2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Medical Diagnosis by Fuzzy Standard Additive Model with Wavelets

Thanh Nguyen, Abbas Khosravi, Douglas Creighton and Saeid Nahavandi
Centre for Intelligent Systems Research (CISR), Deakin University
Geelong Waurn Ponds Campus, Victoria, Australia
E-mail: thanh.nguyen@deakin.edu.au

***Abstract*—This paper proposes a combination of fuzzy standard additive model (SAM) with wavelet features for medical diagnosis. Wavelet transformation is used to reduce the dimension of high-dimensional datasets. This helps to improve the convergence speed of supervised learning process of the fuzzy SAM, which has a heavy computational burden in high-dimensional data. Fuzzy SAM becomes highly capable when deployed with wavelet features. This combination remarkably reduces its computational training burden. The performance of the proposed methodology is examined for two frequently used medical datasets: the lump breast cancer and heart disease. Experiments are deployed with a five-fold cross validation. Results demonstrate the superiority of the proposed method compared to other machine learning methods including probabilistic neural network, support vector machine, fuzzy ARTMAP, and adaptive neuro-fuzzy inference system. Faster convergence but higher accuracy shows a win-win solution of the proposed approach.**

Keywords—fuzzy system; wavelet transformation; medical diagnosis; breast cancer; heart disease

I. Introduction

The automatic diagnosis of heart disease and breast cancer is an important, challenging medical problem. Heart disease affects health and working performance of patients, especially old people. The World Health Organization has assessed globally that 12 million people lose their lives every year because of the heart diseases [1]. The heart disease actually can be detected early by performing a number of medical tests. However, these tests are usually costly and confronted a certain difficulty. An inexpensive solution based on medical history of patients and some simple tests is commonly proposed by heart disease investigators.

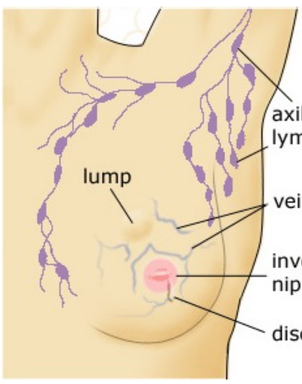


Fig. 1. Early signs of breast cancer (adapted from [2])

Breast cancer is also one of the largest causes of cancer deaths among women. Early prediction of the characteristic of breast lumps (benign or malignant) occurring in patients thus help to determine a suitable treatment for the cancer (Fig. 1).

There have been a number of studies dealing with medical diagnosis in the literature. A comprehensive review of machine learning methods was presented in [3]. Recently, Akay [4] proposed support vector machine (SVM) method combined with feature selection for breast cancer diagnosis. Karabatak and Ince [5] introduced another approach to detecting breast cancer based on association rules and neural network. Alternatively, Marcano-Cedeño et al. [6] presented a novel improvement in neural network training for classifying the breast cancer tumours as benign or malignant. The approach encompasses simulating the biological property of metaplasticity on multilayer perceptron with backpropagation.

On the other hand, Dangare and Apte [7] suggested a system using medical terms such as sex, blood pressure, cholesterol, obesity and smoking attributes to predict the likelihood of patient getting a heart disease.

Bhatla and Jyoti [8] also investigated a number of data mining techniques for automated heart disease prediction systems. Likewise, Sundar et al. [9] proposed a prototype using Naïve Bayes and weighted associative classifier for heart disease diagnosis. Similarly, Lakshmi [10] considered several data mining techniques and constructed a web based user friendly system for predicting heart disease survivability.

Generally, medical diagnosis and prognosis are decision making problems that commonly have uncertainty involved [11]. The use of fuzzy set theory has been emerged in pattern recognition methods for medical diagnosis. To deal with uncertain and high-dimensional medical data, this paper proposes a method using fuzzy SAM and wavelet features for medical diagnosis. To our best knowledge, it is the first application of fuzzy SAM method for medical diagnosis and also the first combination of wavelet features and fuzzy SAM in a classification system. Through this study, we examine and compare performance of fuzzy SAM models with classification methods frequently applied in literature. Experiments are conducted using two medical datasets to make sure conclusions driven out of this study are valid and general.

The rest of the paper is organized as follows. The next section presents the proposed combination wavelet-SAM method. Other machine learning methods are briefly presented in Section III for the sake of comparisons. Section IV is devoted for experimental results, which are followed by concluding remarks.

II. Fuzzy SAM with Wavelet Features

A. Fuzzy SAM

The fuzzy SAM was introduced by Kosko [12]. The fuzzy system F: $R^n \rightarrow R^p$ comprises m if-then rules, which is able to universally approximate continuous and bounded measurable functions in the compact domain. Any kind of the if-part fuzzy sets $A_j \subset R^n$ can be selected. The same goes for the then-part fuzzy sets $B_j \subset R^p$ because the fuzzy SAM employs only the centroid c_j and volume V_j of B_j to calculate the output $F(x)$ from the vector input $x \in R^n$ [13].

