Classification of Epilepsy Seizure Phase using Interval Type-2 Fuzzy Support Vector Machines

Udeme Ekong, H.K. Lam, Bo Xiao, Gaoxiang Ouyang, Hongbin Liu, K.Y. Chan and Sai Ho Ling

Abstract

An interval type-2 fuzzy support vector machine (IT2FSVM) has been proposed to deal with a classification problem which aims to classify between three epileptic seizure phases (seizure-free, pre-seizure and seizure). The input data is from the electroencephalogram (EEG) signal obtained from 10 patients at Peking University Peoples Hospital. Three sets of EEG signals from the different seizure phases were collected where 112 2-second 19-channel EEG epochs for each patient were extracted for each dataset. Feature extraction is then carried out to reduce this to a feature vector of 45 elements which is then used as the input of classifier. The performance of the IT2FSVM classifier is measured based on its recognition accuracy for each of the epileptic seizure phases. Three traditional classifiers (Support Vector Machine, k-Nearest Neighbour and naive Bayes) are used for comparison purposes. The IT2FSVM classifier is able to show superior learning capabilities with the original data when compared to other classifiers. In order to gain an appreciation of the level of robustness of the classifiers, the original EEG dataset is contaminated with Gaussian white noise at levels of 0.05, 0.1, 0.2 and 0.5. The results obtained from simulations show that the IT2FSVM classifier outperforms the other classifiers under the original dataset and also shows a high level of robustness when compared to other classifiers with white gaussian noise applied to it.

Index Terms

Udeme Ekong, H.K. Lam, Bo Xiao and Hongbin Liu are with the Department of Informatics, King's College London, Strand, London, WC2R 2LC, United Kingdom. e-mail: {udeme.ekong, hak-keung.lam, bo.xiao, hongbin.liu}@kcl.ac.uk.

Gaoxiang Ouyang is with state Key Laboratory of Cognitive Neuroscience and Learning, School of Brain and Cognitive Sciences, Beijing Normal University, No.19, XinJieKoWai St., HaiDian District, Beijing, 100875, P. R. China. email: ouyang@bnu.edu.cn

Kit Yan Chan is with the Department of Electrical and Computer Engineering, Curtin University, Perth, Australia. e-mail: kit.chan@curtin.edu.au.

Sai Ho Ling is with the Centre for Health Technologies, Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW, Australia. e-mail: steve.Ling@uts.edu.au.

Manuscript received 2015.

Classification, Epilepsy, Fuzzy Support Vector Machine, Interval Type-2 Fuzzy Sets

I. Introduction

A classification problem can be best illustrated when an object or group of objects have to be assigned into a pre-defined group or class where the assignment is made based on a number of observed features/attributes pertaining to that particular object. Classification is a very important field of research due to the advantageous nature that a classifier of high generalization ability would have in the economical, industrial and medical field [1] just to name a few. As a result of this, extensive research has been carried out over the years and this has resulted in a large number of applications e.g., classification of different investment or lending opportunities as acceptable or unacceptable risk [2], hand-writing recognition [3], image classification [4], medical engineering [5] and speech recognition [6].

Generally speaking, the existing traditional methods for classification can be categorized into logic based (e.g decision trees) [2], statistical approach (e.g bayesian classification) [7], instance-based (e.g. nearest neighbor algorithm [8]), perceptron based (e.g single layer perceptrons and neural networks [9], [10]), and support vector machine (SVM) classification [11].

The decision tree method is a prime example of the logic based classification method. Classification is carried out by categorizing the inputs based on the feature values in the input [8]. A drawback of this method is that once the splitting rule makes a wrong decision, it is impossible to return to the correct path and this would therefore result in an accumulation of errors. Bayesian classifier is based on the assumption that equal prior probabilities exists for all classes [7]. The main limitation of the Bayesian classifier is that the posterior probabilities cannot be determined directly [9]. An example of the instance-based method is the k-Nearest Neighbour (kNN) [8] technique which is based on the principle that objects in a data set will generally exist in the neighbourhood of other objects with similar properties. The technique finds the k nearest objects to the particular input and determines its class by looking for the most frequent class label.

The single layer perceptron can be simply described as a component that computes the sum of weighted inputs and then feeds this to the output of the system. A major limitation of the single layer perceptron is that it can only learn linearly separable problems and is therefore incompatible when considering non-linear problems [10]. This problem is solved by the introduction of the Neural

Network (NN). The Neural Network can be divided into 3 distinct segments. The input units which have the primary responsibility of receiving information, the hidden units which contain neurons carry out the input-output mapping and the output units which store the processed results [8]. By determining properly the connection weights and transfer functions, NN can be regarded as a universal approximator [12] which is able to approximate any continuous functions (e.g., hyperplanes) to any arbitrary precision in a compact domain.

The Support Vector Machine (SVM) was first proposed by Vapnik in 1995 [8] as a machine learning model that went on to be applied to various supervised and unsupervised learning applications [13]–[15]. The SVM approach can be split into Support Vector Classification (SVC) which are used for task such as pattern recognition and Support Vector Regression (SVR) which is mainly applicable to time-series applications [13]. The main concept of the SVMs is that of the hyperplane which is used to separate two data classes. The aim of the SVM is to maximize the margin between the hyperplane and the input samples which is being separated by it thereby reducing the generalization error. Data that is difficult to separate on the input space is mapped into a higher dimensional feature space for ease of separation. Computations on the higher dimensional feature space are possible with the use of a kernel function [8]. This feature illustrates a very important trait of the SVM which is its ability to perform well in a high dimensional feature space [16], [17].

The SVM performs structural risk minimization (SRM) which aims to balance the complexity of the model with its ability to accurately fit the input data [8]. This therefore gives the SVM good generalization ability for classification problems as it can simultaneously minimize the empirical risk [11]. The SRM principle is grounded on the fact that the generalization error of the model is bounded by the sum of the empirical error and a confidence interval which is based on the Vapnik-Chervonenkis (VC) dimension [8], a higher classification performance is achieved by minimizing this bound. The SVM also provides a global optimization solution to the problem at hand and therefore provides a more credible output when compared to the neural network which provides a local optimization solution [11]. One of the drawbacks of the SVM method is its sensitivity to outliers that may exist in the input data, this stems from the fact that the same penalty weight is assigned to each data point and an outlier would therefore significantly distort the representation of the input signal and therefore affect the classification performance. Another drawback is seen in the instance when the SVM is applied to a classification problem with an imbalanced data set (where negative data significantly outweighs the positive data) the optimum separating hyperplane in this case can be skewed towards the positive with the consequence being that the SVM could be very ineffective in identifying targets that should be mapped to the positive class [13], [16].

A relatively recent classification method is based on fuzzy logic [18] which is the theory of fuzzy sets used to handle fuzziness/imprecision in datasets. This is done by assigning each variable with membership functions with respect to its relative distance to the class [4], [18]. There are two main types namely type-1 and type-2 fuzzy sets [5], [19], [20]. In type-1 fuzzy sets, the membership values are precise numbers in the range of 0 and 1 whilst the membership grades of a type-2 fuzzy set is a type-1 fuzzy set due to the imprecision in assigning a membership grade. As a result, type-2 fuzzy sets are useful as they offer an opportunity for the modeling of higher level uncertainty in the human decision making process when compared to the type-1 fuzzy set where the membership grade is distinct. In fuzzy logic, classification rules are specified by the user instead of being inherently decided upon by the machine learning method like in the SVM or NN, this means that it is not a black-box method and the decision rules are clearly visible. Fuzzy logic has been combined with the NN and SVM and this gave birth to the Neural Fuzzy Network (NFN) and Fuzzy Support Vector Machine (FSVM) [14]. The NFN works well when the sample data provided is sufficient but suffers from a significantly reduced generalization performance when the sample data is limited, the FSVM however works well even when the sample data is limited and is proven to provide higher generalization performance [14].

