signal, determined by three consecutive sign changes in the slope. The amplitude, duration, and power relative to the surroundings of the peak were combined in two fuzzy REMov concepts, one optimized for REMov events occurred during WA and the other optimized for REMov detection during REM, and thus expressed as two quality indices. The final output considered that REMov are sparse by nature, and a 1–min memory was applied. The RMS power of the EMG channel was determined to establish the MT pattern. The moving window length was dynamically adjusted according to the strength of the EMG signal.

B. Neuro-Fuzzy Classifier (NFC)

The parameters of the NFC were adjusted using supervised learning. Each example of the training set consists of an input vector $X_j = \{x_1, x_2, \ldots, x_N\}$, whose components are the levels of presence of the sleep patterns, and the corresponding desired output vector $D_j = \{d_1, d_2, \ldots, d_C\}$. They summarize the activity of one sleep epoch of 20 s. The number of inputs N is 5; they are normalized in the interval [0,1]. The number of output classes C is 5; usually one of the outputs is a 1, representing the

Layer 1   Layer 2   Layer 3

SD
SS
TH
REMov
MT

A_1 P_1 A_2 P_2 A_3 P_3 A_4 P_4 A_5 P_5

NREM 1
NREM 2
NREM 3&4
REM
WA

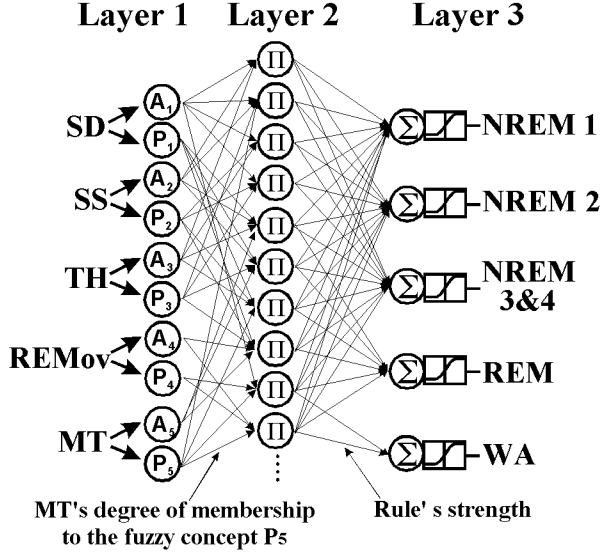MT's degree of membership to the fuzzy concept P$_5$

Rule' s strength

Fig. 2. Neuro-Fuzzy classifier architecture. The inputs are the level of presence detected for the sleep patterns at a given epoch. Layer 1 determines the membership degree to the fuzzy concepts present (P_i) and absent (A_i) for every pattern. In layer 2, the rule strengths are determined using a differentiable T-norm operator, in this case the product Π, which combines the membership degrees determined at layer 1. Each node at layer 3 corresponds to an output class and performs first a linear combination of all the strengths calculated at layer 2. The result is modified by a sigmoid function, establishing a degree of presence of the input vector to the output class at every node. The node with maximum output will determine the class associated to the input vector.

current sleep state as classified by the sleep expert, and the rest are 0 s, except if the current epoch corresponds to the IS class, in which case all outputs are 0 s.

The training method was enhanced to speed up the training process and to avoid local minima convergence. It has two phases. First the system is trained with all possible rules, like the original ANFIS implementation of the NFC [43]. Then a pruning procedure is applied to eliminate the nonrelevant fuzzy rules, extracting relevant rules from a set of data. The parameters are restored to their initial values and the system is retrained with the surviving rules only, determining the final shape of the fuzzy concepts and the final weights for every rule. The result of the training is a NFC capable of classifying input vectors in one out of five sleep states.

1) Architecture of the NFC: Fig. 2 shows the architecture of the ANFIS-based NFC applied to the sleep classification problem. The system can be divided in 3 layers of processing, each with functionally different nodes. The number of input nodes in layer 1 and output nodes in layer 3 was determined according to the structure of the training data for the sleep classification problem, as explained in I-B.

a) Layer 1: Fuzzification: In this layer, the level of presence of the sleep patterns are transformed into fuzzy numbers. The membership degree to the fuzzy concepts "present" (P) and "absent" (A) is evaluated for each input. Each fuzzy concept is represented by a sigmoid

$$\text{Sigm}(a, \theta, x) = \frac{1}{1 + e^{-a(x-\theta)}} \qquad (1)$$

where a is the slope of the sigmoid and θ its center. Sigmoids are suitable for this method because they are monotonic, differentiable and bounded in the [0,1] range. The concept "present" has a positive slope ($a > 0$), whereas the concept "absent" has a negative one ($a < 0$). The output of every node at layer 1 is

$$O^1 f(x_n) = \text{sigm}\left(a_n^1, f, \theta_n^1, f, x_n\right) \qquad (2)$$

where x_n is the nth sleep pattern, $n = 1, \ldots, N$, and $f = 1, \ldots, F(n)$, with $F(n)$ the amount of fuzzy concepts associated to pattern n. In this case, $F(n) = 2$ for all n. The parameters $a_{n,f}^1$ and $\theta_{n,f}^1$ correspond to the slope and center of the fuzzy concept f associated to the input n at layer 1. Their values were adjusted during training.

