Proceedings Paper:
Rubio-Solis, A., Beltran-Perez, C. and Wei, H. (2022) Classification of EEG signals for brain-computer interfaces using a Bayesian-Fuzzy Extreme Learning Machine. In: Jiang, R., Zhang, L., Wei, H.L., Crookes, D. and Chazot, P., (eds.) Recent Advances in AI-enabled Automated Medical Diagnosis. AI4MED 2021 : 2021 International Symposium on Artificial Intelligence for Medical Applications, 19-23 Aug 2021, Virtual Conference (Newcastle upon Tyne, UK). Taylor & Francis, pp. 347-362. ISBN: 9781032008431.

https://doi.org/10.1201/9781003176121-22

Accepted for publication by "Recent Advances in AI-enabled Automated Medical Diagnosis", Edited by Jiang, R., Crookes, D., Wei, H. L., Zhang, L., and Chazot, P. (Accepted 8 August 2021)

Chapter 1
Classification of EEG Signals for Brain-Computer Interfaces using a Bayesian-Fuzzy Extreme Learning Machine

Adrian Rubio-Solis, Carlos Beltran-Perez and Hua-Liang Wei.

Abstract In brain-computer interface (BCI) applications, classification of motor imagery electroencephalogram (EEG) using Extreme Learning Machine (ELM) theory dates back to 2006. Even though, it is relatively new, advances in ELM-based classification have demonstrated to be a robust methodology with strong generalization properties. In this study, a unified framework based on Bayesian and Fuzzy ELM theory referred to as Bayesian-Fuzzy Extreme Learning Machine (BFELM) is developed for EEG signals classification. The proposed methodology is a hybrid approach for the training of a class of Fuzzy Inference Systems (FISs) of Takagi-Sugeno-Kang (TSK). On the one hand, Fuzzy logic theory is applied to handle any bounded non-constant piecewise continuous membership functions (MFs). On the other hand, Bayesian ELM theory is used to calculate the consequent parameters of each fuzzy rule by estimating their likelihood while minimizing training error and improving associated model generalization. Performance comparison of BFELM with other existing ELM methods and Support Vector Machine (SVM) is implemented for the classification of EEG signals using two public data sets. The experimental results confirm the advantages of using a unified framework for an improved classification of EEG data associated with motor imagery (MI) in BCI applications

Adrian Rubio-Solis
Institute for Materials Discovery, University College London (UCL); e-mail: a.rubio-solis@ucl.ac.uk

Carlos Beltran-Perez
Tecnologico de Monterrey, Campus Toluca, Mexico; e-mail: carlos.beltran@tec.mx

Hua-Liang Wei
Automatic Control and Systems Engineering, University of Sheffield; e-mail: w.hualiang@sheffield.ac.uk

1.1 Introduction

A Brain-Interface Computer (BCI) is an emanating technology that allows a direct communication between a human brain and a computer [1, 2, 3, 4]. BCI has been largely used as a technology to exploit the electrical activity in the brain for the diagnosis of neurological diseases, in handicapped patients who have lost their abilities and to understand psycho-physiological processes [5]. EEG is a non-invasive BCI system, in which brain activities are captured with a high temporal resolution, usability, portability and low setup cost [1]. In applications that involve handicapped patients, BCI is frequently used as a system to revive those elementary capabilities by creating an information pathway between the human brain and processing/computing devices. In other words, a BCI system utilises the information from the brain activity in disabled people to assist them while mapping their sensory-motor functions [1, 2].

For the last two decades, the most adopted EEG patterns for BCI development include sensorimotor mu rhythm [7, 8], and beta rhythms (15-30 Hz) [9, 10], recorded from the scalp over the sensorimotor cortex and widely used methods in BCI systems. Mu rhythm usually results from a power change of EEG frequency band between 8-18 Hz. It occurs as a simultaneous neural response at contralateral sensorimotor area during motor imaginery (MI) tasks. Over the last decades, BCI systems have been increasingly developed towards the solution of several problems that involve the control of prosthesis, wheelchair navigation, writing and communication assistance in disabled people [3]. MI is one of the most popular methods in BCI applications that involves carrying out a motor task merely by thinking or imagining [1]. It can be just a simple shifting of hands or legs, and/or closing eyes. Accordingly, MI-based BCI systems have become a prominent solution to recognise the desired commands by classifying MI tasks for deprived people of their motor abilities and for rehabilitation [1, 2]. However, MI signals are highly non-stationary and inevitably contaminated with noise, meanwhile, they strongly depend on subjects [11]. Moreover, the classification of EEG is usually a complex and aperiodic time series, which is the sum of a large number of neuronal membrane potentials [1]. Therefore, a powerful pattern recognition model is required for the implementation of a MI-based BCI system with a high performance [13].

In literature, an extensive number of research efforts have been dedicated to improving EEG feature extraction and classification of MI tasks [1, 4]. Most of these efforts have suggested a two-step approach, where first a process of feature extraction is implemented, and a second step is performed for the classification of the extracted features. Common Spatial Pattern (CSP) and Extreme Learning Machines (ELMs) have been successfully applied to the classification of MI tasks delivering a superior performance over traditional classification approaches such as Support Vector Machine and Multi-layer Perceptron neural networks [11]. On the one hand, CSP has been used as a robust spatial filter for multichannel optimization of EEG to the maximization of the variance of projected signal from one class while to minimize it from another class [12, 13]. On the other hand, ELM has been successfully applied to the classification of these projected signals [11]. For example in [13], a probabilistic framework based on Sparse Bayesian Extreme Learning Machine (SBELM) was suggested to improve the classification of traditional ELM [14]. SBELM was suggested as an improved ELM method that automatically control model complexity, good generalisation properties and exclude redundant hidden neurons by exploiting the advantages of ELM theory and Bayesian learning.