When considering a real world application of the SVM, it is important to account for the difficulty in obtaining a precise measurement of the input data. One of the major disadvantages of the SVM technique was its sensitivity to outliers and noise in the input data, this is due to the fact that the SVM assigns the same penalty cost to each data point, the FSVM is able to solve this problem by assigning membership functions to each data point which vary according to the relative importance of this data point, this therefore helps in reducing the impact of outliers in the input dataset [21].

The application being considered in this paper is the recognition of the phases involved in the onset of an epileptic seizure, the epilepsy signals obtained from the Electroencephalograph (EEG) using real clinical data is subjected to the novel classification technique [22], [23]. The fact that there are multiple features and also the susceptibility of the EEG data to noise results in a very challenging classification problem [24], [25]. The classification technique is designed to differentiate between the 3 seizure phases (seizure-free, pre-seizure and seizure). The early detection of seizure phases is a potentially life-saving application/research field and this is a major motivation for the research being carried out in this paper. The accurate classification/differentiation between the 3 seizure phases would give doctors and other healthcare professionals ample time to be able to prepare for the oncoming seizure. An interval type-2 fuzzy support vector machine (IT2FSVM) is being proposed to deal with this problem. The IT2FSVM will be utilized to differentiate between the 3 seizure phases. The FSVM is proposed due to its superior ability at dealing with uncertainties and unbalanced data [21], this would therefore prove to provide a higher level of recognition accuracy than the traditional SVM and forms the basis for the implementation of this classifier. The classification performance of the IT2FSVM technique will be compared to some traditional classifiers like the kNN technique, SVM and naive Bayes classifier.

This paper is organized as follows. Section II reviews the SVM theory. Section III reviews the interval type-2 fuzzy inference system (IT2FIS). Section IV proposes the IT2FSVM structure with a detailed schematic to illustrate how it functions. Section V introduces epilepsy, data collection and feature extraction. Section VI presents the classification method to deal with the epilepsy seizure phase classification problem. Section VII contains the experimental results obtained from the application of the IT2FSVM method to the epilepsy seizure phase classification problem with a comparison to other existing methods followed by a discussion of the results obtained. Section VIII draws a conclusion.

II. Support Vector Machines

The SVM theory is reviewed in this section, which provides the theoretical background to the development of IT2FSVM. The main objective of the SVM is to create a separating hyperplane such that the distance between the hyperplane and the nearest data point in each class is maximized.

Given a dataset S containing labelled training points

$$(y_1, x_1), (y_2, x_2), \ldots, (y_N, x_N) \qquad i = 1, 2, \ldots, N \tag{1}$$

where vector x_i represents the training point, y_i represents the label and N represents the total number of samples. x_i is assigned to either of two classes and is represented by the class label $y_i \in \{-1, 1\}$. The hyperplane is ideally placed in the middle of the margin between the two classes being separated. The data points that are in close proximity to the margin are the basis of its definition and are known as support vectors (SVs) [8]. In the case of a non-linear function, searching for the optimum hyperplane in the input space proves to be difficult, as a result of this, the input space is mapped onto a higher dimensional feature space. Let $z = \varphi(x)$ represent the feature vector where x is an input vector and $\varphi(x)$ is a transformation function. The hyperplane can then be defined as

$$\omega \cdot z + b = 0 \tag{2}$$

where z is the feature space vector, ω is the weight vector and b is the scalar threshold (bias). The set S can be said to be linearly separable if there exists a combination of $\boldsymbol{\omega}$ and b that satisfy the following inequalities for all elements of the set S.

$$\begin{cases} \omega \cdot z_i + b \geq 1, & \text{if } y_i = 1 \\ \omega \cdot z_i + b \leq -1, & \text{if } y_i = -1, \qquad i = 1, 2, \ldots, N \end{cases} \tag{3}$$

where $z_i = \varphi(x_i)$.

In the case where the set S is not linearly separable for all of its elements, a leeway for some classification violations must be allowed in order to accommodate the elements of the set that are not linearly separable. This problem is dealt with by introducing non-negative slack variables $\xi_i \geq 0$ for the points x_i which do not satisfy (3). (3) is then modified to

$$\begin{cases} \omega \cdot z_i + b \geq 1 - \xi_i, & \text{if } y_i = 1 \\ \omega \cdot z_i + b \leq -1 - \xi_i, & \text{if } y_i = -1, \qquad i = 1, 2, \ldots, N \end{cases} \tag{4}$$

The optimal hyperplane can be obtained as a solution to the constrained optimization problem

$$min \qquad \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}\xi_i \tag{5}$$

subject to

$$y_i(\omega \cdot z_i + b) \geq 1 - \xi_i, \qquad i = 1, 2, \ldots, N \tag{6}$$

$$\xi_i \geq 0 \qquad i = 1, 2, \ldots, N \tag{7}$$

where (5) is the convex cost function, (6) and (7) are the constraints, $\|\cdot\|$ denotes the l^2 norm (i.e.

Euclidean norm) and C is known as the regularization constant. It is the only free parameter in the SVM formulation and can be tuned to find a balance between margin maximization and classification violation. The optimal hyperplane can be found by constructing a Lagrangian multiplier and obtaining the dual formation:

$$min \qquad Q(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j z_i \cdot z_j - \sum_{i=1}^{N} \alpha_i \tag{8}$$

subject to

$$\sum_{i=1}^{N} y_i \alpha_i = 0, \qquad 0 \leq \alpha_i \leq C, \qquad i = 1, 2, \ldots, N \tag{9}$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ represents the vector of the nonnegative langrange multipliers which satisfy the constraints in (5).

A very important theorem for the SVM theory is the Karush-Kuhn-Tucker theorem [26] which states that the solution α_i to (9) satisfies the following conditions:

$$\alpha_i(y_i(\omega \cdot z_i + b) - 1 + \xi_i) = 0, \qquad i = 1, 2, \ldots, N \tag{10}$$

$$(C - \alpha_i)\xi_i = 0, \qquad i = 1, 2, \ldots, N \tag{11}$$

The equalities (10) and (11) suggest that it is only the nonzero values α_i in (8) that satisfy the constraints in (6). The values of x_i that corresponds with the solution α_i are known as support vectors. The instances where x_i corresponds with $\alpha_i = 0$ is correctly classified and is of a significant distance away from the decision margin.

For the construction of the optimal hyperplane $\omega \cdot z + b$, we would require that

$$\omega = \sum_{i=1}^{N} \alpha_i y_i z_i \tag{12}$$

and the scalar bias b should be determined via the Karush-Kuhn-Tucker conditions in (10).

The decision function can then be derived from (3) and (12) to give

$$f(x) = \text{sgn}(\omega \cdot z + b) = sgn\Big(\sum_{i=1}^{N} \alpha_i y_i z_i \cdot z + b\Big) \tag{13}$$

where sgn$(\cdot)$ represents the sign function which extracts the sign (positive or negative) of a real number.

As we have no knowledge of the higher dimensional feature space $\varphi(\cdot)$, carrying out the computation in (8) and (13) would be rendered impossible due to its complicated nature. An advantageous characteristic of the SVM is that it is not necessary to know about the $\varphi(\cdot)$. The problem is alleviated with the aid of a kernel function which has the ability to compute the dot product of the data points in the feature space of z, it is however obligatory for these functions to satisfy Mercer's theorem [27] before they can be used for computing the dot product [21].

$$z_i \cdot z_j = \varphi(x_i) \cdot \varphi(x_j) = K(x_i, x_j) \tag{14}$$

where $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$ is the kernel function which is used for the mapping onto a higher dimensional feature space. The kernel functions may be linear or nonlinear. The nonlinear separating hyperplane can then be found by solving the following equation

$$min \qquad Q(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{i=1}^{N} \alpha_i \tag{15}$$

subject to

$$\sum_{i=1}^{N} y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, \qquad i = 1, 2, \ldots, N. \tag{16}$$

The decision function can then be described as follows:

$$f(x) = \text{sgn}(\omega \cdot z + b) = sgn\Big(\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b\Big) \tag{17}$$

III. Interval Type-2 Fuzzy Inference System (IT2FIS)

Fuzzy inference systems are mainly used to represent the relationship between the input and output variables in systems which are governed by a selection of IF-THEN rules which utilize linguistic labels for the expression of rules and facts. An IT2FIS is a fuzzy logic system where the uncertainty of the membership functions are incorporated into fuzzy set theory, in the circumstance where no uncertainty exists, a type-2 fuzzy set would reduce to a type-1 fuzzy set, and this is identical to the concept of probability reducing to the determinism when the unpredictability is eradicated [28]. In order to distinguish between a type-1 and type-2 fuzzy set, a tilde symbol is placed above the symbol for the fuzzy set, in this case, A would represent a type-1 fuzzy set and $\tilde{A}$ would represent a type-2 fuzzy set [29].

