Designing an Interval Type-2 Fuzzy Logic System for Handling Uncertainty Effects in Brain-Computer Interface Classification of Motor Imagery Induced EEG Patterns

Pawel Andrzej Herman, *Member, IEEE*, Girijesh Prasad, *Senior Member, IEEE* and Thomas Martin McGinnity, *Senior Member, IEEE*

***Abstract*— One of the urgent challenges in the automated analysis and interpretation of electrical brain activity is the effective handling of uncertainties associated with the complexity and variability of brain dynamics, reflected in the nonstationary nature of brain signals such as electroencephalogram (EEG). This poses a severe problem for existing approaches to the classification task within brain–computer interface (BCI) systems. Recently emerged type-2 fuzzy logic (T2FL) methodology has shown a remarkable potential in dealing with uncertain information given limited insight into the nature of the data generating mechanism. The objective of this work is thus to examine the applicability of T2FL approach to the problem of EEG pattern recognition. In particular, the focus is two-fold: *i*) the design methodology for the interval T2FL system (IT2FLS) that can robustly deal with inter-session as well as within-session manifestations of nonstationary spectral EEG correlates of motor imagery (MI), and *ii*) the comprehensive examination of the proposed fuzzy classifier in both off-line and on-line EEG classification case studies. The on-line evaluation of the IT2FLS-controlled real-time neurofeedback over multiple recording sessions holds special importance for EEG-based BCI technology. In addition, a retrospective comparative analysis accounting for other popular BCI classifiers such as linear discriminant analysis (LDA), kernel Fisher discriminant (KFD) and support vector machines (SVMs) as well as a conventional type-1 FLS (T1FLS), simulated off-line on the recorded EEGs, has demonstrated the enhanced potential of the proposed IT2FLS approach to robustly handle uncertainty effects in BCI classification.**

***Index Terms*—Interval type-2 fuzzy systems, brain–computer interface (BCI), electroencephalogram (EEG), uncertainty, pattern recognition, real-time systems.**

I. Introduction

THE capacity to handle uncertain and ambiguous information inherently present in the real-world modeling environments renders fuzzy logic (FL) one of the most commonly exploited soft computing paradigms in pattern recognition [1]. More specifically, fuzzy methodology has enjoyed considerable popularity in the problem domains where multifarious manifestations of uncertainty are particularly difficult to capture analytically. Brain signal pattern recognition, which lies within the focus of this work, serves as the prime instance in this regard. Consequently, the literature reports the abundance of FL based approaches to analysis, interpretation and classification of brain neurophysiologic data, especially electroencephalogram (EEG), e.g. [2]-[7]. This paper addresses a particularly challenging task of discriminating patterns in the EEG signals recorded from subjects performing motor imagery (MI), which has already received some attention in fuzzy pattern recognition [3],[4] [8]-[11]. The outcome of these investigations has intrinsic implications for the broader field of EEG-based brain-computer interface (BCI). The significance of BCI technology is reflected in its contribution to the improvement of living standards for people affected from neuromuscular disorders [12],[13] and, more recently, in its potential to support a range of rehabilitative therapies [14],[15] as well as the growing impact on computer games industry [16]. One of the key challenges in the field of EEG-based BCI is nondeterministic and nonstationary variability of the brain dynamics, reflected in EEG, which renders its interpretation particularly demanding [12],[17]-[19]. Origins of nonstationarities and fluctuations observed in EEG at a range of temporal scales, independent of task-dependent modulations, are numerous, e.g. different states of subject's awareness, varying focus, neurofeedback effects, or even slight changes in electrode positions [20]-[22].

Long-term variability in the EEG correlates of the associated BCI mental task is recognized as one of the key aspect of uncertainty effects detrimental to the inter-session performance of BCI systems [19],[21]-[25]. This problem has been approached with varying degree of success. Shenoy et al. [23] made the first attempt to visualize session-to-session changes in the distribution of EEG features using principal components analysis (PCA) and the class separating hyperplane derived from linear discriminant analysis (LDA). Noticeable inter-session translations and rotations of the

P.A. Herman is with the Dept. Computational Science and Technology, School of Computer Science and Communication, KTH Royal Institute of Technology, SE 100 44 Stockholm, Sweden (e-mail: paherman@kth.se).

G. Prasad is with the Intelligent Systems Research Centre, University of Ulster, Derry BT52 1SA, U.K. (e-mail: g.prasad@ulster.ac.uk).

T.M. McGinnity is with the School of Science & Technology, Nottingham Trent University, NG1 4BU, U.K. (e-mail: martin.mcginnity@ntu.ac.uk).

ellipsoidal clusters of the features representing classes of associated MIs were accounted for by adjusting the parameters of the LDA classifier. Guger et al. [24] performed two-session BCI experiments and reported a significant drop in the BCI performance from the first to the second session. For these reasons it is suggested in [26] that a classifier could be updated at the beginning of each session. A similar strategy was applied in [27], where a neural classifier was re-trained every day and then embedded in the BCI that was used the following day. A more frequent adaptation schedule to overcome challenges associated with inter- and within-session nonstationarity seems plausible using some recent methods in an unsupervised or semi-supervised mode [21],[22].

Motivated by intrinsic potential of FL framework to handle uncertainty effects, there have also been proposed FL system (FLS) based approaches to EEG classification for BCI purposes, e.g. [3],[4],[10],[11],[28]. The promising results reported in these studies have however been obtained on small benchmark data sets with limited capacity to test multi-session performance. Moreover, the literature offers very limited evidence of real-time applicability of fuzzy classifiers in EEG pattern recognition [29][30][58]. The majority of the proposed approaches have been examined retrospectively on previously recorded EEG data without providing much insight into their on-line capabilities and suitability for providing real-time neurofeedback [3],[4],[11]. The specific need for on-line evaluation of BCI systems is largely motivated by the fact that they are supposed to operate in closed loop scenarios and the resultant neurofeedback effects cannot be simulated off-line. In addition, it provides an opportunity to validate the capacity of a BCI system to cope with uncertainty effects manifested on shorter time scales, from trial to trial or even within a trial, which are partially contributed by instantaneous feedback. These effects have mostly been addressed in the realm of on-line adaptive BCI classification [31]-[33]. However, like in the case of adaptive calibration aimed mainly at reducing inter-session nonstationarity, instantaneously adaptive classifiers intended for on-line use are often devised under the assumption of a known feature distribution, which may not be satisfied, thus resulting in lower accuracy. A notable exception is Yoon et al.'s approach [32], which allows for on-line inference and prediction of erroneous or missing target labels under a more flexible non-Gaussian model. Despite promising results reported on limited data in off-line mode only, the real-time on-line evaluation has not yet been performed.

In this work, the problem of both short- and long-term manifestations of uncertainty effects in the interpretation of MI induced EEG patterns for BCI purposes is addressed in the framework of a type-2 FLS (T2FLS) [34],[35]. This recently emerged methodology has been demonstrated to offer enhanced potential for handling uncertainty when compared to classical type-1 FLSs (T1FLSs) in various pattern recognition applications (for review, see [36]). In the context of EEG classification, it is envisaged that the concept of a T2 fuzzy set (T2FS) provides sufficient flexibility for embracing the range of possible EEG dynamical states and nonstationarities within the operating regime under investigation. This line of reasoning motivated the pioneering interval T2FLS (IT2FLS) based approach to brain signal pattern classification originally introduced by Herman et al. [3],[8]. The evaluation presented here primarily aims *i*) to report a more systematic validation of the proposed IT2FLS framework in off-line classification of EEG correlates of the MI of left vs right arm, and *ii*) to verify its applicability to on-line EEG-based BCI with real-time feedback. The focus in the evaluation, carried out on independent sets of data for off-line and on-line cases, is on inter-session performance transfer, particularly relevant for practical BCI applications where frequent re-tuning or adaptation of the system to account for nonstationarity is undesirable. The off-line paradigm allows for an extensive comparison of classification approaches in batch mode whereas the on-line paradigm provides an opportunity to validate the performance of our IT2FLS method in a realistic BCI context with instantaneous feedback effects. In this light, off-line and on-line cases are considered as complementary. In addition, when compared to the lower-scale original study [3], the concept of data-driven IT2FLS design [34],[37], including both rule generation and gradient descent based parameter tuning, is given more attention and subjected to a more detailed analysis. The classification performance of the IT2FLS is compared with the results obtained with popular BCI classifiers [38] such as LDA, kernel Fisher discriminant (KFD) and support vector machines (SVMs), as well as a T1FLS counterpart retrospectively applied to recorded data.

II. Methodology Development

A. Data Description

1) Off-line Discrimination of MI-related EEG Patterns

There are two sets of EEG data utilized in this work. The first data set was obtained from the Institute of Human-Computer Interfaces, Graz University of Technology, and was part of BCI Competition II. The EEG signals were recorded from 3 subjects (S_A, S_B and S_C) in a timed experimental recording procedure where the subjects were instructed to imagine moving the left and the right hand in accordance with a directional cue displayed on a computer monitor (Fig. 1a). Each trial was 8 s in length. A fixation cross was displayed from t=0s to t=3s. The beginning of a trial was marked by an acoustic stimulus at t=2s. Next, an arrow (left or right) was displayed as a cue at t=3s. Therefore the segment of the data recorded after t=3s of each trial is considered as event related and is used for off-line analysis. The recordings were made with a g.tec amplifier and AgCl electrodes over two sessions, each consisting of 140 trials for S_A and 160 trials for S_B and S_C with equal class representations [40]. Two bipolar EEG channels were measured over C3 and C4 locations (two electrodes 2.5cm anterior and posterior to positions C3 and C4) according to the international 10/20 system [41].

The EEGs were sampled at a frequency of 128 Hz and band-pass filtered in the frequency range 0.5–30 Hz. Experimental settings for recording the data under consideration are described in greater detail in [40].

The second EEG data set was acquired at the Intelligent

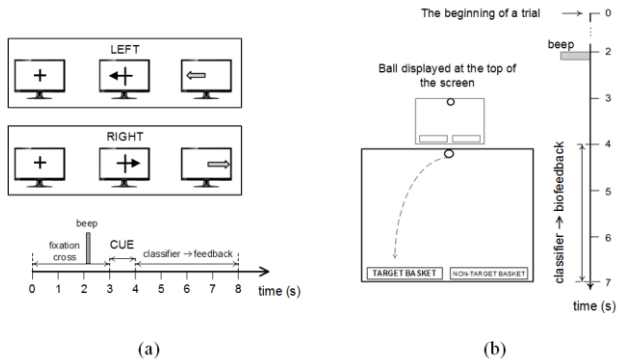

Fig. 1. Data recording in (a) Graz BCI [40] and (b) BCI basket paradigms.