Similarly, other pattern recognition methods that are able to naturally deal with nonlinear and outlier characteristics of EEG signals have been suggested. For instance, in [14] a combined method based on wavelet transformation and Interval Type-2 Fuzzy Logic Systems (IT2-FLSs) called wavelet-IT2FLS was introduced and called. The proposed wavelet-IT2FLS is a higher order fuzzy system more capable of uncertainty handling than traditional ELM and SVM, in which the noisy, nonlinear and outlier-embedded nature of EEG signals can be modelled proficiently by type-2 fuzzy sets (FSs).

In this paper, a new Bayesian Fuzzy Extreme Learning Machine (BFELM) for EEG signal classification in BCI systems is presented. The proposed BFELM is a unified learning approach, in which a probabilistic method based on Bayesian learning and ELM theory is implemented for the training of Fuzzy Inference Systems (FISs) of Takagi-Sugeno-Kang (TSK). On the one hand, Bayesian learning incorporates a priori knowledge in the design of fuzzy rules while the confidence intervals of each consequent in the FIS are analytically determined. Within this context, a BFELM is a simultaneous learning method of antecedent and consequent parts of each fuzzy rule in a FIS, in which fuzziness and probability can work in a collaborative manner rather competitively. Moreover, the proposed BFELM inherits the capabilities of FISs to naturally deal with uncertainty and noisy signals. To validate the performance of the proposed BFELM for EEG signal classification, two public datasets of BCI competitions are used. For feature extraction, CSP is implemented, and the extracted features are feed into the BFELM. To compare the BFELM with other benchmark techniques, traditional ELM, SVM, Multi Kernel ELM (MKELM), Fuzzy ELM (FELM) and Bayesian ELM (BELM) were also implemented.

The rest of this paper is structured as follows. In section 1.2, basic concepts of Bayesian ELM and Fuzzy Inference Systems (FISs) is reviewed, as well as the proposed BFELM is described. In section 1.3, experiments and results are presented, while in section 1.4 the corresponding discussion is provided. Finally, in section 1.5, conclusions and future work are drawn.

1.2 Background Material and Proposed Method

1.2.1 Extreme Learning Machine

Extreme Learning Machine (ELM) is a learning paradigm originally developed to train single-hidden-layer feedforward networks (SLFNs), in which parameters in the hidden neurons are initialized randomly and the output weights are optimized using the Moore-Penrose pseudoinverse. Given a number of $'P'$ distinct samples $D = (\mathbf{x}_p, t_p)$, with each $\mathbf{x}_p$ being a N dimensional vector and t_p as the target scalar output. Hence the goal of ELM is to find a relationship between $\mathbf{x}_p$ and $\mathbf{t}_p$. Standard SLFNs with M hidden nodes and activation $h(\cdot)$ function can be mathematically modelled by:

$$\sum_{k=1}^{M} \beta_k h_k(\mathbf{w}_k; \mathbf{x}_p) = \mathbf{h}(\mathbf{w}_k; \mathbf{x}_p)\boldsymbol{\beta} = y_p, \quad 1 \leq p \leq P \tag{1.1}$$

in which $\mathbf{h}(\mathbf{w}_k; \mathbf{x}_p) = [h_1(\mathbf{w}_1; \mathbf{x}_1), \ldots, h_M(\mathbf{w}_M; \mathbf{x}_M)]$ is the hidden feature mapping, $\mathbf{w}_k = [\mathbf{w}_1, \ldots, \mathbf{w}_N]^T$ is the weight vector a randomly generated parameter of the hidden layer connecting the kth hidden node and the input nodes. The output weight $\beta_p = [\beta_{p1}, \ldots, \beta_{p\tilde{N}}]^T$ is the weight vector connecting the kth hidden node to the nth output. A SLFN with M hidden nodes and activation function $g(\mathbf{x})$ can approximate P samples with zero error means $\sum_{p=1}^{M} \parallel \mathbf{y}_p - \mathbf{t}_p \parallel$. Thus, a matrix representation of Eq. (1.1) is:

$$\mathbf{H} = \begin{pmatrix} h(\mathbf{w}_1; \mathbf{x}_1) & \cdots & h(\mathbf{w}_k; \mathbf{x}_p) \\ \vdots & \vdots & \vdots \\ h(\mathbf{w}_k; \mathbf{x}_p) & \cdots & h(\mathbf{w}_k; \mathbf{x}_p) \end{pmatrix}_{P \times M} \tag{1.2}$$

Where $\mathbf{H}$ is the hidden matrix of an SLFN with respect to the inputs $\mathbf{x}_p$. The target vector is defined by $\mathbf{T} = [t_1, \ldots, t_P]$. The minimum norm least-squares solution of the linear system $\mathbf{H}\beta = \mathbf{T}$ is unique and can be achieved by calculating the Moore-Penrose pseudo-inverse $H^\dagger$ as follows:

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^\dagger \mathbf{T} \tag{1.3}$$

In which, $\mathbf{H}^\dagger$ can be calculated using the orthogonal projection method: $\mathbf{H}^\dagger = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ when $\mathbf{H}^T\mathbf{H}$ is nonsingular, or $\mathbf{H}^\dagger = \mathbf{H}(\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{H}^T$ when $\mathbf{H}\mathbf{H}^T$ is nonsingular. A penalty term can be added to the diagonal of $\mathbf{H}^T\mathbf{H}$ or $\mathbf{H}\mathbf{H}^T$ for regularization purpose. However, the optimum value of this penalty is still subjected the minimization of the validation error. In many real-world applications, the number of hidden nodes is much smaller than the number of training samples $M \ll P$ [**?**]. Hence $\mathbf{H}$ is a non-square matrix, such that one specific value for $\hat{\mathbf{w}}_k$, $\hat{\mathbf{b}}_k$ and $\hat{\beta}_k$ need to be determined.