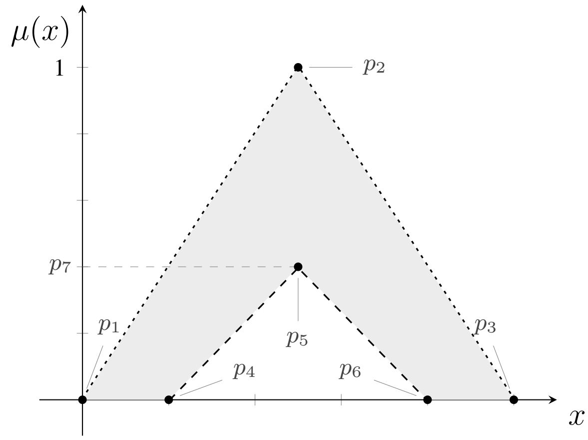

Fig. 1. An example of IT2 membership functions. Dashed line: lower membership function. Dotted line: Upper membership function. Gray area: footprint of uncertainty.

An example triangular IT2FIS membership function is shown in Fig. 1. The shape of the membership function is a triangle, with the dashed lines representing the lower membership function LMF and the dotted line representing the upper membership function UMF. The membership function can either be chosen by the users or it can be designed with the aid of optimization methods such as the genetic algorithm (GA). The shape of the membership function for each input is represented by seven points (p_1 to p_7) which are optimised with the aid of GA. Unlike in the type-1 case where the membership grade is a crisp value, the membership grade in an IT2FIS is an interval. The IT2FIS is then bounded at the two extremes of this interval to give us the LMF and UMF which are both type-1 fuzzy sets. The area between the UMF and LMF is known as the footprint of uncertainty (FOU) which is shown as the gray area in Fig. 1.

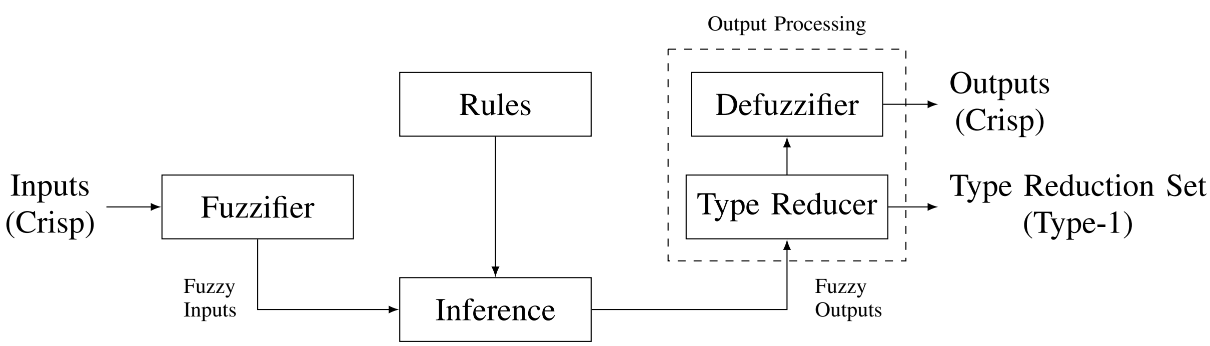

Fig. 2. Block diagram showing the IT2FIS.

Type-2 fuzzy sets are seen to be more prevalent than type-1 fuzzy sets in rule-based fuzzy logic systems as they have a higher level of non-linearity and therefore have the ability to model uncertainties better than the type-1 fuzzy sets with less number of rules. The structure of the IT2FIS detailing the input-output relationship is shown in Fig. 2. The IT2FIS consists of 5 major components [30]: fuzzifier, fuzzy rules, inference engine, type-reducer and defuzzifier. The crisp input is first transformed into fuzzy sets in the fuzzifier block as the rule base is activated by fuzzy sets and not numbers. In the fuzzification stage, when the measurements are perfect the input is modelled as a crisp data set, when the measurements are noisy but stationary it is modelled as an interval type-2 fuzzy set. After the input is fuzzified, the fuzzy input set is then mapped onto the fuzzy output set with the aid of the inference block. This is achieved by quantifying each rule using fuzzy set theory and then using the mathematics behind fuzzy set theory to obtain an output for each rule. The output of the fuzzy inference block would then contain one or more fuzzy output sets. The fuzzy output sets are then converted into a crisp output with the aid of the output processing unit. In an IT2FIS the output processing unit consists of two blocks: the type-reducer and the defuzzifier blocks. In the first step, the IT2 fuzzy output set is reduced to an interval-valued type-1 fuzzy set in a process known as type-reduction.

The most prevalent of these is an algorithm developed by Karnik and Mendel [30] known as the Karnik-Mendel (KM) algorithm which is iterative and very fast in achieving a state of convergence. The second step of ouput processing occurs after type-reduction. In the case of the KM algorithm being used as a type-reducer, the type-reduced set is always confined to a finite interval of numbers, the deffuzifier then obtains the defuzzified value (which is a crisp output) by calculating the average of the upper and lower bounds of this interval.

Given an IT2FIS with n inputs $x_1 \in X_1, x_2 \in X_2, \dots, x_n \in X_n$ to give a singular output $y \in Y$. The rule base for this IT2FIS consists of K IT2 fuzzy rules expressed in the following form [20]:

$$R^k : \text{If } x_1 \text{ is } \tilde{F}_1^k \text{ and } \cdots \text{ and } x_n \text{ is } \tilde{F}_n^k \text{ THEN } y \text{ is } \tilde{G}^k \tag{18}$$

where $k = 1, 2, \dots, K$, $\tilde{F}_n^k$ and $\tilde{G}^k$ represent type-2 fuzzy sets.

The rules are responsible for the mapping of an input space X to an output space Y. Experimentation has shown that the general T2FIS model has high complexity and large computational costs. This has resulted in the development of the IT2FIS which makes the computation simplified. The membership grades for interval fuzzy sets can be portrayed by their lower and upper membership grades of the FOU. The output of the firing strength for an IT2FIS ω_i is represented by a lower and upper bound i.e., $\omega_i = [\underline{\omega}_i, \overline{\omega}_i]$. The defuzzified output is obtained by type reduction which is implemented using the KM algorithm which is shown below [30]:

A. KM Algorithm (Lower Bound)

1) Arrange the lower bound of the output $\underline{x}_i (i = 1, 2, \ldots, n)$ in ascending order and then assign the same labels to them such that $\underline{x}_1 \leq \underline{x}_2 \leq \cdots \leq \underline{x}_n$.
2) Match the weights ω_i with the corresponding $\underline{x}_i$ and reassign the labels to match with the new $\underline{x}_i$ which are now in ascending order.
3) Initialize ω_i, i.e.,

$$\omega_i = \frac{\underline{\omega}_i + \overline{\omega}_i}{2} \text{ where } i = 1, 2, \ldots, n \tag{19}$$

then calculate

$$y = \frac{\sum_{i=1}^{n} \underline{x}_i \omega_i}{\sum_{i=1}^{n} \omega_i} \tag{20}$$

4) Find the pivot point p where $(1 \leq p \leq N - 1)$ such that

$$\underline{x}_p \leq y \leq \underline{x}_{p+1} \tag{21}$$

5) Assign the firing strength as

$$\begin{cases} \overline{\omega}_i, & i \leq p \\ \underline{\omega}_i, & i > p \end{cases} \tag{22}$$

then calculate

$$y' = \frac{\sum_{i=1}^{n} \underline{x}_i \omega_i}{\sum_{i=1}^{n} \omega_i} \tag{23}$$