Systems Research Centre, University of Ulster at Magee Campus using the same g.tec equipment as that used by the Graz BCI group. The EEG data were obtained from five subjects (S_I-S_V) over ten 160-trial (balanced) sessions with a week-long break in between the sessions (five sessions without and five sessions with feedback). Four last feedback sessions for each subject were selected for extensive off-line analysis since the first feedback session resulted in inferior quality of the recorded data with extensive artifact contamination in subjects S_I, S_{II} and S_{IV}. Trials were sampled at a frequency of 125 Hz (band-passed in the frequency range 0.5–30 Hz) and those with artifacts were removed. The subjects were asked to imagine moving the left and the right hand depending on the horizontal location (left/right) of a target basket displayed at the bottom of a monitor screen (Fig. 1b). Each trial was 7 s in length. A ball was displayed at the top of the screen from t=0s to t=3s. In the meantime, at t=2s acoustic stimulus signified the beginning of a trial and then the baskets (target in green and non-target in red) were displayed at t=3s. Also at t=3s the ball started moving to the bottom of the screen. The data epoch recorded after t=3s of each trial was analyzed in this work. The horizontal component of the ball movement was continuously controlled in the original on-line experiments by the subject via the biofeedback mechanism. However, in this case study, the comparative EEG analyses for both data sets were carried out off-line with classification done at the end of a trial, resembling the concept of single MI-related EEG trial discrimination.

2) On-line Classification with BCI Neurofeedback

The EEG data used in the on-line case study were obtained from eight healthy participants (S1-S8, three females) in ten sessions (once a week). The recording paradigm and equipment were the same as in the previous study with the recordings collected at the University of Ulster. The only difference lay in the real-time neurofeedback controlled in this case by the IT2FLS classifier in the event-related segment of the data after t=3s. For these on-line experiments, the BCI was deployed in Simulink for real-time capability. The IT2FLS was implemented in ANSI-C, compiled to a .mex file, and then employed in a Simulink block design with native Simulink components for STFT. The BCI implementation was subjected to the timing constraint so that the feature extraction and classification would be carried out faster than the inverse of the sampling frequency, i.e. 1/128 s, to allow for outputting real-time feedback at the rate of 128 Hz.

There were 80 repetitions of each MI per session. The first three sessions were conducted without feedback to allow the subjects to get familiar with the technology and obtain data to calibrate the BCI. In the next sessions, feedback was provided using the basket paradigm, as outlined above.

B. MI-related EEG Features

The EEG feature space was constructed from the time-frequency (t-f) representation of the signal. This approach rests on the fact that the spectral content of the EEG recorded from bipolar channels over C3 and C4 locations when a subject performs imagination of hand movements displays most relevant changes around μ (8-12Hz) and β (18-25Hz) ranges [42],[43]. When the sensorimotor area of the brain is activated as a result of MI, two phenomena can occur – event-related desynchronization (ERD) that performs a band power attenuation and event-related synchronization (ERS) associated with a band power enhancement of μ and central β oscillations [41],[42]. For the subjects examined here, there was μ-ERD on the contralateral side and central β ERS on the ipsilateral side. The minimalistic setup with two EEG channels was also dictated by the need to validate the utilitarian and practical aspects of our on-line BCI system.

1) EEG Feature Extraction for Off-line Discrimination of MI-related EEG trials

The short time Fourier transform (STFT) was applied to obtain the t-f representation of the EEGs analyzed in this study. The event-related segment of each EEG trial was divided into Gaussian windows depending on the settings of two parameters: window length, *win_len* and the amount of overlap, *ovl*, (Fig. 2a). Next, the Fourier transform was applied within the windows. The EEG features were calculated separately for each time window. The frequency components related to ERD and ERS were merged together by estimating the average spectral power in the two bands, μ and β, for each electrode $f_i^{C3}(\mu \cup \beta)$ and $f_i^{C4}(\mu \cup \beta)$ (i – window index, see Fig. 2b). In consequence, a feature vector, $\boldsymbol{F}$, representing the event-related part of a trial (for C3 and C4 channels) had the dimensionality of $2\times N_{win}$, where N_{win} is the number of windows fitted into the relevant signal epoch, determined by the windowing parameters, *win_len* and *ovl* (c.f. Fig. 2b):

$$\boldsymbol{F} = \left[f_1^{C3}(\mu \cup \beta), ..., f_{N_{win}}^{C3}(\mu \cup \beta), f_1^{C4}(\mu \cup \beta), ..., f_{N_{win}}^{C4}(\mu \cup \beta) \right]. \quad (1)$$

Since the exact frequency bands of ERS/ERD vary from subject to subject, the most reactive frequency bands from which to extract features for the given subjects were tuned by optimizing the CA evaluated with an LDA classifier in a five-fold cross-validation procedure. The windowing parameters, *win_len* and *ovl*, were also optimized in an analogous manner. The resulting number of windows, N_{win}, ranged between 3 and 5 across eight subjects in total in the first case study.

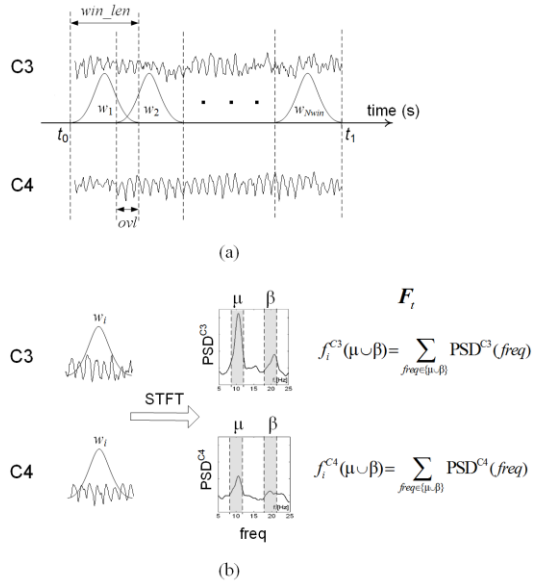

Fig. 2. Spectral feature extraction in the off-line study: (a) Gaussian windowing of the event-related segments (t_0–t_1) in C3/C4; (b) STFT-based quantification of power spectral density (PSD) content in μ and β bands.

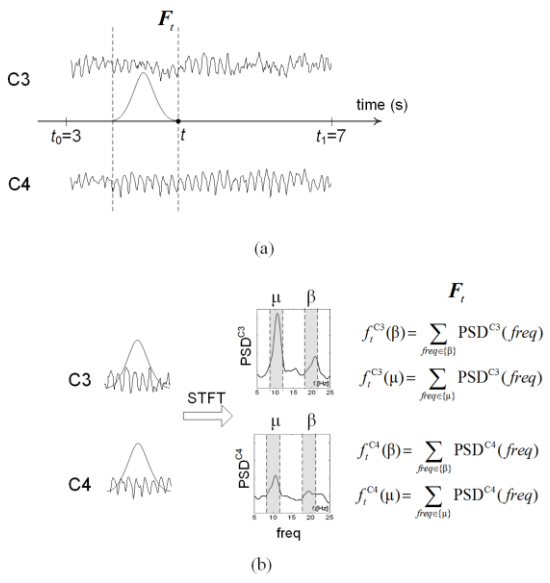

Fig. 3. EEG feature extraction in the on-line BCI: (a) a sliding window with causal feature extraction at each sample *t*; (b) STFT-based quantification of PSD content in μ and β bands (the Euclidean norm of $\boldsymbol{F}_t$ constitutes the input to BCI classifier at each sample *t*).

2) EEG feature Extraction for Continuous On-line BCI Classification

The spectral quantification of ERD and ERS correlates of MI also relied here on STFT. However, there was a considerable difference in how the feature vector was constructed due to the requirement for on-line classification. In particular, feature extraction was carried out at each time point, 3s ≤ *t* ≤ 7s, to provide an instantaneous (within a causal window sliding along a trial on a sample-by-sample basis) estimate of μ and β band power (Fig. 3). Unlike in the off-line classification of full trials, the two spectral components were kept separate in the feature vector as it resulted in better performance on a sample validation set. Consequently, a four-dimensional feature vector $\boldsymbol{F}_t$ (2) was first normalized by its Euclidean norm and then fed to the BCI classifier at time *t*. The window size and the reactive μ and β frequency bands were optimized analogously as in the off-line study case.

$$\boldsymbol{F}_t = \left[f_t^{C3}(\mu), f_t^{C3}(\beta), f_t^{C4}(\mu), f_t^{C4}(\beta) \right]. \qquad (2)$$

C. Fuzzy Classification

The design of the proposed fuzzy classifier was undertaken in two stages. First, an initial structure of a prototype T1FL rule-base was identified and then it was extended to a T2FL rule-base by replacing T1FSs with their T2 counterparts (section II-C-1). In the second stage, the parameters defining membership functions were tuned with a learning algorithm (section II-C-2). The design process is illustrated in Fig. 4.

1) Structure Identification of a Fuzzy Model

The number of the inputs was dictated by the outcome of the feature extraction procedure. An initial rule-base was created to reflect the distribution of the features and their corresponding class assignments in the input-output space due to the supervised nature of the classification problem. There were four schemes of prototype T1FL rule-base initialization in this work: fuzzy c-means clustering (FCM) [44], subtractive clustering [45], mapping-constrained agglomerative (MCA) clustering [39] and its modification discussed in subsection II-C-1. The extension of a prototype T1FL rule-base to the initial T2 fuzzy model is described later in subsection II-C-1b.

1a) Structure identification of a prototype T1 fuzzy model

FCM clustering was first employed as a common approach to fuzzy rule-base identification. The algorithm requires a prior assumption of the number of clusters (from 2 to 10). The input space was clustered and the cluster centers projected on each input dimension served as rule prototypes. The widths of the FSs were calculated as the one-dimensional std.dev. of the subset of the input data points with the membership degree in the corresponding clusters above an arbitrary threshold (0.3–0.8). The consequents were randomized between -1 and 1.

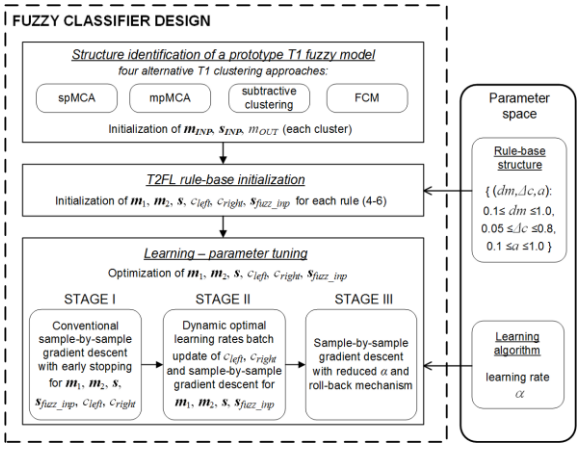

Fig. 4. A flowchart of the data-driven process of the IT2FLS classifier design (rule-base structure identification, IT2FLS initialization and a three-stage parameter tuning process.

Subtractive clustering [45] derived from mountain clustering, proposed by Yager and Filev [46], was also employed in this study. The cluster centres were selected based on the density of data points (feature vectors). The density-related measure had the form of an iterative combination of radial basis functions (radii =0.3–0.5). Analogously to the FCM approach, the neighborhood of each resultant cluster centre was specified to determine the membership status of the clustered data points and then to estimate the corresponding one-dimensional standard deviations. The size of the neighborhood was controlled by a parameter, multiplier of the overall input data spread (0.1–0.5), which facilitated adjustments of the size of overlap between the clusters. The output space assignments were made randomly for the same reasons as in the FCM-based scheme.