1.2.2 Bayesian Extreme Learning Machine

Bayesian Extreme Learning Machine (BELM) was originally introduced to ELM theory to determine the output weight with Bayesian Inference Method (BIM). By using BELM, each observed t_p is assumed to have an independent noise component ϵ_p which is Gaussian distributed with zero mean and variance σ^2, that is $t_p = \mathbf{h}(\mathbf{w}; \mathbf{x}_p)\boldsymbol{\beta} + \epsilon_p$, where, $p(\epsilon_p|\sigma^2) = \mathcal{N}(0, \sigma^2)$. The probabilistic model is then given by:

$$p(t_p|\mathbf{H}, \boldsymbol{\beta}, \sigma^2) = \mathcal{N}(t_p|\mathbf{h}(\mathbf{w}; \boldsymbol{x}_p)\boldsymbol{\beta}, \sigma^2) \tag{1.4}$$

Using all the training sales, the likelihood function can be computed as:

$$\begin{aligned} p(\mathbf{T}|\mathbf{H}, \boldsymbol{\beta}, \sigma^2) &= \prod_{p=1}^{P} p(t_p|\mathbf{H}, \boldsymbol{(}\beta), \sigma^2) \\ &= \prod_{p=1}^{P} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(t_p - \mathbf{h}(\mathbf{w}; \mathbf{x}_p)\boldsymbol{\beta})^2}{2\sigma^2} \right] \end{aligned} \tag{1.5}$$

To penalize large weights, a natural distribution is given by:

$$p(\boldsymbol{\beta}|\alpha) = \mathcal{N}(\boldsymbol{\beta}\mathbf{0}, \alpha^{-1}\mathbf{I}) = \left(\frac{\alpha}{2\pi}\right)^{P/2} \exp\left[-\frac{\alpha}{2}\boldsymbol{\beta}^T\boldsymbol{\beta} \right] \tag{1.6}$$

in which, α is a shared prior, and $\mathbf{I}$ is the identity matrix. As the prior and likelihood functions follow a Gaussian distribution, the posterior is also a Gaussian distribution defined by:

$$p(\boldsymbol{\beta}|\mathbf{T}, \mathbf{H}, \alpha, \sigma^2) = \mathcal{N}(\boldsymbol{\beta}|\mathbf{m}, \mathbf{S}) \tag{1.7}$$

where $\mathbf{m}$ and $\mathbf{S}$ is the mean and covariance respectively, which are be obtained by:

$$\mathbf{m} = \sigma^{-2} \cdot \mathbf{S} \cdot \mathbf{H}^T \cdot \mathbf{T} \tag{1.8}$$

$$\mathbf{S} = (\alpha\mathbf{I} + \sigma^{-2} \cdot \mathbf{H}^T \cdot \mathbf{H})^{-1} \tag{1.9}$$

in which, the posterior distribution of parameters α and σ^2 is $p(\boldsymbol{\alpha}, \sigma^2|\mathbf{T}, \mathbf{H}) \propto p(\mathbf{T}|\mathbf{H}, \boldsymbol{\alpha}, \sigma^2)$. Therefore, the optimal values for the parameters α and σ^2 can be determined with type-II maximum likelihood (ML-II). Such process involves the maximization of the marginal likelihood $p(\mathbf{T}|\mathbf{H}, \alpha, \sigma^2)$ inferred from integral:

$$\int p(\mathbf{T}|\mathbf{H}, \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}|\alpha) d\boldsymbol{\beta} \tag{1.10}$$

By differentiating the marginal log-likelihood function $\log p(\mathbf{T}|\mathbf{H}, \boldsymbol{\beta}, \sigma^2)$ with respect to parameters α and σ^2, their optimal values can be computed by:

$$\alpha^{new} = \frac{M - \alpha \cdot trace[\mathbf{S}]}{\mathbf{m}^T \mathbf{m}} \tag{1.11}$$

$$\sigma^{2,new} = \frac{\sum_{p=1}^{P} (y_p - \mathbf{h}(\mathbf{w}; \mathbf{x}_p)\mathbf{m})^2}{P - M + \alpha \cdot trace[\mathbf{S}]} \tag{1.12}$$

Thus, by initializing α and σ^2, the terms $\mathbf{m}$ and $\mathbf{S}$, are updated iteratively with Eqs. (1.8) – (1.9) and (1.11) – (1.12) until convergence. The new $\mathbf{m}$ can be employed for computing the new output y_{new} when new input data $\mathbf{x}_{new}$ is presented following the distributions:

$$p(y_{new}|\mathbf{h}(\mathbf{w}; \mathbf{x}_{new}), \mathbf{m}, \alpha, \sigma^2) = \mathcal{N}(\mathbf{h}(\mathbf{w}; \mathbf{x}_{new})\mathbf{m}, \sigma^2(\mathbf{x}_{new})) \tag{1.13}$$

where

$$\sigma^2(\mathbf{x}_{new}) = \sigma^2 + \mathbf{h}(\mathbf{w}; \mathbf{x}_{new})^T \cdot \mathbf{S} \cdot \mathbf{h}(\mathbf{w}; \mathbf{x}_{new})$$

Since α is a natural consequence of a Gaussian process, compared to ELM, BELM does not require to include any regularisation term. Hence BELM provides better generalisation properties than traditional ELM.

1.2.3 Fuzzy Extreme Learning Machine

According to ELM theory, a Fuzzy Inference System (FIS) can be interpreted as a SLFN if for a given number of distinct training samples $D = (\mathbf{x}_p, \mathbf{t}_p)$, a model of the FIS with M fuzzy rules is given by [20]:

$$\mathbf{y}_p(\mathbf{x}_p) = \sum_{k=1}^{M} \beta_k G(\mathbf{x}_p, \mathbf{c}_k, a_k) = \mathbf{t}_p, \ p = 1, \ldots, P \tag{1.14}$$