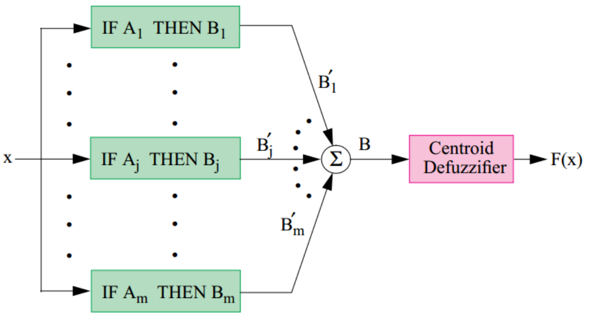
$$F(x)=Centroid\left(\sum_{j=1}^{m} w_j a_j(x) B_j\right)=\frac{\sum_{j=1}^{m} w_j a_j(x) V_j c_j}{\sum_{j=1}^{m} w_j a_j(x) V_j}=\sum_{j=1}^{m} p_j(x) c_j \qquad (1)$$

The graph of an approximand f is covered with m fuzzy rule patches. The form of the fuzzy rule is *"If $X = A_j$ then $Y = B_j$"*. If-part set $A_j \subset R_n$ has the joint membership function a_j: $R_n \rightarrow [0, 1]$ that factors: $a_j(x)=a_j^1(x_1)...a_j^n(x_n)$. Then-part fuzzy set $B_j \subset R_p$ has the membership function b_j: $R_n \rightarrow [0, 1]$ and volume (or area) V_j and centroid c_j [14].

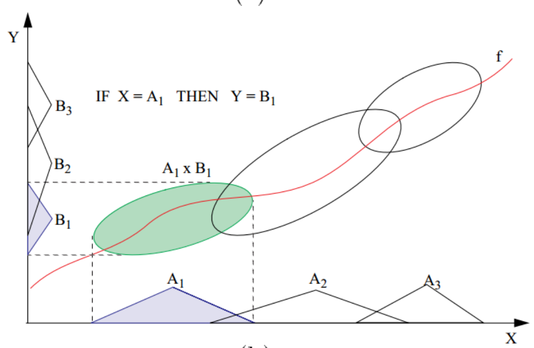
The convex weights:

$$p_j(x)=\frac{w_j a_j(x) V_j}{\sum_{k=1}^{m} w_j a_k(x) V_k} \qquad (2)$$

make the SAM output $F(x)$ as a convex sum of then-part set centroids.



(a)



(b)

Fig. 2. (a) A configuration of SAM. Each input fires each fuzzy rule to some degree to calculate F(x). (b) Fuzzy rules specify patches in the input-output space [15].

Fig. 2 displays the configuration of the additive system and its state-space graph cover. The graph cover induces an exponential rule explosion. A fuzzy system requires on the order of k^{n+p-1} rules to approximate a function f: $R^n \rightarrow R^p$ in a compact domain [15].

Learning is an essential process of SAM to construct a knowledge base that is a structure of if-then fuzzy rules [13]. The SAM learning process includes two basic steps: a) unsupervised learning for constructing if-then fuzzy rules and b) supervised learning for tuning rule parameters.

The supervised learning usually starts from a randomly initialized set of parameters and ends when it meets the predetermined stopping criteria. Proper parameter initialization is of paramount importance as training process costs much time and is often trapped in local minima. The unsupervised learning process helps to initialize parameters of fuzzy rules more skilfully. In this paper, we utilize the adaptive vector quantization (AVQ) clustering method [12] to identify the centres of membership functions (MFs) in the antecedent part and the centroids in the consequent part. The well-separated distribution of the resulting clusters from the AVQ method is useful in identifying the allocation of fuzzy rules in the fuzzy SAM. The AVQ clustering method is briefly summarized in the next section.

B. SAM Unsupervised Learning by the AVQ Clustering

The clustering process uses K quantization vectors to search for fuzzy classes in the learning dataset that cover the unknown function f in the space XY. The K quantization vectors can be initialized randomly. For each data pattern at time t: $z(t) = [x(t)|y(t)]$, the algorithm searches for a fuzzy class that can contain $z(t)$ based on the closest q_j (competitive learning), which is selected based on the following conditions:

$$\|z(t) - q_j\| = \min_{i=\overline{1,k}} \|z(t) - q_i(t)\| \qquad (3)$$

where $\|z\|^2 = z_1^2 + z_2^2 + \cdots + z_n^2$.

Then q_j is updated to be closer to $z(t)$:

$$q_j(t+1) = q_j(t) + \mu_t[z(t) - q_j(t)] \qquad (4)$$

Based on the competitive learning, q_j vectors are updated closer to the fuzzy classes covering the graph of the unknown function f. At the end of training, K quantization vectors q_j obtained reflect the distribution of fuzzy classes of the training data.

Denote the learning dataset as $\{z_t\}$, $t = 1, \ldots, N$, and the local conditional covariance matrix Q_j in pattern class D_j as $Q_j(t) = E[(z-\bar{z})(z-\bar{z})^T|D_j]$, the algorithm is presented as follows:

Step 1. Initialize q_j randomly, $K_j = 0, j = 1, \ldots, K$

Step 2. Consider the learning sample at time t: $z(t) = [x(t)|y(t)]$

Step 3. Search for q_j at time t based on Eq. (3).