In order to reinforce the consistency in the mapping from the input to the output space the MCA algorithm was also employed for the identification of an initial rule-base (the number of seed clusters within the range of 2–10 [39]). This clustering approach allowed for establishing a parsimonious architecture of the fuzzy system. In addition, the MCA scheme has been proven to be robust to noise and outliers that can affect the input-output mapping consistency [39]. However, we observed that the cluster structure obtained with the conventional single-pass (sp) MCA for the EEG features was susceptible to variations in the data ordering. Thus, refinement of the algorithm was necessary. The original spMCA [39] was iterated several times (15–30) with the core input data shuffled and appended with the data points representing means of clusters found in the previous iteration. For each iteration the record of a cluster validity index (described below) was kept. The maximum of the index determined the iteration resulting in the output clusters. This approach, referred to as a multi-pass MCA algorithm (mpMCA), is summarized in Alg.1.

The classification accuracy (CA) rate obtained with an untrained prototype T1 fuzzy classifier devised using cluster means and standard deviations only was used as an index of cluster validity since its correlation with the performance of the trained version of the classifier was observed. This significantly reduced the computational cost of the procedure and allowed for effective selection of the clustering parameters. T1FLS classifier initialization did not require any new parameters and could be directly derived from an output of the clustering algorithm with the same number of rules as clusters. The resulting clusters in the input-output space, defined by the means ($\boldsymbol{m_{INP}}$ and m_{OUT}) and standard deviations ($\boldsymbol{s_{INP}}$, s_{OUT}), determine premise and consequent fuzzy sets of the corresponding T1FL rule-base. Since m_{OUT} corresponds to a crisp class label (-1/1), s_{OUT} is 0 and the consequent part of a T1 fuzzy rule is a singleton centered at m_{OUT}. For easy visualization, an example of the projection of a two-dimensional cluster of data belonging to class C on the axes corresponding to the respective input feature vector components (f_i: $\boldsymbol{F} = [f_1, f_2, .., f_i, .., f_n]$, where n is the number of inputs) and the resulting T1 fuzzy rule (with Gaussian FSs A_i defined by the means $m^{(i)}=m_{INP}{}^{(i)}$ and standard deviations $s^{(i)}=s_{INP}{}^{(i)}$ in the rule antecedent) are illustrated in Fig. 5. This

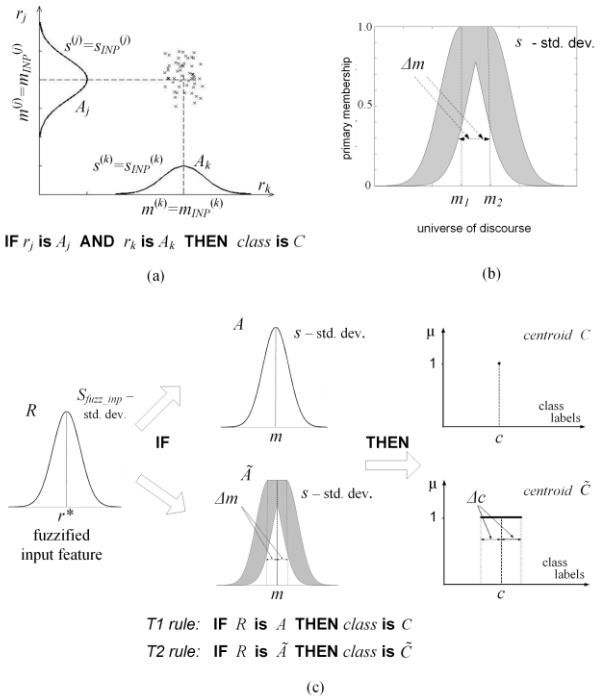

Fig. 5. Fuzzy rule formation, basic components: (a) a two-dimensional cluster in the feature space and the corresponding T1 fuzzy rule; (b) FOU of a Gaussian IT2FS with uncertain mean, $m \in [m_1, m_2]$; (c) an illustrative comparison of a one-rule T2FLS and T1FLS-based classifiers (Δm and Δc define the initial bounds of uncertainty modeled in the system)

method of initializing a T1FL rule-base using a cluster structure of the input data is similar to that employing the FCM and the subtractive algorithm. The only difference lies in setting up the consequents and evaluating the standard deviations of the antecedents. In the FCM and subtractive clustering, the output data space information (class labels) is not exploited and the derived cluster membership degree of the input data points implies the aforementioned heuristic method of estimating the respective clusters' widths.

1b) T2FL rule-base initialization – T1 fuzzy model extension

After initialization of a T1 rule-base, the structure of its T2 counterpart was determined (Alg.1). A template of a Mamdani T2 fuzzy rule exploited in this work is of the following form:

IF R_1 is $\tilde{A}_1$ AND … AND R_n is $\tilde{A}_n$
THEN *class* is C=[c_{left}, c_{right}],

where R_1,..,R_n are the fuzzified components of an input feature vector $\boldsymbol{F}$, and $\tilde{A}_1$,…, $\tilde{A}_n$ denote IT2FSs that model the uncertainty effects in the feature space (antecedent). C is the centroid of the consequent T2FS (interval T1FS for center-of-sets type reduction [34]) representing the class the input feature vector is assigned to. When $\tilde{A}_i$s are replaced by T1FSs and C becomes a crisp centroid of a T1FS, a T2 fuzzy rule reduces to T1 [34]. The inputs to both FLSs are modeled as T1FSs (fuzzification). All FSs are Gaussian to facilitate gradient-based tuning. In particular, the antecedents of the T2 fuzzy rule exploit Gaussian IT2FSs handling uncertainty about the mean, parameterized using the concept of the primary membership function $\mu_A(x)$ [35] defined as

$$\mu_A(x) = \exp(-\frac{(x-m)^2}{2s^2}), m \in [m_1, m_2], \quad (3)$$

where m is an uncertain mean in $[m_1, m_2]$, s is fixed. The 2D domain of support (Fig. 5b) for the secondary membership functions [35] is called the footprint of uncertainty (FOU).

Since the antecedent IT2FSs, $\tilde{A}_i$s, in the proposed T2 fuzzy classifier are described by pairs of the means, $m_1^{(i)}$ and $m_2^{(i)}$, plus the standard deviations $s^{(i)}$ (i=1,..,n) as defined in (3), and the consequents are characterized by two points, c_{left} and c_{right}, additional quantities, $\boldsymbol{\Delta m}$ and Δc, need to be introduced to transform a T1 to a T2 fuzzy rule (Fig. 5c). The antecedents and consequents of each rule were then determined using the following vector formulae:

$$\begin{aligned} &\boldsymbol{m}_1 = \boldsymbol{m}_{INP} - \boldsymbol{\Delta m}, \qquad \boldsymbol{m}_2 = \boldsymbol{m}_{INP} + \boldsymbol{\Delta m}, \\ &\boldsymbol{s} = \boldsymbol{s}_{INP}, \\ &c_{left} = m_{OUT} - \Delta c, \qquad c_{right} = m_{OUT} + \Delta c. \end{aligned} \quad (4)$$

$\boldsymbol{\Delta m}$ was determined with a multiplicative factor dm as follows:

$$\boldsymbol{\Delta m} = dm \cdot \boldsymbol{s}. \quad (5)$$

Finally, $\boldsymbol{s}_{fuzz\_inp}$ used in the description of the fuzzified inputs $R_1,..,R_n$ (Fig. 5c) was set as a scaled vector of the standard deviations of the input features, $\boldsymbol{F}$, in a training set using a scaling factor a:

$$\boldsymbol{s}_{fuzz\_inp} = a \cdot \mathrm{std}(\boldsymbol{F}). \quad (6)$$

Thus the parameters dm, Δc and a, assumed to be homogeneous for the entire rule-base, define the initial bounds of the uncertainty modeled in the system. Their selection was performed in conjunction with the learning process. In particular, the parameters were found using an extensive grid search ($0.1 \le dm \le 1.0$, $0.05 \le \Delta c \le 0.80$, $0.1 \le a \le 1.0$) with the aim of optimizing the performance of the classifier in a five-fold cross-validation setup on the given training session.

It should be mentioned that product t-norm (with meet operator under product t-norm) as part of the FLS specification was employed [34]. As mentioned, the centre-of-sets type reduction was used using Karnik-Mendel algorithm [34], followed by centroid defuzzification in the recall by taking the average of the type-reduced set (interval T1FS).

Alg. 1. Pseudocode of the mpMCA-based rule-base initialization scheme for T1FLS and IT2FLS classifiers

Use the training data as a set of n-dimensional feature vector, $\boldsymbol{F}$, and class label pairs, D_{init},={($\boldsymbol{F}$, class)}:

BEGIN

FOR *iter*=1 TO N_{passes} (N_{passes}=15–30)

- *Run the standard MCA method on reshuffled D_{init} and produce clusters (for each cluster: $\boldsymbol{m}_{INP}$ – cluster center, $\boldsymbol{s}_{INP}$ – cluster spread, and c_{label} – the assigned class).*
- *Build a prototype Mamdani-type T1FLS with Gaussian antecedent FSs and centroid consequents (each cluster k gives a single T1FLS rule, c.f. Fig.5):*

$$\boldsymbol{m}_k = \boldsymbol{m}_{INP_k}, \ \boldsymbol{s}_k = \boldsymbol{s}_{INP_k}, \ \boldsymbol{c}_k = \boldsymbol{c}_{label_k},$$

- *Classify samples in D_{init} using the T1FLS and report the CA as a cluster validity index(iter)*
- IF *cluster validity index(iter)> max CA*
  - *Update max CA = cluster validity index(iter)*

  END_IF
- *Append the data with the cluster centers and their labels:* $D_{init} = D_{init} \cup \left\{ \left(\boldsymbol{m}_{INP_k}, \boldsymbol{c}_{label_k} \right) : k - \text{cluster index} \right\}$.

END_FOR

Select the T1FLS configuration that has maximized CA (T1FLS with max cluster validity index over N_{passes} of the algorithm) and extend it to IT2FLS with Gaussian antecedent FSs (3) according to (4–6) using parameters dm, Δc and a.

END

2) Learning Algorithm

After setting up an initial rule-base, the quantities such as $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, $\boldsymbol{s}$, c_{left}, c_{right} and $\boldsymbol{s}_{fuzz\_inp}$ were tuned for every rule. The learning algorithm followed the concept of gradient descent, one of the most popular techniques used to adapt the parameters of IT2FLSs [47], with the loss function L in the form of the mean square error (7) to be minimized.

$$L = \frac{1}{N}\sum_{i=1}^{N} e_i^2 = \frac{1}{N}\sum_{i=1}^{N} \left(y_i - c_i \right)^2, \quad (7)$$

where N is the number of training samples, y_i is the defuzzified real-valued crisp output of the fuzzy classifier and c_i is the desired class label (-1 or 1) for the i-th sample $(e_i = y_i - c_i)$. In particular, following the conventional gradient descent idea, the aforementioned IT2FLS parameters, $\boldsymbol{P} = \left[\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{s}, c_{left}, c_{right}, \boldsymbol{s}_{fuzz\_inp} \right]$, were updated iteratively sample-by-sample based on the first derivatives of the L function, $\frac{\partial L}{\partial P}$, with a positive learning rate α:

$$\boldsymbol{P}_{l+1} = \boldsymbol{P}_l - \alpha \frac{\partial L}{\partial \boldsymbol{P}} |_{\boldsymbol{P}=\boldsymbol{P}_l}. \quad (8)$$

The training method proposed here consists of three stages and combines two popular approaches in the domain of IT2FLSs, the conventional steepest gradient descent algorithm developed by Liang and Mendel [35] and the method based on the dynamic optimal rate theorem [37]. This hybridization led to more robust and effective searching of a multimodal space for an optimal configuration of the system parameters than conventional Liang and Mendel's approach in our case where generalization properties are emphasized (see section III-B).