An FIS either of Takagi-Sugeno-Kang (TSK) or Mamdani type can be defined by a number of fuzzy rules R^k of the form [**?**, **?**]

$$R^k : \text{IF } x_{p1} \text{ is } A_{1k} \text{ AND } x_{p2} \text{ is } A_{2k} \text{ AND } \ldots \text{ IF } x_N \text{ is } A_{Nk} \text{ THEN } (y_p \text{ is } \beta_k) \tag{1.15}$$

where, $A_{sk}(s = 1, \ldots, N, k = 1, \ldots, M)$ are the fuzzy sets that correspond to the sth input variable x_{ps} in the kth rule. When an FIS employs a TSK inference engine, β_k $(k = 1, \ldots, M)$ is defined by a linear combination of input variables, i.e. $\beta_k = q_{k,0} + q_{k,0}x_1 + \ldots q_{k,N}x_N$, otherwise if the FIS is of Mamdani type, β_k is a crisp value. In Fuzzy Logic System theory (FLS), the degree to which any given input x_{ps} satisfies the quantifier A_{sk} is specified by its Membership Function (MF) $\mu_{A_{ks}}(c_{ks}, a_k)$, where usually a non-constant piece-wise continuous MF is used [19]. By using the symbol $\otimes$ for the representation of fuzzy logic AND operations, the firing strength of the kth fuzzy rule can be computed as [28]:

$$R^k(\mathbf{x}_p; \mathbf{c}_k, a_k) = \mu_{A_{k1}}(x_{p1}, c_{k1}, a_k) \\ \otimes \mu_{A_{k2}}(x_{p2}; c_{k2}, a_k) \otimes \ldots \otimes \mu_{A_{kN}}(x_{pN}; c_{kN}, a_k) \quad (1.16)$$

Each fuzzy rule R^K can be normalised as

$$G^k(\mathbf{x}_p; c_{ks}, c_k) = \frac{R^k(\mathbf{x}_p; c_k, a_k)}{\sum_{k=1}^{M} R^k(\mathbf{x}_p; c_k, a_k)} \quad (1.17)$$

G^k is called fuzzy basis function, where for the pth input-output, each $\mathbf{y}_p$ is:

$$\mathbf{y}_p = \sum_{k=1}^{M} \beta_k G^k(\mathbf{x}_p; c_{ks}, c_k) \quad (1.18)$$

For an FIS with a TSK, consequent parameters are linear combinations of input parameters computed as:

$$\beta_k = \mathbf{x}_{p,e}\mathbf{q}_k^T, \; k = 1, \ldots, M \quad (1.19)$$

For a TSK fuzzy model, $\mathbf{x}_{p,e} = [1 \; \mathbf{x}_p]$ is the extended version of $\mathbf{x}_p$ defined as $\mathbf{x}_{p,e} = [x_0, x_1, \ldots, x_N]$ and each coefficient $\mathbf{q}_k = [q_{k,0}, q_{k,1}, \ldots, q_{k,N}]$. The output weight of SLFN is defined by $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_M]$. For a Mamdani fuzzy model, $\mathbf{x}_{p,e} = [0, \mathbf{x}_p]$, and $\boldsymbol{\beta}_k$ is a single crisp value. In this work, a FIS with a TSK inference is considered. A compact representation for Eq. (1.14) is

$$\mathbf{H}_{TSK}\mathbf{Q} = \mathbf{T} \quad (1.20)$$

in which, $\mathbf{Q}$ is the matrix of coefficients $q_{kj,s}$. If a TSK implication is employed, $\mathbf{H}_{TSK} = [\mathbf{h}_1(\mathbf{c}_k, \mathbf{a}_k; \mathbf{x}_1), \ldots, \mathbf{h}_P(\mathbf{c}_k, \mathbf{a}_k; \mathbf{x}_P)]^T$, where the optimal value for matrix $\hat{\mathbf{Q}}$ is obtained

$$\hat{\mathbf{Q}} = \mathbf{H}_{TSK}^{\dagger} T \quad (1.21)$$

where the vector of firing strengths is $\mathbf{h}_p = (h_{p1}, \ldots, h_{pM})$, and h_{pk} is:

$$h_{pk} = (G^1(\mathbf{x}_p, \mathbf{c}_1, a_1)x_0, \ldots, G^1(\mathbf{x}_p, \mathbf{c}_1, a_1)x_p, \ldots, \\ G^M(\mathbf{x}_p, \mathbf{c}_M, a_M)x_0, \ldots, G^M(\mathbf{x}_p, \mathbf{c}_M, a_M)x_p) \quad (1.22)$$

Matrix $\hat{\mathbf{Q}} = [\hat{\mathbf{q}}_1, \ldots, \hat{\mathbf{q}}_M]^T$. In FELM, the parameters of each MF $(\mathbf{c}_k, a_k)$ are randomly generated, based on this, the consequent parameters β_i are analytically estimated.

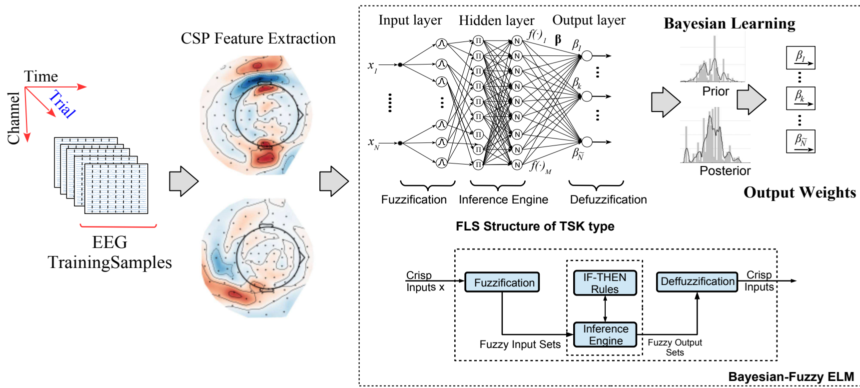

Fig. 1.1: Illustration of the BFELM for EEG signals classification.