b) Layer 2: Rule Strength: Each rule at layer 2 consists of an antecedent and a consequent. The antecedent is the set of conditions that must be satisfied in order to trigger the rule, and the consequent links the rule with an output class. The rule strength reflects how well all the conditions of the antecedent are satisfied. This is performed by a T-norm operation among the membership degree inputs, considering one fuzzy concept (either present or absent) per sleep pattern. The T-norm utilized was the product, although other differentiable T-norms can be used [30], [43]. Before the pruning phase, there was one node for each possible combination of present/absent sleep patterns. The output at the lth node of layer 2 is

$$O_l^2 = \Pi_{n=1}^N O_{m(l,n)}^1(x_n) \qquad (3)$$

where $l = 1, \ldots, R, R = \prod_{n=1}^N F(n)$, is the total number of possible combinations considering one fuzzy concept per input pattern. In this case, since $F(n) = 2$ for all $n, R = 32$. The index $m(l, n)$ indicates which of the $F(n)$ fuzzy concepts associated to input n is considered in the lth rule. It is a function that returns the value of the position (l, n) of a (R $\times$ N) matrix, which is the concept (in this case present or absent) associated to the nth input variable for determining the lth rule strength. The antecedent of each rule has a unique combination of N fuzzy concepts. There are no adaptive parameters in this layer.

c) Layer 3: Classes: In layer 3, there are as many output nodes as sleep classes. Every node performs a weighted sum of all the rule strengths entering to it and the result is modified by a sigmoid function. The output at the cth node is

$$O_c^3 = \text{Sigm}\left(a^3, \theta_c^3, \sum_{l=1}^R W_l^c O_l^2\right) \qquad (4)$$

where a^3 and θ_c^3 are the slope and the center (or bias) of the output sigmoid of class c at layer 3, and W_l^c is the weight associated to rule strength O_l^2 in node c. The slope $a^3 = 5$ was determined empirically and considered a fixed parameter in all output nodes. The weights W_l^c and the bias θ_c^3 are adjusted during training. When a rule is pruned in a certain node c, its

weight is set to 0. Applying equations (1)–(3), (4) can be expressed as

$$O_c^3 = \text{Sigm}\left(a^3, \theta_c^3, \sum_{l=1}^{r} W_1^c \Pi_{n=1}^{N} \text{Sigm}\left(a_{m(l,n)}^1, \right.\right.$$
$$\left.\left. \theta_{m(l,n)}^1, x_n\right)\right) \quad (5)$$

which shows the nonlinear relationship between the input vector of sleep patterns X and its degree of membership to the output class c.

The input vector is classified to the sleep-wake state corresponding to the output node that has the maximum value. Imitating the expert's criteria for visual scoring, if the maximum value is below a threshold of 0.2, the epoch is classified as IS.

2) Parameter Adjustment: Layers 1 and 3 have parameters that can be adjusted to optimize the classification performance of the system. The four parameter types that are adjusted during training are the slope and bias of the fuzzy concepts at layer 1, and the weights and bias in each output node at layer 3. These parameters are adjusted using gradient descent [30], [31], according to

$$\Delta\phi = -\mu \frac{\partial \varepsilon}{\partial \phi} \quad (6)$$

where ϕ is the parameter to be modified, μ is the learning rate, and ε is the error function to be minimized. In this paper, the error function was set as the mean squared error (mse), which is defined as

$$\varepsilon = \frac{1}{P}\sum_{j=1}^{P}\sum_{c=1}^{C}\left(d_c^j(X_j) - O_c^3(X_j)\right)^2 \quad (7)$$

where P is the number of examples in the training set, and d_c^j is the desired output. The learning rate is a constant that determines the step size in the direction that locally minimizes the error function. Each parameter was adjusted applying an individual learning rate. The learning rates $\mu_{a1}, \mu_{\theta1}, \mu_W$, and $\mu_{\theta c}$ used to modify the parameters $a_{n,f}^1, \theta_{n,f}^1, W_l^c$ and θ_c^3, respectively, were set as large as possible to have a fast convergence but without reaching an oscillatory behavior. Using the chain rule of calculus, the update rules for each of the four parameters were obtained, e.g., for $a_{n,f}^1$

$$\Delta a_{n,f}^1 = \mu_{a1} \sum_{j=1}^{P}\sum_{c=1}^{C}\left(d_c(X_j) - O_c^3(X_j)\right)$$
$$\left(O_c^3\left(1 - O_c^3\right)a^3 Y^{c*}(x_n - \theta_{n,f}^1)\right.$$
$$\left.\left(1 - \text{Sigm}\left(a_{n,f}^1, \theta_{n,f}^1, x_n\right)\right)\right) \quad (8)$$

with $Y^{c*} = \sum_{l \in K} W_l^c \Pi_{n=1}^{N} \text{Sigm}(a_{m(l,n)}^1, \theta_{m(l,n)}^1, x_n)$; the set K includes all the rules involving the fuzzy concept being adjusted. Since each input is associated to two fuzzy concepts, the set K has a cardinality of R/2.