6) Check if $y' = y$. If yes, stop and set $\underline{y} = y$, if no, go to step 7
7) Set $y = y'$ and go to step 3

B. KM Algorithm (Upper Bound)

1) Arrange the upper bound of the output $\overline{x}_i (i = 1, 2, \ldots, n)$ in ascending order and then assign the same labels to them such that $\overline{x}_1 \leq \overline{x}_2 \leq \ldots \leq \overline{x}_n$.
2) Match the weights ω_i with the corresponding $\overline{x}_i$ and reassign the labels to match with the new $\overline{x}_i$ which are now in ascending order.
3) Initialise ω_i i.e

$$\omega_i = \frac{\underline{\omega}_i + \overline{\omega}_i}{2} \text{ where } i = 1, 2, \ldots, n \tag{24}$$

then calculate

$$y = \frac{\sum_{i=1}^{n} \overline{x}_i \omega_i}{\sum_{i=1}^{n} \omega_i} \tag{25}$$

4) Find the pivot point p where $(1 \leq p \leq N - 1)$ such that

$$\overline{x}_p \leq y \leq \overline{x}_{p+1} \tag{26}$$

5) Assign the firing strength as

$$\begin{cases} \underline{\omega}_i, & i \leq p \\ \overline{\omega}_i, & i > p \end{cases} \tag{27}$$

then calculate

$$y' = \frac{\sum_{i=1}^{n} \overline{x}_i \omega_i}{\sum_{i=1}^{n} \omega_i} \tag{28}$$

6) Check if $y' = y$. If yes, stop and set $\overline{y} = y$, if no, go to step 7
7) Set $y = y'$ and go to step 3

The defuzzified output of the IT2FIS is

$$y = \frac{\overline{y} + \underline{y}}{2} \tag{29}$$

IV. Interval Type-2 Fuzzy Support Vector Machines (IT2FSVMs)

In this section, the IT2FSVM classifier is introduced. The standard SVM classifier is used for this hybrid classification mechanism which involves the merging of an IT2FIS with an SVM to form the IT2FSVM. The IT2FSVM can be characterized as a multiple-input-single-output classifier. The ability of the IT2FIS to handling non-linear data makes it very complementary to the SVM in solving difficult problems.

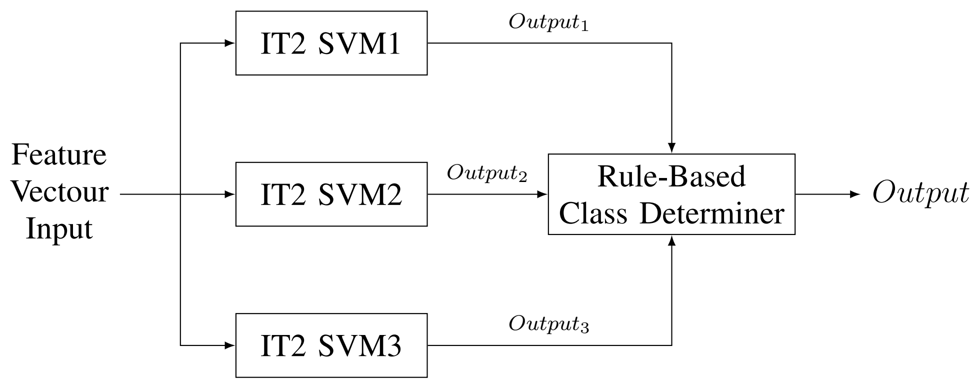

Fig. 3. Block diagram of IT2FSVM.

The overall IT2FSVM architecture is shown in Fig. 3. As the hyperplane can only separate between 2 classes, multiple SVMs will be necessary in a case where there are more than 2 classes in a classification problem. For the application in this paper which is to differentiate between the epileptic seizure stages, multiple SVMs will be needed as there are three classes (seizure-free, pre-seizure and seizure). There are three IT2 SVM blocks in the diagram which are used to individually separate between the seizure phases. IT2 SVM 1 separates between the seizure-free and pre-seizure phases with the label "-1" indicating the input data belongs to the seizure-free class and label "1" indicating the input data belongs to the pre-seizure class. IT2 SVM 2 separates between the seizure-free and seizure phase with the label "-1" indicating the input data belongs to the seizure-free class and label "1" indicating the input data belongs to the seizure class. IT2 SVM 3 separates between the pre-seizure and seizure phase with the label "-1" indicating the input data belongs to the pre-seizure class and label "1" indicating the input data belongs to the seizure class. The output labels of the three IT2 SVM blocks are presented in $Output_1$ to $Output_3$ which are then subjected to a rule-based class determiner in order to determine what the final classification would be.

The rule based class determiner system for selecting the final classification output for the IT2FSVM

Case	$Output_1$	$Output_2$	$Output_3$	Final Class ($Output$)
1	-1	-1	-1 or 1	1
2	1	-1 or 1	-1	2
3	-1 or 1	1	1	3
4	1	-1	1	3
5	-1	1	-1	3

TABLE I
TABLE SHOWING THE IF-THEN RULES USED BY THE RULE BASED CLASS DETERMINER SYSTEM. TABLE SHOWING THE IF-THEN RULES. CLASS 1: SEIZURE-FREE, CLASS 2: PRE-SEIZURE, CLASS 3: SEIZURE

is shown in Table. I. The final class is a whole number between 1 and 3 where "1" representing the seizure-free phase, "2" representing the pre-seizure phase and "3" representing the seizure phase.

The IT2FSVM block consists of a feature vector input, 3 fuzzy rules each consisting of two SVMs associated with the lower and upper membership functions and a defuzzification block which is used to produce the final crisp output. The original EEG input data had a 19 $\times$ 100 vector input and feature extraction is used to reduce it to a 45-input feature vector which is used as the input of the IT2SVM. More details about feature extraction and feature will be provided later on.

The final output is obtained by combining a number of SVMs with the aid of fuzzy rules (which determine the number of SVMs) and membership grades or weights (which depict the impact that a particular fuzzy rule would have on the final output). There is no limit to the number of fuzzy rules that can be applied in this instance but an increase in the number of fuzzy rules would lead to a slower convergence of training and also a higher computational cost of the system. In this paper, there are 3 fuzzy rules employed to implement the IT2FSVM. The membership grade is obtained from the membership function which is defined by the user and the shape of the membership function is a triangle as shown in Fig. 1. The shape of the membership function is represented by the points p_1 to p_7 which are then optimized with the aid of GA.

Referring to Fig. 3, we have three IT2 SVMs. Each IT2 SVM is governed by the following rules:

$$R^j : \text{If } \|x\| \text{ is } \tilde{F}^j \text{ THEN } y \text{ is } \tilde{G}^j, j = 1, 2, 3 \tag{30}$$

where $\tilde{F}^j$ is defined as an IT2 triangular membership function as shown in Fig. 1 and $\tilde{G}^j$ is a singleton with $\underline{\text{SVM}}_{jk}$ as LMF and $\overline{\text{SVM}}_{jk}$ as UMF, k = 1, 2, 3, denoting the number of IT2FSVMs in Fig.

2. $\underline{\text{SVM}}_{jk}$ and $\overline{\text{SVM}}j_k$ are two SVMs with the output $\underline{Out}_{jk}$ and $\overline{Out}_{jk}$ defined by the following hyperplanes:

$$\underline{Out}_{jk} = \underline{\omega}_{jk} \cdot z + \underline{b}_{jk} = \sum_{i=1}^{N} \overline{\alpha}_{ijk} y_i K(x_i, x) + \underline{b}_{jk} \tag{31}$$

$$\overline{Out}_{jk} = \overline{\omega}_{jk} \cdot z + \overline{b}_{jk} = \sum_{i=1}^{N} \overline{\alpha}_{ijk} y_i K(x_i, x) + \overline{b}_{jk} \tag{32}$$

where j = 1, 2, 3 denotes the j-th (lower or upper) SVM in Fig. 2 and k = 1, 2, 3 denotes the k-th IT2 SVM in Fig. 2. The $Output_k$ of the IT2 SVM k can then be obtained by the KM algorithm introduced in Section III. The rule-base class determine will round it off to "1" or "-1" before making the final class decision.