1.2.4 Bayesian-Fuzzy ELM for EEG Classification

Bayesian-Fuzzy Extreme Learning Machine (BFELM) is a unified learning framework based on Bayesian learning regression, Fuzzy Logic Systems (FLSs) theory and ELM to the training of Fuzzy Inference Systems (FISs) of either TSK or Mamdani type. In this study, BFELM is applied to the learning of the consequent parts of a FIS of Takagi-Sugeno-Kang (TSK) that can handle any bounded nonconstant piecewise continuous membership function (MF). As illustrated in Fig. 1.1, given a set of training samples $(\mathbf{x}_p, t_p),\ p = 1, \ldots, P$, where $\mathbf{x}_p = [x_{p1}, \ldots, x_{pN}]^T \in \mathbf{R}^N$ and the corresponding class labels $t_p = \{-1, 1\}$ denotes the class of MI task. Assume each of the training samples is a feature vector obtained by using Common Spatial Pattern (CSP) as detailed in [21]. BFELM aims at finding an optimal hyperplane that maximises the separating margin between the two classes. Given certain MF g and its parameters $(\mathbf{c}_k, \mathbf{a}_k)$, and rule number M for EEG signal classification, BFELM is formulated as an iterative learning algorithm that is implemented in two steps as described in Algorithm 1.

I) Inference of the posterior distribution of the consequent parameters $\mathbf{q}_k$ of the TSK FIS using a Gaussian function with mean $\mathbf{m}_{q,t}$ and covariance $\mathbf{S}_{q,t}$ (Eq. 1.23 and 1.24).
II) With the evidence procedure, estimate the two hyperparameters α_t and σ_t^2 as described in Equations 1.25 and 1.26.

Thus, from algorithm 1, fuzzy consequent parts $\mathbf{q}_k$ are optimized iteratively by means of ML-II [23] or Evidence procedure [24].

Algorithm 1: Bayesian-Fuzzy ELM for EEG Classification

Result: optimized α, σ^2, $\mathbf{S}_q, \mathbf{m}_q$

Initialize: α, σ^2, $\mathbf{c}_k$, $\mathbf{a}_k$, ϵ, M;

$t = 0$;

while *While* $||\boldsymbol{m}_{q,t+1} - \boldsymbol{m}_{q,t}|| > \epsilon$ **do**

  t = t+1;

  Calculate;

$$\mathbf{m}_{q,t} = \sigma^{-2} \cdot \mathbf{S}_{q,t} \cdot \mathbf{H}_{TSK}^T \cdot \mathbf{T} \tag{1.23}$$

$$\mathbf{S}_{q,t} = (\alpha \mathbf{I} + \sigma^{-2} \cdot \mathbf{H}_{TSK}^T \cdot \mathbf{H}_{TSK})^{-1} \tag{1.24}$$

  Update;

$$\alpha_t = \frac{M - \alpha \cdot trace[\mathbf{S}_{q,t}]}{\mathbf{m}_{q,t}^T \mathbf{m}_{q,t}} \tag{1.25}$$

$$\sigma_t^2 = \frac{\sum_{p=1}^{P} (y_p - \mathbf{h}_p(\mathbf{c}_k, \mathbf{a}_k; \mathbf{x}_p)\mathbf{m}_{q,t})}{P - M + \alpha \cdot trace[\mathbf{S}_{q,t}]} \tag{1.26}$$

end

The iterative process for the proposed BFELM is stopped when the difference of the norm $\mathbf{m}_{q,t}$ between successive iterations has fallen below a given value ϵ [22]. Therefore, relevant CSP features are automatically selected for the subsequent classification [13]. Given a test sample $\hat{\mathbf{x}}$, the prediction distribution of the consequent parts of an FIS can be defined as:

$$p(\hat{y}|\alpha, \sigma^2, \mathbf{h}(c_k, \hat{a}_k; \mathbf{x})) = \int (\hat{y}|\hat{\mathbf{q}}_k, \sigma^2, \mathbf{h}(c_k, a_k; \hat{\mathbf{x}}))p(c_k, a_k; \hat{\mathbf{q}}_k|\alpha, \sigma^2, \mathbf{y}))d\hat{\mathbf{q}}_k \tag{1.27}$$

where $\mathbf{h}(c_k, a_k; \hat{\mathbf{x}}) = [h_1(c_k, a_k; \hat{\mathbf{x}}), \ldots, h_M(c_k, a_k; \hat{\mathbf{x}})]$ is again Gaussian with mean $\hat{\mathbf{m}}$ and variance $\hat{\sigma}^2$ defined as:

$$\hat{\mathbf{m}}_q = \mathbf{h}(c_k, \hat{a}_k; \mathbf{x})\mathbf{m}_q \tag{1.28}$$

$$\hat{\sigma}^2 = \sigma^2 + \mathbf{h}(c_k, \hat{a}_k; \mathbf{x})\mathbf{S}_q\mathbf{h}(c_k, \hat{a}_k; \mathbf{x})^T \tag{1.29}$$

The test sample is then classified using the criterion:

$$\hat{\mathbf{y}} = \begin{cases} -1, & \hat{\mathbf{m}} < 0 \\ +1, & \hat{\mathbf{m}} > 0 \end{cases} . \tag{1.30}$$

Note: compared to other learning approaches for the training of FISs, the hyperparameter α can be viewed as a regularization parameter in Eq. (1.24) that naturally results from the Gaussian process.

1.3 Experimental Study

In this section, the performance of the proposed BFELM on two public MI EEG datasets is evaluated and compared to other existing ELM-based techniques. Description of suggested datasets as well as experimental setup and evaluation is also provided in this section.

1.3.1 EEG Data Acquisition

In this work, two public data EEG data sets are used to study the proposed BFELM [6, 17, 18]. The first data set is available from BCI competition IV data set IIb. The EEG data was collected from nine different individuals using three bipolar channel (C_3, C_z and C_4) with a sampling rate of $250Hz$ for the discrimination of two classes (left-hand and right-hand motor imagery - (MI)). The electrooculogram (EOG) was recorded using three monopolar electrodes. They were recorded from each subject in five sessions [18]. For comparison purposes, in this study only the sets that correspond to B0103T, B0203T, $\ldots$, B01903T are used for training the proposed BFELM. Each subject was required to complete 160 trials (half for each class of MI). For each trial, the subject received visual guidance to perform MI task for a period of time of 4.5 seconds.

The second data set corresponds to BCI competition III data set IVa. Such data set was collected at a sampling rate of $100Hz$ from 188 electrodes from five different subjects, namely: "aa", "al", "av", "aw" and "ay". Each subject was required to complete 280 trials for the imagination of two tasks, i.e. right hand or foot movements. Each subject completed 140 trials for each MI task for a period of time of 3.5 seconds.