In the first training stage, the conventional steepest descent was applied with learning rates, α, reduced by factor of 2 every ten epochs. Their initial values were part of the model

identification with the grid search ($0.05 \leq \alpha \leq 0.20$). The update equations for the parameters $\boldsymbol{P}$ were derived from (8) as explicitly shown in [34],[35],[48].

An algorithm based on the dynamic optimal rate theorem was applied in the second phase with a view to speeding up the convergence of tuning the consequents of the IT2FLS. It involved the combination of sample-by-sample training of standard deviations $\boldsymbol{s}_{fuzz\_inp}$ and antecedent parameters $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, $\boldsymbol{s}$ with a batch update of consequents c_{left}, c_{right}, using the same first derivative based formulae as in the first learning stage (conventional gradient descent [34],[35],[48]). The learning rates for $\boldsymbol{s}_{fuzz\_inp}$, $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, $\boldsymbol{s}$ were kept constant at the level inherited from the first learning stage whereas the optimal learning rates, $\alpha_{left}^{(\mathrm{opt})}$ and $\alpha_{right}^{(\mathrm{opt})}$, for c_{left} and c_{right}, were calculated every epoch according to the formulae derived in [37] based on the stability criterion for Lyapunov function defined as L^2. The batch updates for c_{left}, c_{right} at epoch j were:

$$c_{left}(epoch_{j+1}) = c_{left}(epoch_j) - \alpha_{left}^{(opt)}(epoch_j)\frac{\partial L}{\partial c_{left}}|_{c_{left}=c_{left}(epoch_j)} \qquad (9)$$
$$c_{right}(epoch_{j+1}) = c_{right}(epoch_j) - \alpha_{right}^{(opt)}(epoch_j)\frac{\partial L}{\partial c_{right}}|_{c_{right}=c_{right}(epoch_j)}$$

with the rates estimated at each epoch (index j is dropped for clearer notation) according to [37]:

$$\alpha_{left}^{(opt)} = 4N\frac{\boldsymbol{r}^{\mathrm{T}}\boldsymbol{r}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{l} - \boldsymbol{l}^{\mathrm{T}}\boldsymbol{r}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{r}}{\boldsymbol{l}^{\mathrm{T}}\boldsymbol{l}\boldsymbol{r}^{\mathrm{T}}\boldsymbol{r} - \left(\boldsymbol{l}^{\mathrm{T}}\boldsymbol{r}\right)^2}, \quad \alpha_{right}^{(opt)} = 4N\frac{\boldsymbol{l}^{\mathrm{T}}\boldsymbol{l}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{r} - \boldsymbol{l}^{\mathrm{T}}\boldsymbol{r}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{l}}{\boldsymbol{l}^{\mathrm{T}}\boldsymbol{l}\boldsymbol{r}^{\mathrm{T}}\boldsymbol{r} - \left(\boldsymbol{l}^{\mathrm{T}}\boldsymbol{r}\right)^2}, \qquad (10)$$

where $\boldsymbol{e} = \left[e_1, e_2, ..., e_N\right]$ is the column error vector for all N samples, and the column vectors $\boldsymbol{l}$, $\boldsymbol{r}$ are defined as [37]: $\boldsymbol{l} = \boldsymbol{Q}_l^{\mathrm{T}}\boldsymbol{Q}_l\boldsymbol{e}$ and $\boldsymbol{r} = \boldsymbol{Q}_r^{\mathrm{T}}\boldsymbol{Q}_r\boldsymbol{e}$, with $\boldsymbol{Q}_l = \left[\boldsymbol{q}_l^{(1)}, ..., \boldsymbol{q}_l^{(N)}\right]$ and $\boldsymbol{Q}_r = \left[\boldsymbol{q}_r^{(1)}, ..., \boldsymbol{q}_r^{(N)}\right]$ consisting of N (for each sample) so-called left and right, respectively, rule firing strengths column vectors, $\boldsymbol{q}_l$ and $\boldsymbol{q}_r$, obtained in the T2 fuzzy inference (type reduction using an iterative Karnik-Mendel method [49]).

In the third step, the same conventional gradient descent algorithm as in the first stage was utilized to fine tune IT2FLS with the learning rates reduced by a factor of 5. The resulting setup of the system parameters was accepted only if the classifier's training accuracy improved in comparison with the outcome of the second stage. Otherwise, the parameter configuration was rolled back. This last learning phase led to the significant enhancement of the IT2FLS performance on the validation set (20% of data available for setup) in over 60% of cases. The early stopping criterion was applied on the validation set at all stages of the learning algorithm in order to enhance generalization capabilities.

An analogous learning algorithm was developed for a T1FLS classifier (with dynamical optimal training for consequents [50]) in order to carry out a fair comparative analysis. The efficacy of the last phase of the hybrid learning algorithm was at a comparable level to that for the IT2FLS.

Alg. 2. Pseudocode of the proposed three-stage hybrid algorithm for gradient descent based IT2FLS optimization.

BEGIN

Stage I:

FOR *epoch*=1 TO max_epochs_{I} (=100)
- FOR *iter*=1 TO $N_{samples}$ in D_{trn}
  - *Update* $\boldsymbol{P}$ *with learning rate* α_I *according to (8)*
  END_FOR
- *Estimate IT2FLS valid accuracy,* CA_{vld} *(epoch), on* D_{vld}
- IF $CA_{vld}(epoch) < CA_{vld}(epoch\text{-}1)$
  - $\boldsymbol{P}(epoch) = \boldsymbol{P}(epoch\text{-}1)$
  - GOTO *Stage II*

  END_IF
- *Reduce learning rate* α_I *every few epochs*

END_FOR

Stage II:

FOR *epoch*=1 TO max_epochs_{I} (=50)
- FOR *iter*=1 TO $N_{samples}$ in D_{trn}
  - *Update* $\boldsymbol{P}$ *with learning rate* α_{II} *according to (8)*

  END_FOR
- *Estimate optimal* $\alpha_{left}^{(opt)}$, $\alpha_{right}^{(opt)}$ *for* c_{left}, c_{right} (10)
- *Perform batch update of* c_{left}, c_{right} *according to (9)*
- *Estimate IT2FLS valid accuracy,* CA_{vld} *(epoch), on* D_{vld}
- IF $CA_{vld}(epoch) < CA_{vld}(epoch\text{-}1)$
  - $\boldsymbol{P}(epoch) = \boldsymbol{P}(epoch\text{-}1)$
  - GOTO *Stage III*

  END_IF

END_FOR

Stage III:

FOR *epoch*=1 TO $max\_epochs_{\mathrm{III}}$ (=15)
- FOR *iter*=1 TO $N_{samples}$ in D_{trn}
  - *Update* $\boldsymbol{P}$ *with learning rate* α_{III} *according to (8)*

  END_FOR
- *Estimate IT2FLS valid accuracy,* CA_{vld} *(epoch), on* D_{vld}
- IF $CA_{vld}(epoch) < CA_{vld}(epoch\text{-}1)$
  - $\boldsymbol{P}(epoch) = \boldsymbol{P}(epoch\text{-}1)$
  - GOTO *Rollback_Condition*

  END_IF

END_FOR

Rollback_Condition:

IF $CA_{vld} < CA_{vld}^{II}$ (CA_{vld}^{II} *is* CA_{vld} *at the end of Stage II)*
- *Roll back* $\boldsymbol{P}$ *updates in Stage III, return* $\boldsymbol{P}$ *from Stage II*

END_IF

END

3) Other BCI Classifiers

The IT2FLS and T1FLS designed in this work were compared to four other classifiers commonly used in the domain of MI-based BCI (e.g. [23],[38]) – LDA, KFD with a Gaussian kernel and two SVM methods, linear ($\mathrm{SVM}_{\mathrm{lin}}$) and nonlinear with a homoscedastic Gaussian kernel ($\mathrm{SVM}_{\mathrm{Gauss}}$). The implementation of SVM classifiers rested on a quadratic programming problem solver adopted from the SVM-KM toolbox [51]. Regularization parameters for SVMs as well as the variance of the Gaussian kernel for $\mathrm{SVM}_{\mathrm{Gauss}}$ and KFD were identified by maximizing the performance for the validation set, as for the fuzzy classifiers.

D. Performance Measures

The aim of the classification procedure was simply to assign the entire EEG trials (off-line study case) or their short snippets (continuously sliding windows in the on-line study) to classes of the associated mental tasks – imagination of either left or right hand movement. Since the BCI classifiers produced a real-valued output, the class assignment was made by thresholding the output. Then, the proportion of correct MI class assignments determined the CA. Although it was straightforward to quantify the overall performance for full trial recognition in the off-line case study, in continuous mode of operation of an on-line BCI system it became more arbitrary as the CA could be measured as a function of time. Here, the time course of the CA was ultimately quantified by its maximum value, mCA (cf. Fig. 9). The corresponding time point is referred to as classification time (cf. Fig. 9) and is relevant to BCI performance – the shorter this time is, the faster BCI communication can be. This communication aspect of BCI is reflected in the information transfer rate (ITr), measured in bits per minute, which serves as the upper bound for a practically achievable communication throughput of the BCI system [12]. Additionally, the information content of the real-valued output of the on-line BCI classifier, $y(t)$, in relation to the target label of MI class was estimated using the mutual information (MInf) measure [52]:

$$\mathrm{MInf}(t) = 0.5\log_2\left(\frac{2\,\mathrm{var}\left(y(t)\right)}{\mathrm{var}\left(y(t)\,|_{LEFT}\right)+\mathrm{var}\left(y(t)\,|_{RIGHT}\right)}\right), \quad (11)$$

where $\mathrm{var}\left(y(t)\,|_{LEFT/RIGHT}\right)$ is the class dependent variance of the classifier's responses at time t. MInf offers more insight into the credibility of the output class assignments (labels).

III. Results and Discussion

A. Analysis of Uncertainty Effects in the Session-To-Session Feature Distribution

We first investigated the nature of session-to-session variations in the EEG feature distributions. To this end, an intuitive and illustrative approach was adopted by representing the features in the reduced PCA space. In particular, PCA was performed on the EEG features extracted from data obtained in one session as the reference. Then the features from the consecutive sessions were projected according to the new set of directions of the largest variance. For visualization, only the first two components accounting for at least 70% of the total variance were used (Fig. 6).