The update rules were applied iteratively in an off-line (batch) learning mode. After presenting all the examples,

which is called a training epoch,[1] the parameters were updated according to $\phi_{t+1} = \phi_t + \Delta\phi_t$, where t denotes the tth training epoch.

At layer 1, the slopes of the fuzzy concepts were initialized at $a_{n,f}^1 = \pm 8$. The biases of the fuzzy concepts were initialized in the midpoint of its range, according to

$$\theta_{n,f}^1 = \frac{\max_j\{x_n^j\} - \min_j\{x_n^j\}}{2} \quad (9)$$

where x_n^j is the nth component of example j in the training set. Both concepts, present and absent ($f = 1$ or 2, respectively), were initialized with the same bias and the same absolute value of the slope. At layer 3, the output sigmoid biases θ_c^3 were initialized at 0.5 and the rule weights W_l^c were initialized with random numbers in the [0,1] interval. This originated a small variability among different simulations.

The training of the parameters had two phases: before pruning (B.P.) and after pruning (A.P.). In the B.P. phase, the goal was to identify and eliminate the nonrelevant rules for each output. The system was trained for 500 training epochs and then the least relevant rules were pruned. In the A.P. phase, the goal was to establish the optimal parameter values considering only the surviving rules. At the beginning of this phase, the parameters were reset to the same initial values used at the beginning of the B.P. phase. The system was trained again without the pruned rules, until the classification performance of the validation set stopped improving or the system reached a maximum of 500 training epochs. For determining if the performance had stopped improving, the average agreement percentage of the last 50 epochs was compared to the average of the previous 50 epochs. The whole training process is called one simulation.

3) Modifications Introduced to the Original ANFIS System:

a) Pruning mechanism: The ANFIS classifier proposed in [43] considered all possible rules formed by combining one fuzzy concept per input pattern. The aim of the proposed pruning mechanism is to eliminate the least relevant rules without producing a significant deterioration in the classification performance.

The relevancy of the lth rule for a given input vector X is measured as the contribution it makes to the output of the corresponding node c. Using (5), the rule contribution (RC) can be expressed as the weight of the rule W_l^c multiplied by its strength

$$\text{RC}_l^c = W_l^c \Pi_{n=1}^{N} \text{Sigm}\left(a_{m(l,n)}^1, \theta_{m(l,n)}^1, x_n\right). \quad (10)$$

For each output node c, the pruning mechanism stores the contribution of every rule each time a training vector is associated to class c. After a training epoch, the average contribution of every rule in each output node is calculated. These average contributions are normalized dividing them by the highest average contribution in each output node. A rule contributing rarely to a node would present a small average contribution, because even if its contribution is high once or a few times, it would be averaged with several low ones. Rules that always have a weak strength, i.e., they map an empty area of the input space, would

---

[1]We use the term *epoch* as one frame of sleep data and the term *training epoch* as one presentation of all the examples when training the NFC.

also have a low average contribution. After the B.P. training, if the normalized average contribution of a certain rule for output node c is below a pruning threshold, the rule is eliminated for that output node by setting its weight to 0. The pruning threshold value is determined iteratively (iteration steps of 1% of normalized average contribution) as the largest value, hence the smallest set of surviving rules that does not significantly deteriorate the classification, measured by not degrading the results of the validation set in more than 0.5%. As a consequence of pruning the irrelevant rules, the classifier greatly simplifies its architecture. Moreover, because of the fuzzy nature of the input variables, the surviving rules have a direct linguistic meaning, expressing in simple terms the criteria applied to classify each sleep class.

b) Weighted Error: The parameter adjustment was originally driven to minimize the mse However, this could produce a suboptimal classification performance because the expert-classified output for each example was binary (one for the corresponding class and zero for the rest), whereas the system-classified output corresponded to the node with the maximum value. All the parameters were updated proportionally to the output error, which is given by

$$e_c^j = d_c^j(X_j) - O_c^3(X_j). \tag{11}$$

Sometimes adjusting the system to obtain a closer-to-one output of an already well classified example could decrease the mse more than what fixing a wrongly classified example would do. As a result, the training had a less than optimal performance.

In order to improve the algorithm performance, the update rules were modified, amplifying the output error for the classes that showed poor classification performance during training, thus increasing their influence when adjusting the parameters, i.e.,

$$\bar{e}_c^j = e_c^j \frac{\text{Pe}_c + 0.1}{\min_k\{\text{Pe}_k\} + 0.1} \tag{12}$$

where $\bar{e}_c^j$ is the weighted error obtained for the example j at node c, e_c^j is the error calculated as in (11), Pe_c is the percentage of misclassification in class c and Pe_k ($k = 1, \ldots, 5$) is the percentage of misclassification of class k. The 0.1s were arbitrarily added to avoid zeros when there is no error in a class, and chosen small enough to maintain an error amplification proportional to the class error-minimum class error quotient while class results remained unbalanced. The weighted error produced not only an improvement in the classifier homogeneity, i.e., the classes presented a more even classification performance, but also a faster convergence and a better global agreement rate.