1.3.2 Experimental Setup and Evaluation

For comparison purposes, first data preprocessing was performed on the raw EEG data. For each trial, data was band-pass filtered between 4-40Hz using a fifth-order Butterworth filter. Next, the dimension of EEG signals was reduced using Common Spatial Pattern (CSP), a widely used technique for feature selection in the classification of MI-based BCIs [21]. Finally, to dis-

Table 1.1
Classification accuracy (%) obtained by SVM, ELM, MKELM, FELM and BFELM, for competition IV, data set IIb.

Subject	SVM	ELM	MKELM	SBELM	FELM	**BFELM**[1]	**BFELM**[2]
B0103T	76.1	76.8	77.5	77.6	77.7	77.9	77.6
B0203T	56.9	59.8	65.4	61.8	62.3	63.1	63.2
B0303T	50.9	51.5	54.3	54.1	54.2	54.2	54.3
B0403T	98.7	98.7	99.3	99.0	99.1	99.1	99.0
B0503T	83.2	84.1	84.6	84.3	84.5	84.6	84.7
B0603T	67.6	68.3	69.5	69.2	69.3	69.5	69.3
B0703T	82.9	84.2	86.8	84.7	84.9	85.2	85.0
B0803T	87.0	87.5	89.9	89.0	89.3	89.6	89.3
B0903T	81.3	82.9	83.7	83.5	83.6	83.6	83.4
Average	76.0±15.3	77.0±14.8	79.0±14.0	78.1±14.3	78.3±14.1	78.5±14.1	78.4±14.2
Time (s)	16.4± 0.1	2.12±0.2	6.8±0.6	3.7±0.4	2.8±0.3	3.8±0.1	3.9±0.2

[1] BFELM with Gaussian Membership Function
[2] BFELM with Cauchy Membership Function

Table 1.2
Classification accuracy (%) obtained by SVM, ELM, MKELM, FELM and BFELM, for competition III, data set IVa.

Subject	SVM	ELM	MKELM	SBELM	FELM	**BFELM**[1]	**BFELM**[2]
aa	80.6	82.6	83.3	82.9	82.8	83.0	83.1
al	97.7	97.9	98.5	98.2	98.1	98.3	98.3
av	69.3	70.0	71.4	70.6	70.7	70.9	71.1
aw	89.7	90.2	91.3	90.7	90.9	91.0	91.1
ay	91.7	92.4	93.0	92.6	92.5	92.9	92.8
Average	85.8±10.7	86.6±11.3	87.5±10.8	87.0±10.9	87.0±11.1	87.2±10.7	87.3±10.6
Time (s)	26.3± 0.1	3.1±0.1	10.1±0.2	7.1±0.2	3.2±0.1	7.3±0.1	7.2±0.1

[1] BFELM with Gaussian Membership Function
[2] BFELM with Cauchy Membership Function

criminate the the filtered EEG signals, a number of different techniques were implemented. To validate the performance of the proposed BFELM, a comparison study with other existing techniques was implemented. In this study, seven algorithms are compared:

1) SVM: Support Vector Machine [5].
2) ELM: Extreme Learning Machine [25].
3) MLKELM: Multilayer Kernel Extreme Learning Machine [17].
4) SBELM: Sparse Bayesian Extreme Learning Machine [13].
5) FELM: Fuzzy Extreme Learning Machine [26, 27, 28].
6) BFELM[1]: Bayesian-Fuzzy Extreme Learning Machine with Gaussian MFs.
7) BFELM[2]: Bayesian-Fuzzy Extreme Learning Machine with Cauchy MFs.

For each algorithm, a 5×5 cross-validation was implemented. The experimental setup for each model involves the regularization parameter 'C' for SVM; number fuzzy rules and hidden neurons for FELM and the proposed BFELM and ELM respectively. For MKELM Gaussian and polynomial kernels are selected [17]. Table 1.1 and 1.2 summarizes the average classification performance of ten random experiments obtained by different learning algorithms for BCI competition IV data set IIb and BCI competition III data set IVa respectively. In both tables, the computational time was also compared among the seven methods under MATLAB R2016a on a laptop with 2.7 GHz CPU (16 GB RAM). As it can be observed from both tables, in general ELM-based techniques outperform the performance accuracy of SVM. In particular, MLKELM yields the highest accuracy for datasets IIb and IVa on almost all subjects.

For the classification of dataset IIb, the proposed BFELM with Gaussian MFs not only achieves a similar performance to that produced by a MKELM, but also improves the model accuracy of traditional SBELM. From Table 1.1, it can be seen that the implementation of an BFELM represents an accuracy improvement over ELM, FELM and SBELM on subjects $B0103T$, $B0203T$, $B0703T$ and $B0803T$. For subjects $B0103T$, $B0503T$, $B0603T$, the proposed BFELM improved the performance achieved by an MKELM. In general, from table 1.1, the proposed BFELM either with Gaussian or Cauchy MSs achieves an accuracy 78.5% and 78.4% correspondingly, an improvement of 1.02% over ELM and 1.03% over SVM.

For data IVa, the BFELM with a Gaussian and Cauchy MFs achieved a mean accuracy of 87.2% and 87.3%, an improvement of 0.1% with respect to ELM and SVM. As described in table 1.2, BFELM provides the highest accuracy among BELM-based methods with a similar performance to that obtained by an MKELM. In terms of the computational load required to train each model, it can also be observed from table 1.2, the incorporation of TSK inference engine does not implies a significant increase over traditional SBELM. Moreover, while the training time for an MKELM is about $10.1s$ to produce an accuracy of 87.5%, the training time of a BFELM is $7.3s$ to achieve an accuracy of 87.3%. This represents a decrease of 28% for the mean computational training. As illustrated in table 1.2, the proposed BFELM approach represents an improvement on each subject over ELM, SBELM, FELM and SVM.