The shift between the means representing the same MI from one session to another is noticeable. Although this effect can vary in size, most feature sets in the study exhibit considerable inter-session changes comparable to those shown for S_{III} and S_V (Fig. 6). Moreover, there does not seem to be any clear deterministic mechanism explaining the direction of the data translation. This incoherent nature of inter-session changes in the feature distribution gave rise to the problematic issue of uncertainty, and resulted in rather limited session-to-session classification performance (Fig. 6), especially with a linear method, LDA. The fuzzy classifiers, especially IT2FLS, produced more accurate rates as they handled variability in local cluster structures.

B. Analysis of Fuzzy Rule-Base Initialization and Learning Schemes in Off-line Study

To begin with, a comparative analysis of different initialization (mpMCA, spMCA, subtractive and FCM clustering; see section II-C-1) and learning (our three-stage hybrid and conventional gradient descent method [35]) schemes for IT2FLS and T1FLS rule-bases was conducted on an arbitrarily selected subset of three subjects and two sessions from the off-line study. Firstly, ten runs of five-fold cross-validation on the training data set (the first one of the two selected sessions) were carried out. Next, the same data was utilized to tune/train fuzzy classifiers and the subsequent recording session for each of the three subjects considered in this analysis served as a test data set. This one-pass training-test procedure was repeated ten times to examine the consistency of different combinations of initialization and learning schemes. Table I presents the mean CAs and the average standard deviations over ten runs across the chosen subjects. Although it is hard to draw any statistically significant conclusions from Table I, several valuable observations can be made. The outcome of this analysis is

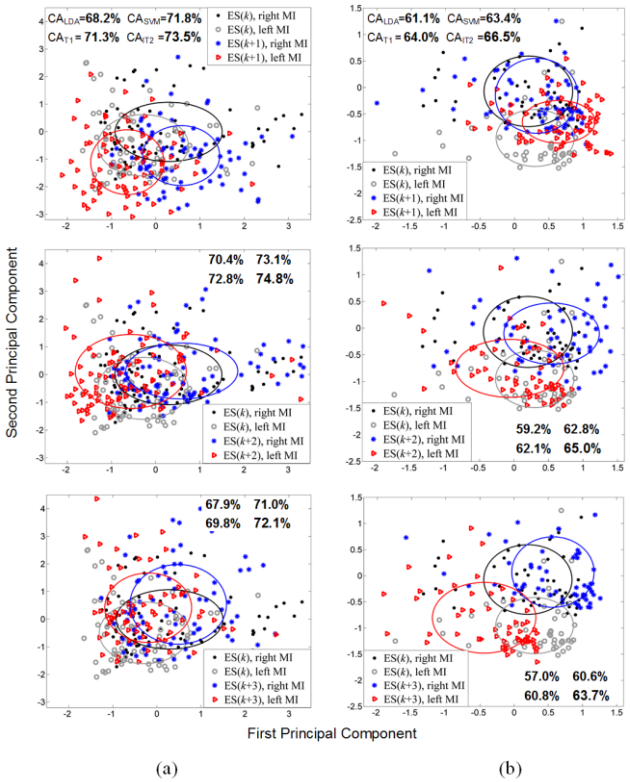

Fig. 6. The distribution of EEG spectral features in a two-dimensional normalized PCA space (over 70% of variance explained in the reference experimental session, ES(k)), complemented with ellipses surrounding the MI class means and radii corresponding to the standard deviations, in four consecutive sessions ES(k)...ES(k+3) for (a) S_{III} and (b) S_V in the off-line study. The top row illustrates a pair of sessions: ES(k) and ES(k+1); middle row: ES(k) and ES(k+2); bottom row: ES(k) and ES(k+3). The results of the corresponding session-to-session test classifications are also included in the panels (LDA, Gaussian SVM, T1FLS, IT2FLS).

influential particularly for the IT2FLS-based BCI classifier. At first, the MCA-based clustering algorithms were examined in the framework of T1FLS and IT2FLS. As expected, mpMCA considerably reduced the inter-run variance and improved the performance of the resultant classifiers, both T1FLS and IT2FLS, when compared to spMCA.

Then, two unsupervised methods often employed in the classical T1FLS structure identification, FCM and subtractive clustering, were validated for both fuzzy classifiers. They resulted in a comparable level of performance of the target fuzzy classifiers but at the cost of higher run-to-run variations and slower convergence. The need to specify a priori the number of clusters for FCM was also considered inconvenient. In conclusion, it was decided to rely on the fast and robust mpMCA implementation for the rule-base initialization. As for the hybrid learning approach, it tended to result in better generalization performance for both fuzzy classifiers than the classical algorithm [35]. On average, it also converged faster with the lower number of training epochs, by a factor of ~1.4 for IT2FLS and ~1.2 for T1FLS (data not shown).

C. Off-line Classification of MI-related EEG Patterns

In the first place, the overall efficacy of the classifiers was estimated and, most importantly, the selection of initial conditions (*dm*, *Δc*, *a* and learning rates for IT2FLS and T1FLS) was made using five-fold cross-validation scheme run ten times on training session data (Table II, column five-fold cross-valid.). The optimized values of initial parameters were employed in the second type of experiments to examine inter-session performance transfer of the BCI classifiers by tuning them on a training session data set (with 20% of data for validation) and testing in a single pass on the subsequent sessions, acting as test data sets. In order to investigate a temporal character of the evolution of the classifiers' performance over the test sessions, all possible combinations of training-test data sets maintaining their sequential order were evaluated for all subjects. In other words, from four sessions, six test and three training CA rates were obtained. The selection of training-test session pairs is shown in Fig. 7.

The overall results in this experimental setup were grouped into three categories reflecting a temporal relationship between a training and a test session. Next, they were averaged within these categories resulting in four mean CAs for every subject. Test results produced from three training-test pairs, session I–II, II–III and III–IV, constituted the first category (Cat.I). The second one (Cat.II) consisted of test CAs from experiments involving session pairs: I–III and II–IV, and the third category (Cat.III) with the largest temporal gap between a training and a test data set was composed of one CA result obtained with a classifier trained on session I and tested on session IV. There were only the first and the second category of the experimental results evaluated for subjects S_A, S_B and S_C since only two data sets were available, session I and II. The last three columns of Table II report the results in the respective categories averaged across the subjects.

The CA results reported in Table II were processed independently in each column in the statistical framework of one-way analysis of variance (ANOVA) with repeated measures. No significant differences between the mean CAs were found for the results obtained with the cross-validation approach, which demonstrates the overall within-session performance of the classifiers without accounting for inter-session uncertainty effects. Special attention was drawn to session-to-session test results presented in the last three columns of Table II. They reflect the capability of the classifiers to effectively account for the inherent inter-session variability of the MI induced EEG patterns. Hence, they are particularly relevant to the BCI research community. The ANOVA carried out on these sets of CA rates revealed statistically significant ($p<0.05$) differences in the classifiers' performances. Tukey's honestly significant difference

TABLE I
COMPARATIVE ANALYSIS OF INITIALIZATION AND LEARNING SCHEMES FOR IT2FLS AND T1FLS ON A SUBSET OF TWO SESSIONS AND THREE SUBJECTS

	Three-stage hybrid approach			Conventional gradient descent		
	Session I	Session II		Session I	Session II	
	CA ± std.dev. [%] [a]		N_{rules} [b]	CA ± std.dev. [%] [a]		N_{rules} [b]
	five-fold cross-valid.	test		five-fold cross-valid.	test	
IT2FLS						
mpMCA	68.0 ± 6.4	68.3 ± 1.1	7.0	67.4 ± 8.4	67.0 ± 2.0	6.7
spMCA	66.9 ± 7.4	64.9 ± 4.6	5.9	67.3 ± 9.3	64.0 ± 4.7	5.9
FCM	68.7 ± 8.7	67.3 ± 2.1	6.0	66.3 ± 8.9	65.3 ± 2.8	6.0
subtractive	69.0 ± 7.7	67.9 ± 2.0	7.0	67.8 ± 8.3	66.8 ± 2.4	7.2
T1FLS						
mpMCA	67.3 ± 6.7	64.9 ± 1.2	6.7	66.8 ± 9.1	64.4 ± 1.7	6.8
spMCA	65.9 ± 7.9	62.4 ± 5.3	5.5	65.5 ± 9.3	58.5 ± 2.6	6.0
FCM	68.0 ± 8.8	65.3 ± 2.9	6.0	66.1 ± 9.1	59.4 ± 3.7	6.0
subtractive	68.2 ± 7.7	65.1 ± 2.3	7.9	66.9 ± 8.1	64.2 ± 3.2	8.0

[a] CA and std.dev. were estimated across ten repetitions for both five-fold cross-validation and test evaluations.
[b] N_{rules} stands for the average number of rules.

TABLE II
MULTI-SESSION PERFORMANCE OF THE FUZZY CLASSIFIERS, SVMS, KFD AND LDA ACROSS EIGHT SUBJECTS [A]

Classifier	Training Session	Test Cat.I	Test Cat.II	Test Cat.III
	CA ± std.dev. [%]			
	five-fold cross-valid.	single-pass tests (inter-session)		
IT2FLS	71.2 ± 8.4	**73.4** ± 9.0 [b]	**64.8** ± 6.7	**65.4** ± 6.7
T1FLS	70.4 ± 8.3	71.8 ± 9.1	63.6 ± 6.3	63.9 ± 7.5
LDA	**71.5** ± 8.4	67.5 ± 9.3	61.8 ± 8.0	60.7 ± 7.1
KFD	70.9 ± 8.9	69.3 ± 8.8	61.9 ± 8.6	60.6 ± 7.7
SVM_{lin}	71.1 ± 9.3	69.8 ± 9.9	61.7 ± 7.5	60.3 ± 6.4
SVM_{Gauss}	71.0 ± 9.3	69.7 ± 9.8	61.8 ± 7.0	60.4 ± 6.9

[a] Eight subjects: S_A, S_B, S_C and S_I – S_V (c.f. section II-A-1).
[b] Best results in their categories are marked in bold.