4) State Duration Algorithm: A state duration algorithm was created in order to filter the epoch by epoch classification obtained by the NFC and simulate a minute-by-minute classification, as introduced in I-A. The state duration algorithm transforms the 20–s based classification into a 1-min-based one, according to expert-based empirically developed criteria.
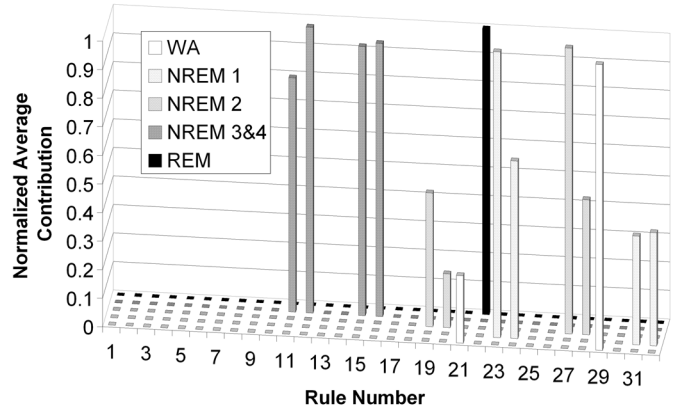

Fig. 3. Average contribution of the 15 extracted (nonpruned) rules for sleep classification. The relevance of each rule was given by its normalized average value. All rules matched the expert rules except rules number 31 and 32 for NREM 1, which were additional rules established by the NFC. For NREM 3 and 4, 4 expert rules were pruned due to lack of training examples with meaningful contribution. These 15 rules were combined and expressed as the 5 rules shown in Table III, recursively applying that, if two rules for the same output class differ only in the fuzzy concept associated to one pattern, they were merged into a single rule declaring that pattern irrelevant.

TABLE III
FUZZY RULES OBTAINED BY THE NFC FOR SLEEP CLASSIFICATION

Input Pattern	NFC Rules				
	R_{WA}	$R_{NREM\ 1}$	$R_{NREM\ 2}$	$R_{NREM\ 3\&4}$	R_{REM}
SD	A	A	A	P	A
TH	X	X	X	A	P
SS	A	A	P	X	A
REMov	P	A	A	A	P
MT	P	X	X	X	A

A: absent, P: present, X: irrelevant. E.g. the extracted rule for sleep stage NREM 2 established that pattern SS must be present, and patterns SD and REMov absent. The other patterns TH and MT are irrelevant for stage NREM 2.

III. RESULTS

The average agreement between the NFC and the sleep expert was determined using 20 simulations, with different initial values for the rule weights set randomly. In all simulations, the same rules survived the pruning, with different parameter values.

A. Fuzzy Rule Extraction

Every fuzzy rule in the sleep classification problem associates a 5-dimensional fuzzy region in the input space with a degree of membership to an output class. This degree of membership was determined by the multiplication of the rule strength and the rule weight, modified by the corresponding output sigmoid. The pruning algorithm identified and eliminated the least relevant fuzzy rules for every class, according to their average contribution as defined in Section II-B3). As a result of applying these criteria, it turned out that any rule whose normalized average contribution was lower than a threshold of 7% was pruned, because it did not contribute significantly to the classification (cumulatively less than 0.5% measured on the validation set). Fig. 3 shows the surviving rules with their average contribution for each output class; 145 rules were pruned, leaving a rule base

TABLE IV
CLASSIFICATION RESULTS OF THE NFC

Training

State	Epochs	Agrmnt.	%Agrmnt.
WA	685	643.3 ± 2.2	93.9 ± 0.3
NREM 1	642	539.7 ± 1.2	84.1 ± 0.2
NREM 2	739	659.4 ± 1.6	89.2 ± 0.2
NREM 3&4	775	679.2 ± 3.2	87.6 ± 0.4
REM	592	530.5 ± 1.7	89.6 ± 0.3
IS	44	0.2 ± 0.7	0.3 ± 1.5
Total	3477	3052.2 ± 2.6	87.8 ± 0.1

NFC \ Exp	WA	NREM 1	NREM 2	NREM 3&4	REM
WA	643.3 ± 2.2	21.0 ± 0.0	0.0 ± 0.0	6.9 ± 0.4	14.0 ± 0.2
NREM 1	23.0 ± 0.0	539.7 ± 1.2	31.8 ± 1.5	14.8 ± 3.5	36.9 ± 1.3
NREM 2	7.0 ± 0.0	18.2 ± 1.2	659.4 ± 1.6	74.1 ± 1.9	0.0 ± 0.0
NREM 3&4	4.2 ± 0.4	50.1 ± 1.2	38.9 ± 0.4	679.2 ± 3.2	7.6 ± 1.5
REM	7.6 ± 2.0	13.0 ± 0.0	9.0 ± 0.0	0.0 ± 0.0	530.5 ± 1.7

Validation

State	Epochs	Agrmnt.	%Agrmnt.
WA	211	194.1 ± 1.4	92.0 ± 0.7
NREM 1	90	82.0 ± 0.0	91.1 ± 0.0
NREM 2	62	48.0 ± 0.0	77.4 ± 0.0
NREM 3&4	267	228.9 ± 2.5	85.7 ± 0.9
REM	48	38.0 ± 0.0	79.2 ± 0.0
IS	13	0.0 ± 0.0	0.0 ± 0.0
Total	691	591.0 ± 3.5	85.5 ± 0.5