Finally, in Fig. 1.2, the effect of varying the number of fuzzy rules on the average accuracy of all subjects using 80% as training for the proposed BFELM models is presented. As illustrated in Fig. 1.2, an increase in the number of fuzzy rules does not necessarily implies an increase in the final accuracy. It can be observed, degraded accuracy of BFELM occurs when either using a large or small number of fuzzy rules. For data IIb, the optimal number of fuzzy rules for a BFELM either with Gaussian or Cauchy MFs is produced by using 30 fuzzy rules. In contrast, for data IVa, the optimal number of fuzzy rules using a BFELM with Gaussian MFs can be achieved using between 40−50 rules, while for a BFELM with Cauchy MFs, the highest accuracy can be obtained with 30 or more fuzzy rules.

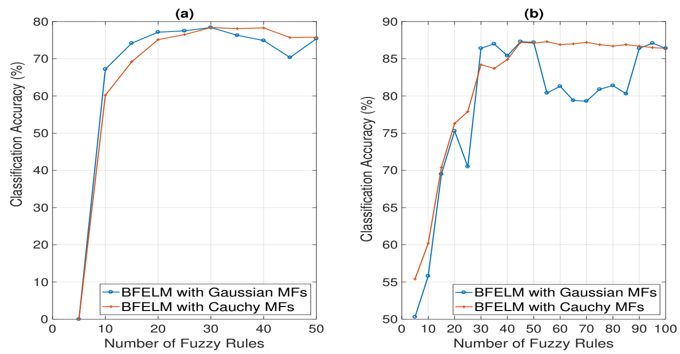

Fig. 1.2: Average accuracy of ten experiments for the BFELM using a different number of fuzzy rules for the classification of (a) competition IV, data IIb and (b) competition III, dataset IVa.

1.4 Discussion

This section provides a performance analysis as well as its pros and cons for MI classification between the proposed BFELM, and a number of literature techniques such as MKELM, SVM, BELM, FELM and ELM. From Table 1.1 and 1.2, it can be observed that the proposed BFELM either with Gaussian MFs or Cauchy MFs produce comparable results to that obtained by using a Multi-Kelnel ELM (MK-ELM) and in general it is more efficient than SVM, BELM, ELM and FELM. Generally speaking, similar to all ELM-based approaches, the suggested BFELM presents universal approximation properties for continuous functions [17, 20]. Moreover, the proposed BFELM also inherits the ability of ELM learning for tuning-free of the hidden layer weights (consequent part in the design of FISs) while delivering an improved generalisation performance. The results presented in this study demonstrate the following:

a) The aforementioned BFELM is a probabilistic method that provides an effective approach to build the confidence intervals of an FIS of TSK type while eliminating the need to incorporate a regularisation parameter.
b) Since the hyperparameters α and σ_t^2 are a consequent of a Gaussian process, the training of a FIS using the proposed BFELM does not require a regularisation term. Therefore, by applying BFELM, the FIS naturally provides an improved generalisation performance compared to traditional ELMs.

Table 1.3
Pros and Cons of SVM, ELM, MKELM, FELM and the proposed BFELM.

Method	Pros
SVM	• Good generalisation properties
	• Can deal with high dimensional data
ELM	• Tuning-free for the hidden layer weights
	• High efficiency and good generalization performance
BELM	• Not need to include a regularisation parameter
MKELM	• Not need to determine the number of hidden units
	• Can explore nonlinear features
	• Can combine multiple kernels
BFELM	• Unifies the concept of fuzziness and probability
	• Good generalisation performance
	• New advances in Fuzzy theory and Bayesian learning may be implemented
	• Less number of hidden units than MKELM to provide a similar performance
	• Consequent part is a linear combination of each hidden activation.
Method	**Cons**
SVM	• Parameter selection is data dependent
ELM	• Need to determine the number of hidden units
	• Need to determine a regularisation parameter to improve generalisation
BELM	• Iterative method
MKELM	• Need to specify balance between kernels
	• Computationally burden increases as data size increases
	• Can combine multiple kernels
BFELM	• Iterative method, hence its training may result more expensive than ELM
	• Number of fuzzy rules needs to be tuned

c) The proposed BFELM is a probabilistic model for TSK FISs, in which fuzziness and probability work well for the classification EEG signals in a collaborative manner. Moreover, new advances in either theory may be implemented under appropriate conditions.

Finally, the main Pros and cons for the proposed BFELM and other techniques for the classification of the datasets IIb and IVa are described in table 1.3.

1.5 Conclusions

In this paper, a Bayesian probabilistic method based on ELM theory for the construction of of TSK Fuzzy Inference Systems (FISs), in which fuzziness and probability can work in a collaborative manner is presented. The proposed method called Bayesian Fuzzy Extreme Learning Machine (BFELM) is a unified learning algorithm based on Bayesian learning regression and Fuzzy Inference Systems theory. On the one hand, Bayesian learning allows the introduction of a priori knowledge while the confidence intervals of each consequent in the Fuzzy Inference System are analytically determined. On the other hand, simultaneous learning of antecedent and consequent part is achieved, where Fuzziness and Probability can work complementary rather than competitively for the classification of MI EEG tasks.

To evaluate the performance of the proposed BFELM, two public datasets about BCI competition IV dataset IIb and BCI competition III, data set IVa are suggested. To compare the performance of a BFELM, other techniques such as traditional ELM and SVM as well as Fuzzy ELM (FELM), Bayesian ELM (BELM) and Multi Kernel ELM (MKELM) have been also implemented. As described in our results, it was demonstrated that the proposed BFELM shows similar performance to that provided by a MKELM and better than traditional SVM, BFELM, FELM and ELM. It can also be concluded that the associated computational training time is approximately 28% less expensive than MKELM.

It can also be concluded that the proposed learning framework inherits the capability of Extreme learning Machines for universal approximation of continuous functions as well as to randomly select the parameters of antecedent of each fuzzy rule in the FIS, and analytically determine their consequent. Future work includes the incorporation of new advances not only from Bayesian theory, but also from the design of higher FISs for the classification of MI EGG signals.