Step 4. Update quantization vectors q and distance vector K:

If $i = j$: $q_i(t+1) = q_i(t) + \mu_t[z(t) - q_i(t)]$

$$Q_i(t+1) = Q_i(t) + \mu_t[(z(t) - q_i(t))(z(t) - q_i(t))^T - Q_i(t)] \qquad (5)$$

If $i \neq j$: $q_i(t+1) = q_i(t)$

$$Q_i(t+1) = Q_i(t) \tag{6}$$

Step 5. If $t < N$ then $t = t + 1$, back to step 2.

Step 6. End.

The AVQ clustering method is applied to initialize parameters of fuzzy SAM. We organize the corresponding input and output data into a unique observation of $p + 1$ dimensions where p is the number of inputs and one output corresponding to the class being assigned. Denote x_i is the ith organized observation ($i = 1, \ldots, N$), x_i is presented as follows:

$$x_i = [input_i^1, input_i^2, \ldots, input_i^p, output_i] \tag{7}$$

where $input_i^j$ is the jth input of the ith observation and $output_i$ is the output of the ith observation. By clustering the sample of N observations having the above format, we are able to derive the K resulting clusters corresponding with K fuzzy rules of the fuzzy SAM. Since the AVQ clustering is completed, centres of the resulting clusters are assigned to centres of the MFs. The centres of the output of each rule will be assigned equal to the output value of the corresponding cluster. The widths of the MFs of each rule are initialized equal to the standard deviation of the data.

The *Sinc* membership function $\sin(x)/x$ recommended as the best shape for a fuzzy set in function approximation is used to construct if-then fuzzy rules [15]. The j^{th} sinc set function (Fig. 3) centered at m_j and width $d_j > 0$ is characterized as below:

$$a_j(x) = \sin\left(\frac{x - m_j}{d_j}\right) \Big/ \left(\frac{x - m_j}{d_j}\right) \tag{8}$$

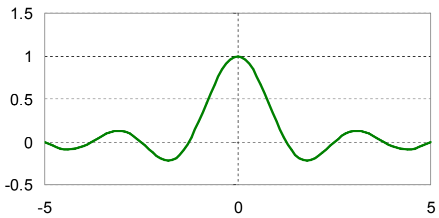


Fig. 3. Sinc membership function in 1-D case with centre m = 0 and width d = 0.4 [16]

C. SAM Supervised Learning

Fuzzy rule parameters are adjusted using a supervised learning process. The supervised gradient descent can adjust all parameters of the SAM model [13, 17]. The aim is to minimize the squared error:

$$E(x) = \frac{1}{2}\left(f(x) - F(x)\right)^2 \tag{9}$$

of the function approximation. Both the vector function $f: R^n \to R^p$ and the approximated function F have p components. The f function is the form $f(x) = \left(f_1(x), \ldots, f_p(x)\right)^T$. On the other hand the F encompasses $F(x) = \left(F_1(x), \ldots, F_p(x)\right)^T$. Let ξ_j^k denote the k^{th} parameter in the membership function a_j. Then the chain rule induces the gradient of the error function with respect to ξ_j^k, with respect to the then-part set centroid $c_j = \left(c_j^i, \ldots, c_j^p\right)^T$, and with respect to the then-part set volume V_j [14].

A gradient descent learning rule for a SAM parameter is as follows:

$$\xi(t+1) = \xi(t) - \mu_t \frac{\partial E}{\partial \xi} \tag{10}$$

where μ_t is the learning rate at iteration t.

Generally, there are two ways to adjust parameters. Batch form refers to the update process that occurred when all training samples have completely passed through the system. Incremental form refers to the update that occurred as soon as a sample was processed. With significantly nonlinear data, incremental adjustment often proves effective and more stable, and it is therefore applied in this study.

The momentum technique is also integrated so as to enhance the convergent speed of the parameter tuning process [18]. The learning expression with momentum is as follows:

$$\xi(t+1) = \xi(t) - \mu_t \frac{\partial E}{\partial \xi} + \varepsilon.\Delta\xi(t) \tag{11}$$

where ε is the momentum coefficent.

D. Wavelet Transformation (WT) Combined with SAM

Fuzzy systems in general or SAM in particular normally faces a big challenge in training if there are many inputs of the data. The curses of dimensionality of the fuzzy SAM were specifically investigated in [19]. In general, high-dimensional data would decline convergence speed and thus performance of the fuzzy SAM system. Therefore, there must be a need of a dimension reduction or feature selection tool that may be implemented before the fuzzy SAM is executed. This is particularly important as the medical data are usually assembled in high dimension.

WT is one of popular methods for reducing number of dimensions in datasets. With datasets having dimensions reduced, fuzzy systems would demonstrate more powerful ability in function approximation and classification. The proposed methodology for medical diagnosis in this study is diagrammed in Fig. 4.