NFC \ Exp	WA	NREM 1	NREM 2	NREM 3&4	REM
WA	194.1 ± 1.4	7.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
NREM 1	10.0 ± 0.0	82.0 ± 0.0	12.0 ± 0.0	0.0 ± 0.0	6.0 ± 0.0
NREM 2	0.0 ± 0.0	1.0 ± 0.0	48.0 ± 0.0	38.2 ± 2.5	0.0 ± 0.0
NREM 3&4	0.0 ± 0.0	0.0 ± 0.0	2.0 ± 0.0	228.9 ± 2.5	4.0 ± 0.0
REM	6.9 ± 1.4	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	38.0 ± 0.0

Test

State	Epochs	Agrmnt.	%Agrmnt.
WA	621	602.0 ± 0.0	96.9 ± 0.0
NREM 1	187	134.0 ± 0.0	71.7 ± 0.0
NREM 2	678	507.3 ± 10.8	74.8 ± 1.6
NREM 3&4	616	561.4 ± 1.2	91.1 ± 0.2
REM	250	179.5 ± 0.9	71.8 ± 0.4
IS	17	4.0 ± 0.0	23.5 ± 0.0
Total	2369	1988.2 ± 9.4	83.9 ± 0.4

NFC \ Exp	WA	NREM 1	NREM 2	NREM 3&4	REM
WA	602.0 ± 0.0	11.0 ± 0.0	5.9 ± 0.7	4.0 ± 0.0	10.0 ± 0.0
NREM 1	2.0 ± 0.0	134.0 ± 0.0	42.0 ± 0.0	16.0 ± 0.0	49.0 ± 1.1
NREM 2	1.0 ± 0.0	24.0 ± 0.0	507.3 ± 10.8	31.6 ± 1.2	4.6 ± 0.5
NREM 3&4	0.0 ± 0.0	15.0 ± 0.0	117.7 ± 10.9	561.4 ± 1.2	7.0 ± 0.0
REM	16.0 ± 0.0	3.0 ± 0.0	5.2 ± 0.9	0.0 ± 0.0	179.5 ± 0.9

Average and standard deviation results obtained after 20 simulations for the training, validation and test sets. Left, average agreement in number of epochs and percentage. Right, confusion matrix between the NFC and the expert.

of only 15 relevant rules. These rules were later combined in a reduced set of 5 rules as explained below.

To enhance the human understandability of the extracted knowledge, when two meaningful rules belonging to the same output class differ only in the fuzzy concept associated to one pattern, they were merged into a single rule declaring that pattern irrelevant. This means that if one rule required concept n-Present and the other required concept n-Absent while matching all other conditions for the same class, the level of presence of pattern n was irrelevant in discriminating whether the input vector corresponded to the class or not. Applying it recursively, individual rules for each output class were merged, forming the extracted rule base for sleep classification shown in Table III. Thus, for interpretation purposes, the final set is reduced to only 5 rules.

B. Sleep Pattern Classification

The average results are shown in Table IV for the training, validation and test sets. In each simulation, the parameters of the NFC were adjusted using the training set; nonrelevant rules were pruned and the system was re-trained until the overall agreement of the validation set did not improve anymore. To illustrate the difference between a minute-based classification and a frame by frame classification, Table V shows the agreement percentage achieved for the train, validation and test sets with and without applying the 1 min state duration algorithm. Also this table illustrates the differences in agreement achieved with and without pruning the irrelevant rules.

TABLE V
AVERAGE RESULTS WITHOUT AND WITH APPLYING THE RULE-PRUNING
MECHANISM AND THE S.D.A.

	Without S.D.A.		With S.D.A.	
	non-pruned	pruned	non-pruned	pruned
Training	86.2 ± 0.2 %	85.9 ± 0.1 %	88.0 ± 0.2 %	87.8 ± 0.1 %
Validation	83.6 ± 0.7 %	83.9 ± 0.3 %	85.1 ± 0.8 %	85.5 ± 0.5 %
Test	82.1 ± 0.2 %	82.6 ± 0.2 %	83.6 ± 0.3 %	83.9 ± 0.4 %

The results are presented as the percentage of agreement between the system and the expert. The pruning mechanism greatly reduced the rule base size without significant deterioration of the results, whereas the S.D.A. moderately enhanced the average agreement.

For comparison purposes, the same sleep data set was classified with an α-cut classifier [29]. The level of presence of the sleep patterns was reduced to either present or absent. The thresholds (α-cuts) were adjusted to achieve at least 80% of expert agreement in all sleep classes with the training set. The expert criteria of Table I were directly applied as crisp rules on the binarized sleep patterns, to achieve sleep classification. The 1-min class duration algorithm was also applied. The *expert-guided* α-cut classifier rendered agreement percentages of 87.0%, 85.2%, and 83.8% for the training, validation and test sets respectively, similar to the results obtained with the *self-trained* NFC shown in Table IV.