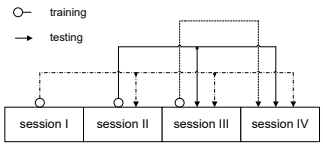

Fig. 7. Illustration of the inter-session experimental design for four-session data (lines connect training sessions with the corresponding test sessions).

criterion (at the significance level of α=0.05) was used to conduct the post test simultaneous comparison of the mean CAs in each category of the test results. In consequence, IT2FLS was found to outperform LDA, KFD, SVM_{lin} and SVM_{Gauss} for all categories of the test CA results (Table II, columns Test). When T1FLS was considered, the only non-significant difference (α=0.05) was detected in the second category of the test results, i.e. when the classifiers were trained on the session directly preceding the test session. The T1FLS classifier was shown then to perform better than LDA, which delivered the poorest CA rates (Table II, column Test – cat. II). Overall, the proposed fuzzy approaches proved to possess better capabilities in accounting for the inter-session uncertainty effects in MI-related EEG discrimination. Although the differences between the corresponding mean CA rates delivered by IT2FLS and T1FLS classifiers were not shown to be statistically significant in the ANOVA framework, the superiority trend of the IT2FLS approach was found consistent for every category of the presented results across all the subjects. This clearly demonstrates the potential of the designed IT2FLS in offering enhanced robustness against the variability of brain dynamics reflected in EEG. The ability of the IT2FLS classifier to learn inter-trial uncertainty

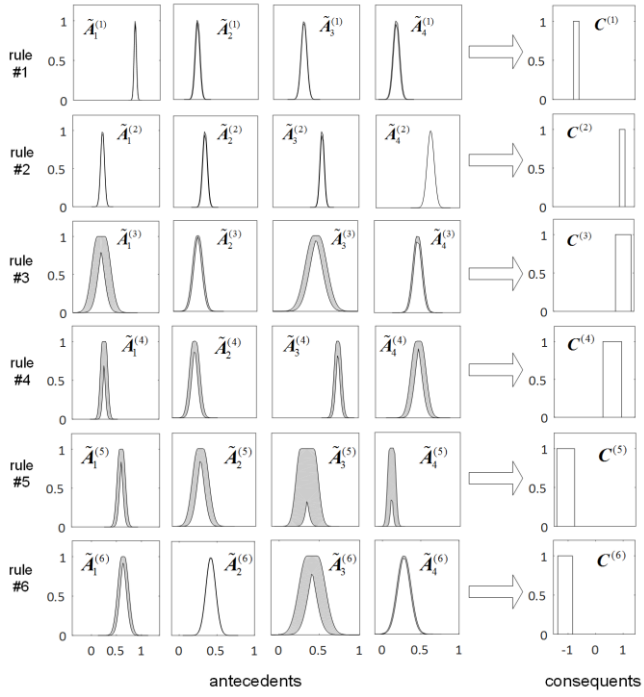

Fig. 8. Fuzzy rule base structure of the IT2FLS classifier trained on session 8 for subject S6. The order of antecedent T2FSs, $\tilde{A}_k^{(r)}$ (k – antecedent index, r – rule index), matches the structure of the input $\boldsymbol{F}$ (c.f. (4)). The consequents, $C^{(r)}$, are T2 centroids in the form of interval T1FSs.

in the EEG feature patterns over a training session and encapsulate it within the FOU for use on subsequent test sessions is considered to be crucial in this regard.

The computing time involved in the IT2FLS optimization was comparable to that of the T1FLS classifier due to rather small rule-base sizes used in the given application (c.f. Table I). Attempts to increase the number of rules in the FLS led to significant deterioration of their performance. It was also observed that the computational time required for training larger IT2FLSs grew exponentially in relation to that of analogous T1FLSs. Finally, a concise structure of the FLSs devised in this work facilitates potential approaches to interpretation of the classification rules.

D. On-line BCI Classification with IT2FLS

In the case study involving on-line BCI classification where the IT2FLS was employed to control the real-time neurofeedback, the evaluation was concentrated on the last 5 sessions (out of 7 feedback sessions; sessions no. 5–10), for which the classifier remain unchanged with the exception of two subjects, S1 and S5. In their case, the IT2FLS had to be re-trained prior to sessions no. 8 and 9, respectively, following poor performance in the preceding session (training error ~40%) due to problems with the recording equipment with negative impact on the subjects' concentration. For the classifier's re-calibration, the recordings collected in that weaker session for a given subject (no.7 and 8) were combined with those obtained in the previous session (no.6 and 7, respectively) to form a training/validation data set. For all the subjects, the process of initial calibration of the classifiers was performed on the data collected during the two first feedback sessions (no.4–5), which were not considered in the evaluation. The training/validation procedure with the selection of initial parameters was analogous to computations performed in the off-line case study. In parallel with the evaluation of the on-line IT2FLS, other classifiers, i.e. LDA, KFD, SVM and T1FLS, were applied retrospectively to the recorded EEG data to simulate feedback in the off-line setting for post hoc comparative analysis. Their training, validation and testing followed the same schedule as the on-line IT2FLS. For this reason, the results of the comparative analyses should be treated with caution. After all, the on-line classifier directly benefits from the subject's adaptation during BCI sessions.

An example of an IT2FLS rule-base of an on-line BCI classifier is shown in Fig. 8. This six-rule FLS was designed for S6 on session 8 data. Varying levels of uncertainty captured by antecedents in different rules are worth noting. Further, the scheme for evaluating on-line BCI performance in a session is exemplified in Fig. 9. It depicts the time course of the CA during the event-related part of a trial. A steady nature of the CA increase is worth noting.

The average BCI performance, in terms of mCA, ITr and MInf, for each subject (S1-S8) across five neurofeedback sessions is shown in Fig. 10. As can be seen, the average performance of the IT2FLS is consistently higher for each subject, though with different effect sizes, when compared to the other BCI classifiers simulated off-line retrospectively on

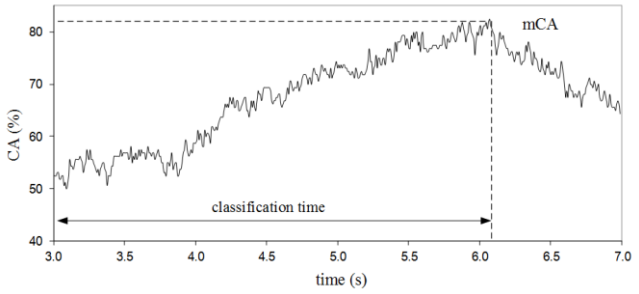

Fig. 9. The time course of the CA estimated over a set of trials within a recording session no.8 for S6 with the markings of the mCA and the optimal classification time.

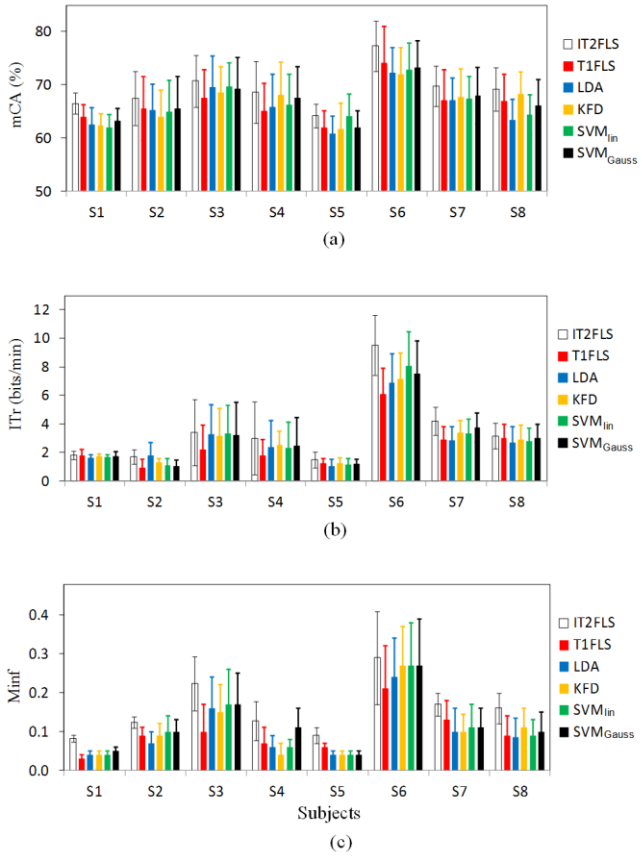

Fig. 10. Comparison of the average on-line BCI performance of the IT2FLS classifier with the off-line performance of retrospectively simulated T1FLS, LDA, KFD, SVM_{lin} and SMV_{Gauss} in terms of (a) the CA, (b) the ITr, and (c) the MInf. The mean values and std. dev. (error bars) were estimated for each subject, S1-S8, across their last five neurofeedback sessions.

the recorded data. It is worth noting that the performance level varies across the subjects, which is particularly evident in the evaluation of the ITr and MInf. They account for the speed of communication and the cross-trial distribution of the real-valued outputs produced by the classifiers, respectively. The dominance in the performance of subjects S6 and S3 should be mentioned in this regard.

The grand average (across the population of eight subjects) performance indicators, shown in Table III, confirm that the proposed IT2FLS compares favorably with other methods tested in this study. The results were tested with ANOVA and, after the null hypothesis was rejected ($p<0.01$), the Tukey-Kramer test [53] (at the significance level of $\alpha=0.05$) was used

TABLE III
BCI PERFORMANCE (MCA, ITR, MINF) OF THE IT2FLS, T1FLS, LDA, KFD, SVM_{LIN} AND SVM_{GAUSS} CLASSIFIERS, AVERAGED OVER FIVE BCI NEUROFEEDBACK SESSIONS FOR EIGHT SUBJECTS

Classifiers	mCA ± std.dev.	ITr ± std.dev.	MInf ± std.dev.
On-line IT2FLS[a]	**69.2** ± 3.8[b]	**3.6** ± 2.6	**0.16** ± 0.07
T1FLS	66.5 ± 3.7	2.5 ± 1.6	0.10 ± 0.05
LDA	65.8 ± 3.7	2.8 ± 1.8	0.10 ± 0.07
KFD	66.6 ± 3.6	2.9 ± 1.9	0.10 ± 0.08
SVM_{lin}	66.5 ± 3.4	2.9 ± 2.2	0.11 ± 0.08
SVM_{Gauss}	66.9 ± 3.5	3.0 ± 2.1	0.12 ± 0.07

[a]The IT2FLS was used to control neurofeedback in on-line BCI whereas the remaining classifiers were simulated off-line on previously recorded data.
[b]The mean values and std. dev. were estimated across subjects S1–S8.

for post-hoc pair-wise comparisons between the classifiers. The parametric analysis was conclusive only for the mCAs since the conditions underlying parametric tests for ITr and MInf results were not met. Instead, nonparametric Kruskal-Wallis test was employed at $\alpha=0.05$. However, the comparison of the ITr and MInf statistics turned out to be inconclusive due to nonsignificant outcome of the test.

Overall, the IT2FLS was shown to deliver statistically significant higher mCAs than the rest of classifiers. In addition, a planned comparison between the two fuzzy classifiers, IT2FLS and T1FLS, was conducted for each subject (across five feedback sessions) using Student's t-test. The superiority of the enhanced apparatus of the IT2FLS approach for handling uncertainty effects in the classification of MI induced EEGs for all the performance measures and subjects was clear ($p<0.05$), though the fact that T1FLS could only be tested post-hoc off-line has to be highlighted.

Finally, the study participants were asked to comment on the facilitating role of the neurofeedback provided through on-line IT2FLS classification and the majority expressed positive opinions. Subject S5 was an exception as he complained about the inconsistency and low correlation of the BCI responses with his mental efforts (MI). The rest however emphasized rather smooth nature of the BCI output controlling the horizontal displacement of the ball in the basket. The retrospective qualitative analysis of the continuous neurofeedback signal generated by the IT2FLS was consistent with the subjects' opinions. Also, there were no problematic issues with the real-time constraints of the IT2FLS implementation and for the clear majority of the subjects no extra calibration between the recording sessions was needed (except the initial setup before session no.6, i.e. the first evaluation session).