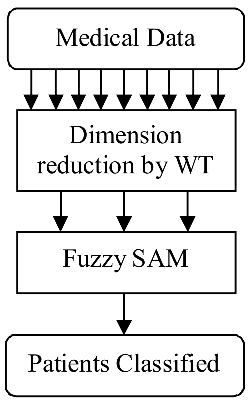


Fig. 4. The proposed combination between WT and Fuzzy SAM

The following presents a summary of WT and its usage in this paper. WT represents a signal in a time-frequency fashion [20]. Once the wavelets (the mother wavelet) $\varphi(x)$ is fixed, translations and dilations of the mother wavelet can be formed $\left\{\varphi\left(\frac{x-b}{a}\right), (a,b) \in \mathrm{R}^{+} \times \mathrm{R}\right\}$. It is convenient to take special values for a and b as $a = 2^{-j}$ and $b = 2^{-j}k$ where j and k are integers. One of the simplest wavelets is the Haar wavelet, which has been used in various applied mathematics. Haar functions can uniformly approximate any continuous function. Dilations and translations of the function φ, which is $\varphi_{jk}(x) = const.\, \varphi(2^{j}x - k)$, define an orthogonal basis in $L^2(R)$. This means that any element in $L^2(R)$ may be represented as a linear combination of these basis functions. The scaling function in Haar wavelet is simply unity on the interval [0,1) as $\phi(x) = 1\ (0 \le x < 1)$. Quiroga et al. [21] employed a four-level decomposition using Haar wavelets for spike sorting. The wavelet coefficients are then selected by the Lilliefors modification of a Kolmogorov-Smirnov test for normality. In this paper, we employ the similar WT procedure as in [21] for dimension reduction applied to medical data.

III. Brief Descriptions of Comparable Classification Methods

In this section, we briefly describe comparable classification methods for the sake of comparisons with the proposed wavelet-SAM method. Methods we select as benchmarks for comparisons are Probabilistic Neural Network (PNN), Fuzzy ARTMAP (FAM), Support Vector Machine (SVM) and Adaptive Neuro-Fuzzy Inference System (ANFIS). As SVM [22] and ANFIS [23] are well-known in the literature, we just scrutinize below PNN and FAM.

A. Probabilistic Neural Network

A PNN, which was introduced by Specht in the early 1990s [24], is a feedforward, one-pass training neural network. It was resulted from the Bayesian network along with the Kernel Fisher discriminant analysis statistical algorithm. A PNN is primarily a classifier that maps inputs to numerous classifications. Operations in PNN are structured into a multilayered network comprising input layer, pattern layer, summation layer and output layer. The PNN employs Parzen probability distribution function (pdf) estimators. The estimated pdf asymptotically approaches the true pdf as the training set size increases given that the true pdf is smooth and continuous. The weighting function commonly used is the spherical Gaussian radial basis functions centred at each training vector. The likelihood of an unknown vector pertaining to a given class can be formularized as:

$$f_i(x) = \frac{1}{(2\pi)^{p/2}\sigma^{p} M_i} \sum_{j=1}^{M_i} exp \frac{-(x-x_{ij})^T (x-x_{ij})}{2\sigma^2} \qquad (12)$$

where i is the class number, j is the pattern number, x_{ij} is the jth training vector from class i, x is the test vector, M_i is the number of training vectors in class i, p is the dimension of vector x, σ is the smoothing factor, and $f_i(x)$ is the sum of multivariate spherical Gaussians centred at each of the training vectors x_{ij} for the ith class pdf estimate [25].

Classification decisions are made based on the Bayes optimal decision rule as follows:

$$d(x) = C_i \text{ if } f_i(x) > f_k(x) \text{ for } k \neq i \qquad (13)$$

where C_i is the class i.

B. Fuzzy ARTMAP

FAM is a supervised clustering algorithm, which can be regarded as one of the leading neural networks for classification [26]. A FAM comprises two fuzzy ART modules, i.e. fuzzy ART$_a$ and ART$_b$, interrelated through a map field (Fig. 5).

The fuzzy ART$_a$ executes clustering in the input space of data whilst the fuzzy ART$_b$ carries out clustering in the output space of the target data. Each fuzzy ART model consists of three layers [27]:

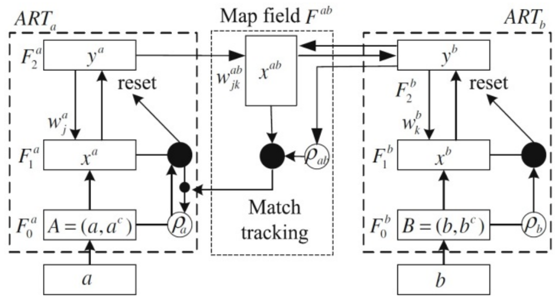


Fig. 5. Fuzzy ARTMAP structure [28]

- A normalization layer F_0 that complement-code an M-dimensional input vector a to a 2M-dimensional vector A: $A = (a, a^c) = (a_1, \ldots, a_M, 1-a_1, \ldots, 1-a_M)$.
- An input layer F_1 which receives A.
- A recognition layer F_2 that encodes prototypes of input patterns and is able to create new nodes when necessary.