IV. DISCUSSION AND CONCLUSION

Our proposed neuro-fuzzy classifier achieved an expert agreement of about 84% on an independent and unedited test set in the classification of sleep-wake states and stages, as shown in Table IV. This performance is comparable both to the

inter-expert agreement percentages described in the literature [9] and the level of agreement achieved by an α-cut classifier applying the expert rules. It is a remarkable result that the system can discover a set of fuzzy concepts and fuzzy rules that give a "human-understandable" interpretation of a classification scheme learned from examples. The pruning method showed to be robust in presence of noise, allowing a clear discrimination between relevant and irrelevant rules (data not shown) provided a sufficient number of training examples. The only a priori information used by the NFC was the amount of concepts associated to the patterns of each sleep-wake state or stage and the number of output classes; this was needed to compare the extracted rules with the expert's. Alternatively, clustering methods can be implemented to assess an optimal number of fuzzy concepts per input and output.

Table III shows the fuzzy rules extracted by the system. Except for slight differences, the NFC rediscovered the crisp rules stated by the expert (Table I). The only differences are the roles of TH in the classification of NREM 1 and NREM 3 and 4. In the former case, the system established TH as irrelevant, since there were many epochs of NREM 1 in the training set where the pattern detection algorithm didn't show a significant presence of TH. In the latter case, only the TH-Absent rules survived the pruning, because there weren't enough training examples in which TH coexisted with SD to make a meaningful contribution to the class NREM 3 and 4.

Table V allows comparing the agreement percentages achieved for the training, validation and test sets without and with the 1–min state duration algorithm (S.D.A.) and the rule pruning mechanism. The S.D.A. improved the agreement in less than 2%. Comparing the results obtained with all the rules (B.P.) and only with the rules which survived the pruning, it can be established that eliminating non relevant rules did not produce a statistically significant decrease in the classification performance, after removing more than 90% of the rules; it even improved the classification agreement for the test set, although not significantly. This can be interpreted as that pruning not only provides interpretability but also may increase the generalization ability of the NFC.

The amount of noise in the data can be estimated as the complement of the agreement percentage achieved with the training set without pruning. There were agreements of 86.2% and 88% without and with applying the 1-min state duration algorithm respectively. Thus, about 14% of the training epochs were noisy, including imperfections in the automated pattern detection system as well as inconsistencies in the expert classification.

The main drawback of the NFC is that its complexity grows exponentially with the number of inputs and concepts associated to them. This is known as the "curse of dimensionality" [30]. However, in most medical applications, physicians and experts in general make decisions based on only a few parameters that can have an also limited number of characteristics or concepts associated, since humans are also limited by the same curse. Typically, if more variables are involved, intermediate decision points are used. Therefore, applications of the NFC could be implemented for modeling and automating many human classification tasks. The ability to extract rules from examples and to

select them by pruning, together with the flexibility to achieve a good performance, makes the NFC an adequate tool to automate medical diagnosis procedures.

REFERENCES

[1] T. Penzel and R. Conradt, "Computer based sleep recording and analysis," *Sleep Med. Rev.*, vol. 4, pp. 131–148, 2000.

[2] A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques, and scoring system for sleep stages of human subjects," in *UCLA, Brain Research Institute/Brain Information Service*, Los Angeles, CA, 1968.

[3] S.-L. Himanen and J. Hasan, "Limitations of rechtschaffen and kales," *Sleep Med. Rev.*, vol. 4, pp. 149–167, 2000.

[4] G. Klösch, B. Kemp, T. Penzel, A. Schlögl, P. Rappelsberger, E. Trenker, G. Gruber, J. Zeitlhofer, B. Saletu, W. M. Herrmann, S. L. Himanen, D. Kunz, M. J. Barbanoj, J. Roschke, A. Varri, and G. Dorffner, "The SIESTA project polygraphic and clinical database," *IEEE Engineering in Medicine and Biiology Mag.*, vol. 20, no. 3, pp. 51–57, May-Jun. 2001.

[5] M. Hirshkowitz, "Standing on the shoulders of giants: The standardized sleep manual after 30 years," *Sleep Med. Rev.*, vol. 4, pp. 169–179, 2000.

[6] C. Guilleminault and M. Souquet, , R. Korobkin and C. Guilleminault, Eds., "Sleep states and related pathology," in *Advances in Perinatal Neurology*. New York: Spectrum, 1979, vol. 1.

[7] L. Curzi-Dascalova and M. Mirmiran, "Manual of methods of recording and analyzing sleep-wakefulness states in preterm and full-term infants," presented at the INSERM, Paris, France, 1996.

[8] R. Ariagno, K. Bentele, M. R. Bardini, R. Brouillette, O. Bruni, M. J. Challamel, L. Curzi-Dascalova, Y. Dalmaz, G. Donzelli, J. M. Dubru, I. Fagioli, R. Ferber, G. Ficca, P. Franco, G. Franccois, C. Gaultier, J. Gingras, J. Groswasser, C. Guilleminault, T. Hoppenbrouwers, R. Horne, K. Hunt, G. Jorch, A. Kahn, M. Katz-Salomon, T. Keens, H. Lagercrantz, M. Lecendreux, C. Marcus, J. Milrad, M. Mirmiran, E. Mitchell, Y. Navelet, P. Peirano, R. Piumelli, P. Plouin, C. Poets, F. Ravet, M. Rocca-Rivarola, M. Ronchetti, P. Salzarulo, B. Taylor, B. Thach, M. Urschitz, M. F. Veccherini, P. Villa Maria, H. Wulbrand, and H. Zotter, "The scoring of arousals in healthy term infants (between the ages of 1 and 6 months)," *J. Sleep Res.*, vol. 14, pp. 37–41, 2005.