References

1. Chaudhary, S., Taran, S., Bajaj, V. and Sengur, A., 2019. Convolutional neural network based approach towards motor imagery tasks EEG signals classification. IEEE Sensors Journal, 19(12), pp.4494-4500.
2. Khosla, A., Khandnor, P. and Chand, T., 2020. A comparative analysis of signal processing and classification methods for different applications based on EEG signals. Biocybernetics and Biomedical Engineering, 40(2), pp.649-690.
3. Al-Saegh, A., Dawwd, S.A. and Abdul-Jabbar, J.M., 2021. Deep learning for motor imagery EEG-based classification: A review. Biomedical Signal Processing and Control, 63, p.102172.
4. Ren, W. and Han, M., 2019. Classification of EEG signals using hybrid feature extraction and ensemble extreme learning machine. Neural Processing Letters, 50(2), pp.1281-1301.
5. Guler, I. and Ubeyli, E.D., 2007. Multiclass support vector machines for EEG-signals classification. IEEE transactions on information technology in biomedicine, 11(2), pp.117-126.
6. Al-Saegh, A., Dawwd, S.A. and Abdul-Jabbar, J.M., 2021. Deep learning for motor imagery EEG-based classification: A review. Biomedical Signal Processing and Control, 63, p.102172.
7. Pfurtscheller, G. and Neuper, C., 2001. Motor imagery and direct brain-computer communication. Proceedings of the IEEE, 89(7), pp.1123-1134.
8. Spatial–temporal discriminant analysis for ERP-based brain– computer interface

9. Two-stage frequency recognition method based on correlated component analysis for SSVEP-based BCI
10. Zhang, Y., Guo, D., Li, F., Yin, E., Zhang, Y., Li, P., Zhao, Q., Tanaka, T., Yao, D. and Xu, P., 2018. Correlated component analysis for enhancing the performance of SSVEP-based brain-computer interface. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 26(5), pp.948-956.
11. She, Q., Chen, K., Ma, Y., Nguyen, T. and Zhang, Y., 2018. Sparse representation-based extreme learning machine for motor imagery EEG classification. Computational intelligence and neuroscience, 2018.
12. Park, C., Looney, D., ur Rehman, N., Ahrabian, A. and Mandic, D.P., 2012. Classification of motor imagery BCI using multivariate empirical mode decomposition. IEEE Transactions on neural systems and rehabilitation engineering, 21(1), pp.10-22.
13. Jin, Z., Zhou, G., Gao, D. and Zhang, Y., 2020. EEG classification using sparse Bayesian extreme learning machine for brain–computer interface. Neural Computing and Applications, 32(11), pp.6601-6609.
14. Huang, G.B., Zhu, Q.Y. and Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing, 70(1-3), pp.489-501.
15. Nguyen, T., Khosravi, A., Creighton, D. and Nahavandi, S., 2015. EEG signal classification for BCI applications by wavelets and interval type-2 fuzzy logic systems. Expert Systems with Applications, 42(9), pp.4370-4380.
16. Soria-Olivas, E., Gomez-Sanchis, J., Martin, J.D., Vila-Frances, J., Martinez, M., Magdalena, J.R. and Serrano, A.J., 2011. BELM: Bayesian extreme learning machine. IEEE transactions on neural networks, 22(3), pp.505-509.
17. Zhang, Y., Wang, Y., Zhou, G., Jin, J., Wang, B., Wang, X. and Cichocki, A., 2018. Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces. Expert Systems with Applications, 96, pp.302-310.
18. Tangermann, M., Müller, K.R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., Leeb, R., Mehring, C., Miller, K.J., Mueller-Putz, G. and Nolte, G., 2012. Review of the BCI competition IV. Frontiers in neuroscience, 6, p.55.
19. Martinez-Hernandez, U., Rubio-Solis, A., Panoutsos, G. and Dehghani-Sanij, A.A., 2017, July. A combined adaptive neuro-fuzzy and bayesian strategy for recognition and prediction of gait events using wearable sensors. In 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-6). IEEE.
20. Rubio-Solis, A., Panoutsos, G., Beltran-Perez, C. and Martinez-Hernandez, U., 2020. A multilayer interval type-2 fuzzy extreme learning machine for the recognition of walking activities and gait events using wearable sensors. Neurocomputing, 389, pp.42-55.
21. Wang, Y., Gao, S. and Gao, X., 2006, January. Common spatial pattern method for channel selelction in motor imagery based brain-computer interface. In 2005 IEEE engineering in medicine and biology 27th annual conference (pp. 5392-5395). IEEE.
22. Gu, X. and Wang, S., 2018. Bayesian Takagi–Sugeno–Kang fuzzy model and its joint learning of structure identification and parameter estimation. IEEE Transactions on Industrial Informatics, 14(12), pp.5327-5337.
23. Berger, J.O., 2013. Statistical decision theory and Bayesian analysis. Springer Science & Business Media.
24. MacKay, D.J., 1995. Probable networks and plausible predictions-a review of practical Bayesian methods for supervised neural networks. Network: computation in neural systems, 6(3), p.469.
25. Liang, N.Y., Saratchandran, P., Huang, G.B. and Sundararajan, N., 2006. Classification of mental tasks from EEG signals using extreme learning machine. International journal of neural systems, 16(01), pp.29-38.
26. Rubio-Solis, A. and Panoutsos, G., 2014. Interval type-2 radial basis function neural network: a modeling framework. IEEE Transactions on Fuzzy Systems, 23(2), pp.457-473.

27. Hernandez-Hernandez, R.A., Martinez-Hernandez, U. and Rubio-Solis, A., 2020, July. Multilayer Fuzzy Extreme Learning Machine Applied to Active classification and Transport of objects using an Unmanned Aerial Vehicle. In 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-8). IEEE.
28. Rubio-Solis, A., Martinez-Hernandez, U. and Panoutsos, G., 2018, July. Evolutionary extreme learning machine for the interval type-2 radial basis function neural network: A fuzzy modelling approach. In 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-8). IEEE.