V. Absence Epilepsy

Epilepsy, which is characterized with its ability to instantiate recurrent seizures (an interruption of normal brain functions) which are unforeseen in nature is a very common and significant neurological disorder caused by a sudden discharge of cortical neurons [22], [23]. Epileptic seizures are classified as either partial (involving focal brain regions) or generalized (where it involves a widespread region of the brain across both hemispheres) [31]. The length of time for the seizure occurrence varies from a few seconds up to a minute with some of the effects including momentary lapse of consciousness for the sufferer of the seizure [31]. A complete loss of consciousness occurs when the epileptic activity involves both the cortical and subcortical structures of the brain and this occurrence is known as an absence seizure.

The unexpected nature of these seizures has proven to have an adverse effect on the quality of life for those who are suffering from them. The impact is most prevalent in the formative stages of a childs life as we see an increase in the requirements for special education and also a higher incidence of below-average school performance [23], [32]. It also proves life-threatening in situations where the sufferer is isolated at the time of its occurrence and there is no experienced or medical help on hand to alleviate the situation. Therefore having an accurate understanding or predictive model for the pre-seizure phase (the transition towards an absence seizure occurrence) is a very vital task as it would provide the sufferers and their carers enough notice of the upcoming seizure so they could prepare themselves and dampen the impact of the seizure occurrence.

Absence seizures can be best characterized by the spike-and-wave discharges (SWDs) which are as a result of synchronized oscillations in the thalamocortical networks of the brain [33], [34]. The classification process of EEG signals consists of two main parts which are feature extraction and classification. In the literature, there are a wide range of available feature extraction methods which range from the traditional methods to the non-linear methods. Traditional methods include the fourier transform and also spectral analysis whilst the non-linear methods include Lyapunov exponents [23], [35], correlation dimension [23] and similarity [36]. After feature extraction has been implemented to the raw data, the extracted features are then used and applied to the pre-determined classification technique. There are a wide range of classification techniques for EEG classification in the literature, examples of these include the artificial neural network [37], [38] and also the neuro-fuzzy systems [39]–[41].

For this particular problem of accurately classifying and thereby predicting the onset of an epileptic seizure, the extracted features are applied to various classifiers (kNN, naive Bayes, SVM and FSVM) with the main aim of being able to recognize and distinguish between the 3 seizure phases (seizure-free, pre-seizure and seizure phase). The raw data obtained for the simulations being carried out were obtained from the Peking University by the aid of 10 patients who were suffering from absence epilepsy, their ages ranging from 6 to 21 years old. The study has been approved by the ethics committee of Peking University Peoples Hospital and the patients all signed documents in consent of their clinical data being used for research purposes. The EEG data (which was sampled at a frequency of 256 Hz with the aid of a 16-bit analogue-to-digital converter and then filtered within a frequency band of 0.5 to 35 Hz) was recorded by the Neurofile NT digital video EEG system using a standard international 10-20 electrode placement (Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz and Pz).

There are 3 sets of EEG signals which are extracted from the 3 seizure phases (seizure-free, pre-seizure and seizure) to obtain 112 2-second 19-channel EEG epochs from 10 patients for each dataset. The timing of the onset and offset in the SWDs were identified by a neurologist and these SWDs were identified to be large amplitude 3-4Hz discharges with a spike-wave morphology typically lasting above a second in duration. The criteria for determining the different seizure phases are that there is an interval between the seizure-free phase and beginning of the seizure phase which is greater than 15 seconds, the interval is between 0 to 2 seconds before the occurrence of the seizure and that the interval occurs during the first 2 seconds of the absence seizure. A more detailed description of the procedure for data collection can be found in [42], [43].

A. Feature extraction

The feature extraction procedure is very vital in the classification process as it obtains the relevant characteristics and information from a large dataset (EEG signals in this instance), this has the knock-on effect of simplifying the dataset and also reducing the effect of redundant data points that have little or no effect in the classification of the dataset. This is a very important step in improving the performance of the classifier as classification is easier when the classifier is subject to fewer data points.

For the EEG case being undertaken, there are 19 columns (19 channels) of signal output. The 19 columns represent signals that were drawn from 19 EEG sensors with each column containing 100 sample points. The purpose of the feature extraction being carried out here is to extract the relevant feature points from the 19×100 dataset and thereby reducing the dimensionality.

Research into the existing literature provides evidence to suggest that the 19 channels of the EEG data vary in importance with regards to classification. It was observed that some of the channels have a lesser impact on the classification of the EEG and the exclusion of these channels has been investigated in [24], [25]. Both studies have discovered that some of the electrodes (F3, Fz, F4, C3 and Cz) are the most significant ones for the classification between the seizure-free and seizure patients and the remaining electrodes are found to have relevant information for the classification between the different seizure phases.

We utilize the relevant channels for classification (i.e channel 1, 2, 3, 4, 5, 6, 11, 12, 13, 14) for the simulations carried out in this paper. For each of the channels, a feature vector containing the time-domain and frequency-domain components of the dataset is created [38]. The first part of the feature vector comprises of computations in the time-domain such as the standard deviation, second order norm, third order norm, fourth order norm, absolute sum, maximum value and minimum value of the 100 sample points from each channel. The second part is comprised of computations in the frequency domain such as the mean frequency, maximum frequency, minimum frequency, standard deviation of frequency, windowing filtered mean frequency and windowing filtered maximum frequency of each chosen channel will form the second part of the feature vector.

A problem that arises from these computations would result in a large vector which would be difficult to classify. This is solved by implementing principal component analysis (PCA) to reduce the number of dimensions in the feature vector. After this dimensionality reduction method has been implemented, we finally have 45 points which form the feature vector. This feature vector is then applied to the pre-determined classifiers.

VI. METHOD

A classifier based on the proposed IT2FSVM structure has been implemented for the classification of the 3 seizure phases with the aid of the feature vectors obtained from the feature extraction method applied in the preceding section. The structure of the IT2FSVM consists of 3 IT2 SVM blocks that are used to distinguish between the 3 seizure phases. Fig. 3 shows the overall structure of the FSVM classifier which consists of 18 45-input-single-output SVMs (6 for each of the IT2 SVM blocks). The need for 3 sets of SVM machines to distinguish between 3 classes of data stems from the fact that the SVM can only separate between 2 classes at any given time.

There are 3 fuzzy rules for each of the IT2 SVM blocks. The parameters of the triangular membership functions, i.e., p_1 to p_7, as shown in Fig. 1 are optimized with the aid of GA which has the ability to influence the shape of the membership functions. The GA optimization is performed to maximize the recognition accuracy using 70% of dataset as the training samples. The rest 30% of dataset are used as the test samples. The lower and upper membership functions for SVM Block 1 to 3 after training are shown in Figs. 4 to 6. The membership grade is represented on the y-axis and the normalized inputs are represented on the x-axis. The normalized input denoted as x_{norm} is calculated as follows:

$$x_{norm} = \overline{x}_1{}^2 + \overline{x}_2{}^2 + \ldots + \overline{x}_N{}^2 \tag{33}$$

where

$$\overline{x}_i = \frac{x_i}{\max(x) - \min(x)}, i = 1, 2, \ldots, N, \tag{34}$$

x_i is the i-th element of x, $\min(x)$ and $\max(x)$ denote the minimum and maximum value of the elements in x, respectively.