[9] R. Norman, I. Pal, C. Stewart, J. Walsleben, and D. Rappaport, "Interobserver agreement among sleep scorers from different centers in a large dataset," *Sleep*, vol. 23, pp. 901–908, 2000.

[10] R. Ferri, P. Ferri, R. M. Colognola, M. A. Petrella, S. A. Musumesi, and P. Bergonzi, "Comparison between the results of an automatic and a visual scoring of sleep EEG recordings," *Sleep*, vol. 12, pp. 354–362, 1989.

[11] R. Agarwal and J. Gotman, "Computer-assisted sleep staging," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 12, pp. 1412–1423, Dec. 2001.

[12] A. Besset, , M. Billiard, Ed., "L'Analyse Automatique du Sommeil," in *Le Sommeil Normal et Pathologique*, 2nd ed. Paris, France: Masson, 1998, pp. 126–133.

[13] C. Guilleminault, , M. Billiard, Ed., "L'Analyse Automatique du Sommeil," in *Le Sommeil Normal et Pathologique*, 2nd ed. Paris, France: Masson, 1998, pp. 3–11.

[14] B. H. Jansen and B. M. Dawant, "Knowledge-based approach to sleep EEG analysis-a feasibility study," *IEEE Trans. Biomed. Eng.*, vol. 36, no. 5, pp. 510–518, May 1989.

[15] H. J. Park, J. S. Oh, D. U. Jeong, and K. S. Park, "Automated sleep stage scoring using hybrid rule-and case-based reasoning," *Comput. Biomed. Res.*, vol. 33, pp. 330–349, 2000.

[16] J. C. Principe, S. K. Gala, and T. G. Chang, "Sleep staging automaton based on the theory of evidence," *IEEE Trans. Biomed. Eng.*, vol. 36, no. 5, pp. 503–509, May 1989.

[17] J. R. Smith, , F. H. Lopes de Silva, W. Storm, v. Leeuwen, and A. Rémond, Eds., "Automated analysis of sleep EEG data," in *Handbook of Electroencephalography and Clinical Neurophysiology*. Amsterdam, The Netherlands: Elsevier, 1986, vol. 2, pp. 131–147.

[18] J. C. Principe and J. R. Smith, "Samicos-a sleep analyzing microcomputer system," *IEEE Trans. Biomed. Eng.*, vol. BME-33, pp. 935–941, 1986.

[19] E. Huupponen, A. Värri, S. L. Himanen, J. Hasan, M. Lehtokangas, and J. Saarinen, "Optimization of sigma amplitude threshold in sleep spindle detection," *J. Sleep Res.*, vol. 9, pp. 327–334, 2000.

[20] S. Kubicki, L. Höller, I. Berg, C. Pastelak-Price, and R. Dorrow, "Sleep EEG evaluation: A comparison of results obtained by visual scoring and automatic analysis with the oxford sleep stager," *Sleep*, vol. 12, pp. 140–149, 1989.

[21] S. Park, J. C. Principe, J. R. Smith, and S. A. Reid, "TDAT—Time domain analysis tool for EEG analysis," *IEEE Trans. Biomed. Eng.*, vol. 37, pp. 803–811, 1990.

[22] S. Roberts and L. Tarassenko, "Analysis of the sleep EEG using a multilayer network with spatial organization," *IEE Proceedings F*, vol. 139, pp. 420–425, Dec. 1992.

[23] I. Gath, C. Feuerstein, and A. Geva, "Unsupervised classification and adaptive definition of sleep patterns," *Pattern Recognit. Lett.*, vol. 15, pp. 977–984, 1994.

[24] T. Watanabe and K. Watanabe, "Noncontact method for sleep stage estimation," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 10, pp. 1735–1748, Oct. 2004.

[25] M. K. Kiymik, M. Akin, and A. Subasi, "Automatic recognition of alertness level by using wavelet transform and artificial neural network," *J. Neurosci. Meth.*, vol. 139, pp. 231–240, 2004.

[26] P. Piñero, P. Garcia, L. Arco, A. Álvarez, M. M. Garcia, and R. Bonal, "Sleep stage classification using fuzzy sets and machine learning techniques," *Neurocomputing*, vol. 58–60, pp. 137–1143, 2004.

[27] A. Flexer, G. Gruber, and G. Dorffner, "A Reliable probabilistic sleep stager based on a single EEG signal," *Artif. Intell. Med.*, vol. 33, pp. 199–207, 2005.

[28] D. P. White and T. J. Gibb, "Evaluation of a computerized polysomnographic system," *Sleep*, vol. 21, pp. 188–196, 1998.

[29] P. A. Estévez, C. M. Held, C. A. Holzmann, C. A. Perez, J. Heiss, J. P. Pérez, M. Garrido, and P. Peirano, "Polysomnographic pattern recognition for automated classification of sleep-waking states in infants," *Med. Biol. Eng. Comput.*, vol. 40, pp. 105–113, 2002.