By propagating A from F_1 to F_2 the responses of each node j in F_2 is calculated based on a choice function.

$$T_j = \frac{|A \wedge w_j|}{\alpha + |w_j|} \qquad (14)$$

where α is the choice parameter, w_j is the weight of node j. The winning node J is selected with the highest response based on the winning-take-all strategy. The winning node is then checked with a vigilance threshold by the following formula:

$$\frac{|A \wedge w_J|}{|A|} \ge \rho \qquad (15)$$

A map field vigilance test is carried out for the two winning nodes from fuzzy ART$_a$ and ART$_b$ to confirm the outcomes. If the test is acceptable, parameters are updated:

$$w_J^{new} = \beta\left(A \wedge w_J^{old}\right) + (1-\beta) w_J^{old} \qquad (16)$$

where β is the learning rate, $\beta \in [0,1]$. Otherwise, it implies that the prediction of the winning prototype of fuzzy ART$_a$ is not matched with the target class in fuzzy ART$_b$. If this circumstance happens, a match-tracking process is commenced to obstruct the current winning node and a new search loop for other winning prototype is produced in fuzzy ART$_a$. If none of the existing prototypes recognizes the input pattern, a new node is generated, and the input is assigned as its prototype pattern [27].

IV. Experimental Results

Two medical datasets used in this research are breast cancer and heart disease datasets. The breast cancer database was acquired from the University of Wisconsin Hospitals, Madison by Dr. William H. Wolberg [29]. The heart dataset was from the M.D. Robert Detrano at V.A. Medical Center, Long Beach and Cleveland Clinic Foundation [30]. Details of each dataset and its experimental settings and results are reported in the following.

A. Breast Cancer Prediction

The dataset contains 699 cases about patients who had undergone surgery for breast cancer. The output values are either 2 or 4 indicating that sleeping cancer lump (benign) or dangerous lump (malignant). Nine other fields are valued from 1 to 10, which are detailed in Table 1. The task is to determine if the detected tumour is benign (2) or malignant (4) given values of nine attributes described in Table 1.

Table 1. The breast cancer dataset attribute description [31]

Attribute	Domain
Clump Thickness	[1, 10]
Cell Size	[1, 10]
Cell Shape	[1, 10]
Marginal Adhesion	[1, 10]
Epithelial Size	[1, 10]
Bare Nuclei	[1, 10]
Bland Chromatin	[1, 10]
Normal Nucleoli	[1, 10]
Mitoses	[1, 10]
Class	{2,4}

Sixteen cases with missing data in any field are excluded from this experiment. In general, the benign cases occupy 65.5% of the whole dataset whilst the rest 34.5% is of the malignant cases. For ease of processing, the output class are transformed into 1 or 2 where 1 is represented for benign and 2 is for malignant. The output value of fuzzy SAM is considered to be benign if it is smaller than 1.5, otherwise it is indicated as malignant. Before running the fuzzy SAM, wavelet transformation is performed to reduce from original 9 inputs to only 3 inputs.

We use five-fold cross validation procedure to deploy the experiments where four folds of data are used for learning and the last fold is used for evaluating the performance. The process is repeated 20 times and the average accuracy is reported in Table 2.

The comparable methods, i.e., PNN, SVM, fuzzy ARTMAP, and ANFIS, are also carried out in the same settings for the sake of comparisons.

In the PNN training with the Matlab toolbox, the smoothing factor σ is set at default value of 0.1. On the other hand, the commonly used Gaussian radial basis function kernel is selected for the SVM training with the scaling factor at 0.5. In the ANFIS, the Sugeno-type inference system is used with the fuzzy c-means clustering for parameter initialization. The number of fuzzy rules is constructed 30 times less than the number of training samples as so is the fuzzy SAM. Both SAM and ANFIS are trained over 100 epochs.

Table 2. Average results of 20 running times

Methods	Accuracy (%)	
	Without Wavelet	With Wavelet
PNN	93.89	93.75
SVM	93.85	96.01
Fuzzy ARTMAP	94.91	95.57
ANFIS	93.11	95.68
Fuzzy SAM	94.32	97.26

Table 2 reports results of five methods: PNN, SVM, Fuzzy ARTMAP, ANFIS and Fuzzy SAM. Each method is deployed with and without wavelet features. The column "Without Wavelet" shows results obtained when running methods on the original 9 features. The column "With Wavelet" referred to cases that each method runs with 3 wavelet features.

It is seen that fuzzy SAM with wavelet features obtains the highest accuracy at 97.26% whilst SAM without wavelet features just obtains 94.32% of the accuracy. The wavelet transformation obviously boosts the performance of fuzzy SAM as it helps SAM to alleviate the curse of dimensionality. In general, wavelet features help improve accuracy of most of the machine learning methods. However, SAM benefits from dimension reduction the most with near 3% of improvement. SVM, Fuzzy ARTMAP and ANFIS also show a mediocre enhancement in accuracy when using wavelet features. In contrast, PNN does not show the advantage of using wavelet features.