E. Robust Approaches to Handling Nonstationarity Effects in EEG-based MI BCI

The most prevalent BCI approaches to the problem of short- as well as long-term EEG nonstationarity fall into two major categories. The objective of methods in the first group is to identify stable aspects of the EEG dynamics and thus enhance the model's invariance capabilities. For example, Blankertz et

al. [25] developed an invariant version of well-established common spatial patterns EEG features, which improved cross-session performance by 5% per subject in a small two-session study. The second group of approaches relies on adaptive mechanisms in the presence of nonstationary phenomena. It has been proposed to adapt a BCI classifier relying on supervised [57] or unsupervised learning techniques [21],[31]. Unsupervised approaches with covariate shift adaptation methods in the first place attract growing attention due to their practical value. Sati et al. [54] reported the increase of the average CA from 73.9% to 78.6% as a result of applying covariate shift minimisation along with common spatial patterns and LDA. The nine-subject study was carried out on a five-session data set. However, the relevant inter-session evaluation was limited to pair configurations with the first training session and the remaining four test sessions. Furthermore, Vidaurre et al. [31] confined their comparative analysis of supervised and unsupervised LDA adaptation routines just to multiple independent pairs of calibration and feedback sessions. In various tests, the performance of the better unsupervised approach ranged from ~72% to 85% on average. Apart from the adaptation of BCI classifiers, there have also been attempts to adapt the feature space with the aim of minimizing the mismatch between different recording sessions [22],[55]. Arvaneh et al. [22] proposed a data space adaptation technique that could be used in either supervised or unsupervised mode with single or continuous trial-by-trial adaptation, and demonstrated its superior performance over non-adaptive methods with the average accuracy of ~72% in a two-session evaluation.

Although the aforementioned off-line BCI evaluations with or without simulated feedback have clearly dominated the literature, it is desirable that EEG nonstationarity effects for BCI systems should be examined in on-line studies with real-time feedback to fully understand their scope and the challenging nature. In early work Brunner et al. [56] trained a linear classifier on the off-line calibration data to later provide on-line feedback in consecutive MI BCI sessions without making any adaptations. As a result, they reported CA oscillating between 49% and 54% for the first two sessions and 60-67% for the third one. In a later study, Vidaurre et al. [57] tested on-line BCI performance of a non-adaptive LDA classifier trained on non-feedback session data. The on-line CA in two-class MI task did not exceed 60% over multiple sessions. Nonstationary effects were mitigated with an adaptive LDA, allowing the performance to grow to 74-84% [57]. Similarly, high CA rates (77%-87%) were reported in [58] with the on-line use of a probabilistic adaptive classifier in the virtual reality environment supporting MI practice.

F. Related work on FLS approaches to BCI classification

One of the first attempts at classifying MI EEGs for BCI purposes with a FLS was reported to result in the three-data-set average session-to-session CA of 79%, slightly lower than for SVM [4]. A more advanced neuro-fuzzy approach, S-dFasArt, has been evaluated on multi-session 8-channel EEG data recorded from three experienced BCI users each performing three mental tasks [10]. On average, for different training configurations of three-session data, the CAs reported for each subject on the fourth test session were ~83%, 68% and 48%. With additional model boosting the average CA of ~76% was obtained. Overall, it should be noted that these offline studies have been reported on limited (just a few subjects and recording sessions) but good-quality data originating from BCI competition benchmarks. There is very little evidence of deploying fuzzy classifiers in on-line EEG classification. Only recently, an IT2FLS approach to the discrimination of wrist and finger MIs has been validated with on-line, but not instantaneous, neurofeedback [30]. Four binary classifiers were trained on three-day data and tested on-line during two consecutive days. The average CA rate of ~78% was reported for eight subjects when tailored extreme energy ratio features were extracted from 14-channel EEGs. However, the communication throughput of the real-time system was rather limited with ITr of ~1.8 bits/min.

IV. CONCLUSIONS

This paper has investigated the application of an IT2FLS classifier to the MI-related EEG discrimination task. The effectiveness of the proposed fuzzy approach in handling both long- and short-term manifestations of uncertainty effects inherent to BCI was evaluated in both off-line and on-line case studies. Firstly, in the comprehensive off-line analysis conducted on multi-session data, IT2FLS was proven to outperform some state-of-the-art BCI classifiers such as LDA, KFD, linear and Gaussian SVMs. In this regard, a vital role of the proposed fuzzy design methodology should be noted. Additionally, the IT2FLS was found to consistently deliver higher CA rates than T1FLS. The difference was statistically significant in cases where uncertainty manifestations were more challenging to address due to the larger gap between the recording times of training and test sessions.

Secondly, the on-line study demonstrated the feasibility of employing an IT2FLS classifier within a real-time BCI system. It was verified that the proposed system is capable of producing reasonable on-line performance, though for effective utilization in practical applications further progress is desirable. In the comparative retrospective simulations of the other BCI classifiers, including the counterpart T1FLS, tested off-line, the devised IT2FLS classifier was shown to offer more potential in effective handling the uncertainty effects associated with the observed variability of MI induced EEG patterns at various time scales. IT2FLS generated not only higher CA rates but also more informative feedback responses correlated with subjects' MIs, as reflected in MInf. Still, special care has to be taken when interpreting these results due to a different nature of retrospective BCI evaluation in comparison with on-line tests accounting for instantaneous neurofeedback effect. The overall level of BCI performance reported in this study is limited, which can partly be explained by the BCI novice status of the subjects involved in the presented evaluation, relatively infrequent BCI sessions and simple spectral EEG features extracted without any extensive optimization from only two-channel EEG recordings.

An important aspect of the work reported here is the data-driven design of the IT2FLS. The initialization scheme involving the modified mpMCA clustering and the enhanced gradient descent-based learning algorithm were found effective in alleviating the problem of poor initial conditions, sensitivity to initial parameter values and slow convergence, indicated earlier in [53]. The consistency and improved generalization capabilities of the FLSs in the given application are worth noting. Even though the IT2FLS's setup time is short (on the order of a couple of minutes), the system does not have to be retrained very frequently to maintain performance for relatively novice BCI users across recording sessions over the span of even a few weeks.

The major limitation of the proposed method is rather a sizeable amount of data required to tune the antecedent and consequent parameters of the IT2FLS, partly given the non-convex nature of the optimization problem. A mitigating factor is the robustness of the fuzzy classifier with respect to the meta parameters that control rule-base initialization. This implies a less demanding search for their suitable values, still ensuring good generalization even on smaller data sets. Besides, our gradient descent based approach does not support the use of any arbitrary non-differentiable form of T2FSs.

In conclusion, the contribution of this work is three-fold. First, the paper reports on a systematic and comprehensive evaluation of the pioneering T2FLS approach to handling uncertainty effects in multi-session EEG classification. Second, a hybrid method for data-driven design of an IT2FLS classifier has been proposed and validated. Third, to the best of our knowledge, this is the first study that has examined in depth, providing evidence from multi-session experiments on multiple subjects, the applicability of FLS classifiers to on-line EEG-BCI with instantaneous feedback in real-time conditions.

As for further research, it would be desirable to investigate how the proposed IT2FLS could be slowly adapted or trained incrementally to sustain the adequate balance between the existing knowledge rule-base learnt from data and novel trends, especially if they become repetitive, and not merely a consequence of outliers or incidental anomalies. Here, training on multiple preceding sessions appears as a promising direction for further investigation. In addition, it is intended to explore how the uncertainty bounds of the classifier's output can be effectively exploited with the aim of improving the performance of the classifier. Only the central point of the output interval has been utilized during recall and as part of the loss function for training purposes so far. In particular, the problem of initializing uncertainty bounds for the antecedents and consequents of an IT2FLS should be investigated further. Finally, the inherent interpretability of FL methodology could be exploited to examine the nature of fuzzy rules underlying robust BCI classification.

REFERENCES

[1] S. Mitra and S. K. Pal, "Fuzzy Sets in Pattern Recognition and Machine Intelligence," *Fuzzy Sets and Syst*., vol. 156, pp. 381–386, 2005.

[2] A. Aarabi, R. Fazel-Rezai and Y. Aghakhani, "A fuzzy rule-based system for epileptic seizure detection in intracranial EEG," *Clin. Neurophysiol.*, vol. 120, pp.1648–1657, 2009.

[3] P. Herman, G. Prasad and T. M. McGinnity, "Investigation of the type-2 fuzzy logic approach to classification in an EEG-based brain-computer interface," in *Proc. IEEE EMBC*, Shanghai, Sep 2005, pp. 5354-5357.

[4] F. Lotte, A. Lécuyer, F. Lamarche and B. Arnaldi, "Studying the use of fuzzy inference systems for motor imagery classification," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 15, no. 2, pp. 322–324, Jun 2007.

[5] C. Xydeas, P. Angelov, S. Chiao and M. Reoullas, "Advances in classification of EEG signals via evolving fuzzy classifiers and dependant multiple HMMs," *Comp. Biol. Med.*, vol. 36, pp. 1064–1083, 2006.

[6] M. Mahfouf, M. Abbod and D. Linkens, "A survey of fuzzy logic monitoring and control utilisation in medicine," *Artif. Intell. Med.*, vol. 21, pp. 27–42, 2001.

[7] A. Subasi, "Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction," *Comp. Biol. Med.*, vol. 37, no. 2, pp. 227–244, 2007.

[8] P. Herman, G. Prasad and T. M. McGinnity, "Designing a robust type-2 fuzzy logic classifier for non-stationary systems with application in brain-computer interfacing," in *Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC)*, Singapore, Oct 2008.

[9] D. Coyle, G. Prasad and T. M. McGinnity, "Faster self-organizing fuzzy neural network training and a hyperparameter analysis for a brain–computer interface," *IEEE Trans. Syst., Man, Cybern.*, vol. 39, no. 6, pp. 1458–1471, Dec 2009.

[10] J.-M Cano-Izquierdo, J. Ibarrola and M. Almonacid, "Improving motor imagery classification with a new BCI design using neuro-fuzzy S-dFasArt," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol.20, pp.2–7, 2012.

[11] T. Nguyen, A. Khosravi, D. Creighton and S. Nahavandi, "EEG signal classification for BCI applications by wavelets and interval type-2 fuzzy logic systems," *Expert Syst Appl.*, vol. 42, pp. 4370–4380, 2015.

[12] J. R. Wolpaw, N. Birbaumer, D. J. McFarland and G. Pfurtscheller, T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol*., vol. 113, pp. 767–791, 2002.

[13] A. Emery, "Population frequencies of inherited neuromuscular diseases – a world survey," *Neuromuscul. Disord.*, vol. 1, pp. 19–29, 1991.

[14] G. Prasad, P. Herman, D. Coyle, S. McDonough and J. Crosbie, "Applying a brain-computer interface to support motor imagery practice in people with stroke for upper limb recovery: a feasibility study," *J. Neuroeng. Rehab.*, vol. 7:60, 2010. doi: 10.1186/1743-0003-7-60.

[15] M. Bamdad, H. Zarshenas and M. A. Auais, "Application of BCI systems in neurorehabilitation: A scoping review," *Disability and Rehabilitation: Assistive Technology*, vol. 10, pp. 355–364, 2015.