The simulations that have been conducted with the aid of the MATLAB software. The control parameters of the GA are shown in Table II. Different combinations of kernel functions we utilized

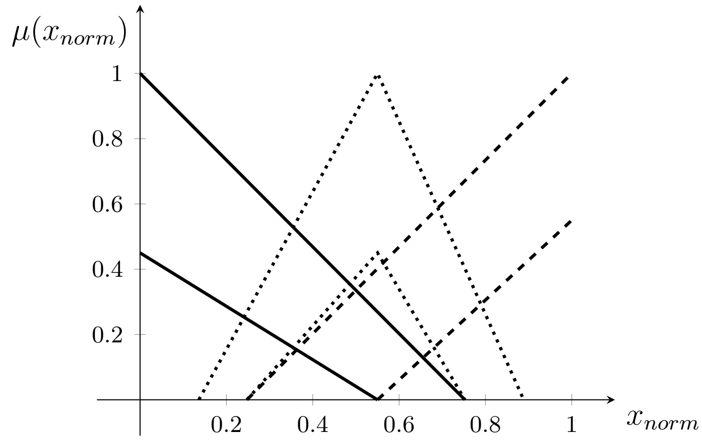

Fig. 4. Membership functions for SVM Block 1. Dotted line: Membership function for rule 1, Straight Line: Membership function for rule 2, Dashed Line: Membership function for rule 3.

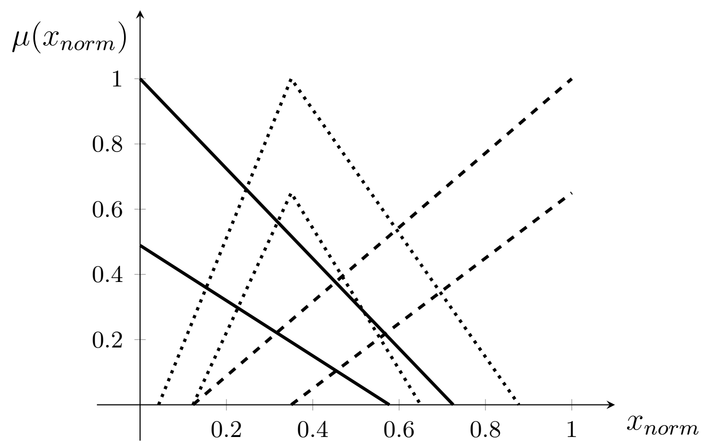

Fig. 5. Membership functions for SVM Block 2. Dotted line: Membership function for rule 1, Straight Line: Membership function for rule 2, Dashed Line: Membership function for rule 3.

in the SVMs. The optimal combination was chosen based on its ability to maximize the recognition accuracy of the classifier. The parameters used for the SVM are as follows. In the IT2 SVM1, there are 6 SVMs used, with all utilizing the RBF kernel function with the width of the RBFs for all 6 of them set to $\sqrt{1/200}$, and the regularization constant $C = 500$. In IT2 SVM2 there are 6 SVMs used, with the polynomial kernel function applied in all cases and the degree of polynomial set to 2, and $C = 5000$. In IT2 SVM3 the kernel function utilised for all SVMs is the quadratic kernel function with $C = 500$.

In order to gain an appreciation of the robustness of the proposed classifier, white Gaussian noise of levels 0.05, 0.1, 0.2 and 0.5 have been added to the original test dataset. Under these noisy conditions, the simulations were carried out 10 times for each of the noise levels and the worst, average, best and

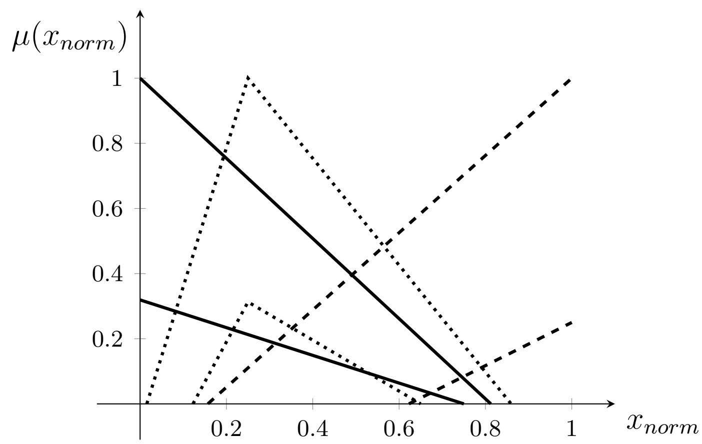

Fig. 6. Membership functions for SVM Block 3. Dotted line: Membership function for rule 1, Straight Line: Membership function for rule 2, Dashed Line: Membership function for rule 3.

Parameter	Value
Number of Iterations	10
Population Size	20
Selection	Stochastic uniform selection function
Elitism	Yes (Best two chromosomes are passed onto the next generation)
Crossover	Scattered Crossover
Crossover Fraction	0.8
Mutation	Gaussian Mutation
Stopping Criterion	It stops when the weighted average relative change in the best fitness function value over 100 generations is less than or equal to 10^{-6}

TABLE II
GA PARAMETERS

standard deviation of recognition accuracy were calculated. This was done to aid fair comparison due to the fact that the noisy data is random in nature and drawing conclusions from a single simulation would not accurately represent the robustness of the classifier to noise.

	Recognition Accuracy (%)			
Classifier	Average	Class 1	Class 2	Class 3
1	99.0510	100.000	97.1400	100.0000
2	86.6667	100.0000	90.0000	70.0000
3	100.0000	100.0000	100.0000	100.0000
4	77.1400	90.0000	41.4333	100.0000

TABLE III
SUMMARY OF TRAINING SAMPLES RECOGNITION PERFORMANCE FOR EEG SIGNAL WITH ORIGINAL DATASET. CLASSIFIER 1: FSVM CLASSIFIER, CLASSIFIER 2: TRADITIONAL SVM CLASSIFIER, CLASSIFIER 3: K-NEAREST NEIGHBOR CLASSIFIER, 4: NAIVE BAYES CLASSIFIER.

	Recognition Accuracy (%)			
Classifier	Average	Class 1	Class 2	Class 3
1	87.7800	100.000	70.0000	93.3300
2	71.1100	90.0000	70.0000	53.333
3	56.6667	96.6700	23.3333	50.0000
4	77.7778	100.0000	33.3333	100.0000

TABLE IV

SUMMARY OF TESTING SAMPLES RECOGNITION PERFORMANCE FOR EEG SIGNAL WITH ORIGINAL DATASET. CLASSIFIER 1: FSVM CLASSIFIER, CLASSIFIER 2: TRADITIONAL SVM CLASSIFIER, CLASSIFIER 3: K-NEAREST NEIGHBOR CLASSIFIER, 4: NAIVE BAYES CLASSIFIER.

	Recognition Accuracy (%)						
Classifier	Worst	Mean	Best	Std	Class 1	Class 2	Class 3
1	62.2200	66.1100	68.8900	0.0211	8.3000	93.0000	97.0000
2	61.1100	66.2200	68.8900	0.0235	11.1333	96.0000	97.3333
3	56.6700	57.8900	58.8900	0.0176	96.0000	25.0000	52.6700
4	77.7778	78.3333	80.0000	0.7857	99.0000	37.8900	100.0000

TABLE V

SUMMARY OF TESTING RECOGNITION PERFORMANCE FOR EEG SIGNAL UNDER DATASET SUBJECT TO NOISE LEVEL OF 0.05. CLASSIFIER 1: FSVM CLASSIFIER, CLASSIFIER 2: TRADITIONAL SVM CLASSIFIER, CLASSIFIER 3: K-NEAREST NEIGHBOR CLASSIFIER, 4: NAIVE BAYES CLASSIFIER.

	Recognition Accuracy (%)						
Classifier	Worst	Mean	Best	Std	Class 1	Class 2	Class 3
1	74.4400	79.4400	85.5600	0.0034	55.3333	66.3333	99.0000
2	66.6700	68.6700	70.0000	0.0126	15.0000	89.0000	98.3333
3	54.4400	56.2200	57.8800	0.0228	92.6700	22.0000	54.0000
4	76.6667	78.8889	82.2222	1.8251	100.0000	33.6667	100.0000

TABLE VI

SUMMARY OF TESTING RECOGNITION PERFORMANCE FOR EEG SIGNAL UNDER DATASET SUBJECT TO NOISE LEVEL OF 0.1. CLASSIFIER 1: FSVM CLASSIFIER, CLASSIFIER 2: TRADITIONAL SVM CLASSIFIER, CLASSIFIER 3: K-NEAREST NEIGHBOR CLASSIFIER, 4: NAIVE BAYES CLASSIFIER.