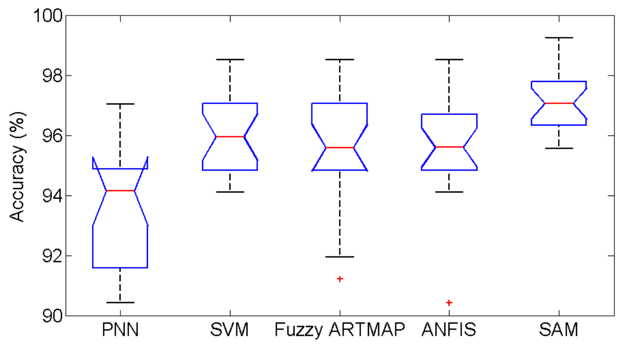


Fig. 6. Box plots for 20 times of trials for each classification method

The box plot in Fig. 6 shows graphical comparisons among Fuzzy SAM versus other methods when using 3 wavelet features. The boxes show the median of the distribution, which is in line with the mean (average) values reported in Table 2. SAM with wavelets demonstrates the highest performance compared to the others. SAM obviously exhibits a consistent and stable performance with the smallest interquartile range (IQR). With the biggest IQR, PNN shows the worst performance among five investigated methods.

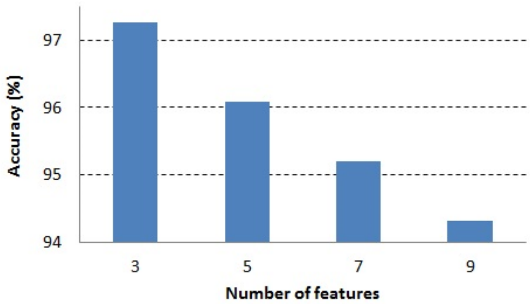


Fig. 7. SAM's performance sensitivity to number of wavelet features

Fig. 7 exemplifies the performance of SAM against different number of wavelet features. As we can see, SAM performance deteriorates drastically when the number of features increases. This fact again confirms the curse of dimensionality in SAM processing. It, in another aspect, shows the effectiveness of dimension reduction by the wavelet transformation technique we applied herein.

B. Heart Disease Prediction

The dataset comprises 303 cases at the Medical Cleveland Centre including 76 variables but only 14 variables are actually used. Descriptions of attributes are presented in Table 3 whilst the attribute domain is summarized in Table 4. Six cases with missing values are excluded before experiments. The "goal" field indicates the occurrence of heart disease in the patient. The task is to detect the occurrence of heart disease in the patient. It is an integer variables ranged from 0 (no presence) to 4. There is 164 cases without the heart disease (valued 0) whilst the rest 139 cases having heart disease with different levels from 1 to 4. Experiments with the Cleveland database have focused to differentiate presence (values 1, 2, 3, or 4) from absence (value 0).

Table 3. Descriptions of attributes of the heart disease dataset [32]

age: age in years	thalach: maximum heart rate achieved
sex: sex (1 = male; 0 = female)	exang: exercise induced angina (1 = yes; 0 = no)
cp: chest pain type, (1: typical angina, 2: atypical angina, 3: non-anginal pain, 4: asymptomatic)	oldpeak: ST depression induced by exercise relative to rest
trestbps: resting blood pressure (in mm Hg on admission to the hospital)	slope: the slope of the peak exercise ST segment, (1: upsloping, 2: flat, 3: downsloping)
chol: serum cholestoral in mg/dl	ca: number of major vessels (0-3) colored by flourosopy
fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)	thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
restecg: resting electrocardiographic results, (0: normal, 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), 2: showing probable or definite left ventricular hypertrophy by Estes' criteria)	num: diagnosis of heart disease (angiographic disease status), (0: < 50% diameter narrowing, 1: > 50% diameter narrowing)

Table 4 elaborates the noisy nature of the heart disease dataset. Fourteen attributes are valued with various ranges.

Table 4. Attribute summary of the heart dataset [32]

Attribute	Domain	Attribute	Domain
Age	[29.0, 77.0]	Thalach	[71.0, 202.0]
Sex	[0.0, 1.0]	Exang	[0.0, 1.0]
Cp	[1.0, 4.0]	Oldpeak	[0.0, 6.2]
Trestbps	[94.0, 200.0]	Slope	[1.0, 3.0]
Chol	[126.0, 564.0]	Ca	[0.0, 3.0]
Fbs	[0.0, 1.0]	Thal	[3.0, 7.0]
Restecg	[0.0, 2.0]	Num	{0,1,2,3,4}

Similar to the previous dataset experiment, we also deploy five-fold cross validation and replicate the process for 20 times before reporting the average results. The results of five machine learning methods are reported in Table 5.