[16] D. Marshall, D. Coyle, S. Wilson and M. Callaghan, "Games, gameplay, and BCI: The State of the Art," *IEEE Trans. Comput. Intellig. and AI in Games*, vol. 5, no. 2, pp. 82–99, Jun 2013.

[17] J.R. Wolpaw, N. Birbaumer, J. Heetderks, D.J. McFarland, P.H. Peckham, G. Schalk, E. Donchin, L.A. Quatrano, C.J. Robinson and T.M. Vaughan, "Brain-computer interface technology: A review of the first international meeting*," IEEE Trans. Rehab. Eng.*, vol. 8, no. 2, pp. 164–173, Jun 2002.

[18] T.M. Vaughan, W.J. Heetderks, L.J. Trejo, W.Z. Rymer, M. Weinrich, M.M. Moore, A. Kubler, B.H. Dobkin, N. Birbaumer, E, Donchin, E.W. Wolpaw and J.R. Wolpaw, "Brain-computer interface technology: A review of the second international meeting," *IEEE Trans.Neural Sys.Rehab.Eng.* vol. 11, no. 2, pp. 94–109, Jun 2003.

[19] D. J. Krusienski, M. Grosse-Wentrup, F. Galan, D. Coyle, K. J. Miller, E. Forney and C. W. Anderson. "Critical issues in state-of-the-art brain-computer interface signal processing." *J. Neural Eng.*, vol. 8, no. 2, pp. 025002, 2011.

[20] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe and K.-R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Proc. Mag.*, vol. 25, no. 1, pp. 41–56, Jan 2008.

[21] Y. Li, H. Kambara, Y. Koike and M. Sugiyama, "Application of covariate shift adaptation techniques in brain-computer interfaces," *IEEE Tran. Biomed. Eng.*, vol. 57, no. 6, pp.1318–1324, Jun 2010.

[22] M. Arvaneh, C. Guan, K. K. Ang and C. Quek, "EEG Data Space Adaptation to reduce intersession nonstationary in brain-computer interface," *Neural Comput.*, vol. 25, pp. 2146–2171, 2013.

[23] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao and K.-R. Müller, "Towards adaptive classification for BCI," *J. Neural Eng.*, vol. 3, pp. 13–23, 2006.

[24] C. Guger, G. Edlinger, W. Harkam, I. Niedermayer and G. Pfurtscheller, "How many people are able to operate an EEG-based BCI?" *IEEE Trans. Neural Sys. Rehab. Eng.*, vol. 11, no. 2, pp. 145–147, Jun 2003.

[25] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin and K.-R. Müller, "Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing," in Advances in Neural Information Processing Systems 20, pp. 113–120, 2008, MIT Press.
[26] G. Krausz, R. Scherer, G. Korisek and G. Pfurtscheller, "Critical decision-speed and information transfer in the Graz Brain-Computer Interface," *Appl Psychophysiol Biofeedback*, vol. 28, pp. 233–240, 2003.
[27] J. del R. Millan, J. Mourino, M. Franze, F. Cincotti, M. Varsta, J. Heikkonen and F. Babiloni, "A local neural classifier for the recognition of EEG patterns associated to mental tasks," *IEEE Trans. Neural Net.*, vol. 13, no. 3, pp. 678–685, May 2002.
[28] R. Palaniappan, R. Paramesran, S. Nishida and N. Saiwaki, "A new brain-computer interface design using fuzzy ARTMAP," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 10, no. 3, pp. 140–148, Sep 2002.
[29] P. Martinez, H. Bakardjian and A. Cichocki, "Fully online multi-command brain-computer interface with visual neurofeedback using SSVEP paradigm," *Comput. Intell. Neurosci.*, vol. 2007, pp. 94561-1–94561-9, 2007.
[30] S. Bhattacharyya, M. Pal, A. Konar and D.N. Tibarewala, "An interval type-2 fuzzy approach for real-time EEG-based control of wrist and finger movement," *Biomed. Sig. Proc. Control*, vol. 21, pp. 90–98, 2015.
[31] C. Vidaurre, M. Kawanabe, P. von Bünau, B. Blankertz and K.-R. Müller, "Toward unsupervised adaptation of LDA for BCIs," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 587–597, Mar 2011.
[32] J. W. Yoon, S.J. Roberts, M. Dyson and J.Q. Gan, "Adaptive classification for brain computer interface systems using sequential Monte Carlo sampling," *Neural Networks*, vol. 22, pp.1286–1294, 2009.
[33] A. Llera, V. Gómez and H. J. Kappen, "Adaptive multiclass classification for brain computer interfaces," *Neural Comput.*, vol. 26, pp. 61108–61127, 2014.
[34] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. USA: Prentice-Hall, 2001.
[35] Q. Liang and J. M. Mendel, "Interval type-2 fuzzy logic systems: theory and design," *IEEE Trans. Fuzzy Sys.*, vol. 8, no. 5, pp. 535–550, 2000.
[36] P. Melin and O. Castillo, "A review on the applications of type-2 fuzzy logic in classification and pattern recognition," *Expert Syst. Appl.*, vol. 40, no. 13, pp. 5413–5423, 2013.
[37] C. H. Wang, C. S. Cheng and T. T. Lee, "Dynamical optimal training for interval type-2 fuzzy neural network (T2FNN)," *IEEE Trans. Syst., Man, Cybern.*, vol. 34, no. 3, pp.1462–1477, Jun 2004.
[38] K-R. Muller, C. W. Anderson and G. E. Birch, "Linear and nonlinear methods for brain-computer interfaces," *IEEE Trans. Neural. Syst. Rehab. Eng.*, vol. 11, no. 2, pp. 165–169, Jun 2003.
[39] J-S. Wang and C. S. G. Lee, "Self-adaptive neuro-fuzzy inference systems for classification applications," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 6, pp. 790–802, Dec 2002.
[40] E. Haselsteiner and G. Pfurtscheller, "Using time-dependent NNs for EEG classification," *IEEE Trans. Neural Sys. Rehab. Eng.*, vol. 8, no. 4, pp. 457–462, Dec 2000.
[41] E. Niedermeyer and F. Lopes da Silva, Electroencephalography: Basic Principles, Clinical Applications and Related Fields, 4th ed., Williams and Wilkins, USA, 1998.
[42] G. Pfurtscheller, C. Brunner, A. Schlögl and F. Lopes da Silva, "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, pp. 153-159, 2006.
[43] G. Pfurtscheller, C. Neuper, "Motor imagery activates primary sensori-motor area in humans," *Neuroscience Letters*, vol. 239, pp. 65–68, 1997.
[44] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. NY, USA: Plenum Press, 1981.
[45] S. L. Chiu, "Fuzzy model identification based on cluster estimation," *J. Intell. Fuzzy Syst*, vol. 2, pp. 267–278, 1994.
[46] R. Yager and D. Filev, "Approximate clustering by the mountain clustering," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 8, pp. 338–358, Aug 1994.
[47] S. Hassan, M.A. Khanesar, E. Kayacan, J. Jaafar and A. Khosravi, "Optimal design of adaptive type-2 neuro-fuzzy systems: A review," *Appl. Soft Comput.*, vol. 44, pp. 134–143, 2016.
[48] J. M. Mendel, "Computing derivatives in interval type-2 fuzzy logic systems," *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 1, pp. 84–98, Feb. 2004.
[49] N. N. Karnik and J. M. Mendel, "Centroid of a type-2 fuzzy set," *Inform. Sci.*, vol. 132, pp. 195-220, 2001.
[50] C. Wang, H Liu and C. Lin, "Dynamic optimal learning rate of a certain class of fuzzy neural networks and its applications with genetic algorithm," *IEEE Trans. Syst. Man Cybern. B*, vol. 31, no. 3, pp. 467–475, Jun 2001.
[51] S. Canu, Y. Grandvalet, V. Guigue and A. Rakotomamonjy, *SVM and Kernel Methods Matlab Toolbox.* Perception Systèmes et Information, INSA de Rouen, France, 2005.
[52] A. Schlögl, C. Neuper, G. Pfurtscheller, "Estimating the mutual information of an EEG-based Brain-Computer Interface," *Biomed Tech.* (Berl). vol. 47, no. 1–2, pp. 3–8, 2002.
[53] G. Prasad, P. Herman, T.M. McGinnity, "Critical observations on type-2 fuzzy logic approach to uncertainty handling in a brain-computer interface design," in *Proc. 11th IPMU*, Paris, 2006, pp. 1356–1363.
[54] A. Satti, C. Guan, D. Coyle and G. Prasad, "A covariate shift minimisation method to alleviate non-stationarity effects for an adaptive BCI," in *Proc. 20th ICPR*, Istanbul, 2010, pp. 105–108.
[55] R. Tomioka, J. N. Hill, B. Blankertz, K. Aihara, "Adapting spatial filter methods for nonstationary BCIs," in *Proc. IBIS*, 2006, pp. 65–70.
[56] C. Brunner, R. Scherer, B. Graimann, G. Supp and G. Pfurtscheller, "Online control of a brain-computer interface using phase synchronization," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2501–2506, Dec 2006.
[57] C. Vidaurre, A. Schlögl, R. Cabeza, R. Scherer and G. Pfurtscheller, "Study of on-line adaptive discriminant analysis for EEG-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 3, pp. 550–556, Mar 2007.
[58] M. Hazrati and A. Erfanian, "An online EEG-based brain-computer interface for controlling," *Med Eng Phys.*, vol. 32, pp.730–739, 2010.

Pawel A. Herman (M'04) received an MSc degree in computer science and electronics from Wroclaw University of Technology, Poland in 2002, a PhD degree from Ulster University, Derry, UK in 2008, and a Docent title from KTH Royal Institute of Technology, Sweden. He is currently an Assistant Professor at KTH Royal Inst. of Technology, Sweden. His research interests are focused on computational neuroscience, brain-like and cognitive computing, brain-computer interfaces, neural information processing, fuzzy systems and machine learning.

Girijesh Prasad (M'98-SM'07) received a BTech degree in electrical engineering from REC Calicut, India in 1987, an MTech degree in Computer Science & Technology from University of Roorkee, India in 1992, and a PhD degree from Queen's University, Belfast, UK in 1997. He is Professor of Intelligent Systems in the School of Computing and Intelligent Systems, Ulster University, UK. His research interests are in computational intelligence, brain modelling, BCI and neuro-rehabilitation, and assistive technology. He has published over 210 research papers in journals, books, and conference proceedings.

T. Martin McGinnity (SMIEEE, FIET) received a First Class (Hons.) degree in physics in 1975, and a Ph.D. degree from the University of Durham, UK in 1979. He is currently Pro Vice Chancellor for Student Affairs and Head of the College of Science and Technology at Nottingham Trent University, UK. Formerly he was Professor of Intelligent Systems Engineering and Director of the Intelligent Systems Research Centre, University of Ulster. He has published over 300 research papers and has attracted over £25 million in research funding. His research interests are focused on computational intelligence, cognitive robotics and biological information processing.