	Recognition Accuracy (%)						
Classifier	Worst	Mean	Best	Std	Class 1	Class 2	Class 3
1	73.3300	78.0000	83.3300	0.0384	94.0000	40.6667	99.3300
2	72.2200	74.6700	80.0000	0.0250	27.6667	79.3333	99.3333
3	50.0000	53.3333	55.7800	0.0207	91.6667	21.6667	54.0000
4	76.6667	79.0000	82.2222	1.8898	99.3333	34.3333	100.0000

TABLE VII

SUMMARY OF TESTING SAMPLES RECOGNITION PERFORMANCE FOR EEG SIGNAL UNDER DATASET SUBJECT TO NOISE LEVEL OF 0.2. CLASSIFIER 1: FSVM CLASSIFIER, CLASSIFIER 2: TRADITIONAL SVM CLASSIFIER, CLASSIFIER 3: K-NEAREST NEIGHBOR CLASSIFIER, 4: NAIVE BAYES CLASSIFIER.

	Recognition Accuracy (%)						
Classifier	Worst	Mean	Best	Std	Class 1	Class 2	Class 3
1	73.3300	78.0000	82.2200	0.0295	86.0000	48.0000	100.0000
2	67.7800	68.6700	70.0000	0.0126	36.3333	76.3333	98.3333
3	54.4444	56.6700	58.8900	0.0236	92.6700	23.0000	54.3333
4	75.5556	78.2222	80.0000	1.5585	99.6667	33.6667	100.0000

TABLE VIII
SUMMARY OF TESTING RECOGNITION PERFORMANCE FOR EEG SIGNAL UNDER DATASET SUBJECT TO NOISE LEVEL OF 0.5. CLASSIFIER 1: FSVM CLASSIFIER, CLASSIFIER 2: TRADITIONAL SVM CLASSIFIER, CLASSIFIER 3: K-NEAREST NEIGHBOR CLASSIFIER, 4: NAIVE BAYES CLASSIFIER.

VII. EXPERIMENTAL RESULTS/DISCUSSION

The proposed IT2FSVM classifier is used to classify between the 3 epilepsy seizure phases using the feature vector that has been obtained by the method detailed in Section V-A. For comparison purposes, 3 traditional classifiers (kNN, naive Bayes and SVM classifiers) are considered. When traditional SVM classifier is considered, they are connected in the classifier structure as shown in Fig. 2, i.e., replacing the IT2SVM with the traditional SVM. For the design of the hyperplane, all three traditional SVMs take the RBF kernel with the width of $\sqrt{1/1400}$ and regularization constant $C = 500$.

The recognition accuracy with respect to the training dataset for all classifiers is given in Tables III and IV. The tables show the training and testing recognition accuracy from the best performed classifiers during the design. It tabulates the worst (among the three classes), best (among the three classes), average (over the three classes) and individual class recognition accuracy for both training and testing dataset.

It can be seen from Tables III that the kNN classifier performs the best in terms of average recognition accuracy of 100%. The IT2FSVM classifier comes in second place with 99.0510% (less than 1% compared with the kNN classifier). This however is not an indication of the kNN being a superior classifier as we see that it suffers from a significant reduction in its average recognition performance when exposed to unseen test data with and without noise as seen in column 3 of Table. IV and column 2 of Tables V-VIII. The 100% average training accuracy seen in Table. III is reduced to 56.6667% in Table.IV when the classifier is subjected to the test data. Another significant impact of this is that the kNN now has an individual testing recognition accuracy of 23.3333% as seen in column 5 of Table.IV when classifying the pre-seizure phase (class 2), this is of significance because the accurate classification of the pre-seizure phase is a core objective in addressing the problem of epilepsy seizure phase classification as this would give the patients the advance warning and therefore sufficient time to prepare for the onset of the seizure. The SVM and naive Bayes come in the third and fourth places. Referring to Table IV, it can be seen that the IT2FSVM classifier outperforms other classifiers in terms of average recognition accuracy for testing dataset, which shows that the IT2FSVM demonstrates an outstanding generalization ability dealing with unseen data. Compared with other classifiers, the average testing recognition accuracy is 10% to 21% higher. It is interestingly observed that the naive Bayes classifier performs the worst to the testing data, which shows that it is very sensitive to unseen dataset. Referring to the worst individual class testing recognition accuracy, IT2FSVM can maintain 70% while other drops to around 23% to 50%.

Tables V to VIII show the testing recognition accuracy for the testing data subject to Gaussian noise levels of 0.05, 0.1, 0.2 and 0.5. The experiments were repeated 10 times for each classifiers. The “Worst” and the “Best” column show the worst and best testing individual class recognition accuracy among the 10 times of experiments. The “Mean” and “Std” column show the mean and standard deviation of the average testing accuracy of the three classes of the 10 experiments. The “Class 1”, “Class 2” and “Class 3” columns show average testing recognition accuracy for classes 1 to 3, respectively, of the 10 experiments.

In general, the recognition accuracy drops for all classifiers when the noise level increases. In most of the cases, the average testing recognition of IT2FSVM and naive Bayes classifiers offer the best result. However, when it is down to the individual class recognition accuracy, especially for higher noise levels (0.1, 0.2 and 0.5), the IT2FSVM performs more stable giving the lowest class recognition accuracy of 40% while other classifiers give the lowest class recognition accuracy ranging from 15% to 36%. Similar to the comment concerning the kNN and its poor performance in accurately classifying the pre-seizure phase (Class 2), it is important to also note that the nave Bayes classifier exhibits a relatively poor ability to classify the pre-seziure phase as we see that the SVM and IT2FSVM provides superior class recognition for the pre-seizure phase in the training, noise-free testing and noise testing of both classifiers. This is a critical difference between these classifiers. The IT2FSVM however shows a superior overall/average recognition accuracy when compared to the SVM and this shows the superiority and suitability of the IT2FSVM for classifying between the three epilepsy seizure phases.

VIII. Conclusion

In the research conducted in this paper, an IT2FSVM is introduced which has been able to show a high level of proficiency in dealing with complicated classification and recognition problems. The IT2FSVM merges the SVM and IT2FIS to create a hybrid classifier with improved performance when compared to the some traditional classifiers. The IT2FSVM classifier has been applied to the epilepsy phase classification problem. The results obtained from the simulations that were carried out show that the IT2FSVM has a better performance than the traditional kNN and naive Bayes and SVM method when the classifier is subjected to the original and uncontaminated input data. The input data was then contaminated with noise in order to test the robustness of the classification methods. The IT2FSVM proved to have a significant level of robustness to noise in the data as there was a relatively smaller impact on the recognition accuracy of the IT2FSVM under noisy data when compared to other classification methods. Further work on this research direction will involve investigating better ways to optimise the membership function, also trying out other IT2FSVM architectures in an effort to improve the overall recognition accuracy.

Acknowledgment

The work described in this paper was partly supported by King's College London and China Scholar Council.

References

[1] P. L. Lanzi, W. Stolzmann, and S. W. Wilson, *Learning classifier systems: from foundations to applications*. Springer Science & Business Media, 2000, no. 1813.

[2] J. Zurada, "Could decision trees improve the classification accuracy and interpretability of loan granting decisions?" in *2010 43rd Hawaii International Conference on System Sciences (HICSS)*. IEEE, 2010, pp. 1–9.

[3] J. Pradeep, E. Srinivasan, and S. Himavathi, "Neural network based handwritten character recognition system without feature extraction," in *2011 International Conference on Computer, Communication and Electrical Technology (ICCCET),*. IEEE, 2011, pp. 40–44.

[4] H. Zhang, W. Shi, and K. Liu, "Fuzzy-topology-integrated support vector machine for remotely sensed image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 3, pp. 850–862, 2012.

[5] C.-F. Juang, R.-B. Huang, and W.-Y. Cheng, "An interval type-2 fuzzy-neural network with support-vector regression for noisy regression problems," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 4, pp. 686–699, 2010.