Table 5. Average results of 20 running times with 5-fold cross validation

Methods	Accuracy (%)	
	Without Wavelet	With Wavelet
PNN	55.06	73.80
SVM	57.25	74.27
Fuzzy ARTMAP	62.58	63.46
ANFIS	73.10	74.90
Fuzzy SAM	54.23	78.19

Because of the noisier dataset, all five methods demonstrate lower performance compared to the previous dataset. In this experiment, it is also seen that SAM benefits the most from the dimension reduction. Without wavelet, fuzzy SAM just obtains the accuracy at 54.23% although it increases to the highest accuracy at 78.19% with 3 wavelet features. The SAM performance improvement of 24% when using wavelet features shows a vital need of dimension reduction for SAM especially in noisy data. Fuzz ARTMAP and ANFIS are not benefited much from wavelet features. On the other hand, PNN and SVM show a relative enhancement by using wavelet features.

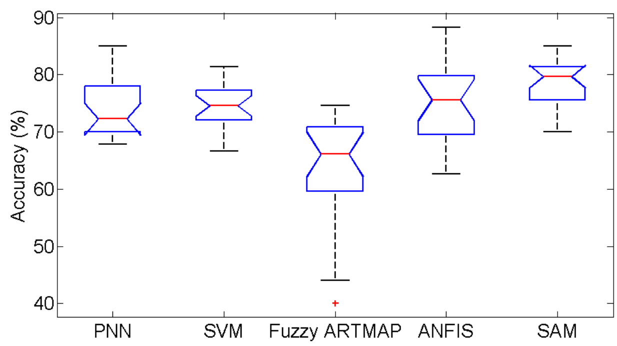


Fig. 8. Box plot comparisons among five methods

As the IQR of SAM is small, it offers a more consistent method compared to the other fuzzy system ANFIS (Fig. 8). Fuzzy ARTMAP shows the worst and most unpredictable results. SVM with wavelets in this experiment illustrates a stable performance though on average it is still lower than the proposed wavelet-SAM.

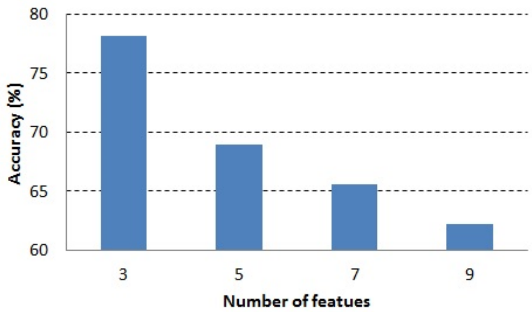


Fig. 9. SAM performance on different number of wavelet features

The sensitivity of wavelet-SAM with the number of selected features is shown in Fig 9. We see similar results to the previous experiment: SAM performance is declined when the number of features increases. Obviously, WT does alleviate the curse of dimensionality of SAM by reducing the number of inputs. SAM needs fewer inputs to achieve higher performance. This on the other hand leads to a substantial computational complexity reduction. The efficiency of the proposed approach wavelet-SAM is thus clearly manifested.

V. CONCLUSIONS

This paper presents a combination of wavelet features with fuzzy SAM for medical diagnosis. Medical data are usually noisy and collected in a high-dimensional format. It is generally a difficult practice to select the most suitable features for a medical diagnosis system. The use of fuzzy system helps to handle the noisiness of the medical data. SAM however has a computational burden in dealing with high-dimensional data. Reducing the data dimensions using wavelet transformation enhances the performance of fuzzy SAM. Experiments carry out not only for the proposed method but also for four other comparable methods, i.e. PNN, SVM, Fuzzy ARTMAP, and ANFIS. Through results of two experiments, we see the dominance of the proposed wavelet-SAM method against the others. Most the investigated machine learning methods show an enhancement when combining with wavelet features. It is thus confirmed that most machine learning methods face a challenge of high-dimensional data. In particular, wavelet features are found most effective when applied to the fuzzy SAM method. Less features but higher performance demonstrates a real double-win solution of the proposed wavelet-SAM for medical diagnosis.

ACKNOWLEDGMENT

This research is supported by the Australian Research Council (Discovery Grant DP120102112) and the Centre for Intelligent Systems Research (CISR) at Deakin University.

REFERENCES

[1] J. Soni, U. Ansari, D. Sharma, & S. Soni, "Intelligent and effective heart disease prediction system using weighted associative classifiers," *International Journal on Computer Science and Engineering, 3(6)*, 2385-2392, 2011.

[2] http://www.breastcancertreatments.co.uk/ Accessed on 1 Nov 2013.

[3] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," *Artificial Intelligence in Medicine, 23(1)*, 89-109, 2001.

[4] M.F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Systems with Applications*, *36*(2), 3240-3247, 2009.

[5] M. Karabatak & M.C. Ince, "An expert system for detection of breast cancer based on association rules and neural network," *Expert Systems with Applications*, *36*(2), 3465-3469, 2009.

[6] A. Marcano-Cedeño, J. Quintanilla-Domínguez, & D. Andina, "WBCD breast cancer database classification applying artificial metaplasticity neural network," *Expert Systems with Applications, 38(8)*, 9573-9579, 2011.