[30] J. S. R. Jang, C. T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Upper Saddle River, NJ: Prentice-Hall, 1996.

[31] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York: Macmillan College, 1994.

[32] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. New York: Academic, 1980.

[33] S. Mitra and Y. Hayashi, "Neuro-fuzzy rule generation: Survey in soft computing framework," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 748–768, May 2000.

[34] W. L. Tung and C. Quek, "GenSoFNN: A generic self-organizing fuzzy neural network," *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1075–1086, Sep. 2002.

[35] T. H. Linh, S. Osowski, and M. Stodolski, "On-line heart beat recognition using hermite polynomials and neuro-fuzzy network," *IEEE Trans. Instrum. Meas.*, vol. 52, no. 4, pp. 1224–1231, Aug. 2003.

[36] J. S. R. Jang and C. T. Sun, "Functional equivalence between radial basis function networks and fuzzy inference systems," *IEEE Trans. Neural Netw.*, vol. 4, no. 1, pp. 156–159, Jan. 1993.

[37] J. S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference systems," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 3, pp. 665–685, May-Jun. 1993.

[38] I. Rojas, H. Pomares, J. Ortega, and A. Prieto, "Self-organized fuzzy system generation from training examples," *IEEE Trans. on Fuzzy Systems*, vol. 8, pp. 23–36, 2000.

[39] S. J. Lee and C. S. Ouyang, "A neuro-fuzzy system modeling with self-constructing rule generation and hybrid svd-based learning," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 3, pp. 341–353, Jun. 2003.

[40] L. Rutkowski and K. Cpalka, "Flexible neuro-fuzzy systems," *IEEE Trans. Neural Netw.*, vol. 14, no. 3, pp. 554–574, May 2003.

[41] M. Fazle Azeem, M. Hanmandlu, and N. Ahmad, "Structure identification of generalized adaptive neuro-fuzzy inference systems," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 5, pp. 666–681, Oct. 2003.

[42] M. Setnes, "Supervised fuzzy clustering for rule extraction," *IEEE Trans. Fuzzy Systems*, vol. 8, no. 4, pp. 416–424, Aug. 2000.

[43] J. S. R. Jang and C. T. Sun, "A neuro-fuzzy classifier and its applications," in *Proc. 2nd IEEE Int. Conf. on Fuzzy Systems*, San Francisco, CA, pp. 94–98.

[44] J. S. Wang and C. S. G. Lee, "Self-adaptive neuro-fuzzy inference systems for classification applications," *IEEE Trans. Fuzzy Systems*, vol. 10, no. 6, pp. 790–802, Dec. 2002.

[45] G. G. Yen and P. Meesad, "An effective neuro-fuzzy paradigm for machinery condition health monitoring," *IEEE Trans. on Syst., Man, Cybern.-B*, vol. 31, pp. 523–536, 2001.

[46] N. Kasabov, "Learning fuzzy rules and approximate reasoning in fuzzy neural networks and hybrid systems," in *Fuzzy Sets Syst.*, 1996, vol. 82, pp. 135–149.

[47] T. P. Hong and C. Y. Lee, "Induction of fuzzy rules and memberships functions from training examples," in *Fuzzy Sets Syst.*, 1996, vol. 84, pp. 33–47.

[48] T. P. Wu and S. M. Chen, "A new method for constructing membership functions and fuzzy rules from training examples," *IEEE Trans. Syst., Man, Cybern.-B*, vol. 29, no. 1, pp. 25–40, Feb. 1999.

[49] D. Nauck and R. Kruse, "A neuro-fuzzy method to learn fuzzy classification rules from data," in *Fuzzy Sets Syst.*, 1997, vol. 89, pp. 277–288.

[50] N. R. Pal and S. Chakraborty, "5-type decision trees for real data systems," *IEEE Trans. Syst., Man, Cybern.-B*, vol. 31, no. , pp. 745–754, Oct. 2001.

[51] E. Cavallaro, S. Micera, P. Dario, W. Jensen, and T. Sinkjaer, "On the intersubject generalization ability in extracting kinematic information from afferent nervous signals," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 9, pp. 1063–1073, Sep. 2003.

[52] S. E. Hussein and M. H. Granat, "Intention detection using a neuro-fuzzy EMG classifier," *IEEE Eng. Med. Biol. Mag.*, vol. 21, no. 6, pp. 123–129, Nov.-Dec. 2002.

[53] H. N. Teodorescu, A. Kandel, and L. C. Jain, *Fuzzy and neuro-fuzzy systems in medicine*. Boca Raton, FL: CRC, 1999.

[54] J. E. Heiss, C. M. Held, P. A. Estévez, C. A. Perez, C. A. Holzmann, and J. P. Pérez, "Classification of Sleep Stages in Infants: A neuro fuzzy approach," *IEEE Eng. Med. Biol. Mag.*, vol. 21, no. 5, pp. 147–151, Sep.-Oct. 2002.

[55] H. H. Jasper, "The ten-twenty electrode system of the international federation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 10, pp. 371–375, 1958.