[6] C. Jen-Tzung, "Linear regression based bayesian predictive classification for speech recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 11, pp. 70–79, Jan 2003.

[7] U. Kumar, S. K. Raja, C. Mukhopadhyay, and T. V. Ramachandra, "Hybrid bayesian classifier for improved classification accuracy," *Geoscience and Remote Sensing Letters, IEEE*, vol. 8, no. 3, pp. 474–477, Nov. 2010.

[8] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," *Informatica*, vol. 31, pp. 249–268, 2007.

[9] G. P. Zhang, "Neural networks for classification: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 4, no. 30, pp. 451–462, Nov 2000.

[10] A. Santillana Fernandez, C. Delgado-Mata, and R. Velazquez, "Training a single-layer perceptron for an approximate edge detection on a digital image," in *2011 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*. IEEE, 2011, pp. 189–193.

[11] C.-T. Lin, C.-M. Yeh, S.-F. Liang, J.-F. Chung, and N. Kumar, "Support-vector-based fuzzy neural network for pattern classification," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 1, pp. 31–41, 2006.

[12] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[13] Y.-H. Liu and Y.-T. Chen, "Face recognition using total margin-based adaptive fuzzy support vector machines," *IEEE Transactions on Neural Networks*, vol. 18, no. 1, pp. 178–192, 2007.

[14] H.-S. Yan and D. Xu, "An approach to estimating product design time based on fuzzy-support vector machine," *IEEE Transactions on Neural Networks*, vol. 18, no. 3, pp. 721–731, 2007.

[15] H. Lam and J. Prada, "Interpretation of handwritten single-stroke graffiti using support vector machines," *International Journal of Computational Intelligence and Applications*, vol. 8, no. 04, pp. 369–393, 2009.

[16] Y. Chen and J. Z. Wang, "Support vector learning for fuzzy rule-based classification systems," *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 6, pp. 716–728, 2003.

[17] X. Yang, G. Zhang, J. Lu, and J. Ma, "A kernel fuzzy C-means clustering-based fuzzy support vector machine algorithm for classification problems with outliers or noises," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 1, pp. 105–115, 2011.

[18] Y. Wang, "Type-2 fuzzy probabilistic system," in *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. IEEE, 2012, pp. 482–486.

[19] S. Greenfield and F. Chiclana, "Fuzzy in 3-D: Contrasting complex fuzzy sets with type-2 fuzzy sets," in *2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*. IEEE, 2013, pp. 1237–1242.

[20] O. Uncu and I. Turksen, "Discrete interval type 2 fuzzy system models using uncertainty in learning parameters," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 1, pp. 90–106, 2007.

[21] C.-F. Lin and S.-D. Wang, "Fuzzy support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 464–471, 2002.

[22] R. S. Fisher, W. v. E. Boas, W. Blume, C. Elger, P. Genton, P. Lee, and J. Engel, "Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ILAE) and the international bureau for epilepsy (IBE)," *Epilepsia*, vol. 46, no. 4, pp. 470–472, 2005.

[23] L. B. and E. J., "Prediction of epileptic seizures," *Lancet Neurol*, vol. 1, pp. 22–30, May. 2002.

[24] O. A. Rosso, A. Mendes, R. Berretta, J. A. Rostas, M. Hunter, and P. Moscato, "Distinguishing childhood absence epilepsy patients from controls by the analysis of their background brain electrical activity (ii): a combinatorial optimization approach for electrode selection," *Journal of neuroscience methods*, vol. 181, no. 2, pp. 257–267, 2009.

[25] F. Amor, S. Baillet, V. Navarro, C. Adam, J. Martinerie, and M. Le Van Quyen, "Cortical local and long-range synchronization interplay in human absence seizure initiation," *Neuroimage*, vol. 45, no. 3, pp. 950–962, 2009.

[26] C. Wu, X. Wang, D. Bai, and H. Zhang, "Fast incremental learning algorithm of SVM on KKT conditions," in *Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 1. IEEE, 2009, pp. 551–554.

[27] C.-T. Lin, C.-M. Yeh, S.-F. Liang, J.-F. Chung, and N. Kumar, "Support-vector-based fuzzy neural network for pattern classification," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 1, pp. 31–41, 2006.

[28] J. M. Mendel, "Type-2 fuzzy sets and systems: an overview," *IEEE Computational Intelligence Magazine*, vol. 2, no. 1, pp. 20–29, 2007.

[29] N. N Karnik and J. M Mendel, "Operations on type-2 fuzzy sets," *Fuzzy sets and systems*, vol. 122, no. 2, pp. 327–348, 2001.

[30] D. Wu and J. M. Mendel, "Enhanced Karnik–Mendel algorithms," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 4, pp. 923–934, 2009.

[31] C. A. E. and M. F., "Brain mechanisms of altered conscious states during epileptic seizures," *Nat Rev Neurol*, vol. 5, pp. 267–76, May. 2009.

[32] B. D. Killory, X. Bai, M. Negishi, C. Vega, M. N. Spann, M. Vestal, J. Guo, R. Berman, N. Danielson, J. Trejo *et al.*, "Impaired attention and network connectivity in childhood absence epilepsy," *Neuroimage*, vol. 56, no. 4, pp. 2209–2217, 2011.

[33] A. Gorji, C. Mittag, P. Shahabi, T. Seidenbecher, and H.-C. Pape, "Seizure-related activity of intralaminar thalamic neurons in a genetic model of absence epilepsy," *Neurobiology of disease*, vol. 43, no. 1, pp. 266–274, 2011.

[34] H. K. Meeren., J. P. Pijn., E. L. V. Luijtelaar., A. M. Coenen., and F. H. L. da Silva., "Cortical focus drives widespread corticothalamic networks during spontaneous absence seizures in rats," *J Neurosci*, vol. 22, pp. 1480–95, Feb. 2002.

[35] L. D. Iasemidis, J. C. Sackellares, H. P. Zaveri, and W. J. Williams, "Phase space topography and the Lyapunov exponent of electrocorticograms in partial seizures," *Brain topography*, vol. 2, no. 3, pp. 187–201, 1990.

[36] M. Niknazar, S. R. Mousavi, S. Motaghi, A. Dehghani, B. V. Vahdat, M. B. Shamsollahi, M. Sayyah, and S. M. Noorbakhsh, "A unified approach for detection of induced epileptic seizures in rats using ECoG signals," *Epilepsy & Behavior*, vol. 27, no. 2, pp. 355–364, 2013.

[37] H. K. Lam, U. Ekong, H. Liu, B. Xiao, H. Araujo, S. H. Ling, and K. Y. Chan, "A study of neural-network-based classifiers for material classification," *Neurocomputing*, vol. 144, pp. 367–377, 2014.

[38] H. K. Lam, U. Ekong, B. Xiao, G. Ouyang, H. Liu, K. Y. Chan, and S. H. Ling, "Variable weight neural networks and their applications on material surface and epilepsy seizure phase classifications," *Neurocomputing*, vol. 149, pp. 1177–1187, 2015.

[39] I. Güler and E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients," *Journal of neuroscience methods*, vol. 148, no. 2, pp. 113–121, 2005.

[40] N. Kannathal, M. L. Choo, U. R. Acharya, and P. K. Sadasivan, "Entropies for detection of epilepsy in EEG," *Computer methods and programs in biomedicine*, vol. 80, no. 3, pp. 187–194, 2005.

[41] E. D. Ubeyli., "Automatic detection of electroencephalographic changes using adaptive neuro-fuzzy inference system employing lyapunov exponents," *Expert Systems with Applications*, vol. 36, pp. 9031–9038, Jul. 2009.

[42] M. Thulasidas, C. Guan, and J. Wu, "Robust classification of eeg signal for brain-computer interface," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, no. 1, pp. 24–29, 2006.

[43] J. O'Keefe and M. L. Recce, "Phase relationship between hippocampal place units and the EEG theta rhythm," *Hippocampus*, vol. 3, no. 3, pp. 317–330, 1993.