[7] C.S. Dangare & S.S. Apte, "Improved study of heart disease prediction system using data mining classification techniques," *International Journal of Computer Applications, 47(10)*, 44-48, 2012.

[8] N. Bhatla & K. Jyoti, "An analysis of heart disease prediction using different data mining techniques," *International Journal of Engineering, 1(8)*, 1-4, 2012.

[9] N.A. Sundar, P.P. Latha, & M.R. Chandra, "Performance analysis of classification data mining techniques over heart disease database," *Intl. Journal of Engineering Science & Advanced Technology, 2(3)*, 2250-3676, 2012.

[10] K.R. Lakshmi, M.V. Krishna, & S.P. Kumar, "Performance comparison of data mining techniques for predicting of heart disease survivability," *Intl. Journal of Scientific and Research Publications, 3(6)*, 1-10, 2013.

[11] P.K. Han, W.M. Klein, & N.K. Arora, "Varieties of uncertainty in health care: A conceptual taxonomy," *Medical Decision Making*, 31(6), 828-838, 2011.

[12] B. Kosko, "*Fuzzy Engineering*," Prentice Hall, 1996.

[13] T. Nguyen & J. Peterson, "Enhancing the performance of the fuzzy system approach to prediction," *IEEE Fifth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, vol. 1, pp. 259-265, 2008.

[14] T. Nguyen, L. Gordon-Brown, & J. Peterson, "The optimal rule structure for fuzzy systems in function approximation by hybrid approach in learning process," *IEEE International Conference on Computational Intelligence for Modelling Control & Automation*, pp. 1211-1216, 2008.

[15] S. Mitaim & B. Kosko, "What is the best shape for a fuzzy set in function approximation?," *Proc. 5th IEEE Int. Conf. Fuzzy Systems (FUZZ-IEEE-96)*, vol. 2, pp.1237-1243, 1996.

[16] T. Nguyen, L. Gordon-Brown, J. Peterson, & P. Wheeler, "Improving convergence of parameter tuning process of the additive fuzzy system by new learning strategy," *WASET International Journal of Computer, Information Science and Engineering*, 2(9), 59 – 63, 2008.

[17] T. Nguyen, A. Khosravi, D. Creighton, & S. Nahavandi, "Epidemiological dynamics modeling by fusion of soft computing techniques," *IEEE International Joint Conference on Neural Networks* (IJCNN), pp. 1-8, 2013.

[18] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural networks*, 12(1), 145-151, 1999.

[19] S. Mitaim & B. Kosko, “The shape of fuzzy sets in adaptive function approximation,” *IEEE Trans. on Fuzzy Systems*, vol. 9, no. 4, pp. 637-656, 2001.

[20] R.A. DeVore & B.J. Lucier, “Wavelets,” *Acta Numerica*, 1(1), 1-56, 1992.

[21] R.Q. Quiroga, Z. Nadasdy, & Y. Ben-Shaul, “Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering,” *Neural Computation*, 16(8), 1661-1687, 2004.

[22] V. Vapnik, “*The Nature of Statistical Learning Theory*,” Springer, New York, 1995.

[13] J.S. Jang, “ANFIS: adaptive-network-based fuzzy inference system,” *IEEE Transactions on Systems, Man and Cybernetics, 23*(3), 665-685, 1993.

[24] D.F. Specht, “Probabilistic neural networks,” *Neural Networks, 3*, 109-118, 1990.

[25] J. Vargas, M. Ferrer, C. Travieso, & J. Alonso, “Off-line signature verification based on high pressure polar distribution,” *The 11th International Conference on Frontiers in Handwriting Recognition*, 2008.

[26] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds, & D.B. Rosen, “Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps,” *IEEE Transactions on Neural Networks, 3(5)*, 698-713, 1992.

[27] S.C. Tan, M.V.C. Rao, & C.P. Lim, “Fuzzy ARTMAP dynamic decay adjustment: An improved fuzzy ARTMAP model with a conflict resolving facility,” *Applied Soft Computing*, 8(1), 543-554, 2008.

[28] Z. Xu, J. Xuan, T. Shi, B. Wu, & Y. Hu, “A novel fault diagnosis method of bearing based on improved fuzzy ARTMAP and modified distance discriminant technique,” *Expert Systems with Applications, 36(9)*, 11801-11807, 2009.

[29] Wolberg, W. H. and Mangasarian, O. L. (1990) Multisurface method of pattern separation for medical diagnosis applied to breast cytology, *Proceedings of the National Academy of Sciences*, U.S.A., 87, 9193-9196.

[30] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J. Schmid, S. Sandhu, K. Guppy, S. Lee, & V. Froelicher, “International application of a new probability algorithm for the diagnosis of coronary artery disease,” *American Journal of Cardiology, 64*, 304-310, 1989.

[31]http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original). Accessed on 1 Nov 2013.

[32] http://archive.ics.uci.edu/ml/datasets/Heart+Disease. Accessed on 1 Nov